**Study protein-protein interaction in Methyl-directed DNA mismatch repair in *E. coli*:
Exonuclease I (Exo I) and DNA helicas II (UvrD)
&
A Minimal Exonuclease Domain of WRN Forms a Hexamer on DNA and Possesses Both 3'-
5' Exonuclease and 5'-Protruding Strand Endonuclease Activities
&
Solving the Structure of the Ligand-Binding Domain of the Pregnane-Xenobiotic-Receptor
with 17β Estradiol and T1317**

Yu Xue

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

Chapel Hill
2008

Approved by

Advisor: Professor Erie Dorothy

Advisor: Professor Matthew Redinbo

Reader: Professor Linda Spremulli

Reader: Professor Susan Lord

Reader: Professor Sompop Bencharit

ii

Abstract

Exonuclease I (ExoI) from *Escherichia coli* is a monomeric enzyme that processively degrades single stranded DNA in the 3′ to 5′ direction and has been implicated in DNA recombination and repair. It functions in numerous genome maintenance pathways, with particularly well defined roles in methyl-directed mismatch repair (MMR). The *Escherichia coli* MMR pathway can be reconstituted *in vitro* with the activities of eight proteins (8). MutS, MutL and MutH are involved in initiation of repair including mismatch recognition and generation of a nick at a nearby GATC sequence (53, 54, 55, 56). The hemimethylated state of GATC sequences immediately following replication serves as a signal to direct repair to the nascent strand of the DNA duplex (57, 58). DNA helicase II and one of several exonucleases (Exonucleas I, Exonuclease VII and RecJ) are required to excise the error-containing DNA strand beginning at the nicked GATC site (34, 35). Restoration of the correct DNA sequence by repair synthesis involves DNA polymerase III holoenzyme and SSB, and the final nick is sealed by DNA ligase (34). To identify interactions with ExoI involved in MMR repair system, we used the yeast two-hybrid system with ExoI as bait. By screening an *E.coli* genomic library, *E. coli* DNA helicase II (UvrD) was identified as a potential interacting protein. UvrD has been shown to be required for DNA excision repair, methyl-directed mismatch repair and has some undefined, role in DNA replication and recombination. In this report, *in vitro*

experiments confirm that UvrD and ExoI make a direct physical interaction that may be required for function of the methyl-directed mismatch repair.

Werner Syndrome is a rare autosomal recessive disease characterized by a premature aging phenotype, genomic instability and a dramatically increased incidence of cancer and heart disease. Mutations in a single gene encoding a 1,432 amino-acid helicase/exonuclease (hWRN) have been shown to be responsible for the development of this disease. We have cloned, over-expressed and purified a minimal, 171-amino acid fragment of hWRN that functions as an exonuclease. This fragment, encompassing residues 70-240 of hWRN (hWRN-N$_{70-240}$), exhibits the same level of 3'-5' exonuclease activity as the previously described exonuclease fragment encompassing residues 1-333 of the full-length protein. The fragment also contains a 5'-protruding DNA strand endonuclease activity at a single-strand/double-strand DNA junction and within single-stranded DNA, as well as a 3'-5' exonuclease activity on single-stranded DNA. We find hWRN-N$_{70-240}$ is in a trimer-hexamer equilibrium in the absence of DNA when examined by gel filtration chromatography and atomic force microscopy (AFM). Upon the addition of DNA substrate, hWRN-N$_{70-240}$ forms a hexamer and interacts with the recessed 3'-end of the DNA. Moreover, we find that the interaction of hWRN-N$_{70-240}$ with the replication protein PCNA also causes this minimal, 171-amino acid exonuclease region to form a hexamer. Thus, the active form of this minimal exonuclease fragment of human WRN appears to be a hexamer. The implications the results presented here have on our understanding of hWRN's roles in DNA replication and repair are discussed.

The pregnane X receptor (PXR) is a nuclear xenobiotic receptor which acts as a molecular sentry that detects potentially toxic foreign chemicals and activates genes to initiate their breakdown and removal. PXR fills this role by its ability to promiscuously bind to a diverse array of structurally distinct ligands which in turn enables it to activate a wide array of genes such as CYP3A, a monooxygenase involved in breaking down greater than 50 percent of all drugs and MDR1, a drug and xenobiotic efflux pump. Activation of PXR has the potentially deadly side effect of causing drug-drug interactions. Crystal structures of the human PXR ligand binding domain (LBD) have revealed a number of unique features which could facilitate PXR's promiscuous binding activity. Chief among these is a very large and highly conformable hydrophobic ligand binding cavity. The overall shapes of the ligand binding cavities of hPXR-LBD without ligand and bound to endogenous compound 17β estradiol and the LXR ligand T1317 are distinct. Several structural features of PXR contribute to the plasticity of its binding cavity including an extended beta-sheet region and two novel helices. One of the novel helices and the extended beta-sheet frames the critical second unique helix. This highly flexible helix, called the pseudo-helix due to its variance from the canonical alpha-helical conformation, adopts distinct orientations in every structure solved and plays the single most important role in adapting the shape of the binding cavity to fit different ligand orientations. The accumulating structural data provides important insights into how PXR detects xenobiotics and endobiotics and may prove useful in structure based drug design.

DEDICATION

In loving memory

To my grandmother,

Peiyu Shi

And to my father,

Maojie Xue

And also, I will dedicate this thesis to my mother, Zhifang Lei, my husband, yudong Liu and my son, Ethan Liu, daughter, Eden Liu for their great supports.

ACKNOWLEDGEMENTS

I would first like to extend my deepest appreciation to my advisors, Dorothy Erie and Matt Redinbo, who handed me these extraordinary research projects and gave me their full supports in my academic and professional pursuits during my stay in their lab. I also want to thank everyone in Erie Dorothy and Matthew Redinbo lab for their friendship, support, and tolerance, especially Glen Ratcliff, Hong Wang, Jill chrencik and Sompop Bencharit for always being there for me.

None of my research would have been possible were it not for the support of our collaborators in UNC and GlaxoSmithKline. Many of them helped us and provided financial assistance for my research from the beginning. I must also recognize the NIH (grant R01-DK62229) for funding the bulk of my research.

I also want to thank my teachers and advisors who I've learned from during my time in Chapel Hill. I also greatly appreciate the time and insight my thesis committee members have given to me over the course of the last few years. I've also benefited from classes with professors like Richard Wolfenden and Linda spremulli whose approach to science inspires my own.

Thank you all very much.

TABLE OF CONTENTS

LIST OF TABLES

xi

LIST OF FIGURES

LIST OF ABBREVIATIONS

MMR   Mismatch repair

UvrD   DNA helicase II

Exo I   Exonuclease I

ABC   ATP-binding cassette

AF   Activation function

Aldh   Aldehyde dehydrogenase

AR   Androgen receptor

ATP   Adenosine triphosphate

CAR   Constitutive androstane receptor

CBP   CREB-binding protein

CREB   cAMP response element binding protein

CV-1   African green monkey kidney fibroblasts

CYP   Cytochrome P450

EC_{50} Median effective concentration

APS   Advanced Photon Source

βME   β-mercaptoethanol

DR   Direct repeat

ER   Everted repeat

Fig   Figure

FXR   Farnesoid X receptor

h   Hour

HAT   Histone acetyltransferase

HRE   Hormone response element

LCA   lithocholic acid

Ki	Inhibition constant
MDR	Multi-drug resistance protein
MRP	MDR-associated protein
OATP2	Organic anion transporting protein 2
PPAR	Peroxisome proliferators-activated receptor
PCR	Polymerase chain reaction
PDB	Protein Data Bank
rmsd	Root-mean-square deviation
RXR	Retinoid X receptor
s	Second
SA	Simulated annealing
SJW	Saint John's wort
SR12813	3,5-di-*tert*-butyl-4-hydroxystyrene-β,β-diphosphonic acid tetraethyl ester
SRC-1	Steroid receptor coactivator-1
SSRL	Stanford Synchrotron Radiation Laboratory
TBP	TATA box binding protein
T_m	Melting temperature
UDP	Uridine diphosphate
UGT	UDP-glucuronysyltransferase
v/v	Volume by volume
w/v	Weight by volulme
XRE	Xenobiotic response element
EDTA	Ethylenediaminetetraacetic acid
Tris	Tris(hydroxymethyl)aminomethane
TEMED	N,N,N',N'-Tetramethylethylenediamine

SDS-PAGE   Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis.

WS   Werner's Syndrome

WRN   The gene defective in WS

hWRN   Human WRN protein

hWRN-N$_{70\text{-}240}$ Residues 70-240 of hWRN

hWRN-N$_{1\text{-}333}$ Residues 1-333 of hWRN

AFM   Atomic Force Microscopy

PCNA   Proliferating Cell Nuclear Antigen

yPCNA   Yeast Proliferating Cell Nuclear Antigen

NTPs   Ribonucleoside triphosphates

NHEJ   Non-homologous End Joining

BLM   Bloom Syndrome

NIEHS   National Institute of Environmental Health Sciences.

Study I: Study protein-protein interaction in Methyl-directed DNA mismatch repair in *E. coli*: Exonuclease I (Exo I) and DNA helicas II (UvrD)

(Under the direction of Dr. Dorothy Erie)

Abstract

Exonuclease I (ExoI) from *Escherichia coli* is a monomeric enzyme that processively degrades single stranded DNA in the 3′ to 5′ direction and has been implicated in DNA recombination and repair. It functions in numerous genome maintenance pathways, with particularly well defined roles in methyl-directed mismatch repair (MMR). The *Escherichia coli* MMR pathway can be reconstituted *in vitro* with the activities of eight proteins (8). MutS, MutL and MutH are involved in initiation of repair including mismatch recognition and generation of a nick at a nearby GATC sequence (53, 54, 55, 56). The hemimethylated state of GATC sequences immediately following replication serves as a signal to direct repair to the nascent strand of the DNA duplex (57, 58). DNA helicase II and one of several exonucleases (Exonucleas I, Exonuclease VII and RecJ) are required to excise the error-containing DNA strand beginning at the nicked GATC site (34, 35). Restoration of the correct DNA sequence by repair synthesis involves DNA polymerase III holoenzyme and SSB, and the final nick is sealed by DNA ligase (34). To identify interactions with ExoI involved in MMR repair system, we used the yeast two-hybrid system with ExoI as bait. By screening an *E.coli* genomic library, *E. coli* DNA helicase II (UvrD) was identified as a potential interacting protein. UvrD has been shown to be required for DNA excision repair,

methyl-directed mismatch repair and has some undefined, role in DNA replication and recombination. In this report, *in vitro* experiments confirm that UvrD and ExoI make a direct physical interaction that may be required for function of the methyl-directed mismatch repair.

Introduction

DNA damage accumulates in cells over time as a result of exposure to exogenous chemicals and physical agents (i.e., polychlorinated biphenyls, dioxin, cigarette smoke, asbestos, ultraviolet light, radon), as well as endogenous reactive metabolites including reactive oxygen and nitrogen species (ROS and NOS). Another source of DNA damage is errors that occur during normal DNA metabolism or aberrant DNA processing reactions, including DNA replication, recombination, and repair. Nucleotide mis-incorporation generates DNA base-base mismatches during DNA synthesis at variable rates, depending on many factors, including the specific DNA polymerases. In general, the replicative DNA polymerases have relatively high replication fidelity, while translesion DNA polymerases, which specifically bypass sites of DNA damage, have lower replication fidelity. DNA damage, if unrepaired, has the potential to generate mutations in somatic or germline cells, which can alter cellular phenotype and cause dysfunction and disease. To prevent such deleterious effects and safeguard the integrity of the genome, cells possess multiple mechanisms to repair DNA damage and thus prevent mutations. One such system is the critical pathway known as DNA mismatch repair (MMR). MMR corrects DNA mismatches generated during DNA replication, thereby preventing mutations from becoming permanent in dividing cells (1-

3). Because MMR reduces the number of replication-associated errors, defects in MMR increase the spontaneous mutation rate (4). Inactivation of MMR in human cells is associated with hereditary and sporadic human cancers (1, 3, 5), and the MMR system is required for cell cycle arrest and/or programmed cell death in response to certain types of DNA damage (6, 7).

Thus, MMR plays a role in the DNA damage response pathway that eliminates severely damaged cells and prevents both mutagenesis in the short term and tumorigenesis in the long term. *Escherichia coli* MMR pathway has been extensively studied and is well characterized both biochemically and genetically. Thus, *E. coli* MMR is a useful and important framework for understanding eukaryotic MMR. *E. coli* MMR requires the following protein components: MutS, MutL, MutH, DNA helicase II (MutU/UvrD), three exonucleases (ExoI, ExoVII, and RecJ), single-stranded DNA binding protein (SSB), DNA polymerase III holoenzyme, and DNA ligase (8, 9). MutS, MutL, and MutH initiate MMR and play specialized biological roles in MMR in *E. coli*. MutS recognizes base-base mismatches and small nucleotide insertion/deletion (ID) mispairs, and thus MutS has been called the "mismatch recognition" protein (3). MutS possesses intrinsic ATPase activity. High-resolution structures of MutS bound to DNA have been determined by X-ray crystallography (10, 11). These structures revealed that MutS binds to a mismatch as a homodimer. Interestingly, the mismatch-binding site is comprised of sequencewise identical but structurally and functionally different domains from the two subunits, indicating asymmetry in the protein-DNA complex. Hence, the MutS homodimer acts as a virtual heterodimer when bound to a DNA mismatch. This characteristic is mimicked by eukaryotic MutS homologs (MSH), which function as

heterodimers instead of homodimers. MMR in *E. coli* is ATP-dependent, and requires the functional MutS ATPase. MutL interacts physically with MutS, enhances mismatch recognition, and recruits and activates MutH. Defects in MutL completely inhibit MMR in *E. coli*. Despite the fact that a functional human MutL homolog, MutLα, possesses an endonuclease activity that is essential for mammalian MMR (12), no hydrolytic activity has been detected in MutL. However, MutL may play a role as a molecular matchmaker that facilitates assembly of a functional MMR complex (3, 13), because it stimulates the loading and the processivity of helicase II (or UvrD) at the MMR initiation site (14, 15). Like MutS, MutL functions as a homodimer and possesses ATPase activity (16). Mutations in the ATP-binding domain lead to a dominant negative mutator phenotype (17). MutL mutants that are defective in ATP hydrolysis but proficient in ATP binding can activate MutH but cannot stimulate MutH in response to a mismatch or MutS, suggesting that ATP hydrolysis by MutL is essential for mediating the activation of MutH by MutS (18). Recent studies show that MutL interacts physically with the clamp loader subunits of DNA polymerase III (19, 20), suggesting that MutL may promote binding of DNA polymerase III to MMR intermediates. These observations suggest that MMR is coupled with DNA replication. In *E. coli*, DNA is methylated at the N6 position of adenine in dGATC sequences. In replicating DNA, the daughter strand is transiently unmethylated, and it is the presence of hemimethylated dGATC sequences that molecularly distinguishes the newly synthesized daughter strand from the parental DNA strand. In MMR, hemimethylated dGATC sites determine the strand specificity of repair. MutH, which recognizes hemimethylated dGATC sequences, functions as a monomer and belongs to a family of type-II restriction endonucleases (21, 22). Upon its recruitment

5

and activation by MutS and MutL in the presence of ATP, MutH specifically incises the unmethylated daughter strand of hemimethylated dGATC (3, 18), and this strand-specific nick provides the initiation site for mismatch-provoked excision. The first step of the MMR pathway is binding of a MutS homodimer to the mismatch. Subsequently, a hemimethylated dGATC site 5′ or 3′ to the mismatch is located and cleaved by the concerted action of MutS, MutL, MutH, and ATP. Three models have been proposed to address how mismatch binding by MutS leads to cleavage of the hemimethylated dGATC site. The strandspecific nick generated by MutH at hemimethylated dGATC is a starting point for excision of the mispaired base. In the presence of MutL, helicase II (UvrD) loads at the nick and unwinds the duplex from the nick towards the mismatch (14), generating single-strand DNA, which is rapidly bound by single-stranded DNA-binding protein (SSB) and protected from nuclease attack (23). Depending on the position of the strand break relative to the mismatch, ExoI (3′→5′ exonuclease), or ExoVII or RecJ (5′→3′ exonuclease) excises the nicked strand from the nicked site (the dGATC site) up to and slightly past the mismatch. The resulting single-stranded gap undergoes repair DNA resynthesis and ligation by DNA polymerase III holoenzyme, SSB, and DNA ligase (3) (Figure 1). These early studies on *E. coli* MMR demonstrate three key features of this important pathway: first, repair is strand specific (i.e., restricted to the newly synthesized DNA strand); second, repair is bi-directional, proceeding 5′→3′ or 3′→5′ from the nick to the site of the mismatch; and third, MMR has broad substrate specificity including base-base mismatches and small ID mispairs. All of these properties require functional MutS, MutL, and MutH. Because the mechanism of MMR is highly conserved throughout evolution, *E. coli* MMR is an excellent model for MMR in eukaryotic cells.

6

The main objective of this study was to better understand the mechanism of Methyl-directed DNA mismatch repair in *E. coli*. To identify interactions within eight proteins involved in MMR, we used the yeast two-hybrid system with Exonuclease I as bait. By screening an *E.coli* genomic library, *E. coli* DNA helicase II (UvrD) was identified as a potential interacting protein with ExoI. UvrD has been shown to be required for DNA excision repair, methyl-directed mismatch repair and has some undefined, role in DNA replication and recombination. ExoI was originally identified as a suppressor of recombination in *recBC* mutants. However, later results suggested that ExoI and other single stranded DNA exonucleases participate in recombination pathways. In this report, *in vitro* experiments confirm that UvrD and exonuclease I make a direct physical interaction that may be required for function of the methyl-directed mismatch repair.

Materials and methods

Strains, enzymes, DNA and nucleotides

Escherichia coli JS4 was from Bio-Rad. *Escherichia coli* HB101 was from New England Biolabs (NEB). *Escherichia coli* BL21 was from Novagen. *Escherichia coli* K-12 was from NEB. HF7c and SFY526, and plasmids pGAD424 and pGBT9 were from the Matchmaker two-hybrid system (Clontech). The primer 5'-TTCGATGATGAAGATACC- 3' was used to sequence the amino-terminal end of fusions in the pGAD series of activation domain vectors. The primer 5'-AAGAGAGTAGTAAC3' was used to sequence the amino-terminal end of fusions in the pGBD series of DNA binding domain vectors. The 3' ends of these primers are located

25 and 27 bp upstream of the EcoRI site, respectively. The carboxyl-terminal end of

fusions in either pGAD, pGBD vectors were all sequenced using the same primer, 5'-

TGAAGTGAACTTGCGGGG3', which is located 22 bp from the Bgl II site.

Construction of two hybrid expression vector

To generate two-hybrid libraries using a set of five different restriction enzymes,

we needed transcription activation domain vectors with a unique ClaI site in the

polylinker. The oligonucleotide primers (1) 5'-

TGCTTTCGAAGCTCCCCHACAGGTGTCCC-

3' and (2) 5'-

TTAAGAATTCCCCGGGGGATCCATCGATGTCGACCTGCAGAGATC-3' were

used to amplify a 1666-bp fragment from the vector pGAD424. Primer 1 contains a

BstBI site in place of the ClaI restriction site in the LEU2 gene of pGAD424. Primer 2

introduces two single-bp insertions and a *Cla*I site into the pGAD424 polylinker. These

changes were incorporated into pGAD424 in two steps. First the PCR product was

digested with BstBI/AflII and the 194bp fragment was ligated into pGAD424 digested

with ClaI/AflII. Next the resulting plasmid and the PCR product were both digested by

Afl II/EcoRI, AflII/SmaI, or AflII/BamHI and each -1.5-kb PCR fragment was subcloned

into the appropriately digested vector. The ClaI site in the LEU2 gene of the resulting

plasmids was destroyed and a new ClaI site was introduced into the polylinker. We

verified that each plasmid is still able to rescue both a *leu2* mutation in yeast and the

leuB6 mutation in *E. coli*. Cloning with BamHI, SmaI, or EcoRI also incorporated zero,

one, or two single base insertions, respectively, into the polylinker. The resulting

plasmids, pGAD-Cl, pGADC2, and pGAD-C3 (Figure l), contain unique ClaI sites in their polylinker regions, each in a different reading frame. The DNA sequence across each polylinker region was verified by sequencing.

Preparation of E. coli genomic DNA

Genomic DNA was prepared from *E. coli* K12. A 2- liter culture grown in LB was harvested by centrifugation at OD_{600} 1 .6, washed once in 500 ml ddH_2O, and resuspended in 30 ml of 1 M sorbitol, 0.1 M EDTA. Zymolyase 20T (ICN Pharmaceuticals) was added to 5 mg/ml and incubated 1 hr at room temperature. The spheroplasts were collected by centrifugation, resuspended in 40 ml lysis buffer (24), and split into two 50-ml Falcon tubes. Twenty milliliters PCI (pheno1:chloroform:isoamyl alchohol, 25:24:1) and 6 ml acid-washed glass beads (0.45-0.55 mm) were added to each tube, and the mixtures were vortexed at top speed for 5 min. The supernatants were recovered and pooled, re-extracted three times with 20 ml PCI, and precipitated with 2 volumes ethanol. The pellet was resuspended in 20 ml 1X TE (10 mM Tris pH 8.0, 1 mM EDTA pH 8.0) and RNase A was added to 50 pg/ml. After 1 hr at 37 ^{0}C, the DNA was extracted three more times with 20 ml PCI to achieve a clear interface and was precipitated with 0.6 volumes of isopropanol. The pellet was resuspended in *5* ml H_2O, and 1 m1 5 M NaCl and 6 ml l3% PEG 8000 were added. The DNA was precipitated 4 hr on ice and spun 10 min at 10,000 rpm in a SS34 rotor (Sorvall). The pellet was resuspended in 3 ml 1 X TE and the final yield was determined by spectrophotometry to be 2.75 mg.

Construction of E. coli genomic two-hybrid libraries:

Vector DNA was prepared by digesting 15 pg each of pGAD-Cl, pGAD-C2, and pGAD-C3 with 40 units of *ClaI* for 6 hr at 37^0C. Digests were treated with Wizard Cleanup resin (Promega) and recovered in 0.1 ml 1 X TE. Each vector digest was treated with 2 units of calf intestinal alkaline phosphatase (CIP, Boerhinger Mannheim) for 30 min at 37^0C, 2 additional units of CIP were added, and the reactions were incubated an additional 30 min at $55 \ ^0$C. The CIP was removed by treatment with Wizard Cleanup resin. Each vector preparation was then incubated with 2 units T4 DNA ligase (Boerhinger Mannheim) at room temperature overnight, placed at 65^0C for 15 min, and run on preparative 0.7% agarose gels to separate linear monomers from circularized and multimeric forms. Each linear vector DNA was recovered into 0.1 ml 1X TE. Insert DNA was prepared by partial digestion of *E. coli* K12 genomic DNA with the enzymes *AciI, MspI, HinPlI* (New England Biolabs), *MaeII* (Boerhinger Mannheim), and *TaqI* (Promega). For each enzyme the optimal concentration for partial digestion was determined in units/pg of DNA, and then five preparative reactions were carried out with enzyme concentrations that bracketed the optimum. Each preparative digest contained 40 pg of genomic DNA in a total volume of 0.2 ml and was incubated for 30 min at the recommended temperature. The reactions were stopped by the addition of 20 µl 0.25 M EDTA.

Partial digests were analyzed by Southern blotting. The 315- bp NczI/XbaI fragment of the genomic DNA was radioactively labeled using the random priming method and used as a probe (25). For each enzyme the digest that produced the most even distribution of partial and complete digestion products was selected and 10 pg of each

digest were size fractionated on 1% agarose gels. DNA from 500 to 2000 bp for *MaeII* and *TaqI*, 500-2500 bp for *AciI*, and 500-3000 bp for *MsfiI* and *HinPlI* was recovered into 0.1 ml 1X TE. Fifteen ligation reactions were carried out (three vectors x five enzyme digests). The vector:insert ratio that resulted in optimal ligation efficiency was empirically determined. Each ligation reaction contained *5* µl of the vector preparation, 40 µl of the insert preparation, and 6 units of T4 DNA ligase in a total reaction volume of 100 µl, and was incubated at room temperature for 9 hr. Ligation reactions were ethanol precipitated and resuspended in 25 µl ddH$_2$0. Ligations were transformed into ultracompetent *Escherichia coli* strain JS4 (Gibco-BRL) by electroporation, using a 1.8 kV pulse, 1 mm gap cuvettes (BTX, Inc.), 20 µl of cells, and 1 µ1 of ligation mix in each transformation reaction. For each of the 15 vector: insert combinations, three to seven transformation reactions were necessary to generate sufficient numbers of transformants. One milliliter of SOC was added to each transformation reaction and incubated 1 hr at 37 ^{0}C. The three to seven transformation reactions from each ligation were pooled and a small aliquot was plated to determine the total number of primary transformants. The remainder of the transformation mixes from each ligation were inoculated into 3 liters of T broth containing 200 pg/ml ampicillin and incubated overnight at 37 ^{0}C resulting in a total of 15 cultures. At OD$_{600}$ 1.5 (-12 hr) cultures were chilled on ice for 30 min. For each of the three different reading frame vectors, five cultures (corresponding to the five restriction enzyme digests) were pooled in appropriate quantities to produce an equal representation of all restriction sites in the final library.

The resulting three pooled cultures, one representing each reading frame, contained- 11 liters each and were harvested by centrifugation. DNA was prepared from

each pellet by large scale alkaline lysis and precipitated with ethanol. The pellets were resuspended in 120 ml 1X TE and treated 1 hr at 37^0C with 10 mg RNase A. Each was extracted five times with 60 ml PCI and ethanol precipitated in the presence of 300 mM NaAc pH 7.0. Pellets were resuspended in 33 ml ddH20, and 7 ml 5 M NaCl and 40 ml 13% PEG 8000 were added. The DNA was precipitated overnight on ice and spun down 10 min at 10,000 rpm (Sorvall). The DNA pellets for libraries Y2HL-C1, Y2HL-C2, and Y2HL-C3 were resuspended in lx TE at a concentration of 1 mg/ml.

Cloning uvrD and Exonuclease I

Vent DNA polymerase was used to amplify the *uvrD* gene by PCR using *E. coli* K12 genomic DNA. Amplified *uvrD* was sequenced and cloned into the *Sma*I site of pGAD424 and pGBT9 to create in-frame translational fusions with the Gal4 transcriptional activation domain and DNA binding domain, respectively. These constructs were designated pGAD424-UvrD and pGBT9-UvrD. In similar fashion, the Exonuclease I gene was amplified by PCR from the genome of *E. coli* strain DH5α and gene was sequenced to make sure the gene sequence is correct before directionally cloned into the *Eco*RI and *Bam*HI sites of pGAD424 and pGBT9 to form the appropriate translational fusions. These constructs were designated pGAD424-ExoI and pGBT9-ExoI.

Detection of reporter gene expression

pGBT9-ExoI and Y2HL were cotransformed into yeast HF7c cells and plated on complete synthetic media lacking leucine, tryptophan and histidine, and supplemented

with 1 mM 3-amino-1,2,4-triazole (3-AT). 3-AT was required to suppress a low level of *HIS3* activation by the pGBT9-UvrD DNA binding domain fusion. Transformation efficiencies were monitored by plating small portions of transformations on complete synthetic media lacking only leucine and tryptophan. Transformants capable of growth in the absence of histidine were colony purified and DNA was extracted as described by the supplier (Clontech). The pGAD424-library plasmid was recovered in each case by transformation into *E.coli* HB101 and selecting for growth on minimal media lacking leucine as described by the supplier. Recovered pGAD424-library plasmids were re-transformed with pGBT9-ExoI into yeast HF7c and SFY526 cells. Confirmation of an interaction was performed by monitoring growth of HF7c on media lacking histidine and the appearance of blue color in SFY526 in the presence of X-Gal at high concentration of 3-AT (10 mM). A spectrophotometric assay for β-galactosidase activity, using the substrate o-nitrophenyl β-D-galacto-pyranoside (ONPG), allowed quantification of experimental and control interactions (26), and was performed as described (Clontech).

Purification of UvrD and ExonucleaseI

Vent DNA polymerase was used to amplify the *uvrD* gene by PCR using pGBT9-UvrD as target. Amplified *uvrD* was cloned into modified pET28 expression vector as HisTev UvrD. His-6-tagged *E. coli* UvrD helicase is expressed in *E. coli* strain BL21 (DE3). Cells were lysed in 20 mM Tris-Cl pH 7.6, 250 mM NaCl, 2 mM TCEP, 5% glycerol, 20 mM imidazole, with Roche EDTA free protease inhibitor cocktail tablets, using a microfluidizer from microfluidics coroperation. The cellular extract was clarified by centrifugation at 17,000 rpm for 60 min at 4 °C, before loading onto a 40ml Ni-NTA

column (Qiagen). After washing extensively with buffer A (20 mM Tris-Cl pH 7.6, 250 mM NaCl, 2 mM TCEP, 5% glycerol, 20 mM imidazole) and buffer B (20 mM Tris-Cl pH 7.6, 250 mM NaCl, 2 mM TCEP, 5% glycerol, 40 mM imidazole), His tagged UvrD was eluted from the column using buffer A containing 250 mM imidazole and then concentrated to 17 ml, loading onto Superdex 200 (GE) pre-equilibrated with buffer D (20 mM Tris-Cl pH 7.6, 150 mM NaCl, 2 mM TCEP, 5% glycerol), UvrD were pooled and diluted 3 fold with buffer E (20 mM Tris-Cl pH 7.6, 2 mM TCEP, 5% glycerol), and then loading onto MonoQ (GE), protein eluted at 180 mM NaCl. Pooled fractions were concentrated to 20 mg/ml and quickly frozen with nitrogen beam.

The gene encoding Exonuclease I was amplified by PCR from pGBT9-ExoI and ligated into the modified pET28 overexpression vector (Novagen) with a N-terminal His tag consisting of six histidine residues and Tev cleavage site. The protein was overexpressed in a fermenter (750 rev min_1, 7.1 l s_1 air) with 4 L LB broth and 0.27 mM ampicillin. The cells were induced at an OD of 1.2 with 1 mM IPTG for 4 h at 289 K (225 rev min_1, 7.5 l min_1 air). The cells were harvested and resuspended in 50 mM Tris–HCl pH 8.0, 5 mM imidazole, 500 mM NaCl, 0.1% Triton X-100, 10% glycerol (+TG buffer) and sonicated for 10 min. The cell lysate was centrifuged (17 000 rev min_1, Beckman JA-20 rotor) for 30 min and the supernatant was applied onto a column of Ni–NTA agarose resin (Qiagen). The column was washed with +TG buffer and -TG buffer (50 mM Tris–HCl pH 8.0, 5 mM imidazole, 500 mM NaCl) until the absorbance was zeroed. The protein was eluted with 250 mM imidazole. Pertinent fractions were pooled and dialyzed overnight against buffer C (50 mM Tris-Cl pH 8.0, 500 mM NaCl, 2 mM TCEP, 10% glycerol, 5 mM imidazole). Tev cleavage was going on overnight and

then the protein sample was loaded onto Superdex 200 (GE) pre-equilibrated with buffer D (50 mM Tris-Cl pH 8.0, 500 mM NaCl, 2 mM TCEP, 10% glycerol), *ExoI* were pooled and dialyzed into 50 mM Tris–HCl pH 8.0, 150 mM NaCl, 1 mM EDTA and 1 mM dithiothreitol (DTT). Finally, ExoI were concentrated to 1 mg/ml and quickly frozen with nitrogen beam.

Pull-down assay

HisTev-UvrD are immobilized on Ni-NTA agarose beads by incubating 1 mg of purified HisTev-UvrD with 1 ml Ni beads for 1 hour at 4°C, which pre-washed with buffer A: 20 mM Tris, pH 7.4, 250 mM NaCl and 20 mM immidazole 4 times and equilibrated in the same buffer A. Wash the beads four times with 20 volumes of buffer A to remove unbound material, resuspend in 1ml of buffer A, and store at 4°C. Incubate 1 mg of ExoI with the immobilized HisTev-UvrD at 4°C rotating for 1 hr. Wash the beads four times with 20 volumes of buffer A. The bound proteins are eluted with 250 mM immidazole in the same buffer and then boiled in sample buffer and visualized using Coomassie blue staining. BSA was used as a control.

Results

Construction of improved two-hybrid vectors

As a first step in creating improved genomic two-hybrid libraries, we needed a set of vectors with a unique *ClaI* cloning site that could be used to generate fusions to a transcription activation domain. The *GAL4* activation domain vector pGAD424 (27) was chosen because of its small size, convenient polylinker sites, and efficient rescue *E. coli 1euB* mutations. Site-directed PCR mutagenesis was used to make several changes in pGAD424. First, the endogenous *ClaI* site present in the *LEU2* marker gene was destroyed. This change did not alter the amino acid sequence of *LEU2,* and the plasmid retained the ability to rescue both *leu2* mutations in yeast and *leuB6* mutations in *E. coli*. Second, a new *ClaI* site was introduced into the center of the polylinker and single base pairs were inserted between the *EcoRI* and *SmaI* sites and between the *SmaI* and *BamHI* sites of the polylinker. By incorporating either zero, one, or two of the single base pair insertions into the final products, we generated a set of three new activation domain vectors, pGAD-C1, *pGAD-C2,* and pGAD-C3. Each contains a unique *ClaI* cloning site in the polylinker, but differs in the translational reading frame of the polylinker sites (Figure 2).

Construction of E. colit genomic two-hybrid libraries

To generate a highly complex set of libraries, we sought to increase the number of restriction enzymes used for the partial digestion of genomic DNA. There are five different commercially available enzymes with 4 bp recognition sequences that produce a 5' overhang with the sequence 5'-CG3': *Aci*I **,** *HinPl*I, *Mae*II, *Msp*I, and *Taq*1

16

("CGenzymes"). Each is compatible for ligation with the *ClaI* site introduced into the polylinkers of the vectors described above. Library inserts were prepared using genomic DNA from *E. coli* K12. The genomic DNA was subjected to partial digestion by each of the CG enzymes. Because the inclusion of as many restriction sites as possible is critical to the quality of the library, partial digests were examined by Southern blotting (Figure 3). This examination was extremely important, as we found that it was impossible to predict the quality of the partial digest by ethidium bromide staining. Southern blotting results demonstrated that digests judged to be partial by ethidium bromide staining actually contained fragment mixes that ranged from nearly complete to nearly uncut. The 315- bp NczI/XbaI fragment of *E. coli* genomic DNA was radioactively labeled using the random priming method and used as a probe (25). For each enzyme the digest that produced the most even distribution of partial and complete digestion products was selected and 10 pg of each digest were size fractionated on 1% agarose gels. DNA from 500 to 2000 bp for *MaeII* and *Taq*I, 500-2500 bp for *AciI*, and 500-3000 bp for *MsfiI* and *HinPlI* was recovered into 0.1 ml 1X TE. The optimized concentration of each enzyme used for partial digestion is 1.8units/ug (Figure 3). After using Southern blots to select the best partial digest for each enzyme, we size-fractionated the selected digests by gel electrophoresis. The selection of small insert sizes was desirable to minimize the production of false positives. However, if the selected insert size is too small, many complete digestion products will not be represented in the library. To avoid excluding completely digested fragments from the library, we examined complete digestions of genomic DNA by ethidium bromide staining. The largest fragment size visible was chosen to be the maximum insert size selected for each enzyme. **As** a result, restriction

fragments from 500 bp to a maximum of 2-3 kb were included in the libraries, with the maximum size depending on which enzyme was used to generate a partial digest. Vector DNA was prepared by digesting pGAD-C1, pGAD-C2, and pGAD-C3 with *ClaI* and treating the cleaved ends with calf intestinal phosphatase to remove the 5' phosphate groups. Dephosphorylated vectors were then religated, and those that remained as linear monomers were purified by electrophoresis. This step eliminated about 90% of the background caused by vector self-ligation. After determining empirically that a 1:8 vector:insert ratio provided optimal ligation efficiency, 15 ligation reactions were performed and each was transformed into *E. coli JS4.* Fifteen separate reactions were required to ligate each of the three different reading frame vectors to each of the five different partial digestions. Production of comprehensive libraries required that a large number of independent clones be obtained from each of the 15 ligation reactions. Because of the practical limitations of recovering this number of colonies from plates, each reaction was inoculated into liquid T broth. To minimize the problem of clone loss due to competition, cultures were harvested while still in the logarithmic phase of growth. The five cultures that represent each reading frame were then mixed together in quantities that would produce equal representation of each individual restriction site in the genome. The mixing produced three large cultures, each representing a different reading frame and containing a comprehensive library of genomic DNA inserts. The detailed procedures for *E. coli* genomic library construction was summarized in Figure 4. DNA was prepared from each of the three libraries, which we call Y2HLC1, Y2HL-C2, and Y2HL-C3. The quality of the two-hybrid libraries has been examined by a variety of methods. First, a large number of independent clones were obtained from each of the 15 transformation

18

reactions. The quality of a library is also dependent on the percentage of library clones that contain an insert. DN A was prepared from 65-75 random clones from each of the three libraries and inserts were analyzed by digestion with the polylinker enzymes *SmaI* and *PstI* followed by gel electrophoresis. A total of 300 clones were analyzed; we found 14 without inserts and an additional 13 that contained small inserts of <400 bp. Thus 96% of the library plasmids contained inserts, and only 5% of those inserts were not within the size range we had selected.

The two-hybrid screen

To identify *E.coli* proteins that potentially interact with ExoI, we constructed an *E.coli* genomic library in the two-hybrid vector pGAD424 as described in Materials and methods. ExoI, expressed as a translational fusion with the Gal4 DNA binding domain in pGBT9, was used as bait. Potential interactions in the yeast two-hybrid reporter strain HF7c were initially selected on complete synthetic media lacking histidine. HF7c contains a *HIS3* gene under control of the *GAL1* promoter which is only expressed when a functional Gal4 protein is reconstituted by an interaction between the activation domain and DNA binding domain fusion proteins. Potential interactors with *Exo*I in pGAD424-library were recovered as described in Materials and methods, and the interaction was confirmed by re-transformation with pGBT9-*Exo*I into HF7c and SFY526. SFY526 contains a *lacZ* reporter gene and its expression can be monitored with the color-producing substrates X-gal and ONPG. One of library clones from Y2HL-C1 that maintained an interaction in HF7c and SFY526 which was dependent on the presence of *Exo*I was sequenced and subjected to a BLAST search of the *E.coli* genome database. It

was found to be 100% identical to a portion of the mismatch repair gene *UvrD* (Figure 5).

Partial sequence of UvrD gene is from 705 to 2162 bp. We totally found 10 clones from

three libraries mainted interactions in HF7c and SFY526 which were dependent on the

prescence of ExoI at 10 mM 3 AT. We only sequence one which still showed interaction

at higher concentration of 3AT (30 mM). Later on we will sequence the other nine

clones.

Full length UvrD and ExoI proteins interact in the two-hybrid system

We were interested in examining the interaction between full length *ExoI* and

UvrD. To accomplish this, *UvrD* was cloned from the *E.coli* genome DNA into

pGAD424 and pGBT9 as described in Materials and methods. The pGAD424-ExoI and

pGBT9-UvrD clones were used to confirm the interaction with ExoI in HF7c and

SFY526. Figure 6 shows that growth of HF7c on complete synthetic media lacking

histidine was dependent on the presence of both *ExoI* and *uvrD* in the two-hybrid vectors.

The absence of *ExoI*, *uvrD* or both resulted in basal levels of reporter gene expression

and, therefore, no growth on media lacking histidine. Identical results were obtained in

the *lacZ* reporter strain SFY526 with regard to b galactosidase expression. In addition,

the interaction was observed regardless of whether UvrD existed as a fusion with the

Gal4 activation domain or the Gal4 DNA binding domain (Figure 6). We next examined

the interaction between full length RecJ (One of the exonclease involved in MMR) and

UvrD, RecJ was cloned from the *E. coli* genome DNA into pGAD424 and pGBT9, the

growth of HF7c on complete synthetic media lacking histidine was dependent on the

presence of both *RecJ* and *uvrD* in the two-hybrid vectors (Data not shown). Result

showed they had interactions like ExoI with UvrD. These data showed UvrD truly interacted with exonuclease in E. coli DNA mismatch repair.

Purified ExoI is specifically retained on a UvrD affinity column

ExoI and UvrD were over-expressed and purified as described in Materials and methods. Affinity chromatography was used to demonstrate a physical interaction between the two purified proteins. UvrD affinity column was generated using Ni-NTA resin. 1 mg of UvrD was coupled to a resin volume of 5 ml. 1 mg of purified ExoI or control protein BSA was loaded onto the column and washed and eluted with 250 mM immidazole as described in Materials and methods. As shown in Figure 7A, a large fraction of the applied ExoI bound the UvrD affinity column and was eluted from the column with 250 mM immidazole washes. There were some in the flowthrough and washes. A control protein BSA failed to bind the UvrD column and was found exclusively in the flow-through and wash fractions (Figure 7B). In addition, a control affinity column was constructed by coupling 1 mg of his tagged 11 β HSD1 to 5 ml resin. 1 mg of ExoI was loaded onto the control column, which was treated in the same manner as the UvrD column. Nearly all of the loaded ExoI was present in the flow-through and 20 mM imidazole wash fractions (data not shown), indicating that ExoI was not retained on the UvrD affinity column by non-specific interactions with the resin or protein. These results support the conclusion that a specific physical interaction exists between UvrD and ExoI.

Discussion

E. coli exonuclease I (ExoI) functions in numerous genome maintenance pathways, with particularly well defined roles in methyl-directed mismatch repair (28). In mismatch repair, incorrect DNA base pair formation triggers cleavage of the non-methylated (newly synthesized) DNA strand in hemimethylated DNA, loading of UvrD to unwind from the nick in the direction of the error, and clearance of the nonmethylated ssDNA through the mispaired element by ExoI or another functionally redundant nuclease ExoVII or RecJ (29, 30). We used the yeast two-hybrid system to search for interacting protein partners that might direct the involvement of ExoI in the methyl-directed mismatch repair and nucleotide excision repair pathways. The two-hybrid screen revealed an interaction between ExoI and helicase II (UvrD), an essential component for methyl-directed mismatch repair (31, 32, 8), future studies aimed at generating point mutants that exhibit defective interactions with UvrD should aid identification of the interaction domain in ExoI. Recently it was shown that ExoI acts on single strand DNA binding protein (SSB)/ssDNA substrates (33) in *In vitro* reconstitution of bacterial mismatch repair reactions. The crystal structure of *E. coli* ExoI and bound form to a peptide comprising the SSB-Ct element complex showed that ExoI in both crystal forms comprises exonuclease (residues 1–201) domain, SH3-like domain (residues 202–352), and helical (residues 360–476) domains (Figure 8). Bounded ExoI crystals revealed features corresponding to the C termini from two SSB-Ct peptides associated with the ExoI surface. Both peptides bind at sites that are 20 Å from the active site. The first peptide binds between the exonuclease and SH3-like domains of ExoI and the second peptide binds entirely the ExoI SH3-like domain. Examination of the SSB-Ct binding site

22

in ExoI has identified surface features that could be useful for similar interaction sites on ExoI's other binding partners, like UvrD. These features include a hydrophobic pocket with a basic residue positioned at the lip of the pocket and flanking basic residues that bind the conserved acidic residues of the SSB-Ct. Interactions with SSB protein are likely to be important for ExoI function. It had also showed RecJ, another functionally redundant nuclease in DNA mismatch repair system, had direct interaction with SSB. In our *in vivo* study we also showed RecJ interact with UvrD. During processive ssDNA degradation, ExoI or RecJ must displace the tightly bound SSB to move ssDNA into their cleft, a process that could be facilitated by specific interactions between the two proteins. Our *in vitro* experiments showed that ExoI made a direct physical interaction with UvrD, however, the mechanism underlying ExoI-UvrD interaction has remained poorly defined. We are currently exploring the functional importance of the ExoI–UvrD interaction for methyl-directed mismatch repair. In Junghoon in Dr. Dorothy Erie lab recetly showed the interaction between UvrD and SSB in DNA mismatch repair. SSB may play a critical role in coordinating the excision repair process in living cells. A plausible mechanism is that SSB interacts with UvrD and that interaction strongly stimulates helicase activity of UvrD at nicked DNA fragments. The interaction of SSB and UvrD induce the conformational change of UvrD or SSB and may then load ExoI or RecJ onto ssDNA sequences, the sequential timing for ExoI interact with UvrD and SSB is unknown. ExoI or RecJ interact with SSB tightly but with UvrD weakly. During degradation of ssDNA, ExoI or RecJ displace the tightly bound SSB to push ssDNA into their cleft. Because UvrD unwinds exclusively in a 3' to 5' direction with respect to the bound DNA strand (37, 38), bidirectional unwinding from a nick necessitates the ability of UvrD to bind to

both strands. The interaction of ExoI or RecJ with UvrD probably determines loading ExoI or RecJ to SSB/ssDNA. If UvrD unwinding occurs on the 3' side of the mismatch (Figure 9A), this excision step is carried out by exonuclease I (which degrades a single strand only in the 3' to 5' direction). If the unwinding occurs on the 5' side of the mismatch, exonuclease VII or RecJ is used to degrade the single stranded DNA in the 5' to 3' direction.

Figure 9 showed a simple model that may account for SSB help recruit ExoI and UvrD to SSB/ssDNA substrates, and also, make a direct interaction with UvrD. The complexes UvrD/ExoI/SSB loads at the nick and unwinds and excises the ss DNA from the nick towards the mismatch very efficiently. If the cleavage occurs on the 3' side of the mismatch (Figure 9A), this step is carried out by exonuclease I (which degrades a single strand only in the 3' to 5' direction). If the cleavage occurs on the 5' side of the mismatch, exonuclease VII or RecJ is used to degrade the single stranded DNA in the 5' to 3' direction (Figure 9B). The gap is filled by DNA polymerase III and DNA ligase.The mismatch repair system displays bidirectional capability (34, 35). Correction of a mismatched base can be directed by a hemimethylated GATC sequence on either the 5' or 3' side of the error. UvrD participates in the excision of the DNA biosynthetic error by unwinding from the nicked GATC site to a location past the error. Interestingly, UvrD preferentially unwinds from the nick toward the mismatch, regardless of whether the nicked GATC sequence is on the 5' or 3' side (35). This occurs despite the fact that helicase II alone catalyzes an unwinding reaction in both directions from a nicked DNA molecule *in vitro* (36). Because UvrD unwinds exclusively in a 3' to 5' direction with respect to the bound DNA strand (37, 38), bidirectional unwinding from a nick

necessitates the ability of UvrD to bind to both strands. It was shown that MutS translocates bidirectionally from a mismatch site in an ATP hydrolysisdependent fashion, creating a looped DNA structure that usually contains the mismatch (39). MutL stimulates the rate of this process and both proteins are found as a complex at the base of the loop structure. The arrival of the MutS–MutL complex at a GATC site activates the MutH-associated endonuclease and allows initiation of unwinding by UvrD from the resulting nick.

The identification of an interaction between ExoI and UvrD may have important consequences for understanding mismatch repair in eukaryotic systems. Defects in eukaryotic mismatch repair result in destabilization of short repetitive sequences and have been linked to various cancers, most notably hereditary colon cancer (40, 41, 42, 43). The error recognition steps of eukaryotic mismatch repair are somewhat more complex than in prokaryotes, but nevertheless occur by a similar mechanism involving homologs of the bacterial MutS and MutL proteins (44, 45). In fact, a number of eukaryotic mismatch repair genes were identified based on their homology to bacterial (Table1). Despite the striking similarities between the initiation steps of prokaryotic and eukaryotic mismatch repair, very little is known about the subsequent steps in the eukaryotic pathway. For instance, an *in vivo* strand discrimination signal has not been discovered, and the mechanisms and proteins involved in excision and repair synthesis remain enigmatic. Of particular significance is the fact that no helicase has yet been identified as a participant in a eukaryotic mismatch repair system. It is possible that eukaryotic mismatch repair does not require a helicase; one or more exonucleases may be solely responsible for the excision step. Four nucleases in MMR exist in yeast systems,

exonuclease 1 (Exo1), Rad27, and the 3' exonucleases intrinsic to Pol ε and Pol δ (46, 47). Currently, Exo1 is most clearly implicated that it interacted with MutS and MutL proteins (48, 49, 50, 51, 52). The discovery of a physical interaction between UvrD and ExoI in *E.coli* may lead to advances in our understanding of mismatch repair in eukaryotic systems.

In summary, we have discoved the direct physical interactions of ExoI and UvrD in DNA mismatch repair and also, defined a model for understanding the recruiting roles of SSB in bacterial genome maintenance. We also discuss the possible surface on ExoI that possibly interact with UvrD. Knowledge of the ExoI/SSB/UvrD should help direct studies into how ExoI interacts with its binding partners and whether (and how) association with UvrD is regulated in cells.

Table I: MMR components and their functions

E. coli	Human	Function
(MutS)₂	hMutSα (MSH2-MSH6)ₐ hMutSβ (MSH2-MSH3)	DNA mismatch/damage recognition
(MutL)₂	hMutLα (MLH1-PMS2)ₐ hMutLβ (MLH1-PMS1) hMutLγ (MLH1-MLH3)	Molecular matchmaker; endonuclease, termination of mismatch-provoked excision
MutH	?ᵦ	Strand discrimination
UvrD	?ᵦ	DNA helicase
ExoI, ExoVII, ExoX, RecJ	ExoI	DNA excision; mismatch excision
Pol III holoenzyme	Pol δ	DNA re-synthesis
	PCNA	Initiation of MMR, DNA re-synthesis
SSB	RPA	ssDNA binding/protection; stimulating mismatch excision; termination of DNA excision; promoting DNA resynthesis
	HMGB1	Mismatch-provoked excision
	RFC	PCNA loading; 3' nick-directed repair; activation of MutLα endonuclease
DNA Ligase	DNA ligase I	Nick ligation

a: Major components in cells.

b: Not yet identified.

Figure legends

FIGURE 1: Mismatch repair. The repairing process begins with the protein MutS

which binds to mismatched base pairs. Then, MutL is recruited to the complex

and activates MutH which binds to GATC sequences. Activation of MutH cleaves the

unmethylated strand at the GATC site. Subsequently, the segment from the cleavage

site to the mismatch is removed by exonuclease (with assistance from helicase II and

SSB proteins). If the cleavage occurs on the 3' side of the mismatch, this step is carried

out by exonuclease I (which degrades a single strand only in the 3' to 5' direction). If

the cleavage occurs on the 5' side of the mismatch, exonuclease VII or RecJ is used to

degrade the single stranded DNA. The gap is filled by DNA polymerase III and DNA

ligase. The distance between the GATC site and the mismatch could be as long as 1,000

base pairs. Therefore, mismatch repair is very expensive and inefficient. Mismatch

repair in eukaryotes may be similar to that in *E. coli*. Homologs of MutS and MutL

have been identified in yeast, mammals, and other eukaryotes. MSH1 to MSH5 are

homologous to MutS; MLH1, PMS1 and PMS2 are homologous to MutL. Mutations of

MSH2, PMS1 and PMS2 are related to colon cancer. In eukaryotes, the mechanism to

distinguish the template strand from the new strand is still unclear.

FIGURE 2: New expression vectors for use in two-hybrid analysis. (a) Structure of the

pGAD-C(x) vectors. Stippled regions indicate the *ADHl* promoter (P) and transcription

termination (T) elements. *GAL4* AD (activation domain) encodes amino acids 768-881.

(b) Sequences of the polylinker regions. Restriction sites are underlined; stop codons are

boxed. *, single base pair insertions introduced during construction. *EcoR* I is present in one reading frame, *Sma* I in two, and all other sites in all three reading frames.

FIGURE 3: Analysis of a partial digestion of genomic DNA using Southern blotting. The panel on the left shows a set of *HinPl*I partial digestion reactions separated by electrophoresis and stained with ethidium bromide. Enzyme concentrations are shown above each lane in units per microgram of DNA. Size markers are labeled on the left. The right panel shows a Southern blot of the same gel probed with a 315-bp fragment of the *AnE2* gene. The fragment size that results from complete digestion is indicated by an arrow.

FIGURE 4: The procedures of *E. coli* K12 genomic library construction

FIGURE 5: Interaction of Exonuclease I and UvrD. Yeast strain HF7c was simultaneously transformed with plasmids pGBT9 (ExoI) and pGAD424-C1 (library), and cells were grown on synthetic medium lacking leucine and tryptophan (panel A) or lacking leucine, tryptophan and histidine (panel B). The ability of cells to grow on medium lacking histidine is dependent on the expression of the his3 gene under the control of a GAL1-responsive promoter.

FIGURE 6: ExoI and UvrD interact in the yeast two-hybrid system. Yeast HF7c cells containing various constructs of pGBT9 and pGAD424 were grown at 30°C on complete synthetic media lacking tryptophan, leucine and histidine, and supplemented with 1 mM

3-AT. Each quadrant on both plates contains cells streaked from a single transformant that was colony-purified. Labels for both panels represent the fusion proteins present in HF7c in the order: DNA binding domain/transcriptional activation domain. Minus (-) represents the absence of ExoI or UvrD from the fusion construct.

FIGURE 7: ExoI was specifically retained on an UvrD affinity column. 1 mg ExoI (A) or BSA (B) were loaded onto a 5 ml Ni-NTA beads containing 1 mg purified Histag UvrD as described in Materials and methods. In both (A) and (B): lanes 1, 2 and 3: flow-through (FT) and washes; lanes 4 and 5, 250 mM imidazole elution. Gels were stained with Coomassie Brilliant Blue.

FIGURE 8: The structure of ExoI. The N-terminal domain (colored yellow) corresponds to the 'exonuclease' or 'proofreading' domain of the DNA polymerases and includes the catalytically essential residues. The second domain (colored green) has a structure akin to an elaborated SH3 domain. The residues toward the C-terminus (colored blue) form a helical region that extends the exonuclease domain.

FIGURE 9: Showed a simple model that may account for SSB help recruit ExoI and UvrD to SSB/ssDNA substrates. If the cleavage occurs on the 3' side of the mismatch (Figure 9A), this step is carried out by complex ExoI/SSB/UvrD (which degrades a single strand only in the 3' to 5' direction). If the cleavage occurs on the 5' side of the mismatch, exonuclease VII or RecJ/SSB/UvrD is used to degrade the single stranded DNA in the 5' to 3' direction (Figure 9B).

Figure 1

Figure 2

Figure 3

We choose this condition

Figure 4

Prepare *E. coli* K12 genomic DNA

↓

Partial digestion of genomic DNA with "GC" enzyme *Aci*I, *HinP*1I, *Mae*II, *Msp*I and *Taq*I. Optimize the concentration of each enzyme by sothern blot

↓

Size fractionated the selected digests by gel electrophoresis

↓

Vectors pGAD-C1, C2, C3 were digested with *Cla*I, treating the cleaved ends with CIP

↓

Linear vectors were purified by electrophoresis

↓

15 independent ligations (Each of the three different frame vectors to each of the five different partial digestions)

↓

Each ligation mix was transformed to *E. coli JS4*

↓

Each reaction was inoculated to LB media, cultures were harvested in mid-log phase

↓

Mix the five cultures that represent each reading frame, produce three large cultures

↓

DNA was prepared from each of the three libraries

Figure 5

Figure 6

Figure 7

Figure 8

272 442 359 354 C 106 417 317 200 47 N

Figure 9

Reference

1. Kolodner RD, Marsischky GT. (1999) *Curr Opin Genet Dev* 9, 89-96.

2. Kunkel TA, Erie DA. (2005) *Annu Rev Biochem* 74, 681-710.

3. Modrich P, Lahue R. (1996) *Annu Rev Biochem* 65, 101-133.

4. Tiraby JG, Fox MS. (1973) *Proc Natl Acad Sci USA* 70, 3541-3545.

5. Lynch HT, de la Chapelle A. (1999) *J Med Genet* 36, 801-818.

6. Li GM. (1999) *Oncol Res* 11: 393-400.

7. Stojic L, Brun R, Jiricny J. (2004) *(Amst)* 3, 1091-1101.

8. Lahue RS, Au KG, Modrich P. (1989) *Science* 245, 160-164.

9. Burdett V, Baitinger C, Viswanathan M, Lovett ST, Modrich P. (2001) *Proc Natl Acad Sci USA* 98, 6765-6770.

10. Obmolova G, Ban C, Hsieh P, Yang W. (2000) *Nature* 407, 703-710.

11. Lamers MH, Perrakis A, Enzlin JH, Winterwerp HH, de Wind N, Sixma TK. (2000) *Nature* 407, 711-717.

12. Kadyrov FA, Dzantiev L, Constantin N, Modrich P. (2006) *Cell* 126, 297-308.

13. Sancar A, Hearst JE. (1993) *Science* 259, 1415-1420.

14. Dao V, Modrich P. (1998) *J Biol Chem* 273, 9202-9207.

15. Guarne A, Ramon-Maiques S, Wolff EM. (2004) *Embo J* 23, 4134-4145.

16. Ban C, Yang W. (1998) *Cell* 95, 541-552.

17. Aronshtam A, Marinus MG. (1996) *Nucleic Acids Res* 24, 2498-2504.

18. Junop MS, Obmolova G, Rausch K, Hsieh P, Yang W. 2001;*Mol Cell* 7, 1-12.

19. Li F, Liu Q, Chen YY *(2007) Mutat Res* doi:10.1016/ j.mrfmmm.2007.07.008.

20. Lopez de Saro FJ, Marinus MG, Modrich P, O'Donnell M. (2006) *J Biol Chem* 281, 14340-14349.

21. Ban C, Yang W. (1998) *Embo J* 17, 1526-1534.

22. Lee JY, Chang J, Joseph N, Ghirlando R, Rao DN, Yang W. (2005) *Mol Cell* 20, 155-166.

23. Ramilo C, Gu L, Guo S (2002) *Mol Cell Biol* 22, 2037-2046.

24. Hoffmanc., S., and F. (2002) *Gene* 57, 267-272.

25. Ausebeif., ., R. Brentr, . Kingstond, . Moorej, . G. Seidman *et al.,* (1989) *Current Protocols in Molecular Biology.* John Wiley and Sons,New York.

26. MILLERJ, . H., (1972) *Experiments in Molecular Genetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

27. Barteip, L., C.-T. Chienr, . Sternglanzd S. Fieldsi, (1993) *Cellular Interactions in Development: A Practical Approach* 153-179.

28. Iyer RR, Pluciennik A, Burdett V, Modrich PL (2006) *Chem Rev* 106, 302–323.

29. Viswanathan M, Burdett V, Baitinger C, Modrich P, Lovett ST (2001) *J Biol Chem* 276, 31053–31058.

30. Burdett V, Baitinger C, Viswanathan M, Lovett ST, Modrich P (2001) *Proc Natl Acad Sci USA* 98, 6765–6770.

31. Caron,P.R., Kushner,S.R. and Grossman,L. (1985) *Proc. Natl Acad. Sci. USA* 82, 4925–4929.

32 Husain,I., Van-Houten,B., Thomas,D.C., Abdel-Monem,M. and Sancar, A. (1985) *Proc. Natl Acad. Sci. USA* 82, 6774–6778.

33 Lu, D., Keck, J. L. (2008) *Proc. Natl Acad. Sci. USA* 105, 9169-9174.

34. Cooper,D.L., Lahue,R.S. and Modrich,P. (1993) *J. Biol. Chem.* 268, 11823–11829.

35. Grilley,M., Griffith,J. and Modrich,P. (1993) *J. Biol. Chem.* 268, 11830–11837.

36. Runyon,G.T., Bear,D.G. and Lohman,T.M. (1990) *Proc. Natl Acad. Sci. USA* 87, 6383–6387.

37. Matson,S.W. (1986) *J. Biol. Chem.* 261, 10169–10175.

38. Georgi-Geisberger,P. and Hoffmann-Berling,H. (1990) *Eur. J. Biochem.* 192, 689–693.

39. Allen,D.J., Makhov,A., Grilley,M., Taylor,J., Thresher,R., Modrich,P. and Griffith, J.D. (1997) *EMBO J.* 16, 4467–4476.

40. Karran,P. and Bignami,M. (1994) *BioEssays* 16, 833–839.

41. Eshleman,J.R. and Markowitz,S.D. (1995) *Curr. Opin. Oncology* 7, 83–89.

42. Jiricny,J. (1996) Mismatch repair and cancer. *Cancer Surv.* 28, 47–68.

43. Umar,A. and Kunkel,T.A. (1996) *Eur. J. Biochem.* 238, 297–307.

44. Kolodner,R. (1996) *Genes Dev.* 10, 1433–1442.

45. Modrich,P. and Lahue,R. (1996) *Annu. Rev. Biochem.* 65, 101–133.

46. Harfe BD, Jinks-Robertson S. (2000) *Annu. Rev. Genet.* 34, 359–99.

47. Schofield MJ, Hsieh P. (2003) *Annu. Rev. Microbiol.* 57, 579–608.

48. Schmutte C, Sadoff MM, Shim KS, Acharya S, Fishel R. (2001) *J. Biol. Chem.* 276:33011–18.

49. Tran PT, Simon JA, Liskay RM. (2001) *Proc. Natl. Acad. Sci. USA* 98, 9760– 65.

50. Tishkoff DX, Boerger AL, Bertrand P, Filosi N, Gaida GM, et al. (1997) *Proc. Natl. Acad. Sci. USA* 94, 7487–92.

51. Schmutte C, Marinescu RC, Sadoff MM, Guerrette S, Overhauser J, Fishel R. (1998) *Cancer Res.* 58, 4537–42.

52. Rasmussen LJ, RasmussenM,Lee B, Rasmussen AK, Wilson DM 3rd. (2000) *Mutat. Res.* 460, 41–52.

53, Su,S.-S. and Modrich,P. (1986) *Proc. Natl Acad. Sci. USA* 83, 5057–5061.

54. Su,S.-S., Lahue,R.S., Au,K.G. and Modrich,P. (1988) *J. Biol. Chem.* 263, 6829– 6835.

55. Grilley,M., Welsh,K.M., Su,S.-S. and Modrich,P. (1989) *J. Biol. Chem.* 264, 1000– 1004.

56. Au,K.G.,Welsh,K.M. and Modrich,P. (1992) *J. Biol. Chem.* 267, 12142–12148.

57. Lu,A.-L., Clark,S. and Modrich,P. (1983) *Proc. Natl Acad. Sci. USA* 80, 4639–4643.

58. Pukkila,P.J., Peterson,J., Herman,G., Modrich,P. and Meselson,M. (1983) *Genetics* 104, 571–582.

STUDY II: A Minimal Exonuclease Domain of WRN Forms a Hexamer on DNA and Possesses Both 3'-5' Exonuclease and 5'-Protruding Strand Endonuclease Activities

(Under the direction of Professor Erie Dorothy and Matt Redinbo)

Abstract

Werner Syndrome is a rare autosomal recessive disease characterized by a premature aging phenotype, genomic instability and a dramatically increased incidence of cancer and heart disease. Mutations in a single gene encoding a 1,432 amino-acid helicase/exonuclease (hWRN) have been shown to be responsible for the development of this disease. We have cloned, over-expressed and purified a minimal, 171-amino acid fragment of hWRN that functions as an exonuclease. This fragment, encompassing residues 70-240 of hWRN (hWRN-N$_{70-240}$), exhibits the same level of 3'-5' exonuclease activity as the previously described exonuclease fragment encompassing residues 1-333 of the full-length protein. The fragment also contains a 5'-protruding DNA strand endonuclease activity at a single-strand/double-strand DNA junction and within single-stranded DNA, as well as a 3'-5' exonuclease activity on single-stranded DNA. We find hWRN-N$_{70-240}$ is in a trimer-hexamer equilibrium in the absence of DNA when examined by gel filtration chromatography and atomic force microscopy (AFM). Upon the addition of DNA substrate, hWRN-N$_{70-240}$ forms a hexamer and interacts with the recessed 3'-end of the DNA. Moreover, we find that the interaction of hWRN-N$_{70-240}$ with the replication protein PCNA also causes this minimal, 171-amino acid exonuclease region to form a hexamer. Thus, the active form of this minimal exonuclease fragment of human WRN

appears to be a hexamer. The implications the results presented here have on our understanding of hWRN's roles in DNA replication and repair are discussed.

Introduction

Werner Syndrome (WS) is an autosomal recessive disease characterized by the early onset of an aged appearance along with the common disorders associated with advancing age (1, 2). These disorders include atherosclerosis, bilateral cataracts, diabetes mellitus, and osteoporosis, as well as an unusually high incidence of tumors of non-epithelial cell origin. WS cells are characterized by chromosomal translocations, defective maintenance of telomeres, elevated rates of homologous recombination, large DNA deletions, and a prolonged S-phase of DNA synthesis (3, 4, 5, 6, 7, 8, 9, 10). *WRN*, the gene defective in WS, consists of 35 exons that encode a 1,432 amino acid protein (hWRN). This protein functions both as a 3' to 5' DNA helicase and as a 3'-5' exonuclease (6, 11, 12, 13). In addition, one group has reported that hWRN is also a 5' to 3' exonuclease (14). The sequence of the helicase domain is homologous to members of the RecQ family of DNA helicases (15). This family includes *Escherichia coli* RecQ, *Saccharomyces cerevisiae* Sgs-1p, *Schizosaccharomyces pombe* Rqh-1p, human RecQL, and the protein associated with Bloom Syndrome, BLM (16, 17, 18, 19, 20). Of the RecQ helicases identified so far, only the WRN protein functions both as a helicase and as an exonuclease. How these two functions are coupled during DNA metabolic events in vivo remains unclear.

The conserved exonuclease motif of WRN is located in the N-terminus, while the RecQ helicase motif is more centrally located with respect to the N- and C-termini. It has been shown that hWRN can efficiently degrade 3' recessed DNA strands of double-stranded or DNA-RNA heteroduplexes (12, 13). hWRN has little activity on blunt-ended DNA, on DNA with a 3' protruding strand, or on single-stranded DNA (5). In the

46

presence of the Ku 70/80 heterodimer, however, the exonuclease activity of hWRN is altered. For example, hWRN is stimulated to degrade blunt-ended and single-stranded DNA substrates in a 3'-5' direction in the presence of Ku (21, 22, 49, 50).   The interaction of hWRN with the Ku heterodimer is mediated by residues 1-50 of hWRN (50).

Information obtained from protein sequence database searches revealed that the exonuclease domain of hWRN is contained within the first 333 amino acids at the N-terminus (5). This region exhibits significant similarity to the 3' to 5' proofreading domain of *E. coli* polymerase I, RNAseD and the nuclease domain of the human polymyositis/scleroderma nuclear autoantigen (23). Within this N-terminal exonuclease region, five amino acids (Asp82, Glu84, Asp143, Tyr212 and Asp216) are proposed to be critical for exonuclease activity (5).

The biological function of the WRN helicase/exonuclease is unclear. It is possible that WRN participates in one or more aspects of DNA replication because abnormalities in both S-phase initiation and transition have been reported in cells harboring WRN mutations (5, 24). Its efficient removal of terminally mismatched nucleotides raises the possibility that WRN may provide a 3' to'5' proofreading activity for DNA polymerases that lack such an activity (25). The findings that WRN interacts with numerous proteins involved in DNA replication, recombination and repair, including PCNA (5, 26), Ku (21, 22, 49, 50), p53 (27), RPA (28, 29), DNA Pol δ and topoisomerase I (30, 31, 32, 33), support the involvement of WRN in these processes, as well.

To gain a better understanding of the functions of the WRN exonuclease, we have performed a detailed analysis of the physical and biochemical properties of a minimal

fragment comprising amino acids 70-240 at the N-terminus of hWRN (hWRN-N$_{70-240}$). We show, through a variety of techniques, that: 1) hWRN-N$_{70-240}$ is capable of hydrolyzing double-stranded DNA with a 3' recessed end in a 3' to 5' direction; 2) hWRN-N$_{70-240}$ exhibits a 5'-protruding strand endonuclease activity and a 3'-5' exonuclease activity on single-stranded DNA; 3) hWRN-N$_{70-240}$ is in a trimer-hexamer equilibrium in the absence of substrate; 4) hWRN-N$_{70-240}$ forms a hexamer on a DNA substrate containing a recessed 3'-end; and 5) the interaction of hWRN-N$_{70-240}$ with PCNA causes hWRN-N$_{70-240}$ to form a hexamer. These results indicate that this minimal exonuclease region of hWRN is active as a hexamer and is capable of functioning both as an exonuclease and an endonuclease. The implications these results have on our understanding of the nuclease functions of full-length WRN are discussed.

Materials and Methods

Materials

The expression vector pET15b and BL21(DE3) cells were obtained from Stratagene. The Ni-NTA resin was from Qiagen. Superdex 200 resin was purchased from Bio-Rad, and DEAE sepharose from Pharmacia. [γ-^{32}P]-ATP was obtained from NEN Life Science Products. Bacteriophage T4 polynucleotide kinase was supplied by New England Biolabs. Deoxyribonucleoside triphosphates were purchased from Promega, and ribonucleoside triphosphates (NTPs) were supplied by GibcoBRL. Oligonucleotides (Table I), purchased in-house (UNC-Chapel Hill), were purified by HPLC. The full-length human WRN cDNA was a gift of Drs. Matthew Gray and George Martin of the University of Washington, Seattle. AFM tips were purchased from Molecular Imaging, Inc., Phoenix, AZ.

Purification of hWRN-N$_{70-240}$

hWRN-N$_{70-240}$ and hWRN-N$_{1-333}$ codons were amplified by polymerase chain reaction (PCR) and cloned into pET15b expression vector to generate pET15b-hWRN-N$_{70-240}$ and pET15b-hWRN-N$_{1-333}$. The point mutation E84A (Glu $\rightarrow$ Ala at amino acid 84) was introduced by site-directed PCR mutagenesis to generate pET15-hWRN-N$_{70-240}$ E84A. The cloned hWRN sequence, exonuclease mutation and cloning in frame to the hexa-His-tag in the expression vectors were verified by DNA sequencing. Recombinant proteins were produced according to the supplier's protocol. Cells were lysed in 20 mM Tris-HCl, pH 7.8, 300 mM NaCl, 5% glycerol, 1 mM phenylmethylsulfonyl fluoride

(PMSF), 1 mM β-ME and disrupted with a sonicator. The lysate was clarified by centrifugation and incubated with Ni-NTA resin at 4 °C for 1 h. The resin was then washed with 20 mM Tris-HCl, pH 7.8, 300 mM NaCl, 5% glycerol and protein was eluted with 20 mM Tris-HCl, pH 7.8, 300 mM NaCl, 5% glycerol, 250 mM imidazole. 20 mM DTT was added into the eluted protein. The protein sample was heated at 37 °C for 15 minutes with 100 mM DTT addition and loaded onto a Superdex S-200 column (80 cm), which was previously equilibrated in buffer A (20 mM Tris-HCl, pH 7.8, 30 mM NaCl, 5% glycerol). Pooled fractions of hWRN-N$_{70-240}$ from Superdex S-200 were finally subjected to anion exchange chromatography (Q-sepharose fast flow). The column was pre-equilibrated in buffer A before loading protein. After washing the column with 1 column volume of equilibration buffer, proteins were eluted by successive washes with three to four column volumes with each of buffer A containing 50, 100, 150, 200 mM NaCl. SDS-PAGE resolution followed by Coomassie Blue staining of proteins indicated that the major protein of the ~22 kDa hWRN-N$_{70-240}$ protein was present in 150 mM NaCl eluate. Pooled hWRN-N$_{70-240}$ fractions with a typical total protein content of 5 – 7 mg were concentrated to 10 mg / ml in the same buffer. The concentrated hWRN-N$_{70-240}$ proteins were resolved by SDS-PAGE and assayed for DNA exonuclease activity. The activity remained stable for at least 1 month at – 80 °C. Protein concentrations were determined by Bradford assay. hWRN-N$_{1-333}$ and hWRN-N$_{70-240}$E84A were purified in a fashion identical to that performed for hWRN-N$_{70-240}$.

DNA Labeling and Annealing

Single-stranded DNA oligomers were labeled with ^{32}P at their 5' ends as described in *Molecular Cloning* (*34*). To form the partial DNA duplex for exonuclease activity assay, the labeled oligomer was mixed with the complementary unlabeled DNA oligomer (1:1) in 50 mM Tris – HCl, pH 7.8, 5 mM $MgCl_2$. The DNA mixture was heated to 100 °C for ten minutes and then allowed to anneal slowly by cooling to room temperature overnight. The resulting partial duplex contained a 3' recessed terminus was ethanol-precipitated two times and resuspended in 100 μl dd H_2O.

Assays for WRN Nuclease Activities

^{32}P-labeled DNA substrate (0.1 pmol) was incubated with 25 fmol of recombinant hWRN-N$_{70-240}$ at 37 °C in a 10 μl reaction mixture containing 40 mM Tris-HCl, pH 7.4, 4 mM $MgCl_2$, 5 mM DTT, 100 μg/ml BSA and 1 mM ATP. The protein was incubated with DNA for increasing lengths of time and DNA hydrolysis was terminated by addition of quench buffer (0.5 mM EDTA and 80% formamide). At the time points indicated (*e.g.,* in Figure 2B: 0, 1, 2, 5, 10, 15, 20, 30 and 40 min), 4 μl of quench buffer were added to 8 μl of the reaction mixture. An aliquot of 6 μl of denaturing loading buffer (76% formamide, 20 mM EDTA, 2% SDS, 3% bromophenol blue, 3% xylene cyanol) was added to 6 μl of the stopped reaction mixture. The mixture was heated at 100 °C for 1 min prior to electrophoresis on a 8 M urea/20% polyacrylamide gel in 1 X TBE (90 mM Tris base, 90 mM boric acid, 1 mM EDTA). Gels were vacuum-dried and reaction products were visualized by autoradiography and quantified by PhosphoImager (Molecular Dynamics, Amersham Pharmacia Biotech, Piscataway, NJ). Assays

involving $hWRN-N_{1-333}$ and $hWRN-N_{70-240}E84A$ were performed identically to those described for $hWRN-N_{70-240}$.

Determination of Oligomerization State of $hWRN-N_{70-240}$

Ni-NTA affinity purified $hWRN-N_{70-24}$ (100 µg/100 µl) was treated with 50 mM EDTA and 100 mM DTT for 15 min at 37 °C and then subjected to gel filtration at 4 °C. The column was Superdex S-200 (80 cm), the flow rate was 0.5 ml/min and fraction size was 2 ml. Protein was eluted with 20 mM Tris-HCl, pH 7.8, 30 mM NaCl and 0.1 mM PMSF. The column was calibrated using ferritin (440 kDa), aldolase (158 kDa), ovalbumin (43 kDa) and ribonuclease A (13.7 kDa). The elution profiles of WRN were examined by SDS-PAGE.

Atomic Force Microscopy

- *Protein and DNA Preparation*

The imaged DNA was prepared using standard PCR techniques. The DNA used was a 1000 bp segment that was PCR amplified from M13 phage DNA and digested by BamHI and Ban II restriction enzymes (named L substrate; Table I). Digested fragments were isolated from 0.8% agarose gels utilizing Qiagen gel extraction kit and then, 100-bp fragment was ligated into the Ban II digested end. Yeast proliferating cell nuclear antigen (yPCNA) was a gift from Dr. Tom Kunkel of NIEHS. $hWRN-N_{70-240}$ was prepared as described above.

- *Imaging by Atomic Force Microscopy (AFM)*

Imaging was performed with Nanoscope IIIa instrument (Digital instrument, Santa Barbara, CA) using Tapping[R] mode in air. Nanosensor Pointprobe[R] non-contact/tapping[R] mode sensors with spring constants of 48 N/m and resonance frequencies of 190 KHz were used for all images. The protein and DNA molecules were deposited onto freshly cleaved mica (Spruce Pine Mica Co., Spruce Pine, NC), washed with deionized distilled water, and dried with a stream of N_2 (gas). To obtain the proper surface coverage, the deposition time was varied from 5 to 60 seconds depending on the protein and DNA concentrations. All images were collected at a scan rate of 3.0 Hz and a scan size of 1 μm. Volume analysis using AFM data was performed using image planefitting, image analysis and volume calculation, as described (35).

- *hWRN-N$_{70-24}$ – yPCNA Binding Reactions*

Binding reactions included 100 nM hWRN-N$_{70-240}$, 400 nM yeast proliferating cell nuclear antigen (yPCNA), 20 mM Tris-HCl, pH 7.8, 30 mM NaCl, 1 mM ATP, 5 mM MgCl$_2$ and 5% glycerol in 20 μl. After incubation for 15 min at room temperature, the reaction mixtures were diluted to 200 μl with binding buffer (20 mM Tris-HCl, pH 7.8, 30 mM NaCl, 5 mM MgCl$_2$ and 5% glycerol) and 20 μl were immediately deposited onto freshly cleaved mica.

- *hWRN-N$_{70-24}$ – DNA Binding Reactions*

Two substrates were used in the binding reactions: one is a short piece of DNA named substrate S with a top strand of 20 nucleotides and a bottom strand of 46 nucleotides (see Table I); the other is substrate L, which contains a 100-base 5'-overhang as described above. Binding reactions included 3-4 nM DNA fragment, 400-500 nM

hWRN-N$_{70-240}$, 20 mM Tris-HCl, pH 7.8, 30 mM NaCl, 5 mM MgCl$_2$ and 5% glycerol in 20 μl. After incubation for 15-20 min at room temperature, the reaction mixtures were diluted to 200 μl in binding buffer (20 mM Tris-HCl, pH 7.8, 30 mM NaCl, 5 mM MgCl$_2$ and 5% glycerol) and 20 μl were immediately deposited onto freshly cleaved mica.

- *Volume Calculations*

The measured AFM volumes, calculated using the equation $V_i = A_i *(M_i - S)$, of each of the proteins were distributed in a Gaussian fashion (35). The volume and its uncertainty for a given protein were taken to be the average and standard deviation of the distribution. The area A_i, total average height M_i, and surface heights are measured using ImageSXM software. Prior to volume analysis, the deposition time that produced optimal surface coverage was determined. The data were filtered to remove false positives of oligomers using major: major axis ratio cut-off values ranging from 2.0 to 2.5. Histogram plots of the volume of the proteins were generated for each cut-off value.

Results

Cloning , Expression and Purification of hWRN-N$_{70-240}$ and hWRN-N$_{1-333}$

The hWRN exonuclease activity resides within the first 333 amino acids at the N-terminus of the protein (5). In aligning the hWRN and mouse WRN sequences (mWRN), we found that the region between amino acids 70 and 240 (Fig. 1) exhibits 83% amino acid identity and includes all five amino acids (Asp82, Glu84, Asp143, Tyr212 and Asp216) predicted to be critical for hWRN exonuclease activity (5). To determine if this region of hWRN was sufficient to function as an exonuclease, we cloned, expressed and characterized a minimal fragment of the WRN exonuclease comprised of residues 70-240 and analyzed it relative to the N-terminal 333-amino acid exonuclease fragment of WRN described previously (5).

Regions of the hWRN cDNA encoding amino acids 70 to 240 and 1 to 333 were amplified by the polymerase chain reaction (PCR) and cloned into pET15b to generate pET15b-hN$_{1-333}$ and pET15b-hN$_{70-240}$. In addition, the point mutation E84A (Glu to Ala at amino acid 84), previously shown to inactivate the exonuclease activity of full-length hWRN (5), was introduced by site-directed PCR mutagenesis to generate pET15b-hN$_{70-240}$E84A. The cloned hWRN-N$_{70-240,}$ hWRN-N$_{1-333}$, and hWRN-N$_{70-240}$E84A regions were confirmed by DNA sequencing and inserted in-frame to allow the production of hexa-His-tagged proteins.

The recombinant hWRN-N$_{1-333}$ and hWRN-N$_{70-240}$ proteins were expressed in BL21 (DE3) pLys cells, and purified by Ni-affinity batch preparation, gel filtration chromatography using a Superdex S-200 column and finally, Q-sepharose fast flow. At the gel filtration step, the recombinant hWRN proteins eluted as a trimer as judged by

apparent molecular weight, and were found to be ~80% pure by SDS-PAGE. After anion

exchange chromatography, proteins were estimated to be >95% pure based on Coomassie

blue stained SDS-PAGE. Purified hWRN-N$_{70-240}$ had an apparent monomeric molecular

weight of 21 kDa, as expected from the calculated molecular weight with the additional

N-terminal 20 amino acids encoded by the hexa-His-tag and vector. Purified hWRN-N$_{1-333}$ and hWRN-N$_{70-240}$ E84A had apparent molecular weights of 40 kDa and 21kDa,

respectively.

hWRN-N$_{70-240}$ is a 3'-5' Exonuclease

We assayed the exonuclease activity of hWRN-N$_{70-240}$ using a 5'-labeled, 3'-

recessed double-stranded DNA substrate composed of a 24-nucleotide top strand and a

60-nucleotide bottom strand which we will refer to as substrate R (see Table I and Figure

2A). As shown in Figure 2, hWRN-N$_{70-240}$ and hWRN-N$_{1-333}$ both catalyze the

exonucleolytic hydrolysis of the 3'-recessed end of this double-stranded DNA duplex in a

3'-to-5' direction. Cleavage of substrate radiolabeled at the 5'-terminus yields time-

dependent accumulation of shorter products. In our hands, both hWRN-N$_{70-240}$ and

hWRN-N$_{1-333}$ function as relatively non-processive 3'-5' exonucleases. In particular,

hWRN-N$_{70-240}$ appears to degrade DNA substrate R efficiently through the G$_6$ region (see

Table I) but not beyond that. No detectable hWRN-N$_{70-240}$ nuclease activity was

observed using the hWRN-N$_{70-240}$E84A construct of human WRN. Similarly, no hWRN-

N$_{70-240}$ nuclease activity was observed with single-stranded DNA alone (*i.e.* not an

overhang from a DNA duplex), blunt-ended DNA, or a DNA duplex containing a 5'-

rececced end (data not shown). These results suggest that hWRN-N$_{70-240}$ exonuclease is

distinguished from the large majority of the known nucleases by its preference for 3'
recessed DNA in a partial duplex.

hWRN-N$_{70-240}$ Appears Trimeric By Gel Filtration Chromatography

Once we established that the minimal 70-240 region of hWRN functioned as an
exonuclease, we next sought to characterize the oligomeric state of this fragment in vitro.
The 1-333 fragment of hWRN was previously reported to be a trimer based on gel
filtration chromatography (5). To determine the oligomeric structure of hWRN-N$_{70-240}$,
we also employed gel filtration chromatography. Purified hWRN-N$_{70-240}$ was re-applied
to the Superdex S-200 column used for protein purification (Figure 3A). A set of standard
proteins was also assayed in separate runs on the same gel filtration column (see
Materials and Methods), and their elution volumes were used to construct a standard
curve. The hWRN-N$_{70-240}$ was found to elute in a peak corresponding to a molecular
mass of 65 kDa (Figure 3A and Figure 3B). As the molecular mass of the recombinant
hWRN-N$_{70-240}$ is ~21 kDa, the oligomerization state of this population of the protein
appears to be trimeric.

hWRN-N$_{70-240}$ is in a Trimer-Hexamer Equilibrium by AFM

Atomic force microscopy (AFM) can be used to image soft samples with
nanometer resolution both in air and in solution (35). We used it to quantify
intermolecular protein and protein-DNA interactions. Because AFM produces
topographical images, it is possible to relate the molecular weight of a protein to its
volume. Previous studies have shown a linear relationship between the measured AFM

57

volume and the molecular weight of the proteins by the equation $v = (1.3 * MW) - 25$, where v is the volume measured by AFM and MW is the molecular mass (35). Furthermore, the expected AFM volume of oligomeric protein complexes can be predicated using this equation. Table II shows the possible hWRN-N$_{70-240}$ complexes considered in this study.

To determine the oligomerization state of this minimal exonuclease fragment of human WRN, hWRN-N$_{70-240}$, it was deposited at 20 nM, images were collected and volume analysis was performed for each set of images. Figure 4A shows an ideal hWRN-N$_{70-240}$ surface coverage for volume analysis. 1135 protein data points were analyzed and a histogram of molecular volumes for hWRN-N$_{70-240}$ is shown in Figure 4B. The distribution has a major peak at ~60 nm^3 with a shoulder at higher volumes. The major peak was fit to a Gaussian function to determine the volume. The observed AFM volume is (60 ± 5) nm^3, which corresponds to a molecular weight of 66 kDa. This mass is consistent with a trimer of hWRN-N$_{70-240}$ (see Table II). The small peak at ~140 nm^3 is consistent with the molecular weight of hWRN-N$_{70-240}$ hexamer (125 kDa). Approximately 10% of the hWRN-N$_{70-240}$ is in the higher association state (hexamer). This result is consistent to the previous report that 90% wild type hWRN existed as trimer in the absence of DNA (5).

hWRN-N$_{70-240}$ Interacts with Trimeric PCNA as a Trimer and a Hexamer

Human PCNA, which is a major component of the DNA replication fork, has been reported to interact with hWRN (5, 26). To determine if PCNA interacts with the minimal

WRN fragment hWRN-N$_{70-240}$, we imaged *Saccharomyces cerevisiae* PCNA (yPCNA) alone and in complex with hWRN-N$_{70-240}$ by AFM. Human PCNA shares 35% sequence identity with yPCNA, and the three-dimensional structures of these two trimeric proteins are highly similar (*36*). We used AFM to determine the oligomerization state of yPCNA and in complex with hWRN-N$_{70-240}$. As expected, yPCNA exists as a trimer with the molecular weight of ~92 kDa (Figure 5A). Inspection of Figure 5A reveals a single peak at ~90 nm^3. With hWRN-N$_{70-240}$ and yPCNA together, two distinct peaks are observed at 150 nm^3 and 235 nm^3 (Figure 5B). These peaks correspond to the molecular weights of 135 kDa and 200 kDa, respectively. Based on Table II, the first peak (135 kDa) can be explained by two species: one is a complex of a trimer yPCNA and a trimer hWRN-N$_{70-240}$ (150 kDa); the other one is a hexamer of hWRN-N$_{70-240}$ (126 kDa). However, as shown above by two methods, hWRN-N$_{70-240}$ is a trimer in the absence of DNA (Figures 3-4). In addition, the disappearance of the peak corresponding to the trimer of PCNA from Figure 5B provides strong evidence that the two peaks present in this figure contain PCNA and hWRN-N$_{70-240}$. Thus, we conclude that the first peak represents trimer PCNA: trimer hWRN-N$_{70-240}$ complexes. The second peak (200 kDa) can be explained by trimer yPCNA: hexamer hWRN-N$_{70-240}$ complexes (213 kDa), or by hexamer yPCNA: trimer hWRN-N$_{70-240}$ complexes (237 kDa). Because there is no evidence for PCNA hexamer formation, we conclude that the second peak is generated by complexes of yPCNA trimers and hWRN-N$_{70-240}$ hexamers. Thus, yPCNA and hWRN-N$_{70-240}$ form both trimer-trimer and trimer-hexamer complexes. Further, the interaction of PCNA with hWRN-N$_{70-240}$ appears to drive the formation of hWRN-N$_{70-240}$ hexamers.

hWRN-N$_{70-240}$ Hexamerizes on DNA

In several previous studies, AFM was used for qualitative analyses of several protein-DNA complexes, including that of Cro protein, RNA polymerase, and Heat shock transcription factor-2 with double-stranded DNA, and the interaction of single-stranded DNA-binding proteins and single-stranded DNA (37, 38, 39, 40). These successful applications of AFM have made it possible to extend the use AFM to the quantitative analysis of DNA-protein interactions. To test whether hWRN-N$_{70-240}$ exists as trimer or as a higher-order oligomer in the presence of DNA, AFM images were obtained in the presence of substrate S, a 3'-recessed double-stranded DNA molecule composed of a 20 nucleotide top strand and a 46-nucleotide bottom strand (Table I, Figure 6A). This DNA substrate is too small, with a maximal expected length of 7 nm, to be visualized by AFM. A statistical analysis reveals that the volume distribution of hWRN-N$_{70-240}$ molecules on DNA substrate S is best explained by the existence of two populations of hWRN-N$_{70-240}$ molecules with two different sizes (Figure 6B): 45% as 60 nm^3 and 49% as 135 nm^3. Based on the volumes in Table II, the smaller and larger populations are considered to be trimers and hexamers, respectively. In the absence of DNA, this hexamer peak is not present (Figure 4B). Because the size of the small DNA substrate S can be neglected in AFM images, we thus conclude that hWRN-N$_{70-240}$ efficiently hexamerizes in the presence of DNA.

We next sought to visualize directly complexes of hWRN-N$_{70-240}$ and DNA. A longer DNA substrate L (1 kb) was used which contained a recessed 3'- end formed by a 100-base 5'-overhang (Table I). This DNA substrate was visualized alone by AFM and revealed molecules of between 305 and 315 nm in length, as expected (Figure 6C). Upon

the addition of hWRN-N$_{70-240}$, it is possible to visualize hWRN-N$_{70-240}$ complexes associating with one end of this DNA substrate (Figure 6D). Based on the activity studies presented above, we conclude that hWRN-N$_{70-240}$ is associating with the recessed 3'-end of the DNA substrate.

These AFM studies also revealed an unexpected observation: upon incubation of hWRN-N$_{70-240}$ with the 1 kb DNA substrate L for 15 minutes with 1 mM ATP, we visualized the accumulation of smaller DNA fragments (*e.g.,* Figures 6E-F). These are not the product of efficient 3'-5' exonuclease activity by hWRN-N$_{70-240}$, as that would produce DNA molecules with single-strand DNA ends that would still appear long by AFM (35). The fragments observed ranged in length from ~300 nm to as small as 37 nm (roughly 900 bp to 110 bp, respectively), with intermediate sizes such as 141 nm and 91 nm (430 bp and 275 bp, respectively) also present. We hypothesized these smaller DNA fragments were the product of a novel, protruding-strand endonuclease activity associated with hWRN-N$_{70-240}$. This activity would involve the cleavage of the intact DNA strand opposite from the recessed 3'-end. A more detailed examination of this activity is described below.

hWRN-N$_{70-240}$ Exhibits a 5'-Protruding Strand DNA Endonuclease Activity

Because AFM images of hWRN-N$_{70-240}$ incubated with DNA substrate L reveal the accumulation of smaller DNA fragments (Figures 6E-F), we suspected that hWRN-N$_{70-240}$ contained a novel nuclease activity. As stated above, we hypothesized that these smaller fragments were the product of the cleavage of the intact DNA strand opposite from the 3'-recessed DNA strand in substrate L. To examine the effect of hWRN-N$_{70-240}$

on a 5'-overhang in a DNA substrate, we employed the same substrate examined in the studies presented in Figure 2, but radiolabeled the 5'-end of the 60-mer, rather than the 5'-end of the 24-mer (Table I; unlabeled 24-mer/5'-^{32}P-labeled 60-mer). As shown in Figure 7, hWRN-N$_{70-240}$ was found to exhibit endonuclease activity at three sites within this substrate. First, a burst of cleavage appears to occur at two sites: at the single-strand/double-strand DNA junction, which generates the 36 base product, and within the single-stranded region of the 5'-protruding strand, which generates the 12 base product. Second, a cleavage occurs in the single-stranded, 5'-overhang region at the T$_{12}$-GAAC junction; this generates the 24 base product. The enzyme also appears to function as a 3'-5' exonuclease on the single-stranded DNA products. This effect is most obvious in the degradation of the 24 base product to generate a ladder of smaller products, although it can also be seen in the transient degradation of the 36 base and 12 base products as well. It is possible that the 24 base product is generated not from an endonuclease cleavage, but from the 36 base product by efficient 3'-5' exonuclease action through the T$_{12}$ region which then slows when it reaches the non-poly-T sequence. Products intermediate in length between 36 and 24 nucleotides are evident in the early time points in Figure 7. In summary, hWRN-N$_{70-240}$ contains an 5'-protruding strand endonuclease activity at a single-strand/double-strand DNA junction and within the single-stranded region of the 5'-protruding strand. hWRN-N$_{70-240}$ also appears to contain a 3'-5' exonuclease activity on the single-stranded DNA products generated from its endonuclease action. The functional implications of these additional activities are discussed below.

Discussion

In our examination of a 171-amino acid exonuclease fragment of the WRN helicase/exonuclease (hWRN-N$_{70-240}$), we have found that this small construct contains 3'-5' exonuclease activity (Figure 2), and forms a hexamer in the presence of DNA (Figure 6). We have further shown that hWRN-N$_{70-240}$ is able to cleave a 5'-protruding strand at the single-strand/double-strand DNA junction and within the single-stranded region of the DNA. The biological importance of the hexamer formation and these additional nuclease activities of hWRN-N$_{70-240}$ is discussed below.

A 333-amino acid N-terminal fragment of the WRN protein containing the exonuclease region was thought be a trimer based on examinations by gel filtration chromatography (*5*). We similarly found that our minimized, 171-amino acid region of the WRN exonuclease (hWRN-N$_{70-240}$) appears to be a trimer by gel filtration (Figure 3). An analysis of hWRN-N$_{70-240}$ by atomic force microscopy, however, revealed that hWRN-N$_{70-240}$ is in a trimer-hexamer equilibrium in the absence of DNA with the trimer being the major species (Figure 4). We further found using AFM that the oligomerization state of hWRN-N$_{70-240}$ is dramatically shifted toward hexamers in the presence of a small DNA substrate S (Table I; Figure 6A-B). In addition, we visualized directly hWRN-N$_{70-240}$ hexamers bound to one end of a 1 kb piece of DNA (Figure 6D). These observations suggest that the active form of hWRN-N$_{70-240}$, and by extension the exonuclease region of full-length WRN, is a hexamer rather than a trimer.

WRN represents the first example of an exonuclease that is a hexamer. The 226-amino acid 5'-3' exonuclease from lambda phage, which participates in phage recombination and double-strand break repair, is a trimer in the x-ray crystal structure

(41). This toroidal structure was determined in the absence of DNA, but was proposed to wrap around a DNA substrate to facilitate highly processive exonucleolytic cleavage of DNA. The WRN exonuclease region may exist as an analogous toroidal trimer in the absence of DNA, and these toroidal trimers may stack to form a WRN exonuclease hexamer in the presence of DNA. Alternatively, the WRN exonuclease hexamer may form a toroid, with the trimeric form of the protein existing as a half-toroid structure.

The oligomerization state of the full-length WRN protein is unknown. Based on its similarity to the BLM protein, which is hexameric (42), WRN may form a hexamer. Indeed, one critical difference between WRM and BLM is the presence of the exonuclease region in WRN, which may impact the oligomerization state of WRN. However, the WRN exonuclease region alone appears to form a hexamer, an observation which supports the suggestion that full-length WRN is a hexamer. Such a conclusion is further supported by the finding that the interaction of hWRN-N$_{70-240}$ with PCNA appears to drive the formation of hWRN-N$_{70-240}$ hexamers (Figure 5). PCNA is composed of a trimer of pseudo-dimers, which together form a toroid ring (43). The observation that interactions with PCNA drive the formation of hWRN-N$_{70-240}$ hexamers suggests that the hWRN-N$_{70-240}$ hexamer structure may be a single toroid composed of two half-toroidal trimers, and that the hWRN-N$_{70-240}$ and PCNA toroids stack. The formation of WRN exonuclease hexamers on PCNA may in turn stabilize the hexameric form of the full-length WRN protein as well.

In examining the interaction of hWRN-N$_{70-240}$ with the 1kb DNA substrate L by AFM, smaller DNA fragments with no evidence of long single-stranded ends were noted (Figures 6E-F). Double-stranded DNA products near 1 kb in size (*e.g.,* 900 bp) were

observed, as well as products of between 430 and 110 bp in length. Based on these observations, we sought to determine if the WRN exonuclease region is able to cleave the 5'-protruding strand near a single-strand/double-strand DNA junction. We found that hWRN-N$_{70-240}$ is capable of cleaving such a 5'-protruding strand at the single-strand/double-strand DNA junction, as well as within the single-stranded region of the 5'-protruding strand (Figure 7). In addition, it appeared that hWRN-N$_{70-240}$ further degraded these single-stranded DNA products using its 3'-5' exonuclease activity (Figure 7). Thus, the activities of hWRN-N$_{70-240}$ can be expanded to include not only a 3'-5' exonuclease acting on the recessed 3'-end of duplex DNA, but also a 5'-protruding strand endonuclease activity and a 3'-5' exonuclease activity on single-stranded DNA.

In previously published reports it was shown that full-length hWRN can cleave a 1 nucleotide 5'-flap imbedded within a DNA duplex; these authors further showed that the interaction of hWRN with the flap endonuclease Fen1 stimulates Fen1's ability to act on an identical substrate containing a 5'-flap (58). In addition, it was recently shown that the interaction of the Ku 70/80 heterodimer with a construct of hWRN encompassing residues 1-388 stimulates this hWRN fragment to degrade single-stranded DNA with a 3'-5' polarity (50). Thus, human WRN appears to harbor the ability to perform multiple nuclease actions depending on the DNA substrate and hWRN's interaction with other proteins. Ku is known to interact with the residues 1-50 of hWRN (50). Perhaps this interaction changes the positioning of these first 50 amino acids of hWRN, expanding the capabilities of the nuclease region of hWRN to include both endonuclease and exonuclease actions on single-stranded DNA. Perhaps we were able to detect this activity without the presence of the Ku heterodimer because hWRN-N$_{70-240}$ lacks these

first 50 amino acids. These first 50 amino acids of hWRN may serve as a regulatory region that controls the nuclease action of the enzyme.

The protruding strand endonuclease function of hWRN suggests an interesting new similarity with the Mre11 nuclease. Mre11 forms a complex with the Nbs1 and Rad50 proteins, and this complex plays important roles in DNA double-strand break repair and meiotic recombination in eukaryotes (44, 45, 46). Mre11 is a 3'-5' exonuclease in the absence of ATP; upon the addition of ATP, however, Mre11 is stimulated to cleave the 3'-protruding strand at a DNA single-strand/double-strand junction (47). WRN is also a 3'-5' exonuclease and cleaves a protruding strand near a DNA single-strand/double-strand junction; in contrast to Mre11, however, WRN cleaves a 5'-protruding strand rather than a 3'-protruding strand. Parallel biological roles for WRN and the Mre11 complex have been suggested previously because genetic experiments indicate the Mre11 complex is required for efficient end joining in *S. cerevisiae* double-strand break repair (48). In addition, WRN has been implicated in the mammalian end-joining pathway through its association with the Ku 70/80 heterodimer (21, 22, 49, 50). Ku70/80 initiates end-joining in eukaryotes by binding to double-strand DNA breaks (51). Mre11 is thought to facilitate the joining of double-strand breaks by removing non-complementary overhangs or damaged bases at broken ends by its ability to combine exonuclease and endonuclease activities (47). Perhaps WRN's largely analogous activities allow it to substitute for Mre11 in this process in human cells. Alternatively, WRN's opposite polarity in protruding strand cleavage (5' vs. 3' for Mre11) suggests that WRN may be called upon to cleave 5'-overhangs to generate

66

double-strand DNA breaks, which can then be processed by the Ku heterodimer and the Mre11 complex.

WRN has also been proposed to function in DNA replication, and its physical interaction with PCNA supports this role (5). The identification of a 5'-protruding strand endonuclease activity associated with the WRN exonuclease region suggests a specific role for WRN in DNA replication. During lagging strand synthesis, regions of Okazaki fragments are displaced by a helicase and the 5'-protruding ends are cleaved by the flap endonuclease Fen1 (52). DNA ligase I seals the nicks remaining at the final stages of lagging strand synthesis (53). Both Fen1 and DNA ligase I are also known to interact physically with PCNA (54, 55, 56, 57); in addition, as stated above, Fen1 and WRN are known to interact via the helicase region of WRN, and this interaction stimulates Fen1's 5'-flap endonuclease activity (58). WRN may be capable of combining a Fen1-like nuclease activity with a helicase function, thereby efficiently coupling these two activities during Okazaki fragment removal. The WRN helicase region could break Watson-Crick base pairs, and the 5'-protruding endonuclease activity of the WRN exonuclease region could cleave the resulting flap. Thus, WRN may be capable of acting in concert with or substituting for Fen1 and other proteins in lagging strand synthesis under particular conditions within the cell.

In summary, our results indicate that the active form of the WRN exonuclease is a hexamer, and that the interaction of the WRN exonuclease region with PCNA aids in the formation of WRN exonuclease hexamers. The formation of WRN exonuclease hexamers may be linked to, or the consequence of, the formation of full-length WRN protein hexamers. We have further found that the WRN exonuclease region exhibits a

5'-protruding strand endonuclease activity. Finally, we propose that the function of WRN in human cells may be to substitute for or aid other helicases and nucleases in the processes of DNA replication and/or double-strand break repair. The action of WRN as an important "back-up" system in DNA replication and repair might provide one explanation for why the symptoms of WS develop over decades, rather than being temporally acute.

Table I. DNA Molecules Employed in these Studies

Length   Sequence

Substrate R:

(1) 24 nt   5'-d (CAGGCACAGGGTCAGGTCGGGGGG) –3'

(2) 60 nt   5'-d (TCTACGCTCTGAGTGACTGACAAGTTTTTTTTTTTTCCCCCCGA

                    CCTGACCCTGTGCCTG) –3'

Substrate S:

(3) 20 nt   5'-d (CGCTAGCAATATTCTGCAGC) –3'

(4) 46 nt   5'-d (GCGCGGAAGCTTGGCTGCAGAATATTGCTAGCGGGAAATCGG

                CGCG) –3'

Substrate L:

1 kb DNA duplex with 100-base 5'-overhang

Table II. Predicted and Observed AFM Volumes for hWRN-N$_{70-240}$ and PCNA Oligomers

Complexes		Molecular Weight (kDa)	Predicated AFM Volume (nm^3)	Observed AFM Volume (nm^3)
Monomer	hWRN-N$_{70-240}$	21	2.3	*not observed*
Dimer	hWRN-N$_{70-240}$	42	29.6	*not observed*
Trimer	hWRN-N$_{70-240}$	63	56.9	68 ± 8
Hexamer	hWRN-N$_{70-240}$	126	138.8	135 ± 7
Monomer	PCNA	29	12.7	*not observed*
Trimer	PCNA	87	88.1	94 ± 7
Hexamer	PCNA	174	201.2	*not observed*
Trimer hWRN-N$_{70-240}$ Trimer PCNA complex		150	170	150 ± 15
Hexamer hWRN-N$_{70-240}$ Trimer PCNA complex		213	251.9	235 ± 11
Hexamer hWRN-N$_{70-240}$ Hexamer PCNA complex		300	365	*not observed*

Figure Legends

Figure 1: Amino acid sequence alignment of residues 1-360 from the N-terminal regions of the human and mouse WRN proteins, generated using Workbench software. Identical amino acids are in red, similar amino acids are in blue, and the five amino acids thought to be critical for exonuclease activity are in bold.

Figure 2: **A**: DNA duplex (substrate R) composed of a ^{32}P 5'-labeled (star) 24-mer with a recessed 3'-end and an unlabeled 60-mer (see Table I). **B**: hWRN-N$_{70-240}$ and hWRN-N$_{1-333}$ exonuclease activities on a DNA substrate with a recessed 3'-end. The lane marked "uncut" contains DNA that has not been treated with wither hWRN-N$_{70-240}$ or hWRN-N$_{1-333}$. The times below each lane corresponds to the length of time DNA substrate R was incubated with an hWRN-N construct at 37 °C prior to the addition of 0.5 mM EDTA and 80% formamide. The nucleolytically fragmented ^{32}P-5'-labeled 24-mer was resolved from undigested DNA by electrophoresis through a 20% denaturing polyacrylamide gel as described (see Materials and Methods).

Figure 3: Analysis of quaternary structure of hWRN-N$_{70-240}$ by gel filtration chromatography. **A**: Elution profile for purified hWRN-N$_{70-240}$ on a Superdex S-200 column. **B**: Elution of the hWRN-N$_{70-240}$ oligomers and molecular weight markers from Superdex S-200 chromatography. The molecular weight markers are ferritin (440 kDa), aldolase (158 kDa), ovalbumin (43 kDa) and ribonuclease A (13.7 kDa). The elution peak for dextran from the column was considered the exclusion volume (V_0). V_t was

determined by the elution of salt from the column. V_e is the elution volume of hWRN-N_{70-240} oligomers (closed square) and molecular weight markers (closed circles).

Figure 4: **A**: AFM image of hWRN-N_{70-240} at 20 nM. The image shows proper surface coverage for volume analysis. hWRN-N_{70-240} was equilibrated at 37 °C for 15 min in 20 mM Tris-HCl, pH 7.8, 30 mM NaCl, 5 mM $MgCl_2$ and 5% glycerol. A deposition time of 30 seconds was used for this concentration of hWRN-N_{70-240}. The scale bar represents 200 nm. **B**: Gaussian fit of the volume histogram for hWRN-N_{70-240} (20 nM). The solid line is the Gaussian fit of the volume data for trimers. The number of proteins under each curve represents that species' population. The number within the parentheses (1135) represents the number of proteins analyzed. Only the trimer distribution has a Gaussian shape, and the fraction of trimers determined by counting the number of proteins under the curve is 83%.

Figure 5: Gaussian fit of the volume histogram for yeast PCNA (yPCNA) and yPCNA:hWRN-N_{70-240} complexes. **A**: yPCNA (40 nM) alone was equilibrated at 37 °C for 15 min in 20 mM Tris-HCl, pH 7.8, 30 mM NaCl, 5 mM $MgCl_2$ and 5% glycerol and deposited to freshly cleaved mica for 30 sec. The solid line is a Gaussian fit of the volume data for yPCNA trimers. **B**: yPCNA and hWRN-N_{70-240} binding in the same buffer as yPCNA alone for 15 min at room temperature prior to deposition on mica. Two distributions represent trimer yPCNA: trimer hWRN-N_{70-240} complexes and trimer yPCNA: hexamer hWRN-N_{70-240} complexes.

Figure 6: Binding of hWRN-N$_{70-240}$ to DNA containing a 3'-recessed end. **A**: Surface plot of an enlarged AFM image of hWRN-N$_{70-240}$–DNA complexes. The scale bar represents 100 nm. The DNA substrate S alone is too small to be visualized by AFM (20-mer:46-mer; 7 nm in length). Arrows indicate hWRN-N$_{70-240}$ monomers, trimers and hexamers. **B**: Gaussian fit of the volume histogram for hWRN-N$_{70-240}$-DNA complexes. The solid line is the Gaussian fit of the volume data for trimers and hexamers. The number of proteins under the trimer curve is 328, and under the hexamer curve is 360. The number within the parentheses (728) is the number of total proteins analyzed. Both the trimer and hexamer distributions have Gaussian shapes. The fraction of trimers is 45%, and that of hexamers is 49%. The predominant hexamer peak is missing in the absence of DNA (see Figure 4). **C**: AFM image of the 1 kb DNA substrate L with 100 base 5'-overhang alone. These molecules measure between 305 nm and 314 nm in length, the expected length of a 900 bp DNA duplex assuming 0.33 nm per bp. **D**: AFM image of hWRN-N$_{70-240}$-DNA complexes formed on the 1 kb substrate L. Bound hWRN-N$_{70-240}$ molecules can be seen at one end of several of the DNA duplexes. Volume analysis indicated that bound hWRN-N$_{70-240}$ molecules are hexamers. **E-F**: AFM images of a small DNA fragments generated by the digestion of DNA substrate L by hWRN-N$_{70-240}$. hWRN-N$_{70-240}$ and substrate L were incubated at 37 °C for 15 minutes in 1 mM ATP before deposition on mica. In E, DNA molecules of 301 nm (912 bp), 299 nm (906 bp), 91 nm (275 bp), 75 nm (227 bp), and 47 nm (142 bp) can been seen. In F, DNA molecules of 289-285 nm (~870 bp), 265 nm (803 bp), 141 nm (427 bp), 112 nm (340 bp), 82 nm (248 bp), 54 nm (163 bp), and 37 nm (112 bp) can been seen. The scale for both E and F is identical.

Figure 7: hWRN-N$_{70-240}$ endonuclease and exonuclease activities on a 5'-protruding DNA strand. The times below each lane indicate the length of time DNA substrate R, which was radiolabeled on 5'-end of the bottom 60 nucleotide strand, and hWRN-N$_{70-240}$ were incubated together at 37 °C prior to the addition of 0.5 mM EDTA and 80% formamide. The nucleolytically fragmented ^{32}P 5'-labeled 60-mer was resolved from undigested DNA by electrophoresis through a 20% denaturing polyacrylamide gel as described (see Materials and Methods). The lane marked "uncut" contains DNA that has not been treated with hWRN-N$_{70-240}$; the position of the 5'-labeled 60-mer is indicated. Three "marker" lanes are also present, which contain standards of 36, 24, 12 10, 8, 6, 4, 3, 2, and 1 nucleotides in length. The sequence of the DNA substrate R employed in these studies is shown; this substrate contains a 24 bp duplex region and a 36 nucleotide 5'-protruding strand. The sites of the initial bursts of endonuclease cleavage that generate the 36 base and 12 base products are indicated (open arrows). The position of the non-transient endonuclease cleavage that generates the 24 base product is shown (closed arrow). In addition, the sites of 3'-5' exonuclease cleavage of the 24 base product, the clearest single-stranded DNA exonuclease activity present in this gel, are also indicated (closed arrowheads).

Figure 1

```
human_WRN_1-360     1 MSEKKLETTAQQRKCPEWMNVQNKRCAVEERKACVRKSVFEDDLPFLEFT
mouse_WRN_1-360     1 -----METTSLQRKFPEWMSMQSQRCATEEK-ACVQKSVLEDNLPFLEFP

human_WRN_1-360    51 GSIVYSYDASDCSFLSEDISMSLSDGDVVGFDMEWPPLYNRGKLGKVALI
mouse_WRN_1-360    45 GSIVYSYEASDCSFLSEDISMRLSDGDVVGFDMEWPPIYKPGKRSRVAVI

human_WRN_1-360   101 QLCVSESKCYLFHVSSMSVFPQGLKMLLENKAVKKAGVGIEGDQWKLLRD
mouse_WRN_1-360    95 QLCVSESKCYLFHISSMSVFPQGLKMLLENKSIKKAGVGIEGDQWKLLRD

human_WRN_1-360   151 FDIKLKNFVELTDVANKKLKCTETWSLNSLVKHLLGKQLLKDKSIRCSNW
mouse_WRN_1-360   145 FDVKLESFVELTDVANEKLKCAETWSLNGLVKHVLGKQLLKDKSIRCSNW

human_WRN_1-360   201 SKFPLTEDQKLYAATDAYAGFIIYRNLEILDDTVQRFAINKEEEILLSDM
mouse_WRN_1-360   195 SNFPLTEDQKLYAATDAYAGLIIYQKLGNLGDTAQVFALNKAEENLPLEM

human_WRN_1-360   251 NKQLTSISEEVMDLAKHLPHAFSKLENPRRVSILLKDISENLYSLRRMII
mouse_WRN_1-360   245 KKQLNSISEEMRDLANRFPVTCRNLETLQRVPVILKSISENLCSLRKVIC

human_WRN_1-360   301 GSTNIETELRPSNNLNLLSFEDSTTGGVQQKQIREHEVLIHVEDETWDPT
mouse_WRN_1-360   295 GPTNTE--------------------------------------------

human_WRN_1-360   351 LDHLAKHDGE
mouse_WRN_1-360       ----------
```

Figure 2.

A.

DNA Substrate

B.

Figure 3.

A

B

Figure 4.

A

hWRN-N$_{70-240}$

200 nm

B

Figure 5.

Figure 6.

A

B

81

C

309 nm

310 nm

1 kb DNA

314 nm

305 nm

1 kb DNA

100 nm

D

Hexamer
hWRN-
N_{70-240}

1 kb DNA+
hWRN-N_{70-240}

100 nm
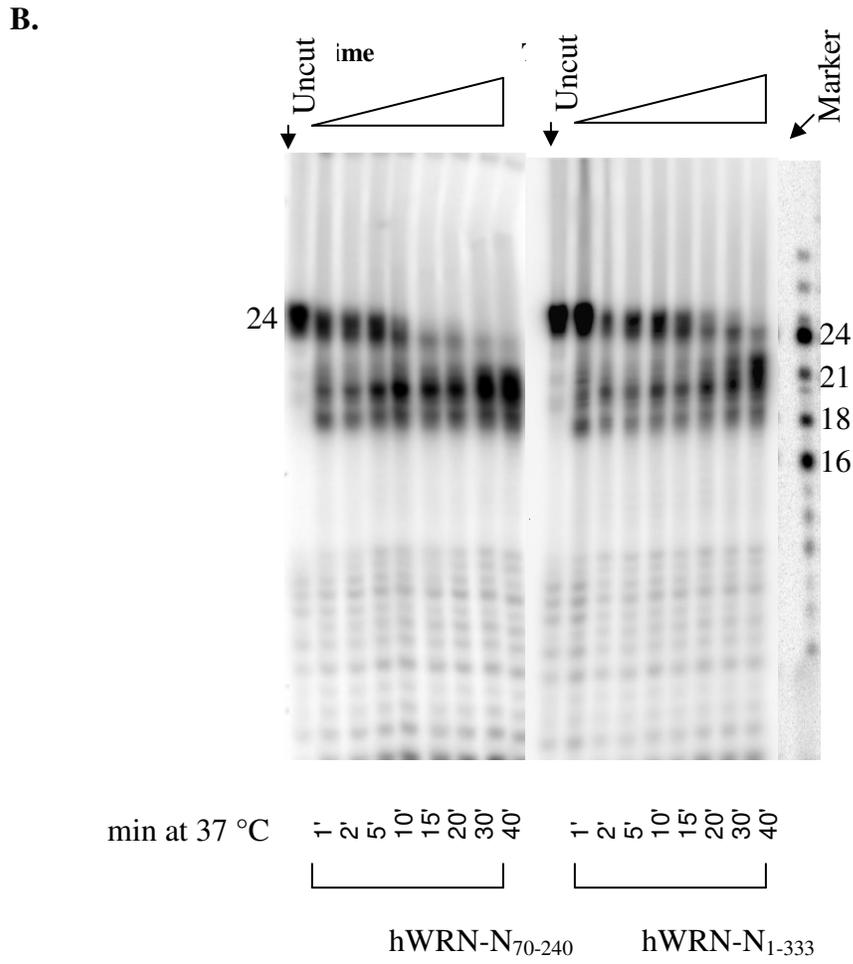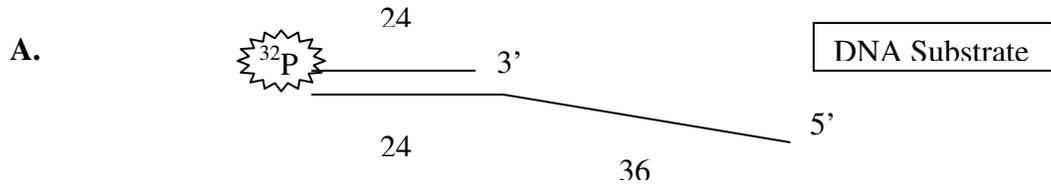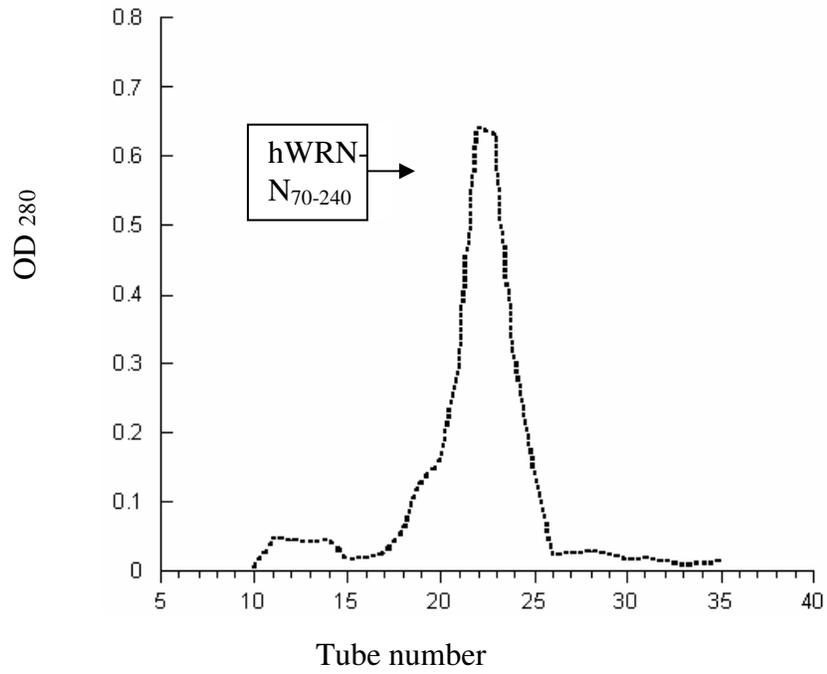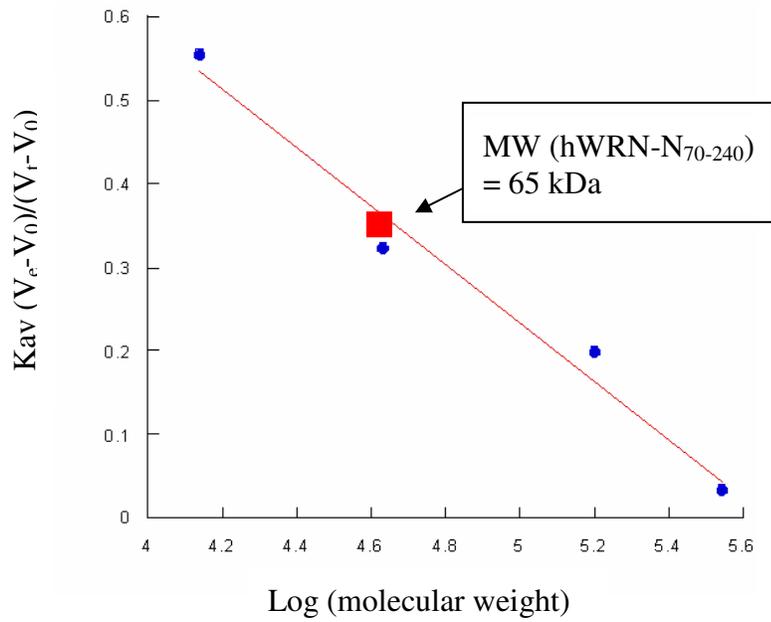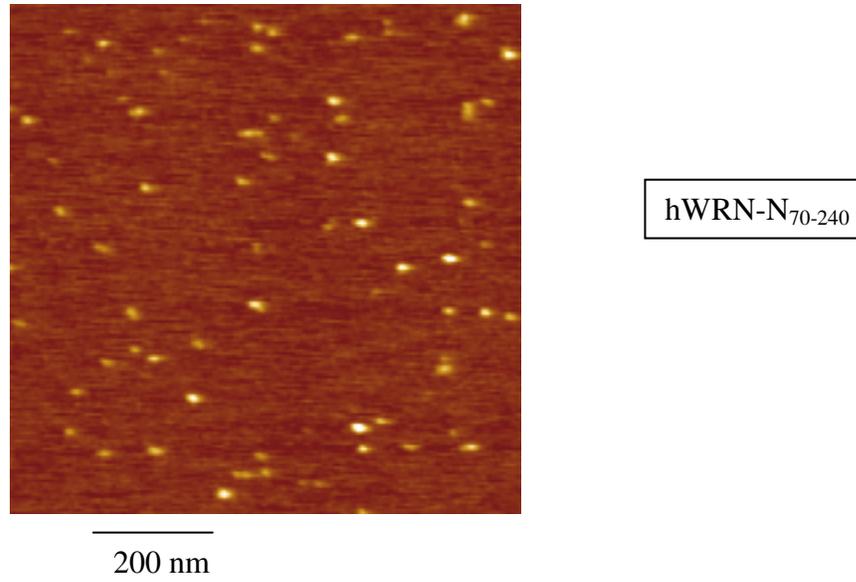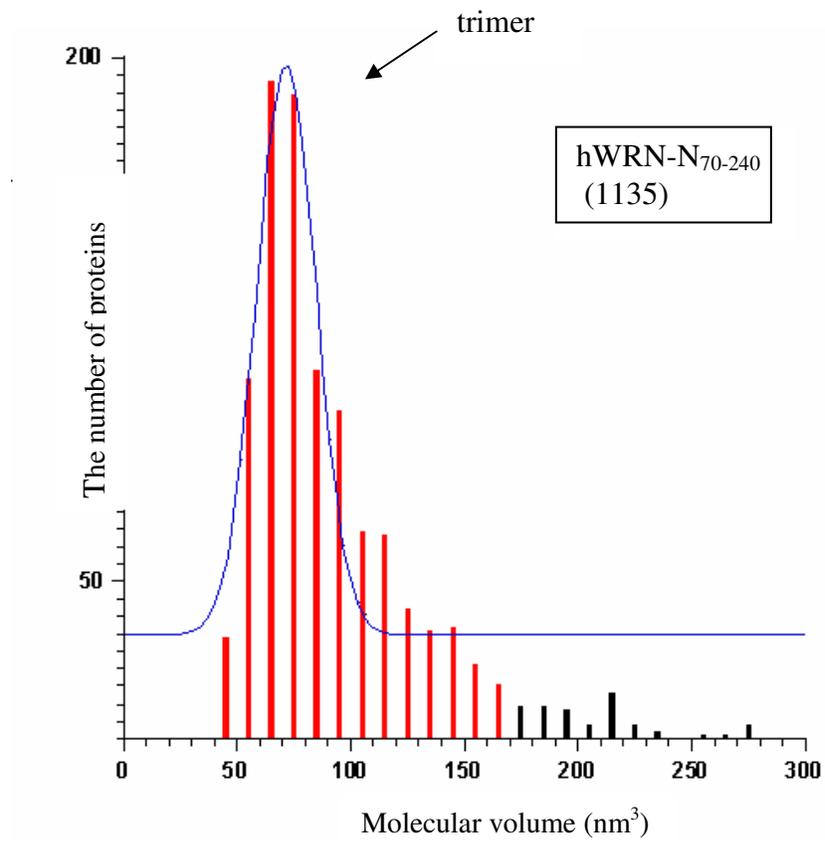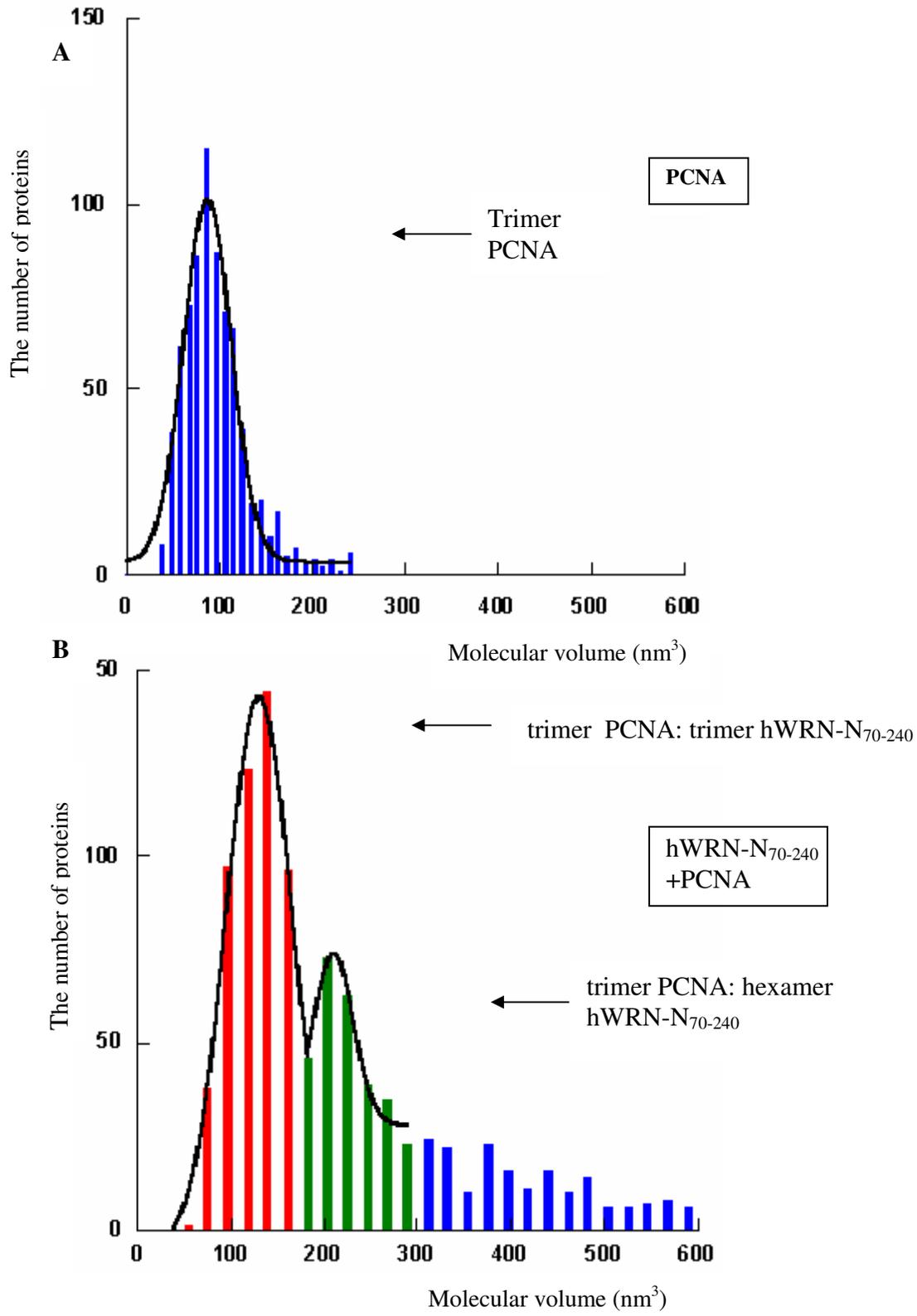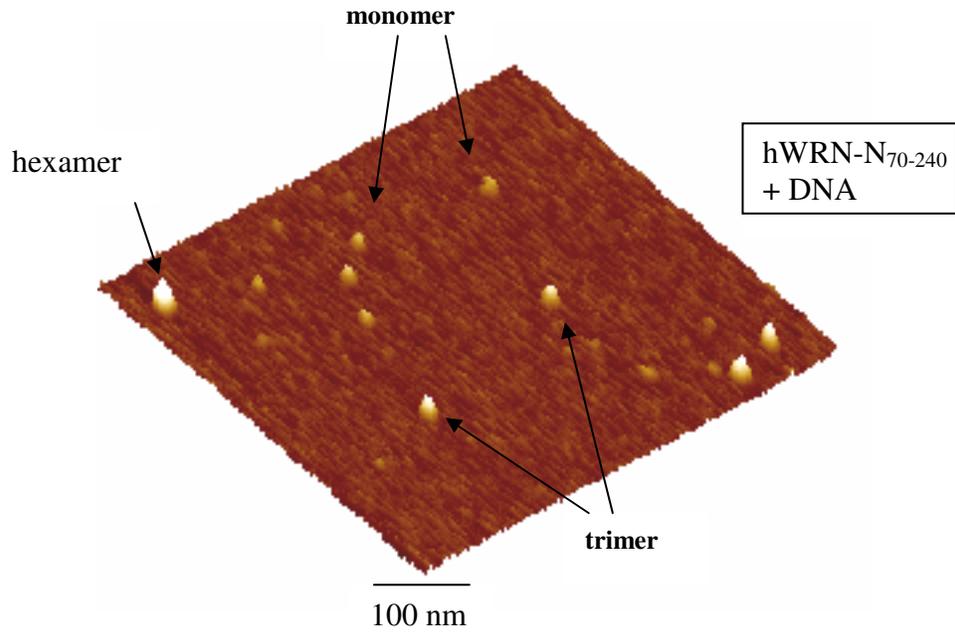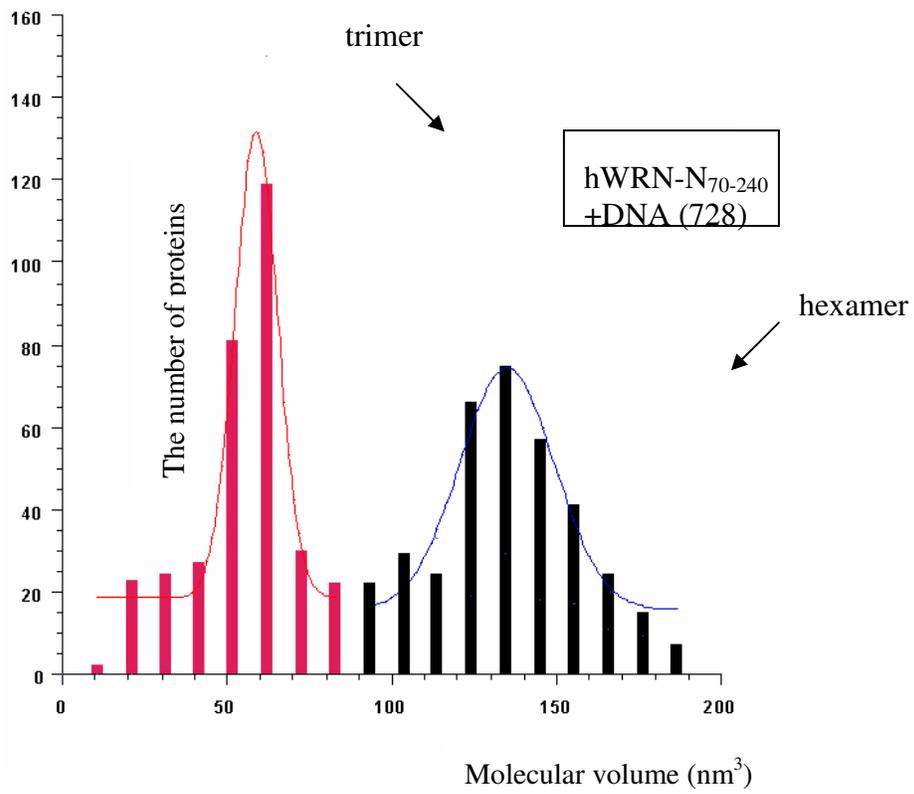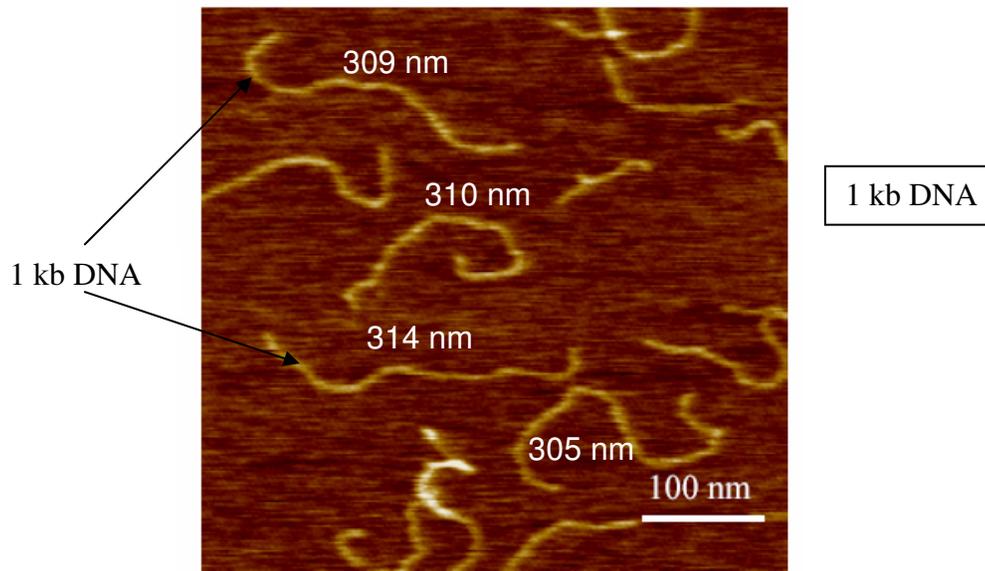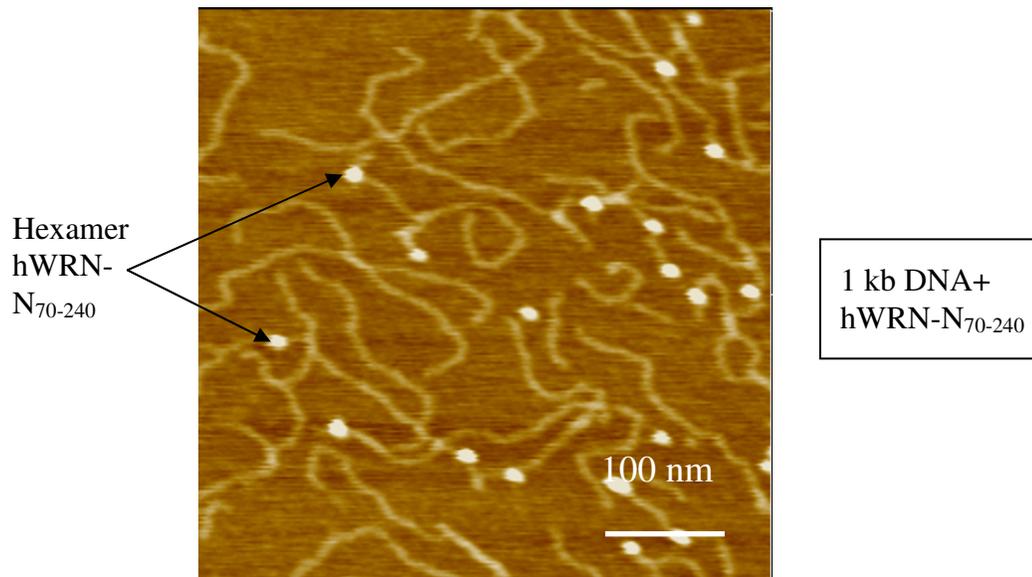
E
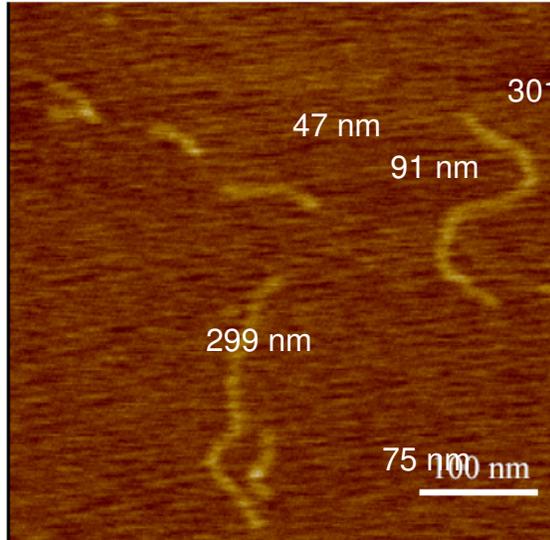
F
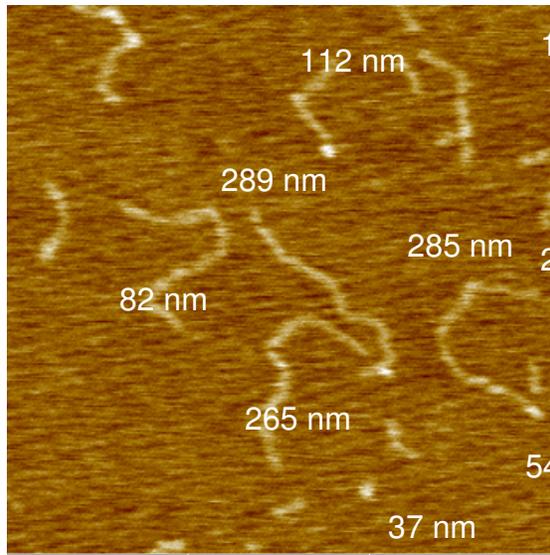
DNA products from incubation of 1kb (310 nm) DNA with hWRN-N$_{70-240}$ for 15 min

Figure 7.

Time

Markers ... Uncut ... Markers

60

36

24

8
6
4
3
2
1

Minutes at 37 °C   1   2   5   10   20   30

24 bp   **36 bases**

```
5'-CAGGCACAGGGTCAGGTCGGGGGG-3'
3'-GTCCGTGTCCCAGTCCAGCCCCCCTTTTTTTTTTTTTGAACAGTCAGTGAGTCTGCCATCT-5'
³²P
```

84
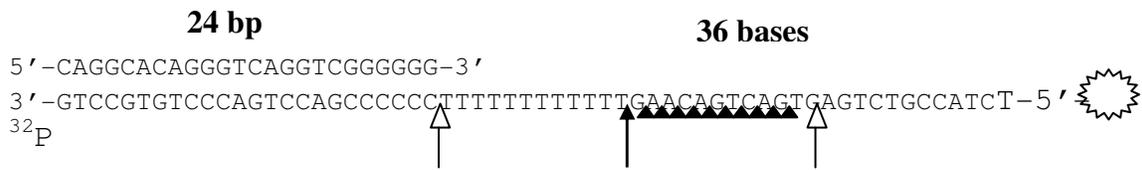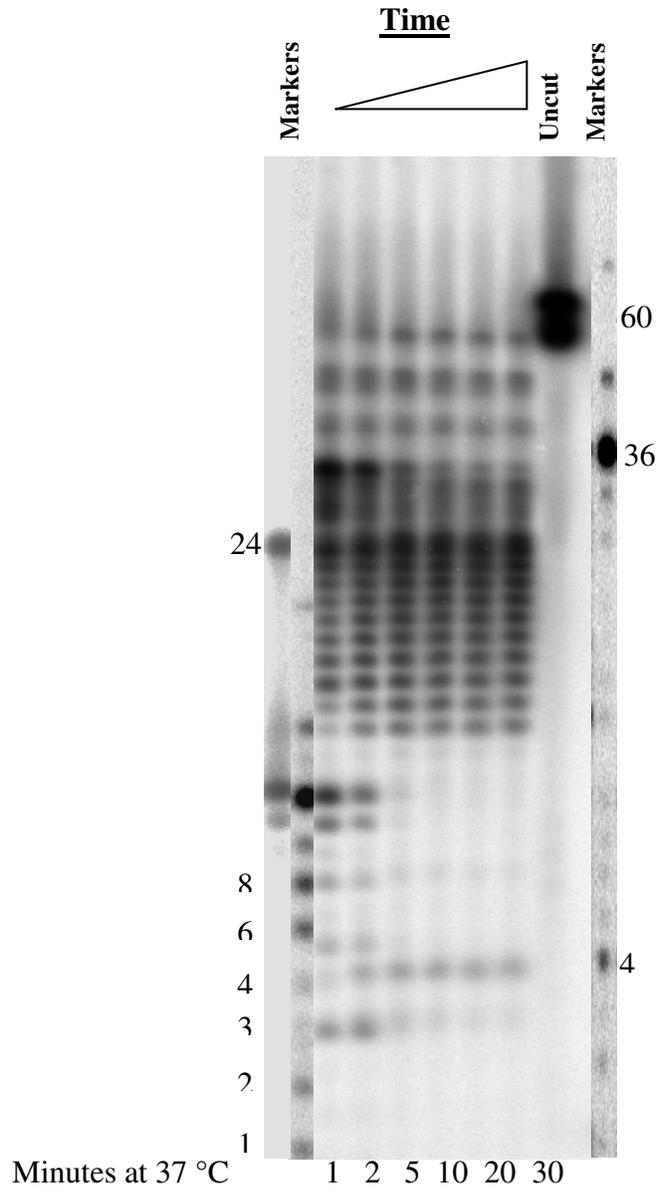
References

1. Goto, M. (1997) *Mechanisms of Ageing & Development 98*, 239-54.

2. Martin, G. M., Oshima, J., Gray, M. D., & Poot, M. (1999) *Journal of the American Geriatrics Society 47*, 1136-44.

3. Fujiwara, Y., Higashikawa, T., & Tatsumi, M. (1977) *Journal of Cellular Physiology 92*, 365-74.

4. Fukuchi, K., Martin, G. M., & Monnat, R. J. (1989) *Proceedings of the National Academy of Sciences of the United States of America 86*, 5893-7.

5. Huang, S., Beresten, S., Li, B., Oshima, J., Ellis, N. A., & Campisi, J. (2000) *Nucleic Acids Research 28*, 2396-405.

6. Kamath-Loeb, A. S., Shen, J. C., Loeb, L. A., & Fry, M. (1998) *Journal of Biological Chemistry 273*, 34145-50.

7. Poot, M., Hoehn, H., Runger, T. M., & Martin, G. M. (1992) *Experimental Cell Research 202*, 267-73.

8. Salk, D., Au, K., Hoehn, H., & Martin, G. M. (1981) *Cytogenetics & Cell Genetics 30*, 92-107.

9. Schulz, V. P., Zakian, V. A., Ogburn, C. E., McKay, J., Jarzebowicz, A. A., Edland, S. D., & Martin, G. M. (1996) *Human Genetics 97*, 750-4.

10. Tahara, H., Tokutake, Y., Maeda, S., Kataoka, H., Watanabe, T., Satoh, M., Matsumoto, T., Sugawara, M., Ide, T., Goto, M., Furuichi, Y., & Sugimoto, M. (1997) *Oncogene 15*, 1911-20.

11. Gray, M. D., Shen, J. C., Kamath-Loeb, A. S., Blank, A., Sopher, B. L., Martin, G. M., Oshima, J., & Loeb, L. A. (1997) *Nature Genetics 17*, 100-3.

12. Huang, S., Li, B., Gray, M. D., Oshima, J., Mian, I. S., & Campisi, J. (1998) *Nature Genetics 20*, 114-6.

13. Shen, J. C., Gray, M. D., Oshima, J., Kamath-Loeb, A. S., Fry, M., & Loeb, L. A. (1998a) *Journal of Biological Chemistry 273*, 34139-44.

14. Suzuki, N., Shimamoto, A., Imamura, O., Kuromitsu, J., Kitao, S., Goto, M., & Furuichi, Y. (1997) *Nucleic Acids Research 25*, 2973-8.

15. Yu, C. E., Oshima, J., Fu, Y. H., Wijsman, E. M., Hisama, F., Alisch, R., Matthews, S., Nakura, J., Miki, T., Ouais, S., Martin, G. M., Mulligan, J., & Schellenberg, G. D. (1996) *Science 272*, 258-62.

16. Ellis, N. A., Groden, J., Ye, T. Z., Straughen, J., Lennon, D. J., Ciocci, S., Proytcheva, M., & German, J. (1995) *Cell 83*, 655-66.

17. Gangloff, S., McDonald, J. P., Bendixen, C., Arthur, L., & Rothstein, R. (1994) *Molecular & Cellular Biology 14*, 8391-8.

18. Nakayama, K., Irino, N., & Nakayama, H. (1985) *Molecular & General Genetics 200*, 266-71.

19. Stewart, E., Chapman, C. R., Al-Khodairy, F., Carr, A. M., & Enoch, T. (1997) *EMBO Journal 16*, 2682-92.

20. Watt, P. M., Louis, E. J., Borts, R. H., & Hickson, I. D. (1995) *Cell 81*, 253-60.

21. Cooper, M. P., Machwe, A., Orren, D. K., Brosh, R. M., Ramsden, D., & Bohr, V. A. (2000) *Genes & Development 14*, 907-12.

22. Orren, D. K., Machwe, A., Karmakar, P., Piotrowski, J., Cooper, M. P. & Bohr, V. A. (2001) *Nucleic Acids Research 29*, 1926-34.

23. Moser, M. J., Holley, W. R., Chatterjee, A., & Mian, I. S. (1997) *Nucleic Acids Research 25*, 5110-8.

24. Takeuchi, F., Hanaoka, F., Goto, M., Akaoka, I., Hori, T., Yamada, M., & Miyamoto, T. (1982) *Human Genetics 60*, 365-8.

25. Shen, J. C., & Loeb, L. A. (2000) *Trends in Genetics 16*, 213-20.

26. Lebel, M., Spillare, E. A., Harris, C. C., & Leder, P. (1999) *Journal of Biological Chemistry 274*, 37795-9.

27. Blander, G., Kipnis, J., Leal, J. F., Yu, C. E., Schellenberg, G. D., & Oren, M. (1999) *Journal of Biological Chemistry 274*, 29463-9.

28. Brosh, R. M., Jr., Orren, D. K., Nehlin, J. O., Ravn, P. H., Kenny, M. K., Machwe, A., & Bohr, V. A. (1999) *Journal of Biological Chemistry 274*, 18341-50.

29. Shen, J. C., Gray, M. D., Oshima, J., & Loeb, L. A. (1998b) *Nucleic Acids Research 26*, 2879-85.

30. Blander, G., Zalle, N., Leal, J. F., Bar-Or, R. L., Yu, C. E., & Oren, M. (2000) *FASEB Journal 14*, 2138-40.

31. Duno, M., Thomsen, B., Westergaard, O., Krejci, L., & Bendixen, C. (2000) *Molecular & General Genetics 264*, 89-97.

32. Pichierri, P., Franchitto, A., Mosesso, P., & Palitti, F. (2000) *Mutation Research 456*, 45-57.

33. Poot, M., Gollahon, K. A., & Rabinovitch, P. S. (1999) *Human Genetics 104*, 10-4.

34. Chaconas, G., Van de Sande, J. H. (1980) *Meth. Enzymol. 65*, 75-88.

35. Ratcliff, G. C., & Erie. D. A. (2001) *Journal of American Chemistry Society 123*, 5632-35.

36. Gulbis, J. M., Kelman, Z., Hurwitz, J., O'Donnell, M., & Kuriyan, J. (1996) *Cell 87*, 297-306.

37. Erie, D. A., Yang, G., Schultz, H. C., & Bustamante, C. (1994) *Science 266*, 1562-6.

38. Kasas, S., Thomson, N. H., Smith, B. L., Hansma, H. G., Zhu, X., Guthold, M., Bustamante, C., Kool, E. T., Kashlev, M., & Hansma, P. K. (1997) *Biochemistry 36*, 461-8.

39. Wyman, C., Rombel, I., North, A. K., Bustamante, C., & Kustu, S. (1997) *Science 275*, 1658-61.

40. Wyman, C., Grotkopp, E., Bustamante, C., & Nelson, H. C. (1995) *EMBO Journal 14*, 117-23.

41. Kovall, R., Matthews, B. W. (1997) *Science 277*, 1824-7.

42. Karow. J. K., Newman, R. H., Freemont, P. S. & Hickson, I. D. (1999) *Current Biology 9*, 597-600.

43. Gulbis, J. M., Kelman, Z., Hurwitz, J., O'Donnell, M. & Kuriyan, J (1996) *Cell 87*, 297-306.

44. Haber, J. E. (1998) *Cell 95*, 583-6.

45. Mirzoeva, O. K. & Petrini, J. H. (2001) *Mol. Cell. Biol. 21*, 281-8.

46. Carney, J. P., Maser, R. S., Olivares, H., Davis, E. M., Le Beau, M., Yates, J. R., 3rd, Hays, L., Morgan, W. F. & Petrini, J. H. (1998) *Cell 93*, 477-86.

47. Paull, T. T. & Gellert, M. (1999) *Genes & Development 13*, 1276-88.

48. Schiestl, R. H., Peters, T. D. (1991) *Proceedings of the National Academy of Science of the United States of America 88*, 7585-89.

49. Li, B. & Comai, L. (2000) *Journal of Biological Chemistry 275*, 28349-52.

50. Li, B. & Comai, L. (2001) *Journal of Biological Chemistry 275*, 9896-902.

51. Liber, M. R. (1999) *Genes Cells 4*, 77-85.

52. Li, X., Li, J., Harrington, J., Lieber, M. R. & Burgers, P. M. (1995) *Journal of Biological Chemistry 27*, 22109-12.

53. Levin, D. S., Mckenna, A. E., Motycka, T. A., Matsumoto, Y., Tomkinson, A. E. (2000) *Current Biology 10*, 919-22.

54. Warbrick, E., Lane, D. P., Glover, D. M., Cox, L. S. (1997) *Oncogene 14*, 2313-21.

55. Li, X., Li, J., Harrington, J., Lieber, M. R. & Burgers, P. M. (1995) *Journal of Biological Chemistry 27*, 22109-12.

56. Chen, V., Chen, S., Saha, P., Dutta, A. (1996) *Proceedings of the National Academy of Science of the United States of America 93*, 11597-602.

57. Levins, D. S., Bai, W., Yao, N., O'Donnell, M., Tomkinson, A. E. (1997) *Proceedings of the National Academy of Science of the United States of America 94*, 12863-8

58. Brosh, R. M., von Kobbe, C., Sommers, J. A., Karmakar, P., Opresko, P. L., Piotrowski, J., Dianova, I., Dianov, G. L., and Bohr, V. A. (2001) *EMBO J.*, *20*, 5791-5801.

STUDY III: Crystal Structure of the PXR-Estradiol Complex Provides Insights into Endobiotic Recognition

(Under the direction of Professor Matt Redinbo)

Abstract

The human nuclear pregnane X receptor (PXR) responds to a wide variety of xenobiotic and endobiotic compounds, including pregnanes, progesterones, corticosterones, lithocholic acids, and 17β-estradiol. In response to these ligands, the receptor controls the expression of genes central to the metabolism and excretion of potentially harmful chemicals from both exogenous and endogenous sources. While the structural basis of PXR's interaction with small and large xenobiotics has been examined, the detailed nature of its binding to endobiotics, including steroid-like ligands, remains unclear. We report the crystal structure of the human PXR ligand binding domain (LBD) in complex with 17β-estradiol, a representative steroid ligand, at 2.65 Å resolution. Estradiol is found to occupy only one region of PXR's expansive ligand binding pocket, leaving a notable 1,000 $Å^3$ of space unoccupied, and to bridge between the key polar residues Ser-247 and Arg-410 in the PXR LBD. Positioning the steroid scaffold in this way allows it to make several direct contacts to αAF of the receptor's AF-2 region. The PXR-estradiol complex was compared to that of other nuclear receptors, including the estrogen receptor, in complexes with analogous ligands. It was found that PXR's placement of the steroid is remarkably distinct relative to other members of the nuclear receptor superfamily. Using the PXR-estradiol complex as a guide, the binding of other steroid- and cholesterol-like molecules was then considered. The results provide detailed insights into the manner in

90

which human PXR responds to a wide range of endobiotic compounds.

Introduction

The pregnane X receptor (PXR; also known as SXR and PAR) plays a central role in xenobiotic detection and the subsequent regulation of genes involved in drug metabolism and excretion (1-4). The receptor has also been shown to respond to key endogenous ligands, including 5-β-prenane-3,20-dione, progesterones, corticosterones, testosterone, prenenolones, lithocholic acids, the steroid-like compound dexamethasone, and 17β-estradiol (4-11). Indeed, PXR was termed the pregnane X receptor because it is activated by a variety of C21 steroids (4). In response to these and other chemicals, PXR regulates the expression of gene products involved in protective cholesterol and bile acid metabolism and processing. PXR is a member of the nuclear receptor (NR) superfamily of ligand-regulated transcription factors and, like most NRs, contains both DNA- and ligand-binding domains (DBD, LBD, respectively) connected by a presumably flexible hinge region(12). PXR functions as an obligate heterodimer with the retinoid X receptor (RXR), and binds to various combinations of direct and everted repeat elements in the regulatory regions of target genes (3). While it does not contain a lengthy activation function 1 (AF-1) region at its N-terminus like other NRs, PXR maintains an intact AF-2 within its LBD, which is stabilized by bound ligand and facilitates the recruitment of transcriptional coactivators (13). Unlike many nuclear receptors, however, PXR exhibits a consistent basal transcriptional activity in the absence of ligand (4, 14).

Crystal structures of the PXR LBD have been reported in its apo (unliganded) state, as well as with small drug-like and herbal ligands (*e.g.,* SR12813, hyperforin) and the large macrolide ligand rifampicin (13-16). These structures have revealed that PXR contains

an expansive and structurally conformable ligand binding pocket capable of changing in shape depending on the nature of its bound ligand. PXR also maintains a ~60-residue insert between α1 and α3 in the well-established LBD fold, which adds a variety of distinct features to PXR including two β-strands that extends the standard two- to three-stranded LBD β-sheet to five strands in PXR (17). The terminal β-strands in these β-sheets interact in an ideal antiparallel fashion in PXR to form a homodimer unique to PXR that has recently been shown to be critical for receptor function (18). It has also been noted that the PXR LBD deviates significantly in sequence across species relatively to other NRs, and these differences have lead to the hypothesis that the PXRs evolved to respond to xenobiotic or endobiotic pressures distinct to each species (4, 19).

While the nature of PXR's interactions with xenobiotics has been well examined structurally, no corresponding structural data exist to date on the interaction of PXR with an endogenous ligand, including steroid-like compounds which appear to be of predominant importance. To address this, we determined the crystal structure of the PXR LBD in complex with 17β-estradiol, a representative steroid, and refined it to 2.65 Å resolution. This structure reveals that PXR positions estradiol in one hemisphere of its ligand binding pocket, bridging key polar regions of the receptor, and directly contacting the AF-2 region. This binding mode is distinct relative to both the estrogen receptor's interaction with the same ligand (20), and the manner in which other NRs contact steroid- or cholesterol-like ligands (21-26). In addition, the data outlined in this paper supports the hypothesis that PXR's evolution has been significantly impacted by the ability to recognize the bile acids unique to different organisms and to play a protective role in

eliminating toxic levels of such compounds. In summary, this structure provides the first scaffold by which the binding of endogenous ligands to PXR can be understood and probed at the molecular level.

Results

Overall Structure. To unravel the structural basis of the recognition of endogenous steroid-like ligands by the nuclear xenobiotic receptor PXR, we determined the crystal structure of the PXR LBD in complex with 17β-estradiol. Repeated attempts to obtain a structure of this complex were hampered by the covalent attachment of the reducing agent dithiothreitol (DTT) to Cys-284 within the ligand binding pocket of PXR, as visualized within several crystal structures (data not shown). Thus, this cysteine side chain was replaced with serine using PCR mutagenesis, which produced a form of the LBD that generated a stable complex with estradiol. Crystals were obtained, and data collected to 2.65 Å resolution at the SER-CAT facility at APS (Argonne, IL). The structure was determined by molecular replacement and refined using torsion angle dynamics to R_{cryst} and R_{free} values of 0.217 and 0.273, respectively (Table I). The PXR LBD conformation is similar to that observed in other PXR-ligand complexes determined to date (13-16), exhibiting the three alpha-helical layers common to nuclear receptor LBDs but with the unique PXR features of an extended β-sheet mediating homodimer formation (Figure 1A). The PXR LBD in the estradiol complex shares 0.5 Å root-mean-square deviation in Cα atom positions with the apo (unliganded) PXR structure (14), and is in the active conformation with regards to its AF-2 surface. Only a few residues adjacent to the ligand binding pocket exhibit small shifts in position between the two structures, including Ser-247 (60° rotation, producing a ~1 Å shift), Cys-284-Ser (60° rotation, ~2 Å shift), Leu-411 (rotamer change for 2.5 Å shift), Met-243 (rotamer change for 1.5 Å shift), and Arg-410 (1 Å shift). Thus, the binding of estradiol to the PXR LBD does not induce large structural changes relative to the unliganded form of the receptor.

95

Estradiol Binding. The use of $|F_{obs}^{Ligand} - F_{obs}^{Apo}|$, ϕ_{calc} maps to identify clearly the positioning of ligands has been effective in past PXR-agonist complex structure determinations (16). Thus, we calculated an $|F_{obs}^{PXR-Est} - F_{obs}^{Apo}|$, ϕ_{calc} map at 2.8 Å resolution; clear electron density at 3σ indicated the position of 17β-estradiol bound within the PXR ligand binding pocket (Figure 1B). It was particularly helpful to observe density for the extracyclic 18-methyl group, which allowed for unambiguous placement of the ligand in a single orientation. Estradiol refined with relatively high thermal displacment parameters (~85 Å^2), although such values have been observed with other ligands bound to PXR (e.g., rifampicin) (15). 17β-estradiol forms hydrogen bonds with two PXR side chains and additional interactions with eight other residues. The 3-hydroxyl group on the steroid A-ring forms a 2.7 Å hydrogen bond with Ser-247, while the 17β-hydroxyl group on the D-ring forms a 2.9 Å hydrogen bond with Arg-410 (Figure 1C). Arg-410 is stabilized by a 2.9 Å hydrogen bond with Ser-208, which itself is 3.5 Å from the oxygen of the ligand's 17β-hydroxyl group. Arg-410 is further stabilized by a 3.0 Å hydrogen bond with the side chain of Glu-321. The side chain of His-407 was placed in two distinct orientations in this structure, and each orientation forms a 3.3 Å van der Waals contact with atoms within the ligand. Thus, polar interactions appear to position the ligand within one region of the ligand binding cavity, which leaves a significant portion of the available room within the pocket unoccupied (Figure 1A).

Non-polar contacts also play a key role in stabilizing 17β-estradiol within the ligand binding pocket of PXR. Phe-420 forms a 4.5 Å edge-to-face contact normal to the conjugated A-ring of the steroid, while the side chain of Phe-429 forms a second edge-to-

face contact more parallel to the A-ring at a distance of 3.5 Å (Figure 1C). The A-ring is further stabilized by 3.4 Å and 3.6 Å van der Waals contacts with Leu-411 and Met-425, respectively, and by a 3.0 Å interaction between the π-orbitals of Phe-251 and the 3-hydroxyl group of the steroid. The B-ring of the steroid is also forms a 3.4 Å van der Waals contact with Met-243. Note that Met-425 and Phe-429, which are directly contacted by estradiol, are located on αAF of the PXR AF-2 surface; thus, these interactions likely help to stabilize the active AF-2 conformation of the receptor. The volume of the human PXR ligand binding pocket without estradiol present is 1,376 $Å^3$, as calculated by the method of CASTp (27). In the presence of the ligand, however, this volume only decreases by 395 $Å^3$, leaving 981 $Å^3$ unoccupied by ligand.

PXR Mutants. To examine the role that individual residues may play in the activation of PXR by steroid-like compounds, single-site mutant forms of the full-length receptor were generated. The activation of a luciferase reporter gene under the control of the cytochrome P450-3A4 promoter was examined in CV-1 cells upon treatment with increasing concentrations of two established PXR agonists, rifampicin and SR12813 (13-15), or 17β-estradiol (Figure 2, Table II). For SR12813 and rifampicin, concentrations between 10 nM and 10 μM were examined. For the weaker agonist estradiol, concentrations up to 100 μM were examined so that EC_{50} values could be calculated (Table II); however, in Figure 2, the complete dose-response curve for estradiol is only shown for the wild-type form of the receptor. Wild-type PXR exhibited the moderate basal activation levels commonly observed for PXR, and EC_{50} values of 0.14, 0.70 and 9.5 μM in response to SR12813, rifampicin, and estradiol, respectively. The binding

affinities of SR12813, rifampicin, and 17β-estradiol to PXR are measured by scintillation proximity assays to be 0.05, 5.2 and 5 μM, respectively (data not shown). Mutating Ser-247, which hydrogen bonds directly to estradiol, to tryptophan produces a form of the receptor that is constitutively active regardless of the presence of agonist. A bulky indole ring in this position likely fills a portion of the ligand binding pocket, and, significantly, would be expected to contact directly Met-425 on the receptor's αAF. Such an interaction would likely stabilize the active conformation of AF-2, leading to reporter gene expression. Recall that in the PXR-estradiol crystal structure, the ligand also directly contacts Met-425. Mutating His-407, which forms van der Waals contacts with estradiol, to glutamine does not significantly impact receptor activity, likely because this polar side chain is expected to form analogous contacts with ligand. It is noted that a slight increase in basal activation of transcription is observed with His-407-Gln forms of full-length PXR; the structural basis of this effect is unclear. Mutating Arg-410, which forms the other direct hydrogen bond with estradiol, with a shorter asparagine side chain does not alter the activation of the receptor by ligands. Asparagine in this position is still capable of forming a hydrogen bond with the 17-hydroxyl group of estradiol. This mutation does eliminate basal, ligand-independent activation of PXR, however. The likely explanation for this effect is that an Arg-410-Asn residue would not be capable for forming a salt bridge with Glu-321. Taken together, these mutations indicate that polar interactions formed between PXR and estradiol play an important role in activation of the receptor by this steroid ligand.

Estradiol Binding by PXR vs. ER. We next sought to compare structurally the binding of estradiol by PXR in relation to the interaction of this steroid hormone to its cognate receptor, the estrogen receptor, another member of the nuclear receptor superfamily. The crystal structure of PXR in complex with 17β-estradiol shares 2.6 Å rmsd over Cα positions with the structure of the estrogen receptor-α (ERα) (20), and the two LBDs share 16% sequence identity. A superposition of the LBDs reveal that, while the overall folds of the proteins are similar, the positions of the bound ligands within the pockets are distinct (Figure 3A). Estradiol in PXR binds closely adjacent to αAF and leaves a portion of the large ligand binding pocket in this receptor unoccupied. In contrast, estradiol is more centrally located in the ERα pocket, is oriented nearly perpendicular to that observed in the PXR complex, and occupies all but ~56 Å^3 of the space in the central pocket. ERα forms three hydrogen bonding and twelve non-polar contacts with estradiol, while PXR forms only two hydrogen bonds and eight van der Waals interactions (Table III). The 3-hydroxyl group on the A-ring of the steroid forms two hydrogen bonds in the ERα complex, with Glu-353 and Arg-394; in PXR, these residues are replaced with the hydrophobic residues Met-250 and Val-291, respectively (Figure 3B). The A-ring is further contacted in ERα by an edge-to-face interaction with Phe-404, which is replaced by Cys-301 in PXR. Thus, contacts analogous to those observed between ERα and the A-ring of estradiol are not possible in PXR. ERα forms a hydrogen bond between His-524 and the 17-hydroxyl group in estradiol; His-524 is replaced by Arg-410 in PXR, which a similar interaction in the PXR-estradiol complex. The D-ring is further contacted by Met-421 in ERα, which is equivalent to Ser-208 in PXR. In summary, only one of three hydrogen bonds formed by ERα with estradiol is conserved in PXR, and

99

only three of twelve van der Waals contacts are maintained (Table III). Indeed, key residues involved in PXR's contact with estradiol are replaced by significantly different side chains in ERα as well. For example, the PXR residues Ser-247, His-407, Phe-420, Phe-429 and Phe-251 are equivalent to alanine, glycine, valine, and two leucine residues, respectively, that do not form analogous contacts in ERα (Figure 3B; Table III). Significantly, however, both complexes with estradiol reveal that direct contacts are formed between the ligand and amino acid side chains on the αAF helices within the AF-2 region (Met-425 in PXR, Leu-540 in ERα), providing a molecular explanation for the ability of these agonists to facilitate gene activation by the receptors.

Insights into PXR Activation by Other Endogenous Ligands. PXR is well established as both a xenobiotic and endobiotic sensor. Endogenous ligands known to activate human PXR-mediated gene expression include pregnanes, progesterone, corticosterone, testosterone, lithocholic acids, and the steroid-like xenobiotic dexamethasone (1, 4, 5, 7-9, 11). We next sought to examine the basis of PXR activation by endogenous compounds by using as a guide the PXR-estradiol structure, the first determined for a steroid scaffold ligand in complex with this receptor. As summarized above, estradiol bridges between two polar regions within the PXR pocket, represented by Ser-247 and Arg-410 (Figure 4). Arg-410 is also proximal to the surface of the receptor, and to a series of other polar side chains that are either adjacent to (*e.g.,* Ser-206 at 3.5 Å) or more distant from estradiol, such as Arg-413, Asp-205, Lys-204, and Asp-230 (Figure 4). Using binding modes analogous to estradiol as a guide, a series of endogenous compounds that are known PXR activators were modeled into the receptor's ligand

binding pocket. Significantly, all compounds capable of maintaining contacts to Ser-247 and the Arg-410 regions were established PXR agonists. For example, 5-β-pregnane-3,20-dione, a ligand for which PXR (the pregnane-X-receptor) is named (4), is capable of forming hydrogen bonds with both Ser-247 and Arg-410. Indeed, particular ligands with 3-keto groups, regardless of the conjugated state of their A-rings, appeared capable of receiving hydrogen bonds from Ser-247 (Figure 4). In the cases of corticosterone, cortisol, 3-keto-lithocholic acid, and 3-keto-7α,12α-dihydroxy-5α-cholanic acid, additional favorable contacts were also likely. For example, the polar side chain of His-407 appeared ideally positioned to interact with hydroxyl or keto oxygens at either the 11 or 12 positions on the steroid scaffolds (see corticosterone, cortisol, and 3-keto-7α,12α-dihydroxy-5α-cholanic acid in Figure 4). In addition, extended side chains at the 17 positions of the steroid are also observed to make favorable contacts with polar residues adjacent to Arg-410, such as those shown for dexamethasone-t-butylacetate and 3-keto-7α,12α-dihydroxy-5α-cholanic acid. 3-β-acetate moieties were also found to be capable of forming polar interactions with Ser-247, which explains the activation of PXR by lithocholic acid acetate and lithocholic acid acetate methyl ester (Figure 4). Similarly, 3-β- and 3-α-hydroxyl groups appeared appropriately positioned to hydrogen bond to Ser-247, helping to explain the agonist character of prenenolones and DHEA (Figure 4). Certain bile acids identified as activators of human PXR were also found to be capable of spanning the distance between Ser-247 and the polar residues adjacent to Arg-410, including glycolithocholic acid and taurolithocholic acid. Taken together, the data presented here provide a structural framework in which endogenous, steroid-like molecules may be understood at the molecular level to be ligands for human PXR.

Discussion

The human PXR LBD, which has been shown to respond to a variety of both endogenous and xenobiotic compounds, shares structural similarity with the NR superfamily LBDs (28), including that of ER. However, in this report, we show that PXR binds to ER's endogenous ligand, 17β-estradiol, in a manner markedly distinct from the estrogen receptor. While estradiol fills the ligand binding pocket of ER, it binds adjacent to the αAF of the AF-2 region in PXR, leaving a significant portion of PXR's large binding pocket unoccupied. Several structural features within each receptor lead to this difference in binding orientation. The central β-strand (β2) in ER's three-stranded β-sheet is dips deeper into the ligand binding pocket in this receptor relative to PXR, and positions Phe-404 such that it can form an edge-to-face contact with the bound estradiol molecule (20). ER also contains a longer α6 and no equivalent to the α2 and β1/ β1' secondary structural elements of PXR; together, these features allow ER to place a large number of side chains snugly within the ligand binding pocket to contact estradiol. In addition, while PXR's pocket is more expansive than that of ER, key estradiol-contacting residues in ER are replaced in PXR. For example, Glu-353 and Arg-394 of ER correspond to Met-250 and Val-291 in PXR, respectively, neither of which is capable of forming central polar contacts to the steroid ligand. Phe-288 and His-407 in PXR, which replace Leu-391 and Gly-521 in ER, respectively, would also clash at ~1 Å from estradiol bound as it is observed in ERα. Taken together, these observations indicate that key structural differences in the LBD folds of ER and PXR allows these two NRs to bind 17β-estradiol in remarkably distinct manners, revealing the plasticity present in the common NR fold.

Several other nuclear receptors bind to steroid-like ligands, including the glucocorticoid receptor (GR), the liver X receptor (LXR) and the constitutive androstane receptor (CAR). We compared the PXR-estradiol complex structure with those of GR-dexamethasone (21), LXR-25-epoxycholesterol (25) and CAR-5β-pregnane-3,20-dione (26) (Figure 5). Similar to ER-estradiol, the GR and LXR complexes reveal that the steroid ligand is bound nearly perpendicular to that observed in PXR. The contacts leading to these binding modes are similar to that described above for ERα, including the packing of several side chains into the ligand binding pockets of GR and LXR to generate receptor-ligand interactions. Recall that GR and LXR are both highly specific for steroid-like ligands, whereas PXR exhibits a much broader ligand-binding profile. In this respect, it was interesting to observe that CAR, another receptor that responds somewhat promiscuously to ligands, positions 5β-pregnane-3,20-dione in a manner somewhat similar to the placement of estradiol in PXR. However, in CAR, 5β-pregnane-3,20-dione does not directly contact αAF of the receptor's AF-2 domain; instead, Asn-165 and Tyr-326, which interact with the ligand, mediate contacts to αAF. In PXR, these residues are replaced by Ser-247 and His-407, respectively, which form both key interactions with estradiol and also facilitate the positioning of the ligand in direct contact with αAF. It is expected that CAR would not respond to estradiol because the planar A-ring of this steroid ligand would clash with aromatic side chains within the CAR binding pocket (26). The farnesoid X receptor (FXR), which responds to bile acids, was shown previously to orient 6-ethyl-chenodeoxylchlolic acid similar to that observed for dexamethasone bound GR, with the distinction that the A-ring faces the opposite direction in FXR (22). Again, this position is distinct relative to the observed estradiol docking adjacent to αAF in the

PXR complex presented here; indeed, the position of the loop after α1 in PXR, which deviates by up to 9 Å relative to the same region in FXR, makes the analogous interaction of bile acids to PXR impossible (not shown). Taken together, these observations highlight the distinct nature of PXR's interaction with steroid-like ligands relative to other members of the NR superfamily.

It has been proposed that bile acids played a key role in driving the evolution of the PXRs in a variety of organisms (8). Indeed, PXR has been shown to respond to toxic levels of bile acids and to coordinate the expression of protective gene products capable of metabolizing and excreting bile acids that may accumulate in hepatocytes (5, 29-38). Schuetz and colleagues recently examined the activation by bile acids of a wide variety of PXRs and PXR-like receptors, including those from human, mouse, rat, dog, cow, pig, rhesus, rabbit, chicken (CXR), the zebrafish PXR, and *Xenopus* BXR, a benzoate receptor (8). They concluded that, in terms of responses to endogenous bile acids present in each organism, the receptors could be grouped in the following fashion: human; dog, pig, rhesus (and, likely, cow); mouse and rat; rabbit; CXR; BXR; zebrafish PXR. Using the human PXR-estradiol complex as a guide, we examined the conservation of key residues in the human PXR ligand binding across these orthologous receptors (Table IV). We found that as the receptors deviate in their responses to distinct bile acids, the conservation of central residues within the binding pocket clearly diminishes. Thus, our structural data support the conclusion that differences in the nature of endogenous bile acids may have played a significant role in the evolution of the PXRs and PXR-like receptors in a wide variety of organisms.

In summary, the data presented here reveals that PXR interacts with the steroid-like molecules by positioning them such that they contact Ser-247, Arg-410, and αAF of the receptor's AF-2 surface. As noted in Figure 2, PXR has a relatively robust basal activation level, which is then significantly enhanced in the presence of ligands. The binding of a ligand like estradiol with a moderate EC_{50} in close proximity to the AF-2 surface likely allows PXR to enhance transcriptional activation without the need for high ligand binding affinity. The placement of steroid-scaffold in PXR is distinct relative to other nuclear receptor-steroid interactions, including that of GR, LXR and the CAR receptor, which shares some overlapping biological functions with PXR. This structure provides a framework by which the interactions of other endogenous steroid- and cholesterol-like molecules can be understood, and their potential physiological roles probed both *in vitro* and *in vivo*.

Materials and Methods

Protein Expression and Purification. The PXR LBD expression construct was engineered as an N-terminal polyhistidine tagged fusion protein with residues 130-434 from the human PXR. The fusion insert was subcloned into the pRSETA expression vector (Invitrogen). Cys-284 within the PXR ligand binding pocket was mutated to serine, to prevent oxidation with DTT during crystallization, using the QuikChange mutagenesis kit (Stratagene, La Jolla, CA) according to the manufacturer's instructions. The mutant was confirmed by sequence analysis. Residues 623-710 of the human SRC-1 gene were subcloned into the bacterial vector pACYC184 along with a T7 promoter (39). The PXR LBD/pRSETA and the SRC-1/pACYC184 plasmids were co-transformed into the BL21(DE3*)* strain of *E. coli*. Ten-liter shake flask liquid cultures containing standard Luria-Bertani (LB) broth with 0.1 mg/ml ampicillin and 0.034 mg/ml chloramphenicol were inoculated and grown at 22 ^{o}C for 20 hours. The cells were harvested by centrifugation (20 minutes, 3500 g, 4 ^{o}C) and the cell pellet was stored at -80°C. The cell pellet was resuspended in 100 ml Buffer A (50 mM Tris-Cl pH7.8, 250 mM NaCl, 50 mM Imidazole pH7.5, 5% glycerol). Cells were sonicated for 3-5 minutes on ice and the cell debris was removed by centrifugation (90 minutes, 20,000g, 4 ^{o}C). The cleared supernatant was loaded on to a 50 ml ProBond nickel-chelating resin (Invitrogen). After washing to baseline with Buffer A, the column was washed with Buffer B containing 75 mM imidazole pH 7.5 and Buffer C containing 75 mM imidazole pH 7.5 and 50 mM NaCl. The PXR LBD/SRC-1 complex was eluted from the column using Buffer D with 250 mM imidazole pH 7.5 and NaCl 50 mM. Column fractions were pooled and subjected to SP cation exchange column (BioRad) pre-equilibrated with buffer containing

20mM Tris-Cl pH 7.8, 50 mM NaCl, 5 mM DTT, 2.5 mM EDTA pH 8.0, 5% glycerol. The column was washed to baseline with the same buffer and fractions containing the PXR/SRC-1 complex were eluted at 400 mM NaCl and pooled and diluted two fold with the dilution buffer (20 mM Tris-Cl pH 7.8, 5mM DTT, 2.5mM EDTA pH 8.0, 5% glycerol). The protein was concentrated using Centri-prep 30K (Amicon) units.

Crystallization. The human PXR ligand-binding domain/SRC-1 complex (hPXR-LBD/SRC-1) was concentrated in the presence of 100-fold molar excesses of 17 β-Estradiol to a final concentration of 4 mg/mL. Crystallization was achieved by hanging-drop vapor diffusion against the following conditions at 22°C: 50 mM imidazole at pH 7.1, 10% 2-propanol, v/v.

Data Collection and Structure Determination. The structure of the Estradiol-bound form of the ligand binding domain of human PXR was determined by molecular replacement using the crystal structure of the apo (unliganded) PXR as a search model (14). Rotation and translation function searches were performed using AMoRe (40); clear solutions for each were obtained in the proper space group, $P4_32_12$. The structure was refined using the torsion angle protocol in CNS with the maximum likelihood function as a target, and included an overall anisotropic B-factor and a bulk solvent correction (41). 10% of the observed data were set aside for cross-validation using the free-R statistic prior to any structural refinement (42). Manual adjustments and rebuilding of the model were performed using O (43) and σA-weighted electron density maps (44). The structure exhibits good geometry (Table I) with no Ramachandran outliers.

Transient Transfections. Mutant forms of full-length PXR were generated using the QuikChange mutagenesis kit (Stratagene, La Jolla, CA) according to the manufacturer's instructions. All mutants were confirmed by sequence analysis. Transient transfection and reporter gene assays were performed as described previously (14-16). CV-1 cells were plated in 96-well plates in phenol red-free Dulbecco's modified Eagle's medium containing high glucose and supplemented with 10% charcoal/dextran treated fetal bovine serum (HyClone, Logan, UT). Transfection mixes contained 5 ng of receptor expression vector, 20 ng of reporter plasmid, 12 ng of β-actin SPAP as internal control, and 43 ng of carrier plasmid. Plasmids for wild-type and mutant forms of human PXR and for the XREM-CYP3A4-LUC reporter, containing the enhancer and promoter of the *CYP3A4* gene driving Luciferase expression, were as previously described (14-16). Transfections were performed with LipofectAMINE (Life Technologies, Inc., Grand Island, NY) essentially according to the manufacturer's instructions. Drug dilutions of estradiol (Sigma, St. Louis, MO) and SR12813 (synthesized in-house) were prepared in phenol red-free Dulbecco's modified Eagle's medium/F-12 medium with 15 mM HEPES supplemented with 10% charcoal-stripped, delipidated calf serum (Sigma, St. Louis, MO) which had previously been heat-inactivated at 62 °C for 35 minutes. Serial drug dilutions were performed in triplicate to generate 11-point concentration response curves. Cells were incubated for 24 hours in the presence of drugs, after which the medium was sampled and assayed for alkaline phosphatase activity. Luciferase reporter activity was measured using the LucLite assay system (Packard Instrument Co., Meriden, CT) and normalized to alkaline phosphatase activity. EC_{50} values were determined by standard methods.

Table I. Crystallographic Statistics for the PXR-Estradiol Complex

Resolution (Å; highest shell)	50-2.65 Å (2.74-2.65)
Space Group	$P4_32_12$
Asymmetric Unit	one molecule
Cell Constants (Å, °)	a = b = 90.9
	c = 84.8
	$\alpha = \beta = \gamma = 90$
Data collection facilities	APS, SER-CAT
Total Reflections	145,600
Unique Reflections	10,736
Mean Redundancy	13.4 (9.5)
R_{sym}* (%; highest shell)	7.0 (21.9)
Wilson B factor (A^2)	50.1
Completeness (%; highest shell)	99.2 (96.7)
Mean I/σ (highest shell)	37.3 (5.0)
R_{cryst}† (highest shell)	21.7 (25.6)
R_{free}‡ (highest shell)	27.3 (35.4)

Table II. EC$_{50}$ Values for Activation by Rifampicin or 17β-Estradiol.

Full-length PXR	Rifampicin	17β-Estradiol
Wild-Type	1.2 ± 0.22 μM	22 ± 3.2 μM
Ser-208-Ala	0.58 ± 0.24 μM	15 ± 6.1 μM
Ser-247-Ala	0.56 ± 0.33 μM	>100 μM
Cys-284-Ser	0.95 ± 0.48 μM	14 ± 7.1 μM
Arg-410-Leu	2.3 ± 1.1 μM	15 ± 4.5 μM
Met-425-Ala + Phe-429-Ala	N.A.*	N.A.*

[*]The Met-425-Ala + Phe-429-Ala variant of full-length PXR exhibited no response to ligands (see Figure 2); thus, no EC$_{50}$ values were calculated for this form of the receptor.

Table III. Comparison of Amino Acids Lining the PXR and ERα Ligand Binding Pockets. Residues listed in bold make polar contacts with 17β-estradiol, while those in parentheses are proximal to the binding pocket but are not observed to contact the ligand.

PXR Residues	Equivalent ERα Residues
(Ser-208)	Met-421
Met-243	Leu-346
Ser-247	Ala-350
(Met-250)	**Glu-353**
Phe-251	(Leu-354)
(Ala-280)	Trp-383
(Phe-281)	Leu-384
(Gln-285)	Met-388
Phe-288	Leu-391
(Val-291)	**Arg-394**
Glu-321	(Asp-426)
(Met-323)	Leu-428
His-407	(Gly-521)
Arg-410	**His-524**
Leu-411	Leu-525
Phe-420	(Val-534)
Met-425	Leu-540
Phe-429	(Leu-544)
(Cys-301)	Phe-404
None[1]	Ile-424

[1]α7 in PXR is one turn shorter at its N-terminus relative to the equivalent helix in ERα; thus, no analogous residue exists in PXR.

111

Table IV. Comparison of Ligand-Binding Pocket Amino Acids Among NRs Related to PXR. Residues in bold are important in contacting 17β-estradiol in the human PXR-estradiol complex reported here. Underlined residues represent a moderate changes, while highlighted residues represent major changes, relative to PXRs isoforms close in function to human PXR. "WY" in the last row corresponds to W223 and Y225, which residue on the sequence insert novel to isoforms close to human PXR and have been shown to be critical to formation of the PXR homodimer and to PXR function.

Residue (Human)	Human	Pig	Cow	Rhesus	Dog	Rabbit	Mouse	Rat	CXR	BXR	Zebrafish
205	D	D	D	D	D	E	D	D	--	--	S
208	**S**	**S**	**S**	**S**	**S**	**T**	**P**	**P**	**S**	**--**	**S**
243	M	I	I	M	M	L	L	L	F	I	F
247	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**V**	**T**
251	F	F	F	F	F	F	F	F	I	I	I
281	F	F	F	F	F	L	F	F	L	A	F
321	E	Q	E	E	E	D	D	D	E	E	D
407	H	H	H	H	H	H	Q	Q	N	H	Y
410	**R**	**R**	**Q**	**R**	**K**	**R**	**Q**	**Q**	**Q**	**Q**	**Q**
411	L	L	L	L	L	L	L	L	I	L	V
413	R	R	R	R	R	R	R	R	H	E	K
420	F	F	F	F	F	F	F	F	M	D	E
425	M	M	M	M	M	M	M	M	L	M	W
WY	WY	WY	WY	WY	QY	WY	WY	WY	--	-Q	DE

Figure Legends

Figure 1. A. Crystal structure of the ligand binding domain of human PXR (PXR LBD, in red, green, grey) in complex with 17β-estradiol (cyan). Note the proximity of the estradiol ligand to αAF in the activation function-2 (AF-2) region of PXR.

B. Electron density from an $|F_{obs}^{PXR-Est} - F_{obs}^{Apo}|$, ϕ_{calc} map at 2.8 Å resolution and contoured at 3σ for 17β-estradiol within the ligand binding pocket of PXR.

C. 17β-estradiol forms hydrogen bonds with Ser-247 and Arg-410 (bold labels) in the PXR ligand binding pocket, as well as van der Waals contacts with several addition residues. The side chain of Arg-410 is also stabilized by a hydrogen bond with Ser-208. Note that the ligand contacts two residues, Met-425 and Phe-429 (labeled in red), located on αAF of the PXR AF-2 surface. The view in this figure is nearly normal to the plane of the estrogen ring system, and rotated roughly 90° about the vertical axis relative to Figure 1A.

Figure 2. Transient transfections in CV-1 cells of wild-type (WT) and mutant forms of full-length PXR. Responses of each form of the receptor to SR12813, rifampicin and 17β-estradiol were measured. Dose responses for estradiol up to 100 μM were measured for all forms of the receptor, but are shown here only for WT PXR.

Figure 3. A. Superposition of the ligand binding domains of PXR (red) and the estrogen receptor-α (ERα, green) in complexes with 17β-estradiol (cyan, yellow, respectively), viewed in the same orientation as Figure 1A.

113

B. Stereoview of ligand binding pockets of PXR and ERα (red, green, respectively) and the interactions they make with 17β-estradiol (cyan, yellow, respectively). The view is the same as that in Figure 3A.

Figure 4. Analysis of the potential interactions between PXR and other endogenous ligands (as well as dexamethasone) predicted by the PXR-estradiol complex reported here. A schematic view of the proximity of key side chains adjacent to estradiol is shown at top (S247, R410, S208), along with several residues near the 17-OH moiety. Known PXR agonists with 3-keto, 3-β-acetate, 3-β-hydroxy, and 3-α-hydroxy groups are all predicted to form favorable interactions with S247; as shown, further favorable interactions may be formed with additional PXR residues, including H407, R413, D205, and K204.

Figure 5. Superpositions of the PXR-Estradiol structure on that of the GR-Dexamethasone (DEX) complex (*top*), the LXR-25-epoxycholesterol (eChol) complex (*middle*), and the CAR-5β-pregnane-3,20-dione (pregnane) complex (*bottom*). The view is the same as that shown in Figure 4, and focusing on the ligand-binding pockets of the receptors.
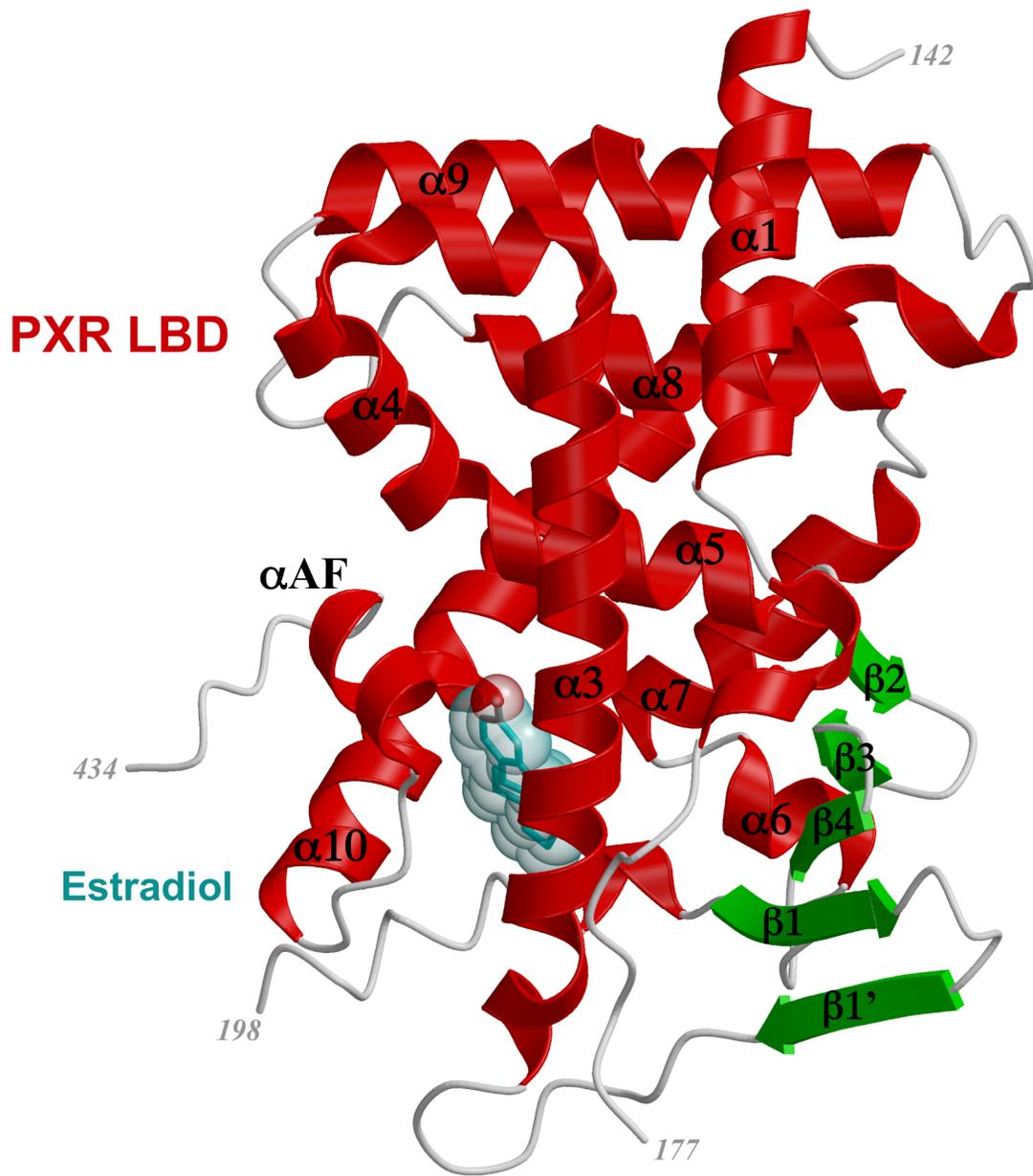
Figure 1A.

Figure 1B.
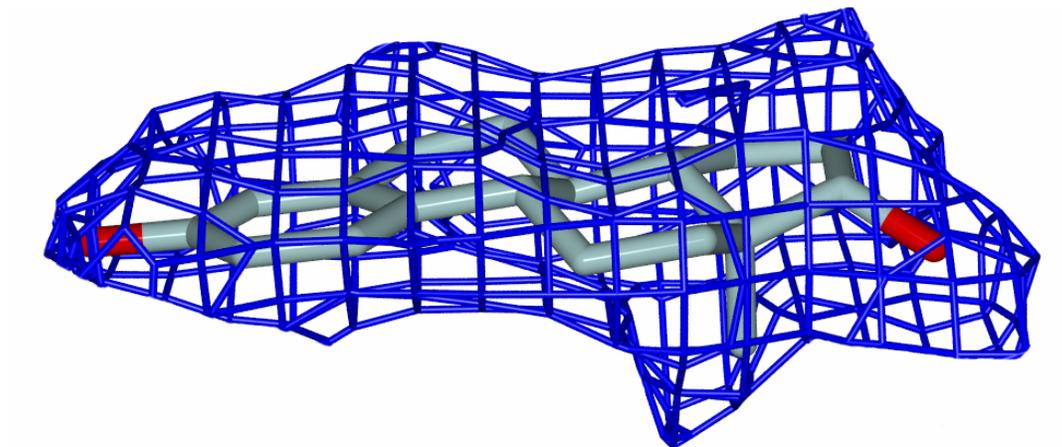
Figure 1C.
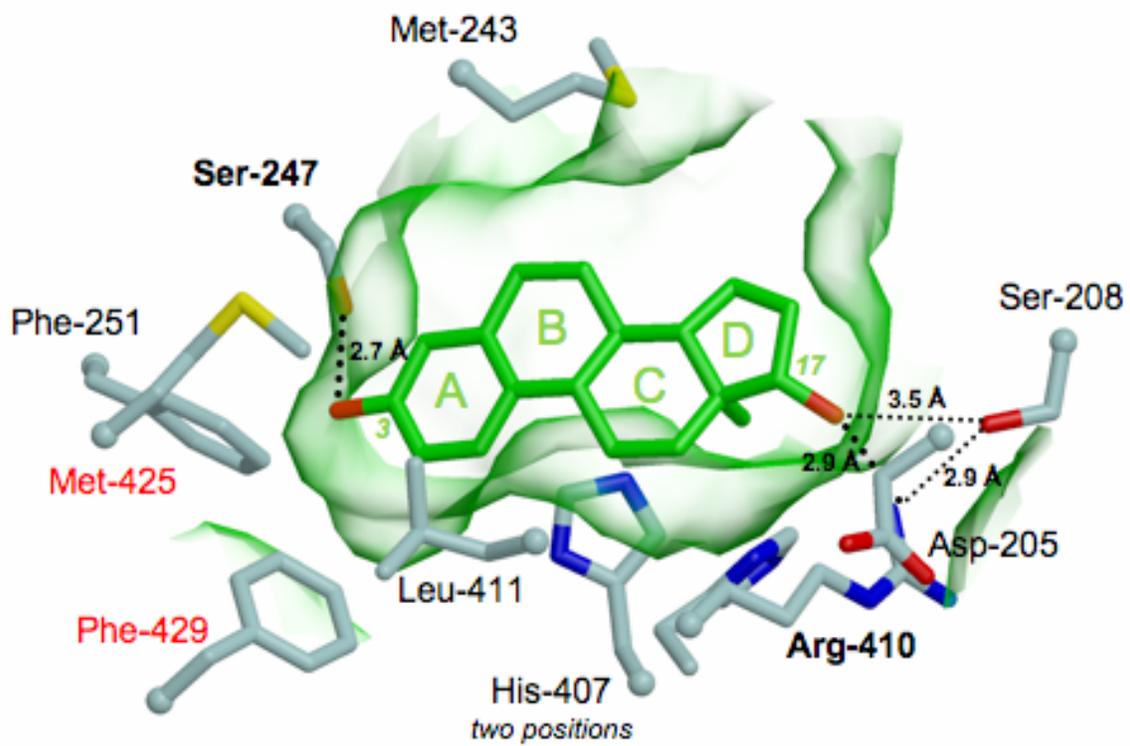
Figure 2.
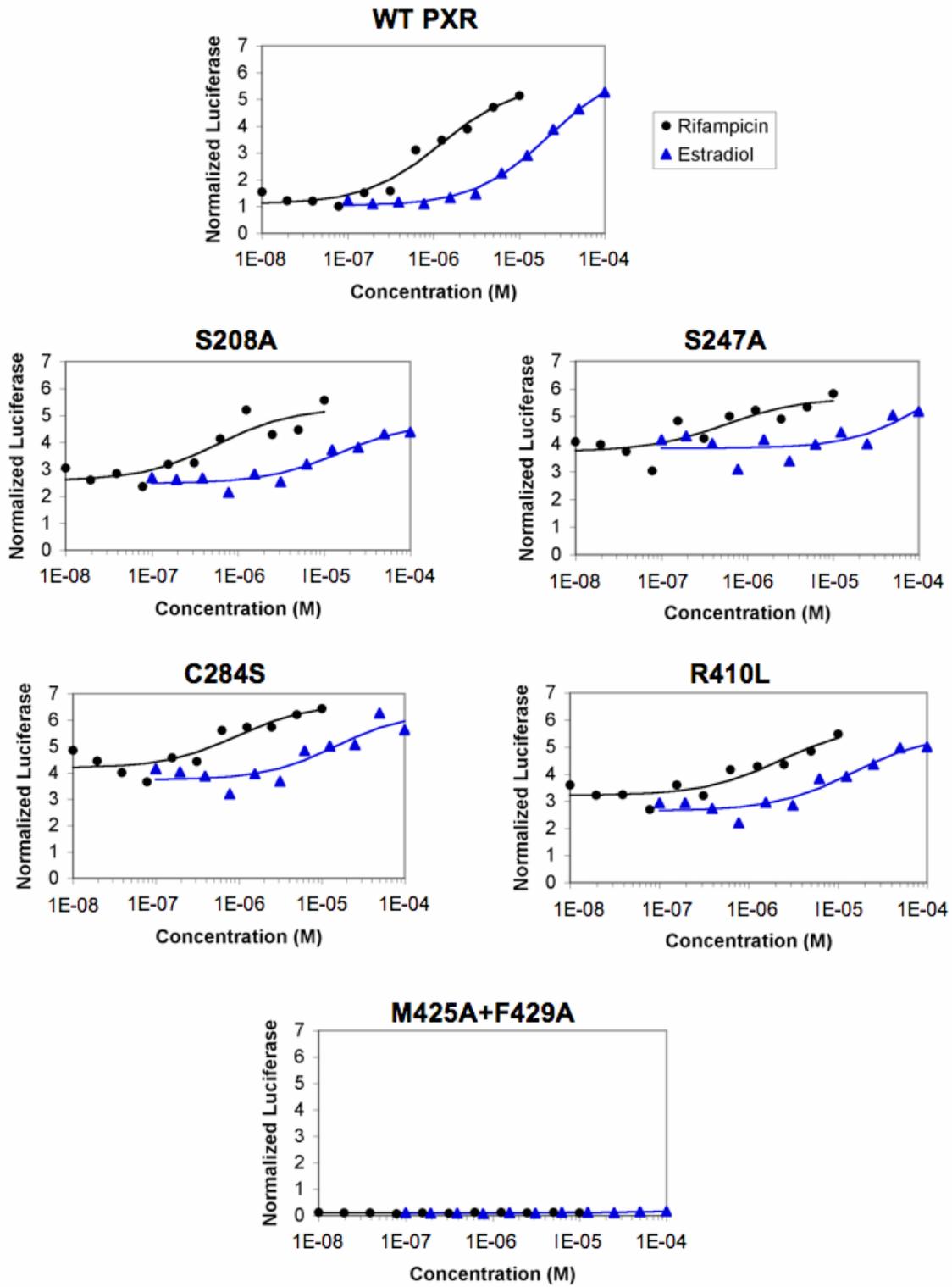
Figure 3A.
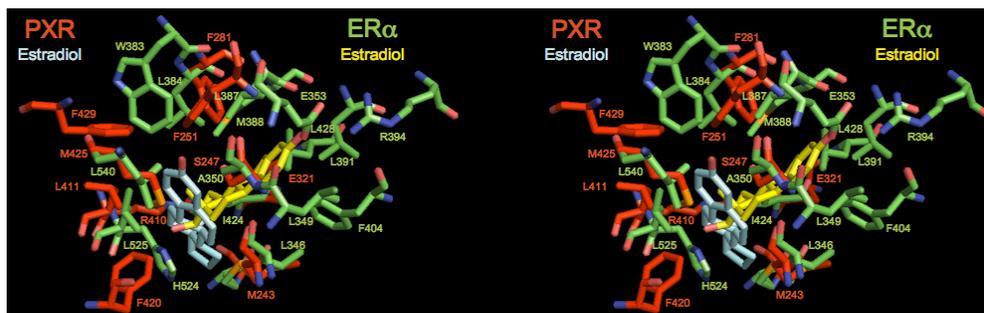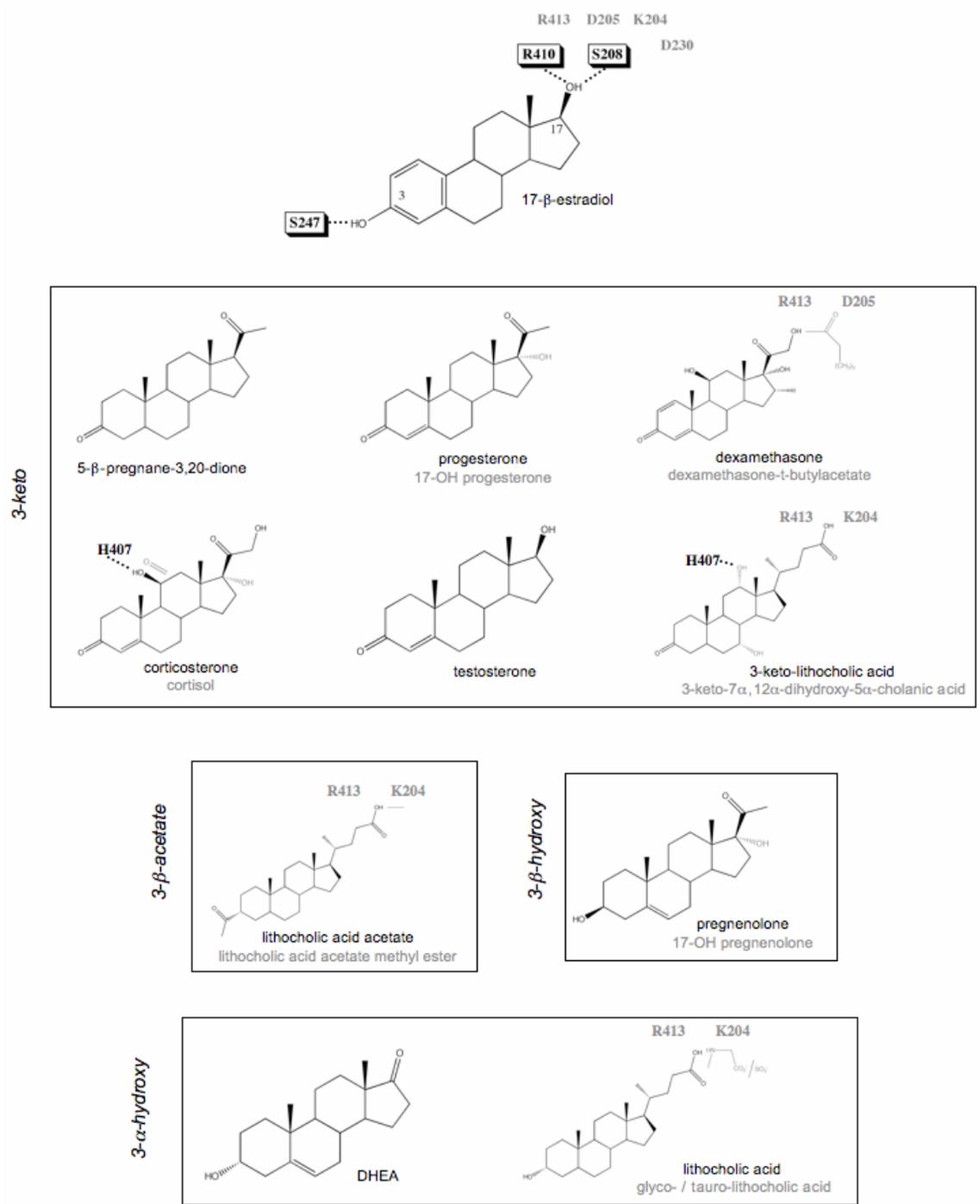
Figure 3B.

Figure 4.

Figure 5.
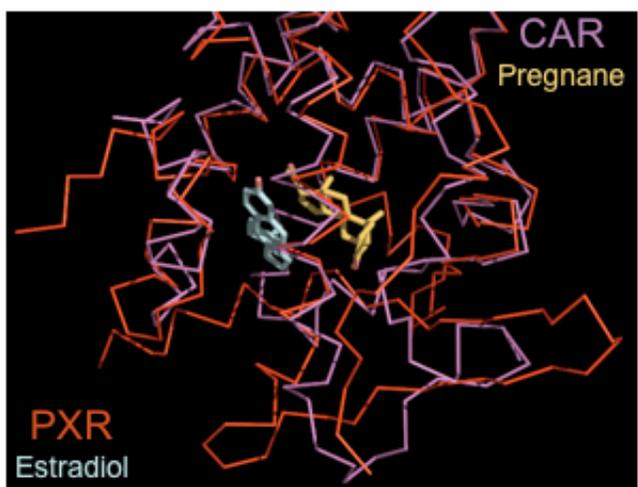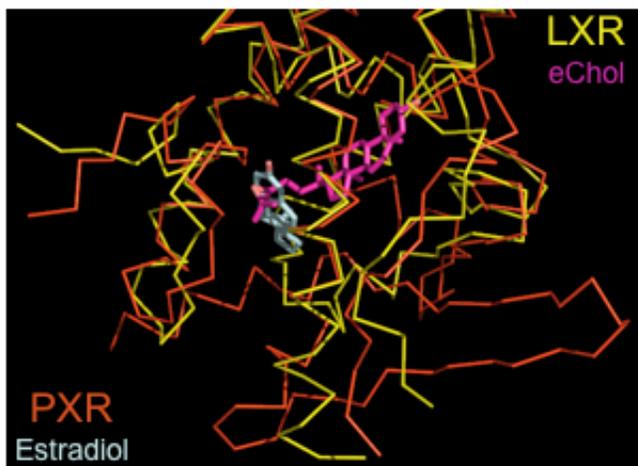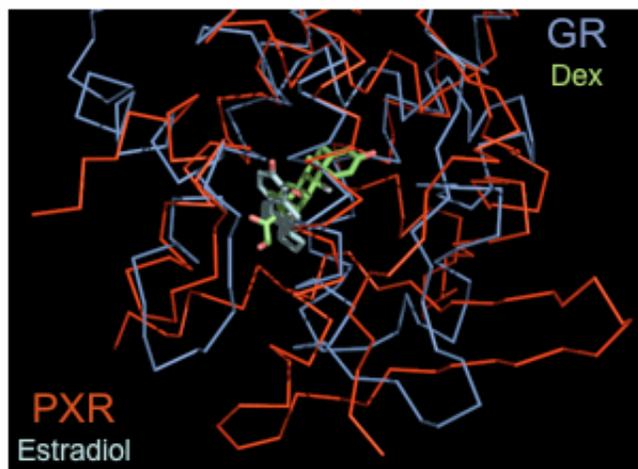
References

1.   Bertilsson G, Heidrich J, Svensson K, Asman M, Jendeberg L, Sydow-Backman M, Ohlsson R, Postlind H, Blomquist P, Berkenstam A (1998) *Proc Natl Acad Sci* 95, 12208-13.

2.   Blumberg B, Sabbagh W, Jr., Juguilon H, Bolado J, Jr., van Meter CM, Ong ES, Evans RM (1998) *Genes & Development* 12, 3195-205.

3.   Goodwin B, Redinbo MR, Kliewer SA (2002) *Annu Rev Pharmacol Toxicol* 42, 1-23.

4.   Kliewer SA, Moore JT, Wade L, Staudinger JL, Watson MA, Jones SA, McKee DD, Oliver BB, Willson TM, Zetterstrom RH, Perlmann T, Lehmann JM (1998) *Cell* 92, 73-82.

5.   Goodwin B, Gauthier KC, Umetani M, Watson MA, Lochansky MI, Collins JL, Leitersdorf E, Mangelsdorf DJ, Kliewer SA, Repa JJ (2003) *Proc Natl Acad Sci* 100, 223-8.

6.   Goodwin B, Hodgson E, Liddle C (1999) *Molecular Pharmacology* 56, 1329-39.

7.   Jones SA, Moore LB, Shenk JL, Wisely GB, Hamilton GA, McKee DD, Tomkinson NC, LeCluyse EL, Lambert MH, Willson TM, Kliewer SA, Moore JT (2000) *Mol Endocrinol* 14, 27-39.

8.   Krasowski MD, Yasuda K, Hagey LR, Schuetz EG (2005) *Mol Endocrinol* 19, 1720-39.

9.   Lehmann JM, McKee DD, Watson MA, Willson TM, Moore JT, Kliewer SA (1998) *J Clin Invest* 102, 1016-23.

10.   Moore LB, Goodwin B, Jones SA, Wisely GB, Serabjit-Singh CJ, Willson TM, Collins JL, Kliewer SA (2000) *Proc Natl Acad Sci* 97, 7500-2.

11.   Moore LB, Parks DJ, Jones SA, Bledsoe RK, Consler TG, Stimmel JB, Goodwin B, Liddle C, Blanchard SG, Willson TM, Collins JL, Kliewer SA (2000) *Journal of Biological Chemistry* 275, 15122-7.

12.   Carnahan VE, Redinbo MR (2005) *Curr Drug Metab* 6, 357-67.

13.   Watkins RE, Davis-Searles PR, Lambert MH, Redinbo MR (2003) *J Mol Biol* 331, 815-28.

14.   Watkins RE, Wisely GB, Moore LB, Collins JL, Lambert MH, Williams SP, Willson TM, Kliewer SA, Redinbo MR (2001) *Science* 292, 2329-33.

15.   Chrencik JE, Orans J, Moore LB, Xue Y, Peng L, Collins JL, Wisely GB, Lambert MH, Kliewer SA, Redinbo MR (2005) *Mol Endocrinol* 19, 1125-34.

16.   Watkins RE, Maglich JM, Moore LB, Wisely GB, Noble SM, Davis-Searles PR, Lambert MH, Kliewer SA, Redinbo MR (2003) *Biochemistry* 42, 1430-8.

17.   Orans J, Teotico DG, Redinbo MR (2005) *Mol Endocrinol* 19, 2891-900.

18.   Noble SM, Carnahan VE, Moore LB, Luntz T, Wang H, Ittoop OR, Stimmel JB, Davis-Searles PR, Watkins RE, Wisely GB, Lecluyse E, Tripathy A, McDonnell DP, Redinbo MR (2006) *Biochemistry* 45, 8579-8589.

19.   Kliewer SA, Goodwin B, Willson TM (2002) *Endocr Rev* 23, 687-702.

20.   Brzozowski AM, Pike ACW, Dauter Z, Hubbard RE, Bonn T, Engstrom O, Ohman L, Greene GL, Gustafsson J-A, Carlquist M (1997) *Nature* 389, 753-758.

21.   Bledsoe RK, Montana VG, Stanley TB, Delves CJ, Apolito CJ, McKee DD, Consler TG, Parks DJ, Stewart EL, Willson TM, Lambert MH, Moore JT, Pearce KH, Xu HE (2002) *Cell* 110, 93-105.

22.   Mi LZ, Devarakonda S, Harp JM, Han Q, Pellicciari R, Willson TM, Khorasanizadeh S, Rastinejad F (2003) *Mol Cell* 11, 1093-100.

23.   Shan L, Vincent J, Brunzelle JS, Dussault I, Lin M, Ianculescu I, Sherman MA, Forman BM, Fernandez EJ 2004 *Mol Cell* 16, 907-17.

24. Suino K, Peng L, Reynolds R, Li Y, Cha JY, Repa JJ, Kliewer SA, Xu HE (2004) *Mol Cell* 16, 893-905.

25. Williams S, Bledsoe RK, Collins JL, Boggs S, Lambert MH, Miller AB, Moore J, McKee DD, Moore L, Nichols J, Parks D, Watson M, Wisely B, Willson TM (2003) *J Biol Chem* 278, 27138-43.

26. Xu RX, Lambert MH, Wisely BB, Warren EN, Weinert EE, Waitt GM, Williams JD, Collins JL, Moore LB, Willson TM, Moore JT (2004) *Mol Cell* 16, 919-28.

27. Binkowski TA, Naghibzadeh S, Liang J (2003) *Nucleic Acids Res* 31, 3352-5.

28. Ingraham HA, Redinbo MR (2005) *Curr Opin Struct Biol* 15, 708-15.

29. Frank C, Makkonen H, Dunlop TW, Matilainen M, Vaisanen S, Carlberg C (2005) *J Mol Biol* 346, 505-19.

30. Guo GL, Lambert G, Negishi M, Ward JM, Brewer HB, Jr., Kliewer SA, Gonzalez FJ, Sinal CJ (2003) *J Biol Chem* 278, 45062-71.

31. Kliewer SA, Willson TM (2002) *J Lipid Res* 43, 359-64.

32. Sonoda J, Chong LW, Downes M, Barish GD, Coulter S, Liddle C, Lee CH, Evans RM (2005) *Proc Natl Acad Sci U S A* 102, 2198-203.

33. Staudinger J, Liu Y, Madan A, Habeebu S, Klaassen CD (2001) *Drug Metab Dispos* 29, 1467-72.

34. Staudinger JL, Goodwin B, Jones SA, Hawkins-Brown D, MacKenzie KI, Latour A, Liu Y, Klaassen CD, Brown KK, Reinhard J, Willson TM, Koller BH, Kliewer SA (2001) *Proc. Natl. Acad. Sci. USA*.

35. Stedman CA, Liddle C, Coulter SA, Sonoda J, Alvarez JG, Moore DD, Evans RM, Downes M (2005) *Proc Natl Acad Sci U S A* 102, 2063-8.

36. Uppal H, Toma D, Saini SP, Ren S, Jones TJ, Xie W (2005) *Hepatology* 41, 168-76.

37. Xie W, Radominska-Pandya A, Shi Y, Simon CM, Nelson MC, Ong ES, Waxman DJ, Evans RM (2001) *Proc Natl Acad Sci U S A* 98, 3375-3380.

38. Zhang J, Huang W, Qatanani M, Evans RM, Moore DD (2004) *J Biol Chem* 27, 49517-22.

39. Nolte RT, Wisely GB, Westin S, Cobb JE, Lambert MH, Kurokawa R, Rosenfeld MG, Willson TM, Glass CK, Milburn MV (1998) *Nature* 395, 137-43.

40. Navaza J, Saludjian P (1997) *Methods Enz.* 276A, 581-594.

41. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) *Acta Crystallographica Section D Biological Crystallography* 54, 905-21.

42. Brünger AT (1993) Acta Crystallogr. *D Biol. Crystallogr.* 49, 24-36.

43. Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) *Acta Crystallogr.* A 47, 110-119.

44. Read RJ (1986) *Acta Crystallogr.* A 42, 140-149.

STUDY IV: Crystal Structure of the PXR-T1317 Complex Provides a Scaffold to Examine the Potential for Receptor Antagonism

(Under the direction of Professor Matt Redinbo)

Abstract

The human pregnane X receptor (PXR) recognizes a range of structurally- and chemically-distinct ligands and plays a key role in regulating the expression of protective gene products involved in the metabolism and excretion of potentially harmful compounds. The identification and development of PXR antagonists is desirable as a potential way to control the up-regulation of drug metabolism pathways during the therapeutic treatment of disease. We present the 2.8 Å resolution crystal structure of the PXR ligand binding domain (LBD) in complex with T0901317 (T1317), which is also an agonist of another member of the orphan class of the nuclear receptor superfamily, the liver X receptor (LXR). In spite of differences in the size and shape of the receptors' ligand binding pockets, key interactions with this ligand are conserved between human PXR and human LXR. Based on the PXR-T1317 structure, analogues of T1317 were generated with the goal of designing an PXR antagonist effective via the receptor's ligand binding pocket. We find that selectivity in activating PXR *vs*. LXR was achieved; such compounds may be useful in addressing neurodegenerative diseases like Niemann-Pick C. We were not successful, however, in producing a PXR antagonist. Based on these observations, we conclude that the generation of PXR antagonists

targeted to the ligand binding pocket may be difficult due to the promiscuity and structural conformability of this xenobiotic sensor.

Introduction

The human pregnane X receptor (PXR; alternatively SXR, PAR) responds to a variety of endogenous and exogenous compounds in liver, intestine and other tissues, and is a key regulator of the expression of genes central to xenobiotic metabolism and excretion (1, 2, 3). PXR is also responsible for an important class of drug interactions caused by the efficient up-regulation of chemoprotective pathways that lead to the elimination of a wide range of therapeutics (4, 5, 6, 7). It has also recently been shown that T0901317 (T1317), the ligand described here in complex with PXR, protects against the development of the neurodegenerative disease Niemann-Pick C in a PXR-dependent fashion within a mouse model of this condition (8).

Like other members of the orphan class of the nuclear receptor (NR) superfamily, PXR contains DNA-binding and ligand-binding domains (DBD, LBD, respectively), acts as a heterodimer with the retinoid X receptor-α (RXRα), and binds to a range of direct- and everted-repeat elements in the regulatory region of target genes (9). Upon association with an activating agonist, a transcriptional coactivator protein like the steroid receptor coactivator-1 (SRC-1) is recruited to the activation function-2 (AF-2) region of the PXR LBD, which facilitates changes in chromatin structure and activation of the basal transcriptional machinery. A leucine-rich LxxLL motif in transcriptional coactivators (where x is any amino acid) has been shown to interact with a groove present in the active orientation of NR LBDs (10, 11). Crystal structures of the human PXR LBD have been determined in complexes with a variety of small (*e.g.,* SR12813, hyperforin) and large (*e.g.,* rifampicin) ligands, and with fragments of the human transcriptional

coactivator SRC-1 (12, 13, 14, 15). These structures have revealed that PXR's ligand binding promiscuity is a function of its large and conformable ligand binding pocket, which is framed in part by sequence elements novel to PXR relative to other NR LBDs. In addition, the PXR LBD forms a unique homodimer mediated by a tryptophan zipper-like motif, and it has been shown that this interface plays a role in receptor function and association with coactivators (16).

The liver X receptor (LXR), another member of the orphan class of NRs that functions as a heterodimer with RXRα, plays an important role in monitoring the levels of oxysterols in hepatocytes and regulates the expression of genes essential for cholesterol homeostasis (17, 18, 19, 20, 21). The LBDs of the two LXR isoforms, α and β, both share 31% sequence identity with the human PXR LBD (and 77% with one another), as well as the conserved overall structural fold common to NR ligand binding domains (22, 23, 24, 25, 26). The PXR LBD deviates from that of LXR in its ~60-residue α1- α3 insert that adds one helix (α2) and two strands (β1, β1') and frame a significantly larger ligand binding pocket relative to LXR (10). The synthetic NR ligand T0901317 (T1317) is an established agonist for LXR, exhibiting robust upregulation of target gene expression (19). The structural basis of T1317 binding to LXR isoform β has been elucidated previously (22, 24, 26).

In addition to their association with transcriptional coactivators, NR LBDs bind to transcriptional corepressors (e.g., NCoR, SMRT) that exert opposite effects on gene transcription by mechanisms that include enhancing the condensed structure of chromatin

(4). Interactions between NRs and corepressors can occur in the absence of ligand, but are enhanced in the presence of antagonizing ligands. A paradigm for this effect is provided by efficient down-regulation of transcription caused by the association of the estrogen receptor (ER) with the established antagonist tamoxifen. It has been shown structurally that corepressors contain an extended leucine-rich region that binds to an inactive LBD conformation distinct in structure from that of an active LBD, particularly in the position of the terminal α-helix in the LBD fold (αAF in PXR) (27). In ER, tamoxifen sterically blocks the active orientation of the terminal helix in that fold, producing a state that preferentially binds to transcriptional corepressors and down-regulates gene expression (28). In the case of PXR, most ligands are found to act as agonists of this receptor, in line with its established role in protecting tissues from potentially harmful chemicals. A small number of antagonists have been described, however, including the ET-743 (29) and the antifungal ketoconazole (30, 31). Indeed, ketoconazole has been shown recently to repress the binding of both transcriptional coactivators and corepressors to PXR (30). The structural basis of the antagonism of PXR by these compounds has not been established, however.

The identification or design of new antagonists of human PXR are desirable because of the important role this receptor plays in drug metabolism, and because the current antagonists exhibit weak EC_{50} values relative to established agonists (29, 30, 31). Repressing the ability of PXR to recognize the presence of therapeutic compounds may allow lower doses to be administered with higher efficacy and fewer side effects. Here we present the 2.8 Å resolution crystal structure of the PXR LBD in complex with

T1317. Using this structure as a guide, and inspired by selective ER modulators produced from ER agonists, we generate T1317 analogues designed to act as PXR antagonists by disrupting the active conformation of the receptor's αAF. We find, however, that compounds based on this scaffold either retain their ability to bind to PXR, and thus serve as agonists, or simply do not bind to the receptor. We conclude that the design of effective pocket-targeted PXR antagonists may be difficult due to the promiscuous and conformable nature of the receptor's ligand binding pocket.

Results

PXR-T1317 Structure. Crystals of the human PXR LBD were grown in the presence of 10-fold molar excess T0901317 (T1317), x-ray diffraction data to 2.8 Å resolution were collected, and the structure was determined and refined to R and R_{free} values of 0.216 and 0.279, respectively (Table I). Two ternary complexes were observed in the asymmetric unit, each containing one human PXR LBD, one orientation of bound T1317, and one fifteen amino acid stretch of the human transcriptional coactivator steroid receptor coactivator 1 (SRC-1, residues 682-696) (Figure 1). The PXR LBD is expressed with an 88-amino acid fragment of SRC-1 (residue 623-710) to improve protein stability. This is the first structure of PXR in which the coactivator fragment remains attached during crystallization. The retention of this fragment is likely due to the relatively high affinity and potent agonist activity of T1317. The 88-amino acid region of human SRC-1 employed contains two leucine-rich NR box motifs (at 633-637, LVQLL, and 690-694. LHRLL). Only the second NR box was observed associated with the AF-2 region of the PXR LBD in both molecules in the crystallographic asymmetric unit. This implies that the sequence around or including LHRLL may have higher affinity for PXR relative to the region around the LVQLL motif.

The PXR LBD in this T1317 complex retains the same overall structure observed in previous PXR structures, sharing, for example, 1 Å root-mean-square deviation (rmsd) over Cα positions with the structure of the apo (unliganded) PXR LBD (15). The LBDs also form a homodimer consistently observed either via crystallographic or, as in this case, non-crystallographic symmetry. This homodimer interaction is mediated largely by

interdigitating aromatic residues from β1' in each monomer, and its formation has been shown to be central to transcriptional activity and coactivator recruitment by PXR (16). Only a small number of shifts in the positions of amino acids that line the ligand binding pocket of PXR were observed between the T1317 and apo structures. For example, Leu-209 and Met-323 undergo a rotamer changes and shifts in position of 6.0 Å. It was also noted that the side chains of both His-407 and Phe-429 shift 1.0 Å toward the bound ligand relative to the apo structure, and in doing so form a 3.8 Å aromatic edge-to-edge van der Waals contact (compared to 5.3 Å for the same atoms in the apo structure). Because Phe-429 is located on the αAF of the PXR's AF-2 region, this interaction likely stabilizes the active form of the receptor during the upregulation of gene transcription.

PXR-T1317 Interactions. T1317 forms three polar and twelve van der Waals contacts with amino acid side chains that line the PXR ligand binding pocket (Figure 2; Table II). His-407 is positioned 2.4 Å from the ligand's hydroxyl group, while the polar groups of the Gln-285 side chain are 2.9 and 3.2 Å from one sulfoxy oxygen and His-327 is 3.3 Å from the other. This is the first time His-327 has been observed within hydrogen bonding distance (generously defined) of a bound ligand in any of the PXR complexes determined to date. Similarly, Tyr-306, which forms an edge-to-face with the free benzyl ring of T1317, has also not been observed to contact ligand in previous structures. The same T1317 benzyl ring forms parallel and edge-to-face aromatic stacking interactions with Phe-288 and Trp-299, respectively. The two CF_3 groups of the ligand form van der Waals contacts with five residues, including a 3.5 Å interaction with Met-425 located on the receptor's αAF that likely helps to stabilize the active conformation of the AF-2

135

surface. The volume of the PXR ligand binding pocket in this structure was measured to be 1,334 Å^3 and it was further found that the T1317 ligand occupied all but 442 Å^3 of that space. In total, the fifteen residues contacted by T1317 is the largest observed for a small ligand in PXR's pocket, but does not exceed the eighteen residues contacted by the large macrolide antibiotic rifampicin (12).

T1317 Binding by LXR vs. PXR. We next compared the structure of the human PXR LBD-T1317 complex to that of the human LXRβ LBD complexed to the same ligand (22, 24, 25, 26). The LBDs exhibited the same overall fold, sharing 2.2 Å rmsd over Cα positions (and 27% sequence identity), although the secondary structural elements present on the α1- α3 insert novel to PXR, including β1, β1', and α2, are not present in the LXR LBD (Figure 3). LXR's ligand binding pocket is roughly half the size of PXR's (650 Å^3) and the observed binding of T1317 within the pocket uses essentially all the available space. The distinct and limited shape of the LXR pocket causes the T1317 ligand to bind in a position rotated by ~30º and shifted by up to 5.3 Å relative to the position observed in the PXR complex. Sixteen LXR side chains contact T1317, two of which form hydrogen bonds with the ligand (His-435 and Thr-316). His-435 in LXR corresponds in both sequence and structure to His-407 in PXR; however, Gln-285, which forms two polar contacts with T1317 in the PXR complex, is replaced by a leucine in LXR (Table II, Figure 4). Thr-316, which forms the second polar interaction in LXR, is related in sequence to Phe-288 in PXR, but corresponds in structure to His-327, albeit shifted in position by ~5.5 Å. The benzyl ring of T1317 forms aromatic contacts with phenylalanines 271 and 329 in LXR, which helps to position the ligand distinctly within

the pocket when compared to PXR. It is also noted that distinct histidine-aromatic interactions are observed in PXR relative to LXR. An edge-to-face aromatic interaction between His-435 and Trp-457 had been noted previously for LXR; this contact is replaced in PXR by a 3.8 Å edge-to-edge interaction between His-407 and Phe-429. A second histidine aromatic contact is observed in PXR: a 3.3 Å face-to-edge interaction between Trp-299 and His-327 (Figure 4). In total, PXR and LXR share one conserved polar contact and eight conserved hydrophobic interactions in their respective complexes with T1317 (Table II). In addition, both receptors directly contact corresponding residues located on the αAF helices of their AF-2 surfaces (Met-425 in PXR, Leu-453 in LXR). LXR apparently further stabilizes its AF-2 region by the formation of the aromatic contact between His-435 and Trp-457, which is located on αAF. Taken together, these observations show that, in spite of differences in pocket shape and ligand orientation, numerous key contacts are conserved between these related receptors, leading to the efficient upregulation of targets genes by T1317.

Design and Examination of Putative Antagonists. Because T1317 appears to bind particularly strongly to PXR, we next sought to design a PXR antagonist using the T1317 structure as a scaffold. It is known that antagonists of other nuclear receptors (*e.g.,* tamoxifen for the estrogen receptor) appear to function by sterically blocking the active position of αAF (28). Thus, we chose to change the structure of the CF_3-containing moiety in T1317, which binds adjacent to αAF in PXR (Figure 2). Eleven analogues were synthesized that included both smaller and larger groups at this position (Figure 5; Scheme 1). These compounds, along with T1317, were tested for their ability to bind to

137

PXR in an *in vitro* ligand-competition assay, and to activate PXR-mediated transcription in transient transfection assays in cultured cells (Table III). Compounds **2-5**, which retained the hydroxyl group and one CF_3 group, all exhibited good binding to and activation of PXR. In contrast, compounds **6** and **7**, which retain the hydroxyl group but not a CF_3, are poor binders and activators of PXR. Similarly, compound **8**, which contains a acetyl group, was the least efficacious compound examined. These data support the importance of the hydrogen bond donated by the T1317 hydroxyl group to the His-407 side chain in PXR, in particular the electron withdrawing character of the CF_3 groups that polarize the hydroxyl group and improve its ability to share its hydrogen atom (25).

Compounds **9-12** were designed to include large adducts adjacent to the hydroxyl and CF_3 groups, with the goal of sterically disrupting αAF position. We were surprised to find, however, that each compound bound well to PXR and served as an effective agonist rather than an antagonist (Table III). For example, the presence of cyclohexyl or benzyl rings (compounds **9**, **10**) were apparently accommodated by the receptor's ligand binding pocket. Indeed, even a benzyl group contained on an extended and rigid two-methylene linker (compound **11**) did not disrupt either ligand binding or receptor activation. These observations suggest that the αAF and AF-2 region of PXR is reasonably conformable and capable of accommodating larger adducts than was previously appreciated. The structural basis of these effects may be based on the mobile α2 region of PXR, which is present on a sequence insert unique to this receptor relative to other members of the nuclear receptor superfamily. This stretch of the PXR LBD structure has been observed

138

to shift in position by several Å between different ligand-bound complexes, and, in the case of the PXR structure bound to the large macrolide antibiotic rifampicin, to become completely disordered (12, 13, 14, 15). The large adducts placed in the T1317 scaffold in compounds **9-12** may be directed toward this region of PXR and may create space for themselves by disrupting the position of α2 rather than αAF. In addition, they could extend from the receptor's pocket into solvent by creating a pore adjacent to Leu-240 (Figure 2). In either case, these observed effects underscore the fundamentally promiscuous nature of PXR in terms of ligand binding – the receptor contains a polypeptide insert that allows its binding pocket to expand and contract, facilitating the productive binding of a wide range of chemical structures.

The activation of gene expression by T1317 analogues was also examined via LXRβ by transient transfection assays in cultured cells (Table III). Similar to PXR, compounds **6-8**, which do not retain the hydroxyl-group proximal to a CF_3 moiety, were poor ligands for LXR, highlighting the importance of the His-435 hydrogen received from the polarized ligand hydroxyl group. In contrast to PXR, however, compounds **9** and **11** were poor agonists of LXR, while compounds **10** and **12** were relatively effective. Considering the structure of the LXRβ-T1317 complex, the benzyl group in **10** may stack upon the aromatic Phe-268 side chain, an interaction not possible with compounds **9** or **11**. For **12**, the flexible nature of its CF_3-rich extension may facilitate productive binding by protruding past Phe-268 into solvent. Taken together, these observations regarding the impact of T1317 analogues on LXR-mediated gene expression reveal that this receptor, with its smaller ligand binding pocket and more narrow agonist profile, is more

139

accommodating to changes in chemical structure than was expected. This highlights the

plasticity present in the conserved nuclear receptor LBD fold, particularly for members of

the former orphan class.

Discussion

T1317 is an efficacious activator of gene transcription mediated by both the nuclear receptors PXR and LXRβ (19). We show, by comparing the 2.8 Å resolution crystal structure of the PXR-T1317 complex to the LXRβ-T1317 complexes reported previously (22, 24, 25, 26), that these two LBDs share some analogous contacts to the ligand, but exhibit key differences as well. The packing of aromatic side chains against the benzyl group of T1317, as well as the distinct placement of the proximal hydrogen-bonding residue Thr-316, cause the ligand to adopt a position in LXR rotated ~30º and shifted by ~5 Å relative to that observed in PXR. Still, nearly half the contacts between protein and ligand are conserved in both complexes, including the close hydrogen bond between a histidine side chain and a polarized hydroxyl group of the ligand. Both LBDs also utilize an interaction between the same histidine side chain and an aromatic residue located on AF-helices of the receptors' AF-2 regions: Trp-457 in LXR and Phe-429 in PXR. Indeed, it was noted in this PXR complex that the presence of the ligand appears to mediate the formation of a direct 3.8 Å aromatic contact between these side chains caused by their shift by 1 Å in position toward the bound T1317. It is likely that the additional stabilization of the active conformation of αAF by this interaction plays in important role in the control of gene expression by both nuclear receptors.

We were surprised both by the difficulty we experienced in designing an antagonist to either receptor, and by the plasticity the LXR receptor exhibits upon derivatization of the efficacious ligand T1317. Given the numerous interactions observed between the T1317 and the ligand binding pockets of both receptors, it was expected that adding bulky

groups to the region adjacent to the hydroxyl and dual CF_3 moieties on the ligand would produce compounds that sterically block the active orientation of the AF-helix (as seen in the estrogen receptor with tamoxifen, for example) (28). Instead, we found for PXR that compounds either bound and were agonists, or appeared incapable of binding to the receptor. In addition, in the cases where reasonably large groups were added to the T1317 scaffold, the region of PXR's ligand binding pocket adjacent to αAF was remarkably amenable to accommodating bulky additional atoms. This is perhaps not surprising for PXR, given its well established promiscuous ligand binding character and the structural flexibility of unique regions (such as α2) in its ligand binding pocket. It would appear that for PXR, in contrast to more rigid and specific steroid receptors like the estrogen receptor, ligands can bridge between rigid portions of the pocket, while more flexible regions of the protein can shift to accommodate a variety of distinct chemical structures.

While the promiscuity of PXR is well known, we were not expecting to find that LXRβ would also show evidence for conformability in accommodating larger T1317 analogues. For example, compounds **10** and **12** were relatively effective agonists for LXR, in spite of the presence of large groups attached adjacent to T1317's hydroxyl moiety (Figure 5). The structural basis of this is likely centered on the region around Phe-268 in the LXR pocket, which may stack with the aromatic group in **10** and provide a pore for the conformable group in **12**. LXR did exhibit more specificity for the types of groups it can accommodate in this position relative to PXR, however, as non-aromatic or longer rigid adducts in **9** and **11**, respectively, that still activated PXR were clearly not agonists for

LXRβ. Thus, the ligand binding domains of numerous members of the nuclear receptor superfamily, particularly the former orphan receptors, are likely to posses some degree of structural flexibility to assist in the accommodation of chemically-distinct ligands. It would appear that PXR, though, with its large and mobile α1-α3 insert, is at the extreme of this flexibility continuum with these unique features that are central to its significant promiscuity.

We conclude that the unique aspects of the PXR ligand binding pocket may make antagonist design particularly difficult for this member of the nuclear receptor superfamily. It is noted that compounds **9** and **11** appear to be selective for PXR relative to LXRβ; as such, they may be useful as leads to address neurodegenerative diseases like Niemann-Pick C (8). However, it may be necessary to look outside the ligand binding pocket of PXR to find sites effective at antagonizing the receptor using small molecules. Candidate sites on the LBD include the AF-2 surface (30) and the PXR homodimer interface, the disruption of which has been shown to impact productive coactivator binding by the receptor (16). An effective PXR antagonist, if identified, may be of significant clinical use to reduce the activation of xenobiotic metabolism pathways during the therapeutic treatment of disease.

Materials and Methods

Protein Expression and Purification. Generation of human PXR LBD in complex with the 88-amino acid fragment of human SRC-1 was accomplished as described previously (12, 13, 14, 15, 16). To prevent the formation of covalent complexes with reducing agent during crystallization, as has been seen with previous structures (data not shown), Cys-284 within the ligand binding pocket of the human PXR LBD was replaced with serine.

Crystallization. The human PXR ligand-binding domain/SRC-1 complex (hPXR-LBD/SRC-1) was concentrated in the presence of 10-fold molar excesses T0901317 (T1317; synthesized in-house) to a final concentration of 5 mg/mL. Crystallization was achieved by hanging-drop vapor diffusion against the following conditions at 22 $^{\circ}$C: 50 mM imidazole at pH 7.1, 10% 2-propanol, v/v.

Data Collection and Structure Determination. The structure of the T1317-bound form of the ligand binding domain of human PXR was determined by molecular replacement using the crystal structure of the apo (unliganded) PXR as a search model (15). Rotation and translation function searches were performed using AMoRe (32); clear solutions for each monomer of the dimer in the asymmetric unit were obtained in the proper space group, P2$_1$2$_1$2$_1$. The structure was refined using CNS with the maximum likelihood function as a target, and included an overall anisotropic B-factor and a bulk solvent correction (33). Ten percent of the observed data were set aside for cross-validation using the free-R statistic prior to any structural refinement (34). Manual adjustments and building of the model (including the placement of the T0901317 ligands and SRC-1

coactivator fragments) were performed using O (35) and σA-weighted electron density maps (36). The structure exhibits good geometry (Table I) with no Ramachandran outliers.

Compound Syntheses. See Scheme 1 for steps involved in compound generation. ^{1}H-NMR spectra were recorded on a Varian Gemini 400 MHz NMR spectrometer. ^{1}H-NMR spectra are reported as chemical shift δ, number of protons, multiplicity (s, singlet; d, doublet; t, triplet; m, multiplet; br s, broad singlet) and coupling constant (J) in Hertz. Electron Spray (ES) or Chemical Ionization (CI) was recorded on a Hewlett Packard 5989A mass spectrometer. Mass spectrometry results are reported as the mass over charge. Purity by HPLC [Luna 20 x 4 mm 3.0 micron C18(2) column, water (+0.1% v/v formic acid)/MeOH (+0.075% v/v formic acid) gradient: 50% MeOH to 100 % MeOH for 5 minutes, holding at 100% MeOH for final 1 minute, flow rate = 2.0 mL/min]. Starting material are either available from commercial sources or via literature procedures. Abbreviations used in the examples below have their accepted meanings in the chemical literature. For example, DCM (dichloromethane), THF (tetrahydrofuran), MeCN (acetonitrile), DMSO (dimethylsulfoxide) and TBAF (tetrabutylammonium fluoride).

Compound 6: 4-Aminobenzyl alcohol (10 g, 80 mmol) and benzenesulfonyl chloride (11.5 mL, 89 mmol) in pyridine were heated at 60^0C overnight. Water was added and extracted with DCM. The organic layer was separated and evaporated to dryness. The residue was purified by silica gel chromatography using ethyl acetate to give 17 g (80%) of intermediate sulfonamide: ^{1}H-NMR (DMSO-d_6) δ: 7.8 (s, 1H), 7.65 (m, 3H). 7.10 (d,

J=8.8Hz, 2H), 7.01 (d, J=8.8Hz, 2H), 5.04 (br s, 1H), 4.45 (s, 2H). The intermediate

sulfonamide (1.2 g, 4.6 mmol), 2,2,2-trifluoroethyltrifluoro-methane sulfonate (1.97 g,

9.1 mmol) and K_2CO_3 (1.25 g, 9.1 mmol) in MeCN (20 mL) was heated at 80^0C

overnight. The solvents were removed under reduced pressure. The residue was purified

by silica gel chromatography using 50% ethyl acetate in hexanes to give **6** (1.0 g, 80%):

^{1}H-NMR (DMSO-d_6) δ: 7.5-7.8 (m, 5H), 7.26 (d, J=8.5 Hz, 2H), 6.99 (d, J=8.5 Hz, 2H),

5.30 (br s, 1H), 4.62 (m, 2H), 4.43 (s, 2H); LC/MS (ES+): *m/e* 346 (M+H), 100% purity.

Compound 8:   Prepared in a similar fashion from 4-aminoacetophenone (5.0 g, 37.0

mmol): ^{1}H-NMR (DMSO-d_6) δ: 7.90 (d, J=8.4 Hz, 2H), 7.56-7.71 (m, 5H), 7.27(d, J=8.4

Hz, 2H), 4.64 (m, 2H), 2.48 (s, 3H), LC/MS (ES+) *m/e* 358 (M + H), 95% purity)

Compound 2: DMSO (5.5 mL, 28 mmol) was added dropwise to 2M oxalyl chloride in

DCM (19 mL)at -78^0C under nitrogen followed by sequential addition of compound **6**

(10 g, 30 mmol) in DMSO:DCM (1:1) (20 mL) and neat triethylamine (75 mL). The

reaction mixture was gradually warmed to room temperature overnight whereupon the

solvents were removed under reduced pressure. The residue was purified by silica gel

chromatography with DCM to give an intermediate aldehyde (8.5 g, 85%): ^{1}H-NMR

(DMSO-d_6) δ: 9.8 (s, 1H), 7.8 (m, 2 H), 7.52-7.81 (m, 5H), 7.38 (m, 2H).

The intermediate aldehyde (5.0 g, 14.6 mmol) and trifluoromethyltrimethylsilane (2.5 g,

17.5 mmol) in anhydrous THF (15 mL) was treated with a catalytic amount of TBAF at

0^0C. The reaction mixture was gradually warmed to room temperature overnight. The

reaction mixture was treated with 1N hydrochloric acid at room temperature for 1h.

Water was added and extracted with DCM. The organic layer was separated and

evaporated to dryness. The residue was purified by silica gel chromatography using 20%

ethyl acetate in hexanes to give **2** (5.0 g, 85%): ^{1}H-NMR (MeOH-d$_4$) δ: 7.65 (m, 1H), 7.58 (d, J=8.6 Hz, 2 H), 7.52 (m, 2H), 7.45 (d, J=8.2 Hz, 2H), 7.11 (d, J=8.6 Hz, 2H), 5.03 (m, 1H), 4.43 (m, 2H); LC/MS (ES+) *m/e* 414 (M+H), 96% purity.

The following compounds were prepared in an analogous fashion:

Compound 12: Treatment of intermediate aldehyde with heptafluoropropyl-trimethylsilane (0.06 mL, 0.29 mmol) and purification by reverse phase chromatography on RP-C18 with 50-100% MeCN in water provided compound **12** (20 mg, 15%) as a racemic mixture: ^{1}H-NMR (MeOH-d$_4$) δ: 7.72 (m, 3H), 7.52 (m, 4H) 7.17 (d, J=8.6 Hz, 2H), 4.41 (m, 2H), 2.22 (m, 1H), 1.99 (m, 1H), 1.38 (m, 2H), 0.91 (m, 3H); LC/MS (ES+) *m/e* 456 (M+H), 98% purity.

Compound 3: Compound **3** was prepared in an analogous fashion from compound **8** to given 30 mg (25%) as a racemic mixture: ^{1}H-NMR (MeOH-d$_4$) δ: 7.49-7.65 (m, 7H), 7.10 (d, J=8.6 Hz, 2H), 4.40 (m, 2H), 1.70(s, 3H), LC/MS (ES+) *m/e* 428 (M + H), 98% purity.

Compound 4: Compound **2** (2.5 g, 6.0 mmol), Dess-Martin periodinane (5.2 g, 12.0 mmol) and pyridine (4mL, 48 mmol) in DCM (60 mL) was stirred at room temperature overnight. The reaction was treated with 60 mL of an aqueous solution of NaHCO$_3$:Na$_2$S$_2$O$_3$ (5:1) and stirred at room temperature for 2 h. The organic layer was separated and evaporated to dryness. The residue was purified by silica gel chromatography using 50% ethyl acetate in hexanes to give an intermediate trifluoromethyl ketone (2.2 g, 90%): ^{1}H-NMR (MeOH-d$_4$) δ: 7.46-7.62 (m, 7H), 7.21 (d, J=8.6 Hz, 2H), 4.45 (m, 2H). 1M ethyl magnesium bromide in THF (0.5 mL) was slowly

added to a solution of the above ketone (100 mg, 0.24 mmol) in THF (3mL) at -78^{0}C. The reaction mixture was gradually warmed to room temperature overnight. Saturated aqueous NH_4Cl was added, and the aqueous layer was extracted with ethyl acetate. The organic layer was separated and evaporated to dryness. The residue was purified by reverse phase chromatography on RP-C18 using 50-100% MeCN in water to give **4** (5 mg, 10%) as a racemic mixture: ^{1}H-NMR (MeOH-d$_4$) δ: 7.54-7.71 (m, 7H), 7.11 (d, J=8.6Hz, 2H), 4.45 (m, 2H), 2.21 (m, 1H), 1.98 (m, 1H), 0.87 (m, 3H); LC/MS (ES+) *m/e* 445 (M+H), 98% purity.

The following compounds were prepared in an analogous fashion from the above intermediate trifluoromethyl ketone:

Compound 5: Using 2M n-propyl magnesium bromide in diethyl ether followed by purification by reverse phase chromatography on RP-C18 with 50-100% MeCN in water gave rise to compound **5** (7 mg, 10%) as a racemic mixture: ^{1}H-NMR (MeOH-d$_4$) δ: 7.47-7.65 (m, 7H), 7.11 (d, J=8.6 Hz, 2H), 4.41 (m, 2H), 2.22 (m, 1H), 1.99 (m, 1H), 1.38 (m, 2H), 0.91 (m, 3H); LC/MS (ES+) *m/e* 456 (M+H), 95% purity.

Compound 9: Using 2M cyclohexyl magnesium bromide in diethyl ether (0.13 mL) followed by purification by reverse phase chromatography on RP-C18 with 50-100% MeCN in water provided compound **9** (20 mg, 20%) as a racemic mixture: ^{1}H-NMR (MeOH-d$_4$) δ: 7.43-7.89 (m, 7H), 7.11 (d, J=8.6 Hz, 2H), 4.41 (m, 2H), 2.61-2.05 (4m, 5H), 1.03 (m, 3H), 0.93 (m, 3H); LC/MS (ES+) *m/e* 496 (M+H), 94% purity.

Compound 10: Using 3M phenyl magnesium bromide in diethyl ether (0.09 mL) followed by purification via reverse phase chromatography on RP-C18 with 50-100% MeCN in water gave rise to compound **10** (70 mg, 70%) as a racemic mixture: ^{1}H-NMR (MeOH-d$_4$) δ: 7.33-7.69 (m, 12H), 7.11 (d, J=8.5 Hz, 2H), 4.42 (m, 2H), LC/MS (ES+) *m/e* 490 (M+H), 100% purity.

Compound 11: To a solution of phenylacetylene (83 mg, 0.75 mmol) in THF (2 mL) was added 1.6M n-BuLi in hexanes (0.315 mL, 0.5 mmol) at -78[0]C over 30 minutes. After this time, a solution of the above intermediate trifluoromethylketone (100 mg, 0.25 mmol) in THF (2 mL) was added. The reaction mixture was gradually warmed to room temperature overnight. Saturated aqueous NH$_4$Cl was added, and the aqueous layer was extracted with DCM. The organic layer was separated and evaporated to dryness. The residue was purified by reverse phase chromatography on RP-C18 using 50-100% MeCN in water to give compound **11** (77 mg, 75%) as a racemic mixture: ^{1}H-NMR (MeOH-d$_4$) δ: 7.43-7.80 (m, 12H), 7.17 (d, J=8.6 Hz, 2H), 4.42 (m, 2H); LC/MS (ES+) *m/e* 514 (M + H), 98% purity.

Compound 7: Compound 7 was prepared from **8** (100 mg, 0.28 mmol) and 1M vinyl magnesium bromide (0.84mL, 0.84 mmol) using the procedure described for the preparation of **4**. Purification by reverse phase chromatography on RP-C18 using 50-100% MeCN in water provided **7** (30 mg, 30%) as a racemic mixture: ^{1}H-NMR (MeOH-d$_4$) δ: 7.49-7.67 (m, 5H), 7.32 (d, J=8.4 Hz, 2H), 7.03 (d, J=8.4 Hz, 2H), 4.79 (m, 1H), 4.39 (m, 2H), 1.40(s,2H); LC/MS (ES+) *m/e* 358 (M +H), 98% purity.

Transient Transfections. Transient transfection and reporter gene assays using full-length

human PXR and LXRβ were performed as described previously (12).

Competition Ligand Binding Assay. Competitive ligand binding assays using [N-methyl-

^{3}H]-GW0438X were preformed as described elsewhere (16).

Table I. Crystallographic Statistics for the PXR-T0901317 Complex

Resolution (Å; highest shell)	50-2.8 Å (2.9-2.8)
Space Group	$P2_12_12_1$
Asymmetric Unit	two molecules
Cell Constants (Å, °)	a = 83.9 b = 90.6
	c = 105.7
	$\alpha = \beta = \gamma = 90$
Data Collection Facility	SER-CAT (APS)
Total Reflections	119,110
Unique Reflections	20,332
Mean Redundancy (highest shell)	5.8 (5.4)
R_{sym}* (%; highest shell)	13.5 (46.7)
Completeness (%; highest shell)	97.6 (91.1)
Mean I/σ (highest shell)	25.9 (5.1)
R_{cryst}†	21.6
R_{free}‡	27.9

---

* $R_{sym} = \Sigma|I-<I>| / \Sigma I$, where I is the observed intensity and $<I>$ is the average intensity of multiple symmetry-related observations of that reflection.

† $R_{cryst} = \Sigma||F_{obs}|-|F_{calc}|| / \Sigma|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.

‡ $R_{free} = \Sigma||F_{obs}|-|F_{calc}|| / \Sigma|F_{obs}|$ for 10% of the data not used at any stage of structural refinement.

Table II. Comparison of Residues Contacting T0901317 in PXR and LXR. Amino acids forming hydrogen bonding interactions with ligand are listed in bold, those on the terminal αAF helices of the receptors are underlined, and those not contacting ligand are in italics and parentheses.

PXR	LXR
Gln-285	Leu-313
His-327	(*Ser-357*)
His-407	**His-435**
Leu-209	--[*]
Val-211	--[*]
Leu-240	Phe-268
Met-243	Phe-271
Met-246	Leu-274
Phe-288	**Thr-316**
Trp-299	Ile-327
Tyr-306	(*Tyr-335*)
Leu-411	Val-439
Ile-413	Leu-442
Phe-420	Leu-449
<u>Met-425</u>	<u>Leu-453</u>
(*Cys-284-Ser*)[†]	Met-312
(*Cys-301*)[‡]	Phe-329
(*Leu-206*)	Leu-345
(*Met-323*)	Ile-353
(*Phe-429*)	Trp-457

[*] Residues 209 and 211 are on a sequence insert novel to PXR; thus, no equivalent side chains exist in LXR.

[†] Cys-284 was mutated to serine to improve the crystallization behavior of PXR; see Materials and Methods.

‡ While Cys-301 corresponds in sequence to Phe-329, the PXR side chain Phe-288

overlaps structurally with this LXR phenylalanine; see Figure 4.

Table III. T0901317 Analogues and Their Impact on Human PXR and LXRβ. See

Figure 5 for analogue structures. Compound **1** is T0901317 present in the PXR complex

structure presented here.

Compound	PXR pIC50	PXR pEC50	PXR %max	LXR pEC50	LXR %max
1	7.4	7.9	93	7.0	100
2	5.0	6.0	90	<4.5	19
3	5.4	6.3	104	5.8	50
4	6.6	6.9	101	5.4	73
5	6.5	6.8	96	5.6	27
6	5.1	5.0	23	<4.5	3
7	5.6	5.0	45	<4.5	3
8	5.8	5.0	11	<4.5	3
9	7.6	8.5	103	<4.5	9
10	7.2	7.8	99	5.5	30
11	6.2	6.0	139	<4.5	5
12	7.7	8.0	98	6.2	58

Figure Legends

Figure 1. Crystal structure of the homodimer in the asymmetric unit of the ligand binding domain of human PXR (PXR LBD) in complex with T0901317 (T1317). The PXR LBD in one monomer is rendered in red, yellow and green, and in the other monomer in cyan, magenta and pink. The fragments of the steroid receptor coactivator 1 (SRC-1) are shown in aqua and orange, and the T1317 ligands in magenta and green.

Figure 2. Stereoview of the binding of T0901317 within the ligand binding pocket of the human PXR LBD. Residues making polar contacts are rendered in green, those making hydrophobic interactions are in red, with Met-425 from αAF highlighted in yellow.

Figure 3. Superposition of the human PXR LBD monomer (red) on that of the human LXRβ LBD (gold), with the T0901317 ligands present in both structures shown in magenta for PXR and blue for LXR.

Figure 4. Stereoview comparing the binding of T0901317 to the LBDs of human PXR and LXRβ. Human PXR residues and T1317 ligand are colored as in Figure 1B, except Phe-429 and Met-243, which are shown in white. In LXR, residues forming polar ligand contacts are shown in cyan, while those forming hydrophobic interactions are in gold or, for Leu-452 and Tyr-335, in pink.

Figure 5. Structures of analogues of T0901317 (compound 1) examined for their impact

on human PXR and human LXRβ.
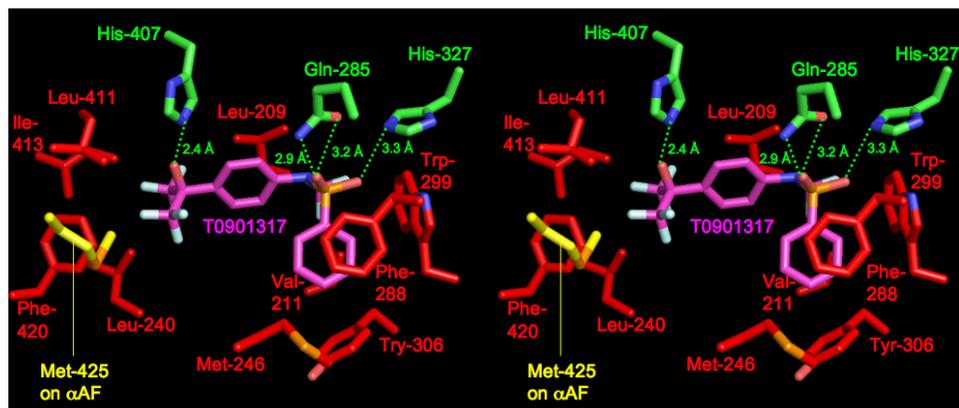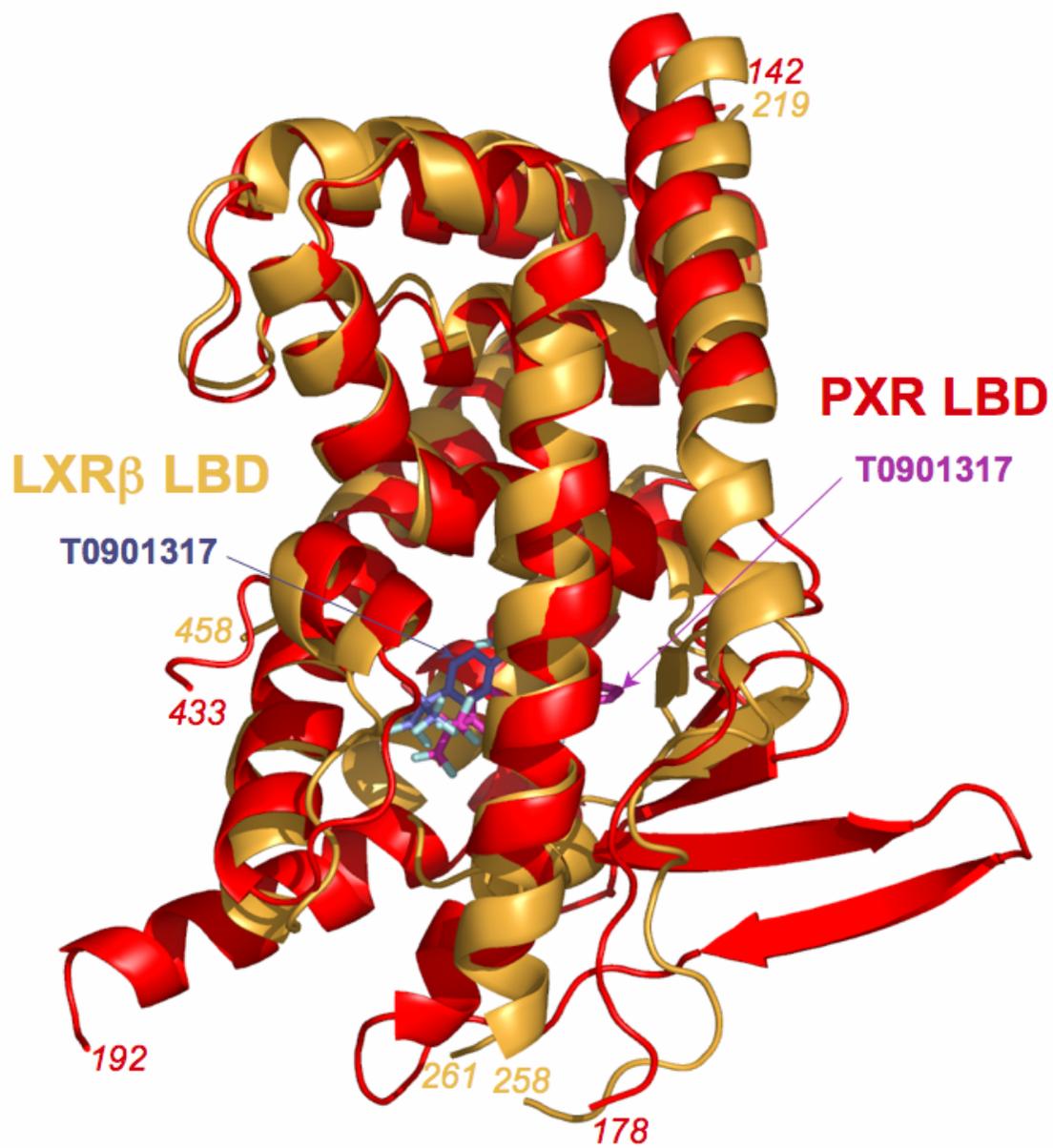
Figure 1.

Figure 2.

Figure 3.

Figure 4.
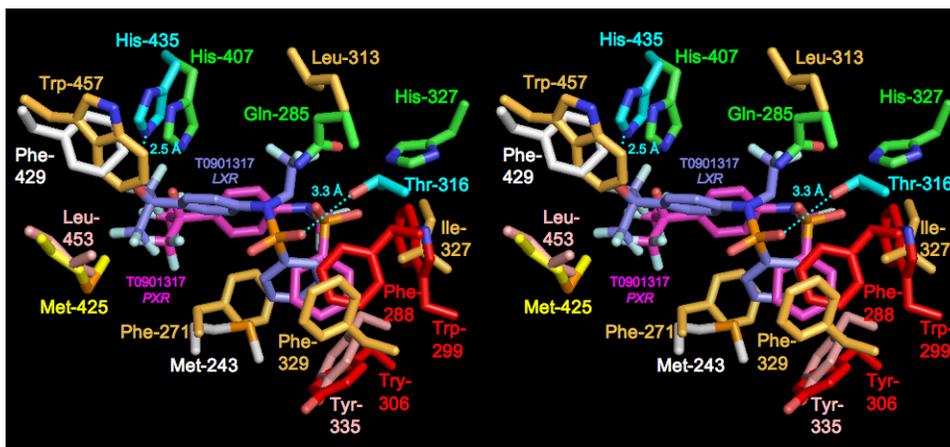
Figure 5.
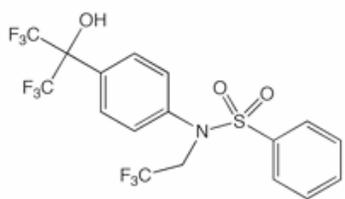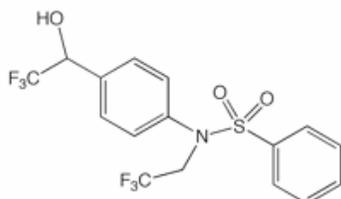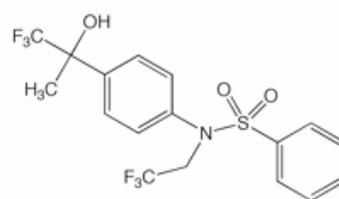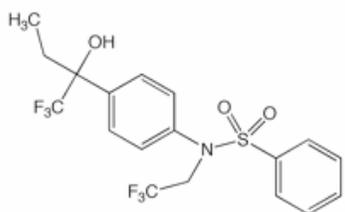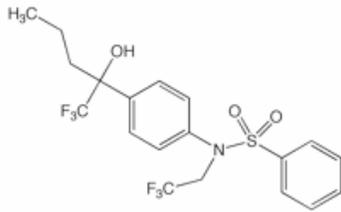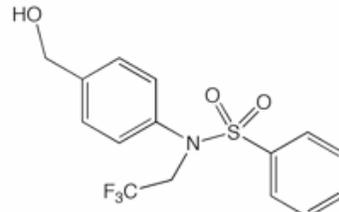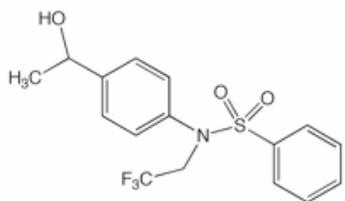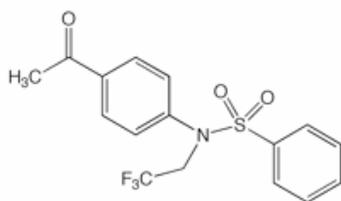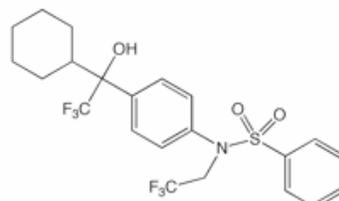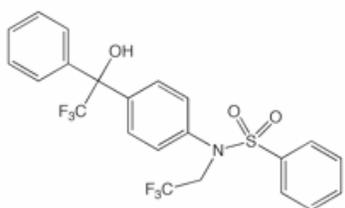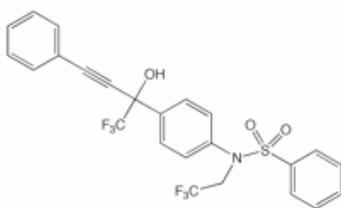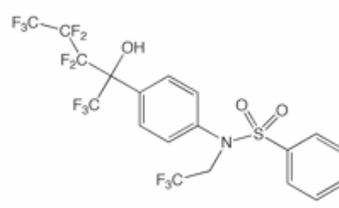
Scheme 1.
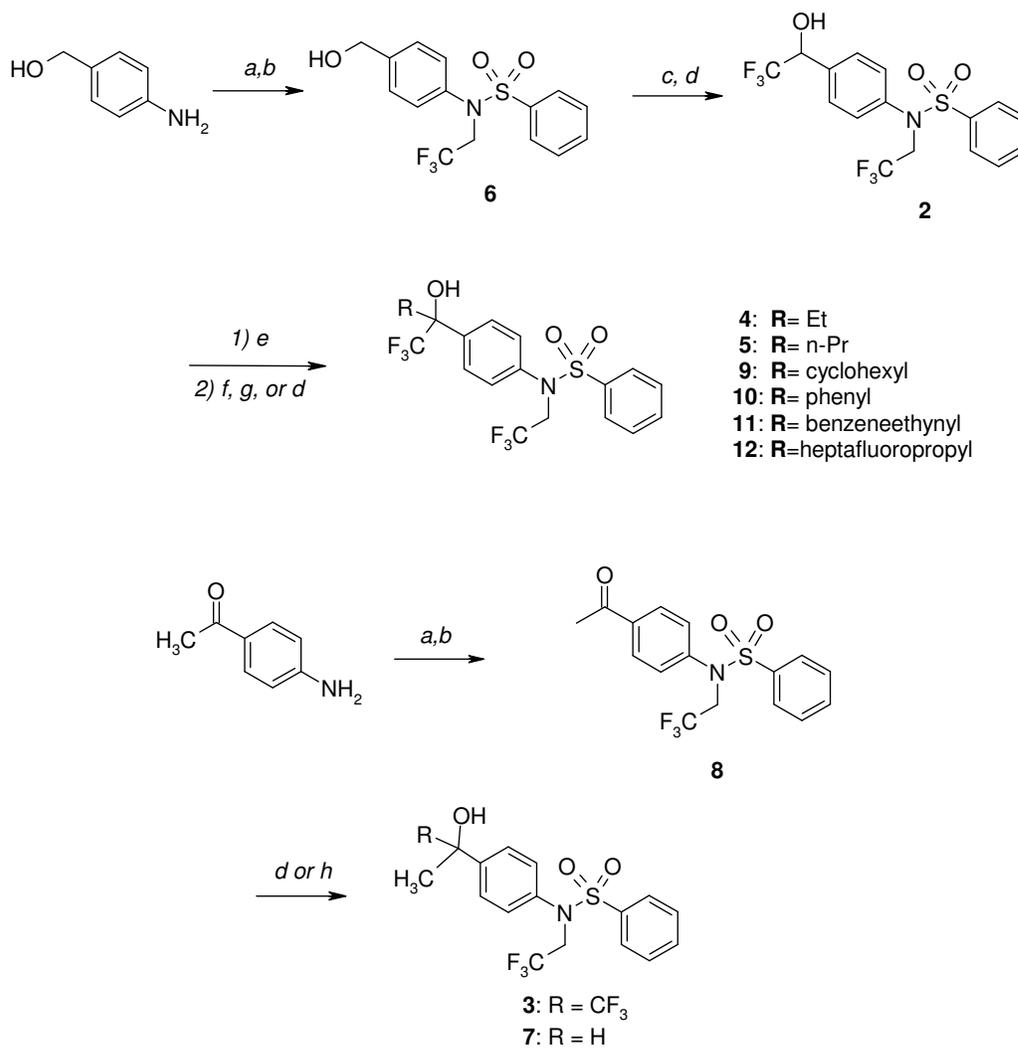
Reaction conditions: (a) benzenesulfonyl chloride, pyridine, 60°C; (b) 2,2,2-trifluoroethyl trifluoromethylsulfonate, K$_2$CO$_3$, MeCN, 80^{0}C; (c) oxalyl chloride, DMSO, Et$_3$N, rt; (d) trifluoromethyltrimethylsilane, TBAF, THF, 50^{0}C; (e) Dess-Martin periodinane, pyridine, DCM, rt; (f) RMgBr, THF, -78^{0}C to rt; (g) phenylacetylene, n-BuLi, THF, -78^{0}C to rt; (h) vinylmagnesiun bromide, THF, -78^{0}C to rt.

References

1.   Bertilsson, G.; Heidrich, J.; Svensson, K.; Asman, M.; Jendeberg, L.; Sydow-Backman, M.; Ohlsson, R.; Postlind, H.; Blomquist, P.; Berkenstam, A. (1998) *Proc Natl Acad Sci U S A* 95, 12208-13.

2.   Blumberg, B.; Sabbagh, W., Jr.; Juguilon, H.; Bolado, J., Jr.; van Meter, C. M.; Ong, E. S.; Evans, R. M. (1998) *Genes & Develop* 12, 3195-205.

3.   Kliewer, S. A.; Moore, J. T.; Wade, L.; Staudinger, J. L.; Watson, M. A.; Jones, S. A.; McKee, D. D.; Oliver, B. B.; Willson, T. M.; Zetterstrom, R. H.; Perlmann, T.; Lehmann, J. (1998) M. *Cell* 92, 73-82.

4.   Carnahan, V. E.; Redinbo, M. R. (2005) *Curr Drug Metab* 6, 357-67.

5.   Lehmann, J. M.; McKee, D. D.; Watson, M. A.; Willson, T. M.; Moore, J. T.; Kliewer, S. A. (1998) *J Clin Invest* 102, 1016-23.

6.   Moore, J. T.; Kliewer, S. A. (2000) *Toxicology* 153, 1-10.

7.   Moore, L. B.; Goodwin, B.; Jones, S. A.; Wisely, G. B.; Serabjit-Singh, C. J.; Willson, T. M.; Collins, J. L.; Kliewer, S. A. (2000) *Proc Natl Acad Sci U S A* 97, 7500-2.

8.   Langmade, S. J.; Gale, S. E.; Frolov, A.; Mohri, I.; Suzuki, K.; Mellon, S. H.; Walkley, S. U.; Covey, D. F.; Schaffer, J. E.; Ory, D. S. (2006) *Proc Natl Acad Sci U S A* 103, 13807-12.

9.   Orans, J.; Teotico, D. G.; Redinbo, M. R. (2005) *Mol Endocrinol* 19, 2891-900.

10.   Ingraham, H. A.; Redinbo, M. R. (2005) *Curr Opin Struct Biol* 15, 708-15.

11.   Nolte, R. T.; Wisely, G. B.; Westin, S.; Cobb, J. E.; Lambert, M. H.; Kurokawa, R.; Rosenfeld, M. G.; Willson, T. M.; Glass, C. K.; Milburn, M. V. (1998) *Nature* 395, 137-43.

12.   Chrencik, J. E.; Orans, J.; Moore, L. B.; Xue, Y.; Peng, L.; Collins, J. L.; Wisely, G. B.; Lambert, M. H.; Kliewer, S. A.; Redinbo, M. R. (2005) *Mol Endocrinol* 19, 1125-34.

13.   Watkins, R. E.; Davis-Searles, P. R.; Lambert, M. H.; Redinbo, M. R. (2003) *J Mol Biol* 331, 815-28.

14.   Watkins, R. E.; Maglich, J. M.; Moore, L. B.; Wisely, G. B.; Noble, S. M.; Davis-Searles, P. R.; Lambert, M. H.; Kliewer, S. A.; Redinbo, M. R. (2003) *Biochemistry* 42, 1430-8.

15. Watkins, R. E.; Wisely, G. B.; Moore, L. B.; Collins, J. L.; Lambert, M. H.; Williams, S. P.; Willson, T. M.; Kliewer, S. A.; Redinbo, M. R. (2001) *Science* 292, 2329-33.

16. Noble, S. M.; Carnahan, V. E.; Moore, L. B.; Luntz, T.; Wang, H.; Ittoop, O. R.; Stimmel, J. B.; Davis-Searles, P. R.; Watkins, R. E.; Wisely, G. B.; Lecluyse, E.; Tripathy, A.; McDonnell, D. P.; Redinbo, M. R. (2006) *Biochemistry* 45, 8579-8589.

17. Janowski, B. A.; Willy, P. J.; Devi, T. R.; Falck, J. R.; Mangelsdorf, D. J. (1996) *Nature* 383, 728-31.

18. Kalaany, N. Y.; Gauthier, K. C.; Zavacki, A. M.; Mammen, P. P.; Kitazume, T.; Peterson, J. A.; Horton, J. D.; Garry, D. J.; Bianco, A. C.; Mangelsdorf, D. J. (2005) *Cell Metab* 1, 231-44.

19. Laffitte, B. A.; Repa, J. J.; Joseph, S. B.; Wilpitz, D. C.; Kast, H. R.; Mangelsdorf, D. J.; Tontonoz, P. (2001) *Proc Natl Acad Sci U S A* 98, 507-12.

20. Peet, D. J.; Turley, S. D.; Ma, W.; Janowski, B. A.; Lobaccaro, J. M.; Hammer, R. E.; Mangelsdorf, D. J. (1998) *Cell* 93, 693-704.

21. Zhang, Y.; Repa, J. J.; Gauthier, K.; Mangelsdorf, D. J. (2001) *J Biol Chem* 276, 43018-24.

22. Hoerer, S.; Schmid, A.; Heckel, A.; Budzinski, R. M.; Nar, H. (2003) *J Mol Biol* 334, 853-61.

23. Jaye, M. C.; Krawiec, J. A.; Campobasso, N.; Smallwood, A.; Qiu, C.; Lu, Q.; Kerrigan, J. J.; De Los Frailes Alvaro, M.; Laffitte, B.; Liu, W. S.; Marino, J. P., Jr.; Meyer, C. R.; Nichols, J. A.; Parks, D. J.; Perez, P.; Sarov-Blat, L.; Seepersaud, S. D.; Steplewski, K. M.; Thompson, S. K.; Wang, P.; Watson, M. A.; Webb, C. L.; Haigh, D.; Caravella, J. A.; Macphee, C. H.; Willson, T. M.; Collins, J. L. (2005) *J Med Chem* 48, 5419-22.

24. Svensson, S.; Ostberg, T.; Jacobsson, M.; Norstrom, C.; Stefansson, K.; Hallen, D.; Johansson, I. C.; Zachrisson, K.; Ogg, D.; Jendeberg, L. (2003) *EMBO J* 22, 4625-33.

25. Williams, S.; Bledsoe, R. K.; Collins, J. L.; Boggs, S.; Lambert, M. H.; Miller, A. B.; Moore, J.; McKee, D. D.; Moore, L.; Nichols, J.; Parks, D.; Watson, M.; Wisely, B.; Willson, T. M. (2003) *J Biol Chem* 278, 27138-43.

26. Farnegardh, M.; Bonn, T.; Sun, S.; Ljunggren, J.; Ahola, H.; Wilhelmsson, A.; Gustafsson, J. A.; Carlquist, M. (2003) *J Biol Chem* 278, 38821-8.

27. Xu, H. E.; Stanley, T. B.; Montana, V. G.; Lambert, M. H.; Shearer, B. G.; Cobb, J. E.; McKee, D. D.; Galardi, C. M.; Plunket, K. D.; Nolte, R. T.; Parks, D. J.; Moore, J. T.; Kliewer, S. A.; Willson, T. M.; Stimmel, J. B. (2002)*Nature* 415, 813-7.

28. Shiau, A. K.; Barstad, D.; Loria, P. M.; Cheng, L.; Kushner, P. J.; Agard, D. A.; Greene, G. L. (1998) *Cell* 95, 927-937.

29. Synold, T. W.; Dussault, I.; Forman, B. M. (2001) *Nat Med* 7, 584-90.

30. Huang, H.; Wang, H.; Sinz, M.; Zoeckler, M.; Staudinger, J.; Redinbo, M. R.; Teotico, D. G.; Locker, J.; Kalpana, G. V.; Mani, S. (2006) *Oncogene* in press.

31. Takeshita, A.; Taguchi, M.; Koibuchi, N.; Ozawa, Y. (2002) *J Biol Chem* 277, 32453-8.

32. Navaza, J.; Saludjian, P. (1997) *Methods Enz.* 276A, 581-594.

33. Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. (1998) *Acta Crystallogr, D* 54, 905-21.

34. Brünger, A. T. (1993) *Acta Crystallogr. D Biol. Crystallogr.* 49, 24-36.

35. Jones, T. A.; Zou, J. Y.; Cowan, S. W.; Kjeldgaard, M. (1991) *Acta Crystallogr. A* 47, 110-119.

36. Read, R. J. (1986) *Acta Crystallogr. A* 42, (140-149).