Structural and Functional Studies of *E. coli* Conjugative Relaxase-Helicase TraI and *Arabidopsis thaliana* Protein Arginine Methyltransferase 10 (PRMT10)

Yuan Cheng

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 Biochemistry and Biophysics

Chapel Hill
2011

Approved by
Advisor: Matthew Redinbo, Ph.D.
Reader: Henrick Dohlman, Ph.D.
Reader: Brian Kuhlman, Ph.D.
Reader: Matthew Wolfgang, Ph.D.
Reader: Kevin Slep, Ph.D.
Reader: Christopher, Thomas, Ph.D.

ABSTRACT

Yuan Cheng

Structural and Functional Studies of *E. coli* Conjugative Relaxase-Helicase TraI and *Arabidopsis thaliana* Protein Arginine Methyltransferase 10 (PRMT10)
(Under the direction of Professor Matthew Redinbo)

TraI, a bifunctional enzyme containing relaxase and helicase activities, initiates and drives the conjugative transfer of the *E. coli* F plasmid. In this work, we investigated the function of the putative RecD-like domain of TraI (284-821). We established that TraI 284-821 can bind nucleotides, but only retained residual ATPase activity likely due to the loss of critical catalytic motifs. We also examined the DNA binding properties of the TraI helicase. We showed that TraI binds to single-stranded DNA (ssDNA) with a site-size of approximately 25 nucleotides and low cooperativity. A double-stranded DNA (dsDNA) binding site was identified within the N-terminal region of TraI (1-858), outside the core helicase motifs of TraI. We further characterized the impacts of ionic strength, nucleotide binding and base composition on TraI-DNA interaction. Finally, we elucidated the solution structure of TraI using small-angle x-ray scattering. Taken together, these data resulted in the assembly of a model for TraI's multi-domain helicase activity.

Protein arginine methyltransferase 10 (PRMT10) regulates flowering-time in *Arabidposis thaliana*. Here, we present the 2.6 Å resolution crystal structure of PRMT10 that reveals significant structural features unique to PRMT10, including a long dimerization arm and distinct accessibility to the active site. Our data also showed that the N-terminal thirty

residues of PRMT10 impact substrate specificity, and that PRMT10 activity was dependent on the sequences distal from the substrate methylation site. We further established that PRMT10 dimerization is required for activity and used structure-based molecular dynamics to indicate how dimerization affects functionally-essential PRMT10 domain motions. Taken together, our results provide substantial insights into the mechanism governing the unique enzymatic function and substrate specificity of PRMT10.

DEDICATION

In loving memory

To my grandmother

Guizhi Su

And also, I will dedicate this thesis to my grandfather Xueliang Cheng, my father, Ruijie Cheng, my mother, Ruixiang, Guan, my wife, Hao Tang and my daughter,

Claire, Ziyun, Cheng

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my advisor, Matthew Redinbo, who gave me the opportunity to work on these extraordinarily interesting projects and provided me with his full support in my academic and professional pursuits. From Matt, I have learned how to become a good scientist and a good leader. Matt has given me a lot of freedom in designing and performing my projects, which helps me to become an independent scientist.

I also want to thank my committee members Dr. Henrick Dohlman, Dr. Brian Kuhlman, Dr. Kevin Slep, Dr. Christopher Thomas and Dr. Matthew Wolfgang for their time and support over the past five years. They have given me a lot of good advice and have been wonderful role models. Thanks also to Dr. Barry Lentz for helping me to overcome all kinds of hardships at the beginning of my graduate study at UNC.

I would like to thank everyone in Redinbo's lab for their friendship and support, especially Joseph Lomino, Monica Fraizer, Jon Edwards, Rebekah Potts, Laurie Betts and Bret Wallace for always being there with me. They are always my good teachers of English and American culture. Thanks also to former members of the Redinbo's lab, Dr. Yue Xue, Dr. Scott Lujan and Dr. Michael Miley for their friendship and support.

None of my research would have been possible were it not for the support of our collaborators, Dr. Xiaofeng Cao, Falong Lu, Dr. Christopher Thomas and Dr. Steven Matson. Thanks also to Dr. Ashutosh Tripathy for his kind assistance in using the Macromolecular Interaction Facility.

TABLE OF CONTENTS

List of Tables ix

List of Figures x

List of Abbreviations and Symbols xii

Chapter 1. Structural and functional studies of multi-domain F Plasmid TraI relaxse-Helicase 1

1.1 Introduction 1

1.2 Materials and methods 4

1.3 Results 13

1.4 Discussion 26

1.5 Figure Legends 32

1.6 Figures and Tables 39

Chapter 2. Structural and functional studies of *Arabidopsis thaliana* protein arginine methyltransferase 10 (PRMT10) 58

2.1 Introduction 58

2.2 Materials and methods 61

2.3 Results 65

2.4 Discussion 77

2.5 Figure Legend 81

2.6 Figures and Tables 86

Chapter 3. Activation of the human nuclear xenobiotic receptor PXR by the reverse transcriptase-targeted anti-HIV drug PNU-142721 99

3.1 Introduction 99

3.2 Materials and methods 100

3.3 Results 102

3.4 Discussion 105

3.5 Figure Legends 109

3.6 Figures and Tables 111

Appendix 117

References 117

List of Tables

Chapter 1

Table 1.1 Sequences of DNA oligomers used in this study ……..39

Table 1.2 ATPase activities of different TraI constructs……... 40

Table 1.3 Thermodynamic parameters of TraI binding to ssDNA oligomers……41

Table 1.4 Thermodynamic parameters of TraI binding to dsDNA oligomers……... 42

Table 1.5 Macroscopic binding constants of different TraI constructs for FL-dsDNA$_{15}$…….. 43

Chapter 2

Table 2.1 Crystallographic data and refinement statistics for PRMT10……..…….....86

Table 2.2 Oligomeric states of PRMT10 mutants……..87

Chapter 3

Table 3.1 Crystallographic statistics for the PXR-PNU142721 Complex …….. 111

List of Figures

Chapter 1

Figure 1.1 Binding of nucleotides to TraI 284-821……44

Figure 1.2 The stoichiometry of TNP-ATP binding to TraI ……..45

Figure 1.3 ATPase and GTPase activities of TraI 284-821.……...46

Figure 1.4 Sequence alignment of TraI orthologs……..47

Figure 1.5 The role of R446 in the ATPase activity of TraI……...48

Figure 1.6 Schematic representations of the domain organization of TraI and the TraI constructs used in this study……49

Figure 1.7 Determination of the stoichiometry of ssDNA binding by TraI……50

Figure 1.8 Determination of binding parameters for TraI binding to ssDNA……51

Figure 1.9 Determination of stoichiometry and binding parameters for TraI binding to dsDNA……52

Figure 1.10 Identification of the location of the dsDNA binding site……53

Figure 1.11 ADP and AMPPNP increased TraI affinity for ssDNA……..54

Figure 1.12 The role of ionic strength and base specificity in TraI DNA binding…….55

Figure 1.13 Solution structure of TraI…… 56

Figure 1.14 Model of TraI association with a dsDNA-ssDNA junction…….....57

Chapter 2

Figure 2.1 Crystal structure of PRMT10……88

Figure 2.2 Structure-based sequence alignment of PRMT10 and other PRMT paralogs……89

Figure 2.3 Dimer formation of PRMT10……90

Figure 2.4 Conserved residues between PRMT10 and other PRMT paralogs……...91

Figure 2.5 Surface representation of various PRMT paralogs……92

Figure 2.6 Methyltransferase activities of different PRMT10 constructs……..93

Figure 2.7 Surface electrostatics of PRMT10 ……..……..….94

Figure 2.8 PRMT10 exhibits a uniquely open conformation……... ..95

Figure 2.9 Conservation of total energy during PRMT10 simulations ……..……96

Figure 2.10 Effects of dimerization on the motion of PRMT10…….97

Figure 2.11 Sequence alignment of PRMT10 orthologs from various plants ……..……. 98

Chapter 3

Figure 3.1 Chemical structures of PNU-142721 and its metabolite PNU-173575……..112

Figure 3.2 Crystal structure of the human PXR-LBD complex with PNU-142721…….113

Figure 3.3 Analysis of the ligand binding pocket of PXR-LBD-PNU complex……..114

Figure 3.4 Modeled structure of human PXR-LBD in complex with PNU-173575……... 115

Figure 3.5 Modeled structure of rat PXR-LBD in complex with PNU-142721…….....116

List of Abbreviations and Symbols

Å	angstrom
A-angle	accessibility angle
ADP	adenosine diphosphate
AF-2	activation function 2
AIDS	acquired immune deficiency syndrome
ATP	adenosine triphosphate
AU	asymmetric unit
CARM1	coactivator-associated arginine methyltransferase 1
CPT	conjugative plasmid transfer
CTD	C-terminal domain
CYP	cytochrome P450
DBD	DNA binding domain
DTT	dithiothreitol
dsDNA	double-stranded DNA
E.coli	*Escherichia coli*
EDTA	ethylenediamine tetraacetic acid
FA	fluorescence anisotropy
FLC	flowering locus c
GAR	glycine and arginine rich
GDP	guanosine diphosphate
GTP	guanosine triphosphate
HIV	human immunodeficiency virus

HPLC	high performance liquid chromatography
IDT	integrated DNA technology
IHF	integration host factor
LBD	ligand-binding domain
LDH	lactate dehydrogenese
MCT	macromolecular competition titration
Min	minimum
Max	maximum
NRTI	nucleoside (or nucleotide) reverse transcriptase inhibitors
NNRTI	non-nucleoside (or non-nucleotide) reverse transcriptase inhibitors
NTD	N-terminal domain
PDB	protein data bank
PEP	phosphoenolpyruvate
PMSF	phenylmethylsulphonyl fluoride
PRMT	protein arginine methyltransferase
PXR	pregnane X receptor
PK	pyruvate kinase
RXR	retinoid X receptor
SAH	s-L-adenosylhomocysteine
SAM	s-L-adenosylmethionine
SAXS	small angle X-ray scattering
SRC-1	steroid receptor coactivator 1
ssDNA	single-stranded DNA

SF	superfamily
WT	wide type
c	concentration
g	maximal binding density
l	pathlength
m	size of DNA binding site
ω	binding stoichiometry
ε	extinction coefficient
℃	celcius degree
D_T	total DNA concentration
K_{int}	intrinsic association constant
K_N	macroscopic association constant
K_d	dissociation constant
P_N	combinatorial factor
P_F	free protein concentration
P_T	total protein concentration

Chapter 1. Structural and functional studies of multi-domain F Plasmid TraI relaxse-Helicase

(This chapter has been published as Cheng Y et al. 2011, *J Bio Chem*, Feb 2)

1.1 Introduction

Conjugative plasmid transfer (CPT) is a central mechanism for the horizontal exchange of genetic material between bacterial cells, as well as for the spread of antibiotic resistance genes and virulence factors (*1-4*). CPT requires both a DNA relaxase and a DNA helicase. The relaxase initiates DNA transfer by cleaving the transferred strand at a specific site within the *oriT*, forming a 5' phosphotyrosine intermediate. Following nicking, the helicase uses the energy from ATP hydrolysis to unwind the plasmid and drive the transfer of DNA into the recipient cell. The relaxase completes plasmid transfer by breaking the covalent phosphotyrosine linkage and releasing the transferred DNA for replication in the recipient (*5-7*).

The relaxase and helicase activities necessary for F plasmid transfer are mediated by a 190-kDa multi-domain protein, TraI (*8, 9*). TraI requires the assistance of a two additional F plasmid-encoded proteins, TraY and TraM, as well as Integration Host Factor (IHF) for conjugative transfer. These four proteins bind to the 500 bp *oriT* site and form a protein complex called the relaxosome (*5*). TraI contains three major functional domains, including the N-terminal relaxase domain (residue 1-309), a central helicase domain (residues 310-1476) and a C-terminal domain (CTD, residues 1477-1756) that is responsible for interactions with TraY and TraM (*10-13*). The region between residues 310-

950 is not fully characterized, although recent work from Schildbach *et al* suggested that it may encode a RecD-like domain (*14*). The covalent linkage between these functional domains is required for efficient conjugation (*15*).

TraI is a superfamily I helicase with ssDNA-dependent NTPase and helicase activities (*16, 17*). The translocation of TraI along DNA has 5' to 3' polarity (*18, 19*). TraI requires a 5' single-stranded overhang for the initiation of DNA unwinding (*18*). In addition, TraI functions as a highly processive helicase as a monomer (*17*). Unlike other helicases, the processivity of TraI helicase requires neither oligomerization nor the presence of a processivity factor. The underlying mechanism of TraI's processivity has remained unclear (*17*).

While the interaction of TraI with DNA is essential to its helicase activity, little is known about the DNA-binding properties of the TraI helicase. Matson *et al.* showed that deletion of the region 309-349 disrupts helicase-associated ssDNA binding and the helicase activity (*10*). Recent work from Schildbach *et al.* suggested that the N-terminal region of TraI helicase (residues 303-844) is important for its interaction with ssDNA (*14*). Numerous fundamental aspects of the interaction of the TraI helicase with DNA, such as the length of DNA bound by TraI, the intrinsic binding affinity, binding cooperativity, dependence of affinity on ionic strength, nucleotide binding, and base specificity, are still unclear. This knowledge is important for the quantitative understanding of the interaction between TraI helicase and DNA, as well as the molecular mechanism and regulation of TraI helicase activity.

Here, we examined the function of the putative RecD-like domain. Our results show that TraI 284-821 can bind nucleotides, but only retains residual ATPase activity likely due

to the loss of critical catalytic motifs. We used fluorescence anisotropy-based equilibrium binding assays to investigate TraI helicase-associated DNA binding properties. Our studies have elucidated the intrinsic binding affinity, binding cooperativity, and the lengths of ssDNA and dsDNA bound by TraI. Also, we identified the location of the dsDNA binding site within TraI. Furthermore, we investigated the impacts of ionic strength, nucleotide binding and base composition on the affinity of TraI for DNA. Finally, we examined the spatial organization of TraI domains in solution. Together, these results advance our understanding of the interaction between TraI helicase and DNA, providing insights into the molecular mechanism by which TraI performs its helicase function.

1.2 Materials and methods

DNA oligomers.

DNA oligomers used in this study were purchased from Integrated DNA Technology (IDT, Coralville, IA). Labeled oligomers have the fluorescent probe 6-FAM™ covalently attached to their 5' end and were purified using HPLC by IDT. Unlabeled oligomers were desalted without further purification. The sequences of DNA oligomers are listed in Table 1.1. All DNA oligomers were dissolved in oligomer annealing buffer containing 50 mM Tris-HCl pH 8.0, 50 mM NaCl and 5 mM $MgCl_2$. To prepare duplex DNA, two complementary DNA oligomers were mixed in a 1:1 molar ratio, heated to 95 ^{o}C and slowly cooled down to 20 ^{o}C with a speed of 1 ^{o}C/ min in a thermocycler.

DNA binding buffers.

Standard DNA binding buffer contained 25 mM Tris-HCl pH 7.5, 100 mM NaCl, 5 mM Mg acetate and 0.1% BSA. Buffer A contained 25 mM Tris-HCl pH 7.5, 50 mM NaCl, 5 mM Mg Acetate and 0.1% BSA. Buffer B contained 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 5 mM Mg Acetate and 0.1% BSA. Buffer C contained 25 mM Tris-HCl pH 7.5 and 0.1% BSA.

Protein constructs, expression and purification

Standard ligation independent cloning techniques, as described by Stols *et al.*, were employed in the construction of expression plasmids encoding the segments of TraI used in this study. *(20)*. The amplified DNA fragments were treated and cloned into empty pMCSG7-Lic-MBP expression vector. The 6-His MBP tag was cleaved off after purification

using tobacco etch virus (TEV) protease. All expression plasmids used in this study were sequence verified.

The expression plasmids were transformed into *Escherichia coli* BL21 (DE3) Gold (Stratagene). Bacteria were grown in LB medium supplemented with 50 μg/ml ampicilin at 37 ^{o}C with shaking. After the OD_{600} reached 0.6, isopropyl β-D-thiogalactoside was added to a final concentration of 0.2 mM and bacteria were grown for another 12 hours at 16 ^{o}C with shaking. Bacteria were harvested and resuspended in loading buffer (50 mM sodium phosphate pH 7.6, 500 mM NaCl, 25 mM imidazole) supplemented with 0.5 mM EDTA, 0.1% Triton X-100, 1 mM PMSF, one tablet of a protease inhibitor cocktail (Roche) and 1 mg/ml lysozyme. After 1 hour of incubation on ice, the resuspended cells were sonicated on ice for 2 min and the lysate was centrifuged at 45,000 × g for 90 min at 4 ^{o}C. The supernatant was passed through a 0.2 μm filter (Millipore) and then loaded onto a 5 ml high performance HisTrap™ column (GE Life Sciences), equilibrated with loading buffer. The column was washed with 100 ml loading buffer before bound protein was eluted with elution buffer (50 mM sodium phosphate pH 7.6, 500 mM NaCl, 500 mM imidazole). The eluted protein was pooled and loaded onto a HiPrep™ 26/10 desalting column (GE Life Sciences) equilibrated with desalting buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM DTT and 0.5 mM EDTA). Fractions containing protein were collected. TEV protease was added to the pooled protein fractions with a ratio of 1:100 (w/w) TEV to TraI. After 16 hours of incubation at 4 ^{o}C, the mixture was reloaded onto a 5 ml HisTrap™ column equilibrated with loading buffer. The flow-through fractions were collected and concentrated in a Centricon YM30 (Amicon) concentrator. Finally, concentrated protein was loaded on a HiLoad™ 16/60 Superdex 200 column (GE Life Sciences) equilibrated with sizing buffer (20 mM Tris-HCl pH 7.5, 100

mM NaCl, 5% Glycerol). Protein containing fractions were concentrated, flash frozen in liquid nitrogen and stored at -80 ^{o}C. Purified protein was >95% pure by SDS-PAGE.

TNP-ATP binding assay

The binding of TNP-ATP to TraI was measured at 22 ^{o}C in spectrofluorometer SPEX Fluorolog 2 as described by weber *et al*. with some modifications (*21*). The binding buffer contains 20 mM Tris-HCl (pH 7.5) and 20 mM NaCl. Indicated concentrations of TraI were titrated with TNP-ATP at a step of 2µl and the samples were equilibrated for 5 min before each measurement. The titration was monitored at λ_{exc}= 410 nm, λ_{em}=547 nm. The slits for excitation and emission were set to 2 nm and 5 nm respectively. All measured data points were corrected for the dilution and the inner filter effects of TNP-ATP using equation 1

$$F = F_{obsd}(V_i / V_o)10^{0.5\varepsilon cl} \quad (1)$$

where F and F_{obsd} represent the corrected and measured fluorescence intensities in arbitrary units; V_i and V_o represent the volume of the sample after the ith titration and the volume of the sample before any titration; ε is the molar absorption of TNP-ATP at the wavelength of 410 nm, c is the concentration of TNP-ATP in the sample, and l is the length of the light path (0.4 cm). V_i/V_o and $10^{0.5\,\varepsilon cl}$ serve to correct the dilution and inner filter effects respectively.

NADH-coupled NTPase assay

The NTPase activity of TraI was measured by the NADH-coupled NTPase assay as described by Kiianitsa *et al. (22)*. In this assay, the ADP produced from the hydrolysis of ATP can be regenerated by the enzyme pyruvate kinase (PK) using phosphoenolpyruvate (PEP) as the substrate. The resulting pyruvate is subsequently converted to lactate by L-

lactate dehydrogenase (LDH) at the expense of the oxidation of NADH, NADH, but not its product NAD, has a strong absorbance at 340 nm. Consequently, the hydrolysis of ATP by TraI can be monitored by measuring the decline of absorbance at 340 nm. The reaction mixture (75 μl) contains 50 mM Tris-HCl (pH 7.5), 10 mM MgAcO, 600 μM NADH, 25 mM NaCl, 25 mM K_2SO4, 0.1% BSA, 700 μM PEP, 3 unit PK and 3.6 unit LDH (Sigma), 40 nM ssDNA M13MP18, 1 μM TraI 284-821. The reaction mixture was added into a 96-well microplate (Corning Inc., NY), and the reaction was triggered by adding indicated concentrations of NTP into the reaction mixture. The absorbance at 340 nm was monitored for 30 min using Pherastar (BMG Labtech, Offenburg, Germany).

Small angle X-ray scattering (SAXS) and data analysis

SAXS data was collected for protein solutions and their matched buffers using the standard SAXS instrument at the beamline 18-ID (Bio-CAT, Advanced Photon Source, Argonne, IL). The experiments were carried out at 16 ^{o}C using an X-ray photon energy of 12 keV. The Mar165 CCD detector with an active area of approximately 160 × 80 mm^2 was used for data collection. The sample-to-detector distances were set to 2.3 m to make scattering vectors, q, range from 0.007 to 0.38 $Å^{-1}$. The scattering vector here is defined as $q = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle. Samples were flowed back and forth during SAXS measurements with a Hamilton programmable syringe pump to minimize radiation damage. Twenty frames of short exposure (2 seconds) were taken on each sample and averaged to improve the signal/noise ratio. Data were reduced using Bio-CAT SAXS data reduction macros installed in the program Igor (WaveMetrics, Portland, OR). Guinier plots, made using the program Igor, were used for detection of aggregation in protein samples.

Guinier approximation $I(q) = I(0) \exp(-q^2 R_G^2/3)$ with the limits $qR_g < 1.3$ was used to determine the radius of gyration, R_g ,and the scattering intensity extrapolated to zero angle $I(0)$. The molecular weight of TraI was determined from $I(0)$ on a relative scale using cytochrome c as a reference. The pair distribution function, *P(r)*, and the maximum dimension of the protein, D_{max}, was computed by the program GNOM (*23*).

Ab initio shape restoration and structural modeling

The low-resolution shape of TraI was restored from the experimental SAXS profile using the program DAMMIN (*24*). The data within the q range of 0.0104 to 0.2129 $Å^{-1}$ were used for the analysis. Forty individual models were generated by DAMMIN (*25*). The twenty models with the best χ^2 were aligned using SUBCOMB and averaged using DAMMAVER (*25*). The averaged model was then filtered based on occupancy and volume using DAMFILT to generate the final model of TraI. The volumetric representation of the SAXS model was prepared by SITUS (*26*).

Crystal structures of the relaxase domain (1-307) and the C-terminal domain (1476-1628) were obtained from Protein Data Bank (PDB) using the accession numbers 1P4B and 3FLD, respectively. Homology models of TraI 309-842 and 846-1473 were constructed by PHYRE (*27*) using the crystal structure of RecD from *E.coli* RecBCD complex (PDB accession number 1W36) as a template. The atomic models were placed into the *ab initio* shape by visual judgment using the program Chimera (*28*). The scattering intensities of the structural model were calculated using CRYSOL (*29*). The goodness of the structural model was assessed by CRYSOL based on the discrepancy (χ^2) between the calculated scattering intensities and the experimental scattering intensities. The orientation and position of

individual domains in the structural model were manually adjusted to minimize the discrepancy.

Direct DNA binding assays

Fluorescence anisotropy (*FA*) based DNA binding assays were employed to study DNA binding by TraI as described by Shildbach *et al.* (*30*) with some modifications. To perform the direct DNA binding assay, a 6FAM-labeled DNA oligomer was mixed with increasing concentrations of protein in DNA binding buffer in 384 well plates (Corning Inc., NY). The fluorescence anisotropy of each well was measured by a Pherastar plate reader (BMG Labtech, Offenburg, Germany) using the excitation and emission wavelengths of 485 nm and 520 nm, respectively. The fluorescence anisotropy (*FA*) was defined as

$$FA = (I_{vv} - I_{vh})/(I_{vv} + 2I_{vh}) \qquad (2)$$

Where I_{vv} and I_{vh} represent the fluorescence intensity signal parallel and perpendicular to the excitation polarization. All experiments were done in triplicate. To calculate the macroscopic DNA binding constant (K_N), normalized data were plotted as average *FA* vs. total protein concentration and fit to equation 2:

$$FA = FA_{\min} + (FA_{\max} - FA_{\min})\left\{\frac{(D_T + x + \frac{1}{K_N}) - \sqrt{(D_T + x + \frac{1}{K_N})^2 - 4D_T x}}{2D_T}\right\} \qquad (3)$$

using nonlinear regression in SigmaPlot 11.0 (Systat Software, Inc.), where *FA*, observed *FA* signal; $FA_{\min}$, *FA* signal in the absence of protein; $FA_{\max}$, *FA* signal in the presence of saturating concentration of protein; D_T, total concentration of 6FAM-labeled DNA; *x*, total

protein concentration; K_N, the macroscopic binding constant. The error bars represent the standard deviation of three replicates.

The macromolecule titration competition (MCT) method

The binding of TraI to unlabeled DNA oligomers was investigated using the macromolecular competition titration (MCT) method as describe by Bujalowski *et al*.(*31*). In brief, a 6FAM-labeled reference DNA oligomer (total concentration, D_{TR}) was mixed with increasing concentration of protein in DNA binding buffer in 384 well plates in the presence of a competing unlabeled DNA oligomer (total concentration, D_{TS}). The fluorescence anisotropy of each well was measured using a Pherastar plate reader with the excitation and emission wavelengths of 485 nm and 520 nm, respectively. The total protein concentration in the presence of a competing DNA oligomer, P_{T1}, was defined as equation 4

$$P_{T1} = \left(\sum \Theta_i\right)_R D_{TR} + \left(\sum \Theta_i\right)_S D_{TS} + P_F \qquad (4)$$

Where $(\sum \Theta_i)_R$, $(\sum \Theta_i)_S$, and P_F represent the binding density of the protein on the fluorescent DNA oligomer, the binding density on the unlabeled competing DNA oligomer, and the free protein concentration, respectively. In the absence of a competing DNA oligomer, equation 4 can be simplified as

$$P_T = \left(\sum \Theta_i\right)_R D_{TR} + P_F \qquad (5)$$

When the same *FA* signal is observed in the absence and in the presence of a competing DNA oligomer, the binding density of the protein on the unlabeled competing DNA oligomer, defined as equation 6, can be derived by solving the set of equation (4) and (5).

$$\left(\sum \Theta_i\right)_S = \frac{P_{T1} - P_T}{D_{TS}} \qquad (6)$$

When $(\sum \Theta_i)_S$ is known, P_F can be obtained by using equation 7:

$$P_F = P_{T1} - \left(\sum \Theta_i\right)_S D_{TS} - \left(\sum \Theta_i\right)_R D_{TR} \qquad (7)$$

The binding stoichiometry is equal to the value of $(\sum \Theta_i)_S$ when *FA* reaches its maximum, as determined from the plot of *FA* versus $(\sum \Theta_i)_S$.

Determination of the binding parameter

The model-independent and thermodynamically rigorous binding isotherm can then be constructed by plotting $(\sum \Theta_i)_S$ against P_F. The binding isotherms were analyzed using Epstein combinatorial approach for binding of a protein to a linear nucleic acid (*32*). The relation between $(\sum \Theta_i)$ and P_F, can be expressed as equation 8 and 9:

$$\left(\sum \Theta_i\right) = \frac{\sum_{k=1}^{g} \sum_{j=0}^{k-1} k P_N(k, j)(K_{\text{int}} P_F)^k \omega^j}{\sum_{k=0}^{g} \sum_{j=0}^{k-1} P_N(k, j)(K_{\text{int}} P_F)^k \omega^j} \qquad (8)$$

$$P_N(k, j) = \frac{(N - mk + 1)!(k-1)!}{(N - mk - k + j + 1)!(k - j)!\, j!(k - j - 1)!} \qquad (9)$$

Where k is the number of protein bound per nucleic acid, g is the maxium number of k, N is the total length of the nucleic acid, m is the size of nucleic acid binding site, K_{int} is the intrinsic binding constant, ω is the cooperativity parameter and j is the number of cooperative contacts formed between protein molecules bound to the nucleic acid. The combinatorial factor $P_N(k, j)$ defines the number of distinct ways that k protein molecules, each of which form j cooperative contacts, can bind to a nucleic acid.

When the nucleic acid can maximally bind one protein molecule, equation 7 and 8 can be simplified as equation 10 or 11:

$$\left(\sum \Theta_i\right) = \frac{(N-m+1)K_{\text{int}}P_F}{1+(N-m+1)K_{\text{int}}P_F} \qquad (10)$$

$$\left(\sum \Theta_i\right) = \frac{K_N P_F}{1+K_N P_F} \qquad (11)$$

Competition binding assay

A 6FAM-labeled DNA oligomer was used as the reference DNA. Protein was mixed with the reference DNA and titrated with increasing concentration of a unlabeled DNA oligomer (the competitor DNA). The decrease in fluorescence anisotropy values with the titration of the competitor DNA was monitored. The IC_{50}, or the concentration of competitor DNA required to displace 50% of the complex formed by protein and reference DNA, was determined by plotting the anisotropy value as a function of competitor DNA concentration and fitting the curve using equation 12:

$$FA = FA_{\text{min}} + \frac{FA_{\text{max}} - FA_{\text{min}}}{1+10^{(\log \text{IC}_{50} - x)}} \qquad (12)$$

1.3 Results

Nucleotide binding to TraI 284-821.

TraI 284-821 has been suggested to have a protein fold similar to that of the RecD helicase from the *E.coli* RecBCD complex (*33*). However, if TraI 284-821 can function as a RecD helicase has remained unknown. Here we examined the nucleotide binding capacity of TraI 284-821 by measuring its interaction with TNP-ATP as described in the materials and methods (Figure 1.1). TNP-ATP, a fluorophore-labeled ATP analog, has been commonly used in the study of protein-nucleotide interactions (*21*). Binding of TNP-ATP to a nucleotide binding site is often accompanied with an increase in TNP-ATP fluorescence, which can be used to monitor the binding process. Our results show that TNP-ATP can bind tightly to TraI 284-821 with a macroscopic association constant (K_N) of 9.8×10^6 M^{-1}.

To examine if TraI 284-821 can bind unlabeled nucleotides, a competitive binding assay was performed using TNP-ATP as a reference. In this assay, TraI 284-821 was titrated with TNP-ATP in the presence of 800 μM unlabeled ATP or ADP (Figure 1.1). If unlabeled nucleotides can compete with TNP-ATP for the nucleotide binding site, the titration curve will shift to a higher TNP-ATP concentration in the presence of unlabeled nucleotides. Indeed, our results show that the presence of unlabeled ATP or ADP shifted the titration curve to the right, or the higher TNP-ATP concentration. More prominent shift was observed in the presence of ATP. These results suggest that unlabeled ATP and ADP can bind to TraI 284-821. Moreover, TraI 284-821 has a stronger affinity for ATP than ADP.

We examined the stoichiometry of nucleotide binding to full-length TraI by titrating TraI at two different concentrations (0.5 and 3 μM) with TNP-ATP as described in the material and methods (Figure 1.2A). The degree of binding was calculated as described by

Bujalowiski *et al.(34)*. The binding stoichiometry of TraI with TNP-ATP was determined by examining the dependence of normalized fluorescence intensity on the binding density from the titration curves (Figure 1.2B). The binding stoichiometry of TraI with TNP-ATP was estimated to be two TNP-ATP molecules per TraI molecule, since the degree of binding equaled two at the maximum normalized fluorescence intensity Given that it is known that the established helicase domain TraI 822-1476 contains one nucleotide binding site, our result suggests the presence of a second nucleotide binding site outside TraI 822-1476. Considering that TraI 1-300 (the relaxase domain) and TraI 1476-1756 (the protein interaction domain) do not bind nucleotides, our results support the notion that TraI 284-821 contains a nucleotide binding site.

The NTPase activity of TraI 284-821

We examined the ATPase and GTPase activities of TraI 284-821 using the NADH-coupled NTPase assay as described in the materials and methods. The ATPase activity of full-length TraI was measured here as a reference. The results are summarized in Table 1.2. Under our assay conditions TraI 284-821 only displayed residual ATPase activity, and had no observable GTPase activity (Figure 1.3A). The catalytic efficiency of TraI 284-821 is only 0.006% of that of full-length TraI (Table 1.2). The ATPase activity of TraI 284-821 dropped for about 95% in the absence of ssDNA (Figure 1.3B), indicating that the ATPase activity of TraI 284-821 is ssDNA-dependent.

To understand why TraI 284-821 only retains residual ATPase activity, we examined the primary sequence of TraI 284-821. Sequence alignment of different TraI orthologs reveals that TraI 284-821 loses the consensus Walker A motif (GxxxGKT/S) of the RecD

helicase that is crucial for catalysis of NTP hydrolysis (Figure 1.4). The key catalytic lysine residue of the Walker A motif could be replaced by the arginine-446 in TraI 284-821. However, our results show that TraI 284-821 R446M displayed wide-type level ATPase activity, suggesting that arginine-446 is not the catalytic residue of TraI 284-821. Our results suggest that TraI 284-821 only retains residual ATPase activity likely due to the loss of critical catalytic motifs.

Binding of TraI helicase to ssDNA: determination of the site-size and DNA binding parameters.

The interaction between protein and DNA can be studied conveniently by quantitative titrations of a fluorescently labeled DNA with the protein of interest. However, the binding parameters obtained in this manner are often inaccurate due to the interference from the fluorescent probe. To understand the intrinsic DNA-binding property of TraI, we have studied the interaction of TraI with unlabeled DNA oligomers using the Macromolecule Competition Titration (MCT) method as described by Bujalowski *et al*. (*31*) (see also materials and methods). This method has been successfully applied to the analysis of the interactions of the PriA and RepA helicases with DNA (*35*, *36*). In brief, a fluorescently labeled DNA oligomer, or the reference DNA, is titrated with TraI in the absence or presence of an unlabeled DNA oligomer, or the competing DNA, whose binding parameters are to be determined. Because the association of the unlabeled ssDNA oligomer with TraI does not give significant anisotropy signal, the presence of the unlabeled DNA oligomer will shift the titration curve to higher TraI concentration as compared to the curve generated in the absence of competitor. Based on the shift of the titration curve, this method allows the determination of the binding stoichiometry of TraI with an unlabeled DNA oligomer. The MCT method

also enables the construction of a model-independent binding isotherm, which can be used for the determination of thermodynamically rigorous binding parameters for the unlabeled DNA oligomer. All protein constructs and DNA oligomers used for this study are listed in Figure 1.9 and Table 1.1.

The interaction of TraI with a series of unlabeled ssDNA oligomers of different lengths was studied using the MCT method (*31*). Titration of 50 nM FL-T_{17} (fluorescently-labeled T_{17}) with TraI in the absence or presence of 200 nM or 350 nM ssDNA oligomers of lengths ranging from 13 to 63 nucleotides were conducted in standard DNA binding buffer (Figure 1.7A and 1.7B). The unlabeled 13-, 15- and 17-mer ssDNA did not cause significant shift of the titration curves to higher TraI concentration, suggesting that TraI cannot bind tightly to unlabeled 13-, 15- and 17-mer under our assay conditions.

The binding stoichiometry of TraI with the 39-mer was determined by examining the dependence of anisotropy values on the binding density from the titration curves (Figure 1.7C). The binding stoichiometry of TraI with the 39-mer was estimated to be one, since binding density equaled one at the maximum anisotropy value of 291. Analogous analyses of the binding stoichiometry were performed for DNA oligomers varying in length from 17 nucleotides to 63 nucleotides. The dependence of maximum stoichiometry on the length of ssDNA oligomers indicated that 51- and 63-mers can accommodate two TraI molecules, while ssDNA oligomers shorter than 45 nucleotides can only accommodate one TraI molecule (data not shown).

To determine the binding parameters of TraI for the ssDNA oligomers, we constructed model-independent binding isotherms for each ssDNA oligomers. Since ssDNA 13-, 15-and 17-mer did not create a significant shift in the titration curve, their binding

isotherms could not be accurately determined. The binding isotherms for the 19-mer and longer oligomers are shown in Figure 1.8A and 1.8B. The macroscopic binding constants (K_N) for 31-mer and below were determined by analyzing the binding isotherms with equation 11 since these DNA oligomers can only accept one TraI molecule; the results are summarized in Table 1.4. The dependence of K_N on the length of ssDNA oligomers (N) can be clearly divided into two phases. K_N increased relatively slowly when N < 25, but faster when N >25. A roughly linear relationship between K_N and N was observed when N ranges from 25 to 31 nucleotides (Figure 1.8C). An explanation for this linear relationship is that TraI experiences multiple potential binding sites on a ssDNA longer than 25 nucleotides, indicating that the ssDNA binding site size of TraI is about 25 nucleotides (*32, 37, 38*). The observed dramatic increase in K_N was caused by a statistical factor arising from the existence of multiple potential binding sites, instead of the formation of more contacts between TraI and ssDNA oligomers. This statistical factor can be defined in terms of the intrinsic binding constant (K_{int}) and the total site-size (m) as (*35, 36*):

$$K_N = (N - m + 1)K_{int} \qquad (13)$$

This empirical linear relationship usually appears only when the length of the DNA oligomer is longer than the total site-size. In contrast, the slow increase of K_N with N when N<25 likely reflects the formation of increasing contacts between TraI and DNA. The intrinsic binding constants (K_{int}) for the ssDNA that bind one TraI molecule were determined by analyzing the binding isotherms in Figure 1.8A using equation 10 under the condition m=25. The results were summarized in Table 1.4. The K_{int} did not change significantly with the length of the ssDNA oligomers, ranging from 1.8 ± 0.5 to $2.6 \pm 0.3 \times 10^7$ M^{-1}. This result supports the estimation of the total site-size to be 25 nucleotides.

Since the total site-size is approximately 25 nucleotides, complete engagement of the ssDNA binding site of two TraI molecules should require ssDNA of at least 50 nucleotides. To determine the cooperativity parameter (ω) for TraI binding to ssDNA, we analyzed the interaction of TraI with the ssDNA 63-mer. The corresponding binding isotherm, shown in Figure 1.8B, was analyzed using the Epstein Equation under the condition m=25 and g=2. The K_{int} and ω were determined to be 6.7 ($\pm$ 0.2) $\times 10^5$ M^{-1} and 33.4 ($\pm$) 2.4, respectively. Our data clearly indicate that TraI binds to ssDNA oligomers with low cooperativity.

Binding of TraI to dsDNA: determination of the site-size and binding parameters.

We next examined the interaction of TraI with unlabeled dsDNA using the MCT method (*31*). Titration of 100 nM FL-dsDNA$_{15}$ with TraI in the absence or presence of 500 nM unlabeled dsDNA oligomers of length ranging from 9 to 27 base pairs in buffer A was performed (Figure 1.9A). Oligomers shorter than 9 base pairs were not studied here because they have low melting temperatures and are unstable at the temperature of the experiment (25 ^oC). The binding stoichiometry of TraI with dsDNA oligomers was estimated from the plot of fluorescence anisotropy versus binding density (Figure 1.9B), as described above. The dependence of binding stoichiometry on the length of dsDNA oligomers indicated that the 19-, 23- and 27-mer dsDNA can accept two TraI molecules, while 11- and 15-mer dsDNA can only bind one TraI molecule (Table 1.5). The fact that a 19-mer, but not a 15-mer dsDNA, can accommodate two TraI molecules indicates that each TraI molecule encompasses at least 8~9 base pairs when forming a stable complex with a dsDNA.

Next, the intrinsic binding constant (K_{int}) was determined by fitting the binding isotherms (Figure 1.9C) with the Epstein Equation under the conditions that the site-size m=9 and the maximum binding density g=1 (for 9-, 11- and 15-mer) or g=2 (for 19-, 23- and 27-

mer). These results are summarized in Table 1.5. The large error in K_{int} for the 9-mer resulted from the low binding density of TraI on the 9-mer under our assay conditions (Figure 1.9C). Our data show that TraI has a similar K_{int} for dsDNA of lengths ranging from 11 to 27 base pairs, supporting the hypothesis that TraI contacts approximately 9 bps when it forms a complex with dsDNA. TraI does not bind dsDNA oligomers in a highly cooperative fashion, as evidenced by ω values ranging from 14.8 ± 2.2 to 38.6 ± 7.7 for dsDNA ranging in length from 19 to 27 base pairs.

Identification of the dsDNA binding site.

To identify the dsDNA binding site, we designed a series of TraI constructs based on the proposed domain organization of TraI (Figure 1.6) and examined their interaction with FL-dsDNA$_{15}$ using the direct DNA binding assay (*30*) (Figure 1.10A). Macroscopic binding constants, K_N, were determined by fitting the titration curves using a one-site binding model (equation 3) and are summarized in Table 1.5. The affinity of TraI 1-1476 and TraI 1-858 for FL-dsDNA$_{15}$ were comparable to that of full-length TraI, suggesting that the dsDNA binding site is located N-terminal to residue 858. TraI 1-330, a construct that contains the relaxase domain, cannot bind FL-dsDNA$_{15}$ under our assay condition (Figure 1.10A). However, the deletion of 1-301 dramatically reduced the affinity of TraI for FL-dsDNA$_{15}$. The binding affinity of TraI 302-1756 for FL-dsDNA$_{15}$ is about 20-fold lower than that of full-length TraI. These results indicate that while the dsDNA binding site is primarily located between residues 302-858, the relaxase domain facilitates the binding of dsDNA.

A helicase-associated ssDNA binding site is located at the N-terminal region of TraI (302-820) (*14*). Therefore, it is possible that the dsDNA binding site identified here overlaps

with the ssDNA binding site. To test this possibility, we performed a competition binding assay using FL-dsDNA$_{15}$ as the reference DNA and the ssDNA oligomer T_{25} as the competitor DNA. Under our experimental conditions, a complex of TraI with FL-dsDNA$_{15}$, but not with T_{25}, can generate a fluorescence anisotropy signal. If T_{25} effectively competes with FL-dsDNA$_{15}$ for the dsDNA binding site of TraI, the total anisotropy value will drop with an increase in the concentration of T_{25}. Indeed, the total anisotropy value decreased with the increase of T_{25} concentration with an apparent IC_{50} of 1.5 μM (Figure 1.10B). This suggests that the binding of T_{25} and FL-dsDNA$_{15}$ by TraI occurs along an overlapping, if not identical binding surface.

Modulation of DNA binding affinity by nucleotides.

We evaluated the effect of nucleotides on the affinity of TraI for unlabeled ssDNA oligomers using the MCT method (*31*). Titration of 50 nM FL-T_{17} with TraI in the absence or presence of T_{25} and in the absence or presence of ADP or the nonhydrolyzable ATP analogue, AMPPNP, was performed in buffer B (Figures 1.14A, 1.14B, 1.14C), and a model-independent binding isotherm was constructed (Figure 1.11D). The K_{int} for T_{25} in the absence of any nucleotides was determined to be 2.8 (± 0.4) × 10^7 M^{-1} by analyzing the binding isotherm using equation 10 under the condition m=25. Since TraI can slowly hydrolyze ATP in the presence of T_{25} under our assay conditions, the nonhydrolyzable ATP analogue (AMPNP) was used instead of ATP. The presence of 1 mM ADP and AMPPNP increased the K_{int} of TraI for T_{25} to 1.2 (± 0.3) × 10^8 and 1.4 (± 0.4) × 10^8 M^{-1}, respectively. These results demonstrate that the binding of ADP and AMPPNP slightly enhance the affinity of TraI for ssDNA.

Effect of ionic strength on DNA binding.

To further define the DNA binding properties of TraI, we examined the effect of ionic strength on the affinity of TraI for ssDNA and dsDNA using the direct DNA binding assay (*30*). Titration of 50 nM FL-T_{25} was performed in buffer C supplemented with increasing concentrations of NaCl: 50, 75, 100 and 125 mM. To avoid the interference of Mg^{2+}, buffer C did not contain magnesium acetate and was supplemented with 1 mM EDTA. Analogous experiments were performed for FL-$dsDNA_{15}$. The affinity of TraI for ssDNA and dsDNA decreased with the increase of NaCl concentration (Figure 1.12A). The K_{int} for FL-T_{25} is 1.1 ($\pm$ 0.3) $\times 10^8$ M^{-1} when [NaCl]=50 mM. A 3.4-fold decrease in K_{int} was observed when [NaCl] was increased to 125 mM NaCl. More dramatic change in K_{int} was observed with FL-$dsDNA_{15}$, with K_{int} decreasing about 36-fold when [NaCl] increased from 50 to 125 mM (Figure 1.12A). Within experimental error, a linear relationship between ln(K_{int}) and ln[NaCl] was detected for both FL-T_{25} and FL-$dsDNA_{15}$, with slopes of -1.2 $\pm$ 0.2 and -3.7 $\pm$ 0.2, respectively, suggesting that the binding of TraI to FL-T_{25} and FL-$dsDNA_{15}$ was accompanied with the net release of about one and four ions respectively. These results suggest that electrostatic interactions are crucial for the stability of TraI-DNA complex. Moreover, electrostatic interactions play a more important role in the interaction of TraI with dsDNA.

Base specificity of DNA binding.

Base specificity in interactions of TraI with unlabeled ssDNA oligomers was addressed using the MCT method (*31*). Titration of 50 nM FL-T_{17} with TraI in the absence or presence of 250 nM unlabeled 25-mer ssDNA (Figure 1.12B). This 25-mer ssDNA contains

a mixture of four types of bases and does not form any secondary structure or duplex under our assay conditions. The presence of the 25-mer ssDNA shifted the titration to higher protein concentration, suggesting competition between the 25-mer ssDNA and FL-T_{17} (Figure 1.12B). A model-independent binding isotherm for the 25-mer ssDNA and subsequent analysis using equation 9 under the condition m=25 gave an intrinsic binding constant, K_{int} = 8.7 (± 0.3) $\times 10^6$ M^{-1} (Figure 1.12C). Analogous titrations of FL-T_{17} in the presence of T_{25} or C_{25} indicate a pronounced shift in titration curves (Figure 1.12B). These data indicate that C_{25} and T_{25} compete more efficiently with FL-T_{17} for TraI than does the 25-mer ssDNA. Binding of TraI to T_{25} and C_{25} were characterized by a K_{int} = 2.5 (± 0.2) $\times 10^7$ and 5.8 (± 0.1) $\times 10^7$ M^{-1}, respectively. These data indicate that TraI prefers to bind ssDNA oligomers containing a single type of base, with a further preference for pyrimidine oligomers. G_{25} was not examined here because it tends to form a cruciform structure and is not easily synthesized. The presence of A_{25} dramatically interfered with the fluorescence anisotropy signal, making the titration curves in the absence or presence of A_{25} incomparable. As a result, the binding constant for A_{25} was not determined by the MCT method.

Spatial organization of the TraI domains.

Although the crystal structures of several domains of TraI are available, there is no known structure of a full-length TraI. To elucidate the spatial organization of the TraI domains, we examined full-length TraI by SAXS, which allows the construction of a low-resolution structural envelope for a macromolecule. Many important structural parameters, such as the radius of gyration (R_g) and the maximum dimension (D_{max}), can also be obtained

by analyzing the SAXS profile. The SAXS experiments were performed on TraI solutions over a concentration range of 0.5-2 mg/ml. The SAXS profile and the linearity of the Guinier region indicated that TraI was well behaved and free of aggregation (Figure 1.13A). The R_g for TraI as obtained by the Guinier approximation was 58.5 ± 0.9 Å (Figure 1.13A inset). The D_{max} of TraI was 220 Å as determined from the $P(r)$ function, calculated using the program GNOM (Figure 1.13B). The $P(r)$ function reflects the probable distribution of inter-atomic distances within the scattering particles. The asymmetric feature of the $P(r)$ function suggested that TraI exists in an elongated shape in solution.

To better define the geometric shape of TraI, we performed *ab inito* shape restoration of TraI using the program DAMMIN (*24*). Forty independent runs were performed and initial parameters were intentionally varied between runs. Independent runs generated very similar structural envelopes, signifying consistency between different runs and reliability of the generated structural envelopes. The forty structural envelopes were ranked based on their respective χ^2. The top twenty envelopes ($\chi^2 \approx 0.34$) were averaged to generate the final SAXS envelope that is approximately 220 Å × 82 Å × 65 Å (Figure 1.13C). The envelope features a highly extended structure with two protuberances at each end of the longest dimension. This structural envelope was used as a guide for the determination of the spatial organization of TraI domains in the following analysis.

In combination with atomic models, the structural envelope generated from SAXS data can be used to determine the domain organization of a multi-domain protein. While crystal structures are available for the relaxase domain (1-236, 267-307) and part of the C-terminal domain (1476-1628), atomic structures are missing for the rest of the protein. We used the program PHYRE (*27*) to generate homology models for these two regions. The

homology models of TraI 309-842 and 846-1473 were successful generated by PHYRE using the crystal structure of *E.coli* RecD (PDB entry, 1W36) as a template. Due to the lack of homology, 56 and 74 residues are missing from the final models of TraI 309-842 and 846-1473 respectively. Also, PHYRE failed to generate a model for TraI 1629-1756 due to the lack of a homologous template. As a result, the region encompassing residues 1629-1756 was not incorporated into the following analysis.

A rigid-body model of TraI was generated by manual placement of TraI domains into the SAXS envelope. The flat and extended shape of the TraI envelope, and the constraints of inter-domain connectivity, allowed relatively little ambiguity in determining the position of the four structural domains of TraI. The crystal structure of the relaxase domain was manually placed into the protuberance at one end of the envelope initially. Then, homology models of the two RecD-like domains were placed into the main body of the envelope, which was the only portion of the envelope that could accommodate the combined size of these two domains. Numerous positions and orientations of these two RecD-like domains were examined to minimize the discrepancy (χ^2) between the calculated scattering intensities and the experimental scattering intensities. The connectivity of contiguous domains was used as a constraint for placement. Finally, the crystal structure of the C-terminal domain was modeled into the protuberance at the other end of the envelope. Part of the envelope at the C-terminal end is empty due to the missing of residues 1629-1756. The rigid-body model of TraI generated here has acceptable geometry and gave a satisfactory χ^2 of 5.0 (Figure 1.13A), especially considering that 292 of 1756 residues are missing from the atomic models. In this model, the four domains of TraI aligned along a single axis, with the relaxase domain and the

C-terminal domain at the opposite ends, and the two RecD-like domains juxtaposed in the middle (Figure 1.13D).

1.4 Discussion

TraI has a long ssDNA binding site.

Most helicases require the presence of a single-stranded tail for the unwinding of duplex DNA. Typical superfamily I helicases, such as UvrD and RecD, can efficiently unwind a duplex DNA with a single-stranded tail of 10~12 nucleotides (*39, 40*). However, TraI requires at least a 27-nucleotides tail for efficient unwinding and cannot unwind duplex DNA if the tail is shorter than 20 nucleotides (*41*). Quantitative analysis in this work suggests that the unusually long single-stranded tail required by TraI for DNA unwinding arises from its unique ssDNA binding property. TraI only weakly binds ssDNA oligomers shorter than 19 nucleotides, potentially explaining why TraI cannot unwind duplex DNA when its single-stranded tail is shorter than 20 nucleotides. Considering that the ssDNA binding site of TraI can encompass up to 25 nucleotides, and that, for optimal helicase activity, the minimal single stranded overhang requirement is 27 nucleotides, it appears that full engagement of the ssDNA binding site by DNA may trigger efficient DNA unwinding by TraI.

TraI binds to ssDNA with low cooperativity.

TraI displayed low cooperativity ($\omega = 33.4 \pm 2.4$) when binding to a ssDNA oligomer that can accept two TraI molecules, indicating that no significant cooperative interactions existed between the two bound TraI molecules. High cooperativity (ω values between 10^2 to 10^5) is often associated with multimeric DNA binding proteins and proteins whose main function is to cover DNA (*42, 43*). Previous studies show that TraI functions as a processive helicase as a monomer *in vitro* (*17*). The low DNA binding cooperativity observed here supports the monomeric nature of TraI. Low cooperativity during DNA binding is also

associated with many other helicases and DNA binding proteins that function as a monomer, including helicase RecQ (*44*) and ssDNA binding protein ICP8 (*45*). It is interesting to note that the low cooperativity during DNA binding might also reflect a physiological adaptation of TraI to function as a single protein in the context of the relaxosome. Many DNA helicases, which function as a part of a large machinery, such as PriA and DnaB, have been reported to display low cooperativity during DNA binding (*34, 46, 47*).

The dsDNA binding site might play a role in the assembly of the relaxosome.

It is of interest to find that TraI can bind dsDNA in the absence of any other protein *in vitro*. The intrinsic binding constant for dsDNA was determined to be about 5×10^5 M^{-1} (Table 1.5), which is significantly lower than that for ssDNA (Table 1.4). The low affinity of TraI for dsDNA might explain why TraI-dsDNA binding was not detected in the study by Matson *et al.*, in which the highest TraI concentration examined was 100 nM (*48*). Our deletion studies show that the dsDNA binding site is located at the N-terminal region (1-858) of TraI and likely overlaps with the helicase-associated ssDNA binding site. Given that the affinity of TraI for dsDNA is at least one magnitude lower than its affinity for ssDNA, dsDNA binding will likely be outcompeted by ssDNA binding when ssDNA, long enough to form stable complex with TraI, is available. Therefore, we consider that this dsDNA binding site will not be engaged during the unwinding of duplex DNA when a long stretch of ssDNA is available. However, this dsDNA binding site might play a role in the assembly of the relaxosome when ssDNA is not available. Biochemical studies indicate that TraI alone can bind to a supercoiled dsDNA that contains the *oriT* site *in vitr*o (*49*). The interaction, however, is sensitive to NaCl concentration and completely inhibited when the NaCl

concentration is above 75 mM (*48*). The binding of TraI to supercoiled dsDNA might be mediated by the dsDNA binding site identified here. Our observation that the affinity of TraI to dsDNA is very sensitive to NaCl concentration explains the sensitivity of relaxosome formation to NaCl concentration.

Spatial organization of TraI domains.

TraI is a bifunctional protein which contains four major structural domains. Understanding the spatial organization of these domains will provide insight into the relative contribution of each domain to TraI function. The SAXS-generated model of TraI reveals that the protein assumes an extended shape in solution with the four structural domains aligning along one axis. The relaxase domain is located at one end of the model and creates a limited number of contacts with the remainder of the protein (Figure 1.13D). This domain configuration provides the relaxase domain significant flexibility that is likely essential for the function of TraI. The two RecD-like helicase domains are modeled into the middle of the SAXS envelope and together form multiple intimate contacts, suggesting they may function as a unit. The CTD is at the opposite end of the model relative to the relaxase domain, consistent with the independent activities of these two domains. The extended conformation of TraI allows for the formation of contacts between each sequential domain of TraI, as well as between TraI and other proteins. This network of contacts may play a role in the regulation of the relaxase and helicase activities of TraI.

A model of the TraI helicase in complex with a DNA substrate containing a dsDNA-ssDNA junction was assembled based on the SAXS envelope (Figure 1.14). In this model, the location of the dsDNA-ssDNA junction was determined based upon the location of the

"pin" domain, or subdomain 1B, of RecD-like domain II, which has been proposed to contact duplex DNA in structural studies of RecD proteins (*50*). The 5' ssDNA overhang binds to the long ssDNA binding groove formed by the two RecD-like domains. Two possible end structures of the 5' ssDNA overhang are presented here. Based on the common belief that one TraI molecule carries out both nicking at *oriT* and ensuing duplex DNA unwinding, the 5' end of the ssDNA will be covalently attached to the relaxase domain. Recall that the catalytic tyrosine residue of the relaxes domain forms covalent linkage with the 5' end of ssDNA upon nicking at *oriT*. Recent discoveries from Schildbach *et al*. that nicking at o*riT* and ensuing duplex DNA unwinding could be performed by two different TraI molecules (*51*) suggests that the 5' end of the ssDNA could also stay unattached during the unwinding. The 3' ssDNA overhang is not bound by the protein in this model, due to the fact that the presence of a 3' ssDNA overhang does not enhance the affinity of TraI for DNA (unpublished data).

The contribution of RecD-like domain I to the processivity of TraI helicase.

TraI functions as a highly processive monomeric helicase (*17*). The helicase domain of TraI contains two RecD-like domains, one with intact helicase motifs (domain II) and one in which the helicase motifs are not present (domain I) (*14*) (Figure 1.6). The N-terminal RecD-like domain I retains ssDNA binding capacity (*14*). Our SAXS model suggests that the two RecD-like dains are closely associated with one another and may form a continuous ssDNA binding groove. Similar domain organization has been observed in the structure of the multimeric *E.coli* helicase RecBCD (*52*). Alone, RecB is a monomeric 3'-5' helicase with limited processivity. Similar to the RecD-like domain I of TraI, RecC has a helicase-like

fold but lacks both helicase activity and key helicase motifs. Together, RecBC, a heterodimer formed by RecB and RecC, has significantly higher processivity than RecD alone (*52*, *53*). The crystal structure of the complex clearly shows that the DNA binding sites of RecB and RecC are brought together in the heterodimer, forming a long and continuous ssDNA binding groove (*52*). The extended binding groove likely facilitates the association of RecB with ssDNA during translocation. We propose that in TraI, RecD-like domain I functions as a processivity domain to assist the motor domain, RecD-like domain II, in a manner analogous to the RecBC complex.

AMPPNP and ADP enhance the affinity of TraI for ssDNA.

Helicases utilize the energy from NTP hydrolysis to unwind nucleic acids. During each cycle of the NTPase reaction, helicases process through a number of distinct nucleotide binding states including unbound, NTP-bound and NDP-bound. In current models, the transitions between nucleotide binding states during the NTPase reaction result in conformational changes in the nucleic acid binding site of the helicase, thereby driving the translocation of the helicase along the nucleic acid (*54*). To understand how the conformational states of TraI are controlled by its nucleotide binding states, we examined the effects of nucleotide concentration on the affinity of TraI for ssDNA. Our data illustrate that the presence of AMPPNP or ADP results in an approximate 3-fold increase in ssDNA binding affinity by TraI. This indicates that TraI exists in two distinct conformational states during each cycle of the NTPase reaction. When the nucleotide binding site is empty, TraI stays in a conformational state that has relatively low affinity for ssDNA (the loose state). The binding of ATP switches TraI into a conformational state that binds ssDNA with higher

affinity (the tight state). At the end of the cycle, TraI returns to the loose state following the release of ADP. The presence of two different conformational states during each NTPase cycle has been observed in most established helicases, while the detailed mechanism can vary. For example, the Rep helicase stays in a tight DNA binding state in the absence of any nucleotide. It switches into a loose DNA binding state upon the binding of ATP. Upon the hydrolysis of ATP, the Rep helicase returns back to the tight state (*55*).

In summary, we have integrated biochemical and structural data to provide the first comprehensive model of the interaction of the TraI helicase with DNA. This model provides insight into the mechanism of the TraI helicase, and in particular suggests how the TraI helicase can achieve such an exceptional processivity through the cooperation of the two RecD-like domains.

1.5 Figure Legends

Figure 1.1 Binding of nucleotides to TraI 284-821. Titration of full-length TraI with TNP-ATP was preformed at 22 ºC in the absence or presence of 800 μM unlabeled ATP or ADP as described in the materials and methods. The binding buffer contained 20 mM Tris-HCl (pH 7.5) and 20 mM NaCl. The solid lines are used to separate the data and do not have any theoretical basis.

Figure 1.2 The stoichiometry of TraI binding to TNP-ATP. A) Full-length TraI was titrated with TNP-ATP at two different concentrations at 22 ℃ as described in the materials and methods. The binding buffer contained 20 mM Tris-HCl (pH 7.5) and 20 mM NaCl. B) Dependence of the normalized fluorescence increase on the degree of binding. The sold line represents a linear computer fit of the data. The maximum relative fluorescence increase is indicated by a dashed horizontal line. The binding stoichiometry was estimated to be two ATP molecules per full-length TraI molecule.

Figure 1.3 The ATPase and GTPase activities of TraI 284-821. A) The ATPase and GTPase activities of TraI were measured at 37 ℃ by the NADH-coupled NTPase assay as described in the materials and methods. The reaction mixture (75 μl) contained 50 mM Tris-HCl (pH 7.5), 10 mM MgAcO, 600 μM NADH, 25 mM NaCl, 25 mM K_2SO4, 0.1% BSA, 700 μM Phosphoenolpyruvate (PEP), 3 unit PK and 3.6 unit LDH, 40 nM ssDNA M13MP18, 1 μM TraI 284-821 and indicated concentrations of ATP/GTP. The kinetic parameters were calculated by fitting the data with the Michaelis-Menten equation and

summarized in Table 1.2 and 1.3 B) The ATPase activity of TraI 284-821 in the presence or absence of 40 nM ssDNA M13MP18.

Figure 1.4 Sequence alignment of TraI orthologs. A) Sequence alignment of different TraI orthologs. Only the sequence around the putative Walker A motif (Motif I) and Walker B motif (Motif II) of TraI 284-821 are shown. The critical residues are higlighgted with stars. B) Comparison of the Motif I and Motif II of TraI 284-821 with the classical Walker A and Walker B motifs. x and h respresent any amino acid and hydrophobic amino acids respectively. The catalytic lysine residue of the classical Walker A motif and the putative catalytic arginine residue (R446) of the motif I are shown in bold and italics.

Figure 1.5 The role of R446 in the ATPase activity of TraI. The ATPase activities of wide-type TraI 284-821 and TraI 284-821 R446M were measured at 37 ℃ by the NADH-coupled NTPase assays as described in the materials and methods.

Figure 1.6 Schematic representations of the domain organization of TraI and the TraI constructs used in this study. The N-terminal relaxase domain contains residues 1-309; helicase domain, 310-1476 and the C-terminal domain (CTD), 1477-1504. The asterisk indicates the location of the catalytic residue K998. RecD-like domain I (303-844) and RecD-like domain II (830-1473) identified by sequence analysis (*14*) are highlighted in dark lines. TraI constructs are shown in open boxes, whose first and last residues are indicated.

Figure 1.7 Determination of the stoichiometry of ssDNA binding by TraI. Titration of 50 nM FL-T_{17} with TraI was performed in the presence or absence of 200 nM T_n (n ranges from 13 to 31) (A) and 350 nM 39-mer, 45-mer, 51-mer and 63-mer ssDNA (B) in standard DNA binding buffer at 25 ^{o}C. The solid lines are used to separate the data and do not have any theoretical basis. C) Dependence of observed anisotropy on the total averaged binding density of TraI on the 39-mer. This plot was constructed based on the titration curve in (B) using the MCT analysis (*31*). The short solid line indicates the maximum value of the observed anisotropy. The stoichiometry was estimated to one TraI molecule per 39-mer from this plot.

Figure 1.8 Determination of binding parameters for TraI binding to ssDNA. *A*) Model-independent binding isotherms of TraI to unlabeled ssDNA T_n (n ranges from 19 to 31). The binding density and the corresponding concentration of free TraI were calculated from the titration curve in Figure 2*A* using the MCT analysis (*31*). The solid lines are computer fits of the data using equation 10 under the condition m=25 or equation 11. The macroscopic binding constants (K_N) and intrinsic binding constants (K_{int}) derived from curve-fitting are summarized in Table 1.4. *B*, Model-independent binding isotherm of TraI to 63-mer ssDNA. The solid line is computer fit of the data using the Epstein combinatorial approach (equation 8 & 9) under the condition m=25 and g=2. The best fit gives $K_{int} = 6.7(\pm 0.2) \times 10^5$ M^{-1} and $\omega = 33.4 \pm 2.4$. *C*, Dependence of K_N on the length of the ssDNA oligomers. The solid line is the linear fit of the data points for T_{25}, T_{27}, T_{29} and T_{31}.

Figure 1.9 Determination of stoichiometry and binding parameters for TraI binding to dsDNA. *A,* Titration of 100 nM FL-dsDNA$_{15}$ with TraI was performed in the absence or presence of 500 nM dsDNA of different lengths (ranging from 9 to 27 residues) in buffer A at 25 ^{o}C. The solid lines are used to separate the data and do not have any theoretical basis. *B*, Dependence of observed fluorescence anisotropy on the total averaged binding density of TraI on the 15-mer dsDNA. The short solid line indicates the maximum value of the observed anisotropy. *C*, Model-independent binding isotherms of TraI to unlabeled dsDNA of different lengths, ranging from 9 to 27 base pairs. The binding density and the concentration of free TraI were calculated from the titration curves in (*A*) using the MCT analysis (*31*). The solid lines are computer fits of the data using the Epstein combinatorial approach (equation 8 & 9). The parameters were m=9 and g=1 for dsDNA$_9$, dsDNA$_{11}$ and dsDNA$_{15}$ and m=9, g=2 for dsDNA19, dsDNA$_{23}$ and dsDNA$_{27}$. The intrinsic binding constants (K_{int}) and cooperativity parameters (ω) derived from computer fits are summarized in Table 1.5.

Figure 1.10 Identification of the location of the dsDNA binding site. A, Titrations of 50 nM FL-dsDNA$_{15}$ with different TraI deletion mutants were performed using the direct DNA binding assay in standard DNA binding buffer at 25 ^{o}C. The solid lines are computer fits of the binding curves using a one-site binding model (equation 3). The macroscopic binding constants, K_N, derived from the computer fits are summarized in Table 1.6. *B*, Competition of dsDNA and ssDNA for the binding of TraI. Titration of a mixture of 100 nM FL-dsDNA$_{15}$ and 100 nM TraI with increasing concentration of 25-mer ssDNA, T$_{25}$, was performed in

buffer A at 25 ^{o}C. The solid line is the computer fit of the curve using non-linear least square regression.

Figure 1.11 ADP and AMPPNP increased TraI affinity for ssDNA. Titration of 50 nM FL-T_{17} with TraI was performed in the absence or presence of 350 nM T_{25} in buffer B (A)), supplemented with 1 mM AMPPNP (*B*) or 1 mM ADP (*C*) at 25 ^{o}C. The solid lines are used to separate the data and do not have theoretical basis. (d) Model-independent binding isotherms of TraI to T_{25} in the absence or presence of 1 mM ADP or AMMPNP. The binding density and the concentration of free TraI were calculated from the titration curves in (A)), (*B*), and (*C*) using the MCT analysis (*31*). The solid lines are computer fits of the data using the Epstein combinatorial approach (equation 10) under the condition m=25. The intrinsic binding constants (K_{int}) for T_{25} in the absence of any nucleotide or in the presence of 1mM ADP or AMPPNP are 2.8 (± 0.4) $\times 10^{7}$, 1.4 (± 0.4) $\times 10^{8}$, and 1.2 (± 0.3) $\times 10^{8}$ M^{-1}, respectively.

Figure 1.12 The role of ionic strength and base specificity in TraI DNA binding. A), Dependence of binding affinity on NaCl concentration. 50 nM FL-T_{25} or FL-dsDNA$_{15}$ were titrated with TraI in buffer C supplemented with different concentrations of NaCl: 50, 75, 100 and 125 mM. The titration curves are not shown here. The intrinsic binding constants, K_{int}, at different NaCl concentrations were determined by fitting the titration curves using the Epstein combinatorial approach (equation 10). The data presented here are averages from three independent experiments. The solid lines are linear least-square fit of the data. The slopes of the computer fits for FL-T_{25} and FL-dsDNA$_{15}$ are -1.2 ± 0.2 and -3.7 ± 0.2 respectively. *B.* Binding of TraI to ssDNA with different base compositions. Titration of 50

nM FL-T_{17} with TraI was performed in the presence or absence of 250 nM N_{25}, T_{25} and C_{25}, in standard DNA binding buffer at 25 ^{o}C. The solid lines are used to separate the data and do not have any theoretical basis. *C*, Dependence of binding density on the concentration of free TraI. The binding density and the concentration of free TraI were calculated from the binding curve in (*B*) using the MCT analysis (*31*). The solid lines are computer fits of the data using the equation 10. The K_{int} for N_{25}, $T_{25,}$ and C_{25}, are 5.4 ($\pm$ 1.1) $\times 10^{6}$, (2.5 $\pm$ 0.2) $\times 10^{7}$, and 5.8 ($\pm$ 0.1) $\times 10^{7}$ M^{-1}, respectively.

Figure 1.13 Solution structure of TraI. A), SAXS profiles of TraI. The theoretical scattering profile of the rigid-body model for TraI (figure 8*D*) is shown by the red dashed line. The inset is the Guinier plot of the experimental scattering profile. The solid lines are least-square linear fits of the data under the condition $R_g \times q < 1.3$. *B, P*(r) plot of TraI. The maximum dimension (D_{max}) is derived to be 220 Å. *C*, DAMMIN model of the TraI shown in volumetric representation. *D*, DAMMIN model of TraI superimposed with the atomic structures of the four structural domains of TraI (blue, the crystal structure of the relaxase domain 1-306, PDB entry 1P4B; yellow, the homology model of the RecD-like domain I 310-844, from this work; pink, a homology model of the RecD-like domain II 845-1476, from this work; green, the crystal structure of TraI 1476-1628, PDB entry 3FLD).

Figure 1.14 Model of TraI association with a dsDNA-ssDNA junction. The relaxase domain is shown in blue, the RecD-like domain I in yellow, the RecD-like domain II in pink, the C-terminal domain (CTD) in green. The two strands of dsDNA are colored in orange and dark green. The four sub-domains (N-terminal domain (NTD), 1A, 2A and 2B) of RecD-like

domain I and the five sub-domains (N-terminal domain (NTD), 1A, 2A, 1B and 2B) of RecD-like domain II are shown and labeled. The known NTP binding site is highlighted in red. In this model, the 5’ ssDNA overhang binds to the ssDNA binding groove mainly formed by RecD-like domain I and RecD-like domain II. The 5’ end is either free or covalently attached to the relaxase domain (*51*). The 3’ ssDNA tail is not bound by TraI. The dsDNA-ssDNA junction contacts subdomain 1B, or the “pin” domain, of RecD-like domain II

1.6 Figures and Tables

Table 1.1 Sequences of DNA oligomers used in this study.

FL-T_{17}	6FAM-5'-TTTTTTTTTTTTTTTTT-3'
25-mer	5'-CACTGACCGTCTGACTGCGATCCGA-3'
39-mer	5'-TCGGATCGCAGTCAGATGGTAAGAGAGACGCATAGATGC-3'
45-mer	5'GCGAACTGTCGAGTCGGCATCCGGATCTAGGGTAACCGGTACTGC-3'
51-mer	5'TCGGATCGCAGTCAGATGGTAAGAGAGACGCATAGATGCTGAGTGA GAGAT-3'
63-mer	5'TCGGATCGCAGTCAGATGGTAAGAGAGACGCATAGATGCTGAGTGA GAGATGCTCAGGTACAG-3'
FL-$dsDNA_{15}$	6FAM-5'-TCG GAT CGC AGT CAG-3'
$dsDNA_9$	5'-TCG GAT CGC-3'
$dsDNA_{11}$	5'-TCG GAT CGC AG-3'
$dsDNA_{15}$	5'-TCG GAT CGC AGT CAG-3'
$dsDNA_{19}$	5'-TCG GAT CGC AGT CAG ACG G-3'
$dsDNA_{21}$	5'-TCG GAT CGC AGT CAG ACG GTC-3'
$dsDNA_{23}$	5'-TCG GAT CGC AGT CAG ACG GTC AG-3'
$dsDNA_{27}$	5'-TCG GAT CGC AGT CAG ACG GTC AGT GAC-3'

6FAM, or 6-carboxyfluorescein, is a fluorescent probe that is covalently attached to the 5' end of DNA oligomers. The subscript indicates the length of the DNA oligomer. For a double-stranded DNA (dsDNA) oligomer, the sequence of the complementary strand is not shown.

Table 1.2 ATPase activities of different TraI constructs

	k_{cat} (S^{-1})	$K_{m,ATP}$ (mM)	$k_{cat}/K_{m,ATP}$ (mM^{-1} S^{-1})	Relative $k_{cat}/K_{m,ATP}$ (%)
WT	506 ± 25	1.6 ± 0.2	316 ± 26	100
284-821	0.014 ± 0.001	0.7 ± 0.08	0.02 ± 0.003	0.006 ± 0.001

Table 1.3 Thermodynamic parameters of TraI binding to ssDNA oligomers of length ranging from 19 to 31 nucleotides in standard DNA binding buffer at 25 ^{o}C

	T_{19}	T_{21}	T_{23}	T_{25}	T_{27}	T_{29}	T_{31}
n	1	1	1	1	1	1	1
$K_N \times 10^7$ (M^{-1})	0.9 ±0.1	1.6 ±0.1	2.5 ±0.1	2.6 ±0.3	6.2 ±0.3	10 ±1	12 ±1
$K_{int} \times 10^7$ (M^{-1})	N/A	N/A	N/A	2.6 ±0.3	2.1 ±0.1	2.1 ±0.1	1.8 ±0.5

n, stoichiometry; K_N, macroscopic binding constant; K_{int}, intrinsic binding constant; N/A, not applicable

Table 1.4 Thermodynamic parameters of TraI binding to dsDNA oligomers ranging from 9- to 27-mer in buffer A at 25 ^{o}C

	$dsDNA_9$	$dsDNA_{11}$	$dsDNA_{15}$	$dsDNA_{19}$	$dsDNA_{23}$	$dsDNA_{27}$
n	1	1	1	2	2	2
$K_{int} \times 10^5$ (M^{-1})	9.9 ±7.9	4.7 ±0.2	5.1 ±0.2	5.2 ±0.5	6.2 ±0.3	5.4 ±0.4
ω	N/A	N/A	N/A	18.1±3.8	38.6 ±7.7	14.8 ±2.2

n, maximum stoichiometry; K_{int}, intrinsic binding constant; ω, cooperativity parameter;

Table 1.5 Macroscopic binding constants of different TraI constructs to FL-dsDNA$_{15}$ in standard DNA binding buffer at 25 ^{o}C

TraI constructs	$K_N \times 10^6$ (M^{-1})
1-1756	1.6 ± 0.1
1-1476	0.78 ± 0.7
1-858	1.1 ± 0.1
1-330	< 0.02
302-1756	0.08 ± 0.01

K_N, macroscopic binding constant

Figure 1.1 Binding of nucleotides to TraI 284-821

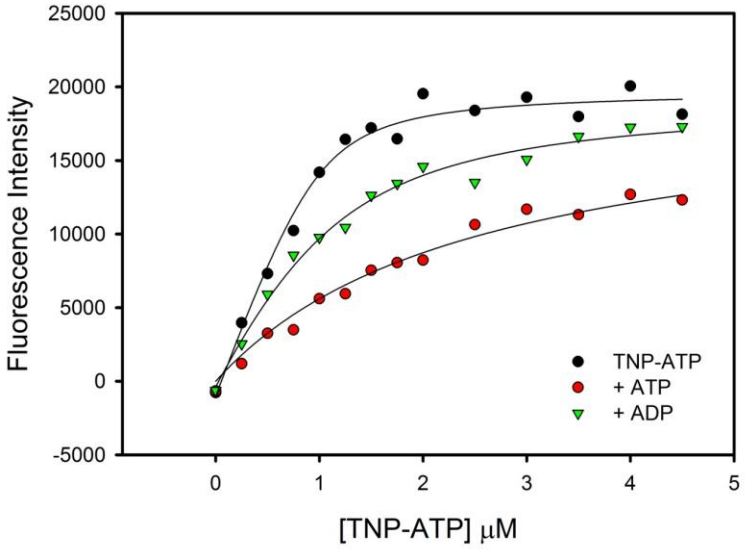


Figure 1.2 The stoichiometry of TNP-ATP binding to full-length TraI

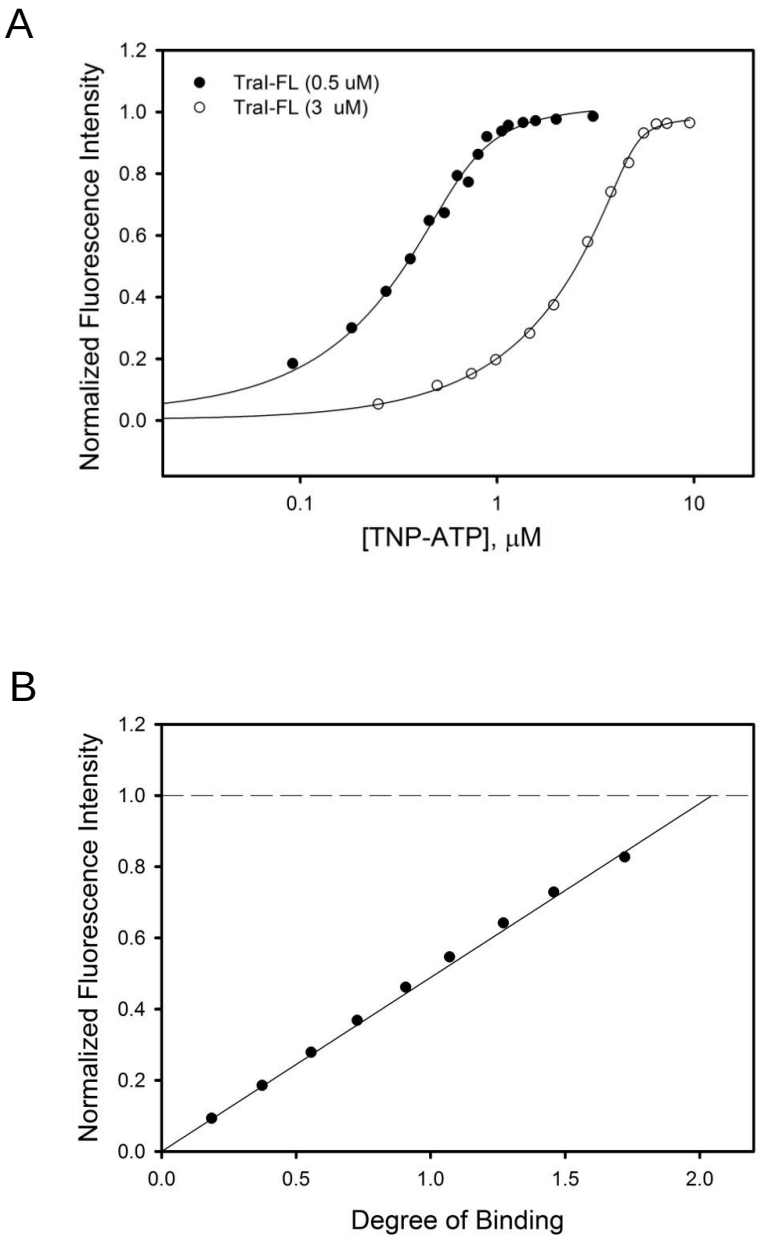
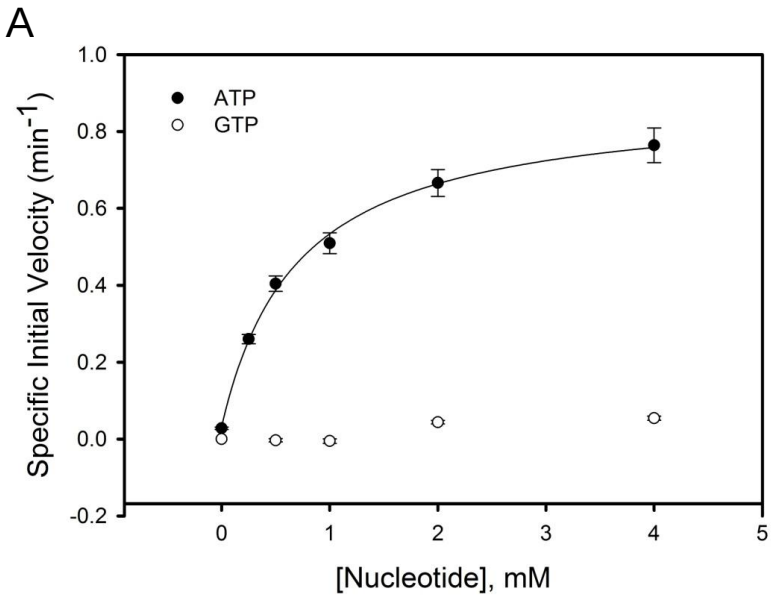
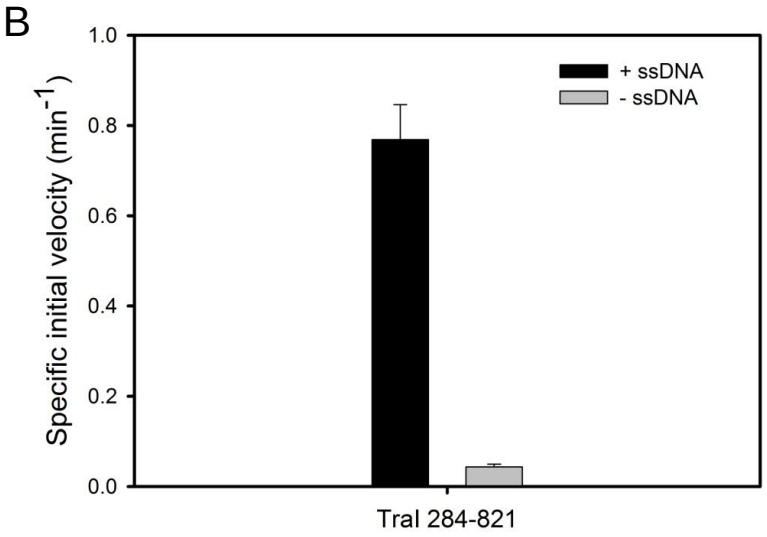


Figure 1.3 The ATPase and GTPase activities of TraI 284-821

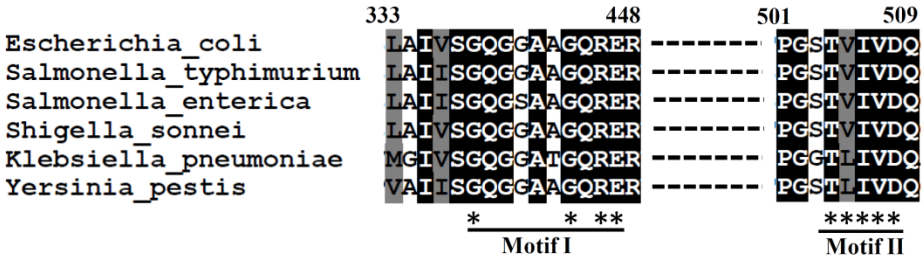




]

Figure 1.4 Sequence alignment of TraI orthologs

A



B

	Walker A/Motif I	Walker B/Motif II
Classical	GxxxxG***K***T/S	hhhhD
TraI 284-821	GxxxxxGQ***R***E	TVIVD

Figure 1.5 The role of R446 in the ATPase activity of TraI

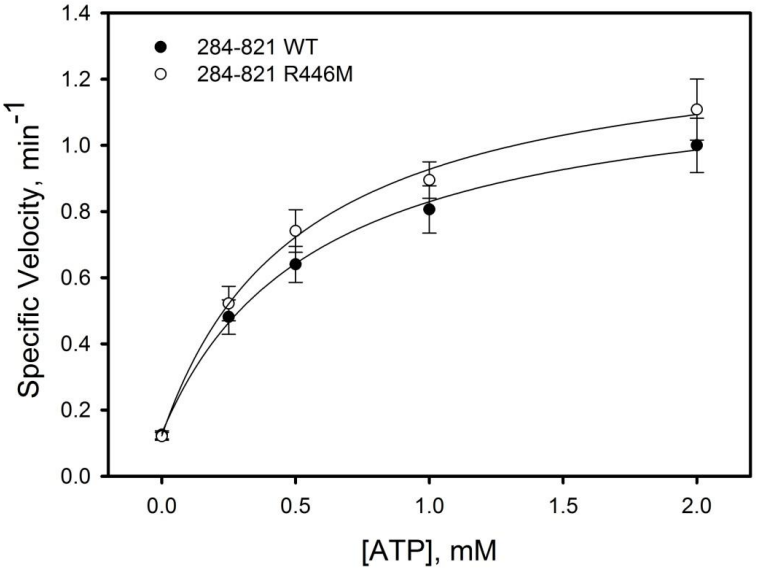


Figure 1.6 Schematic representations of the domain organization of TraI and the TraI constructs used in this study

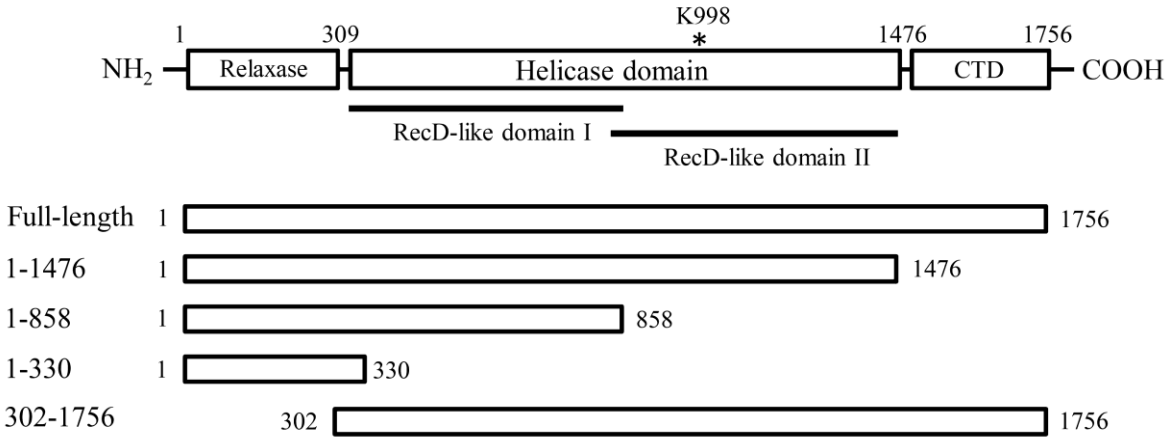


Figure 1.7 Determination of the stoichiometry of ssDNA binding by TraI

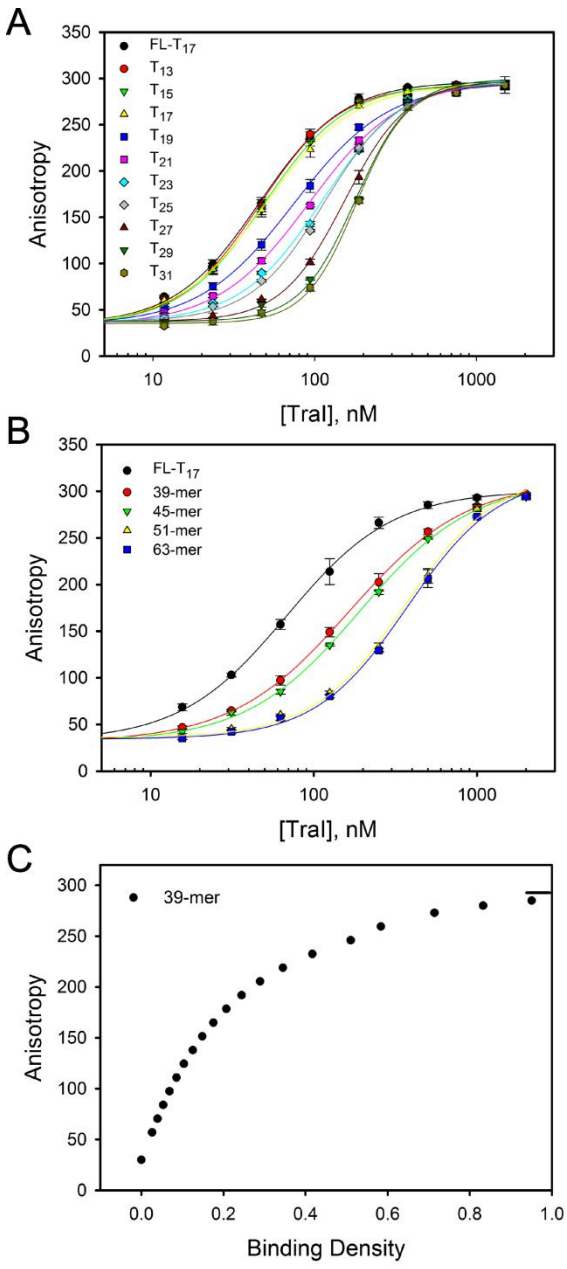


Figure 1.8 Determination of binding parameters for TraI binding to ssDNA

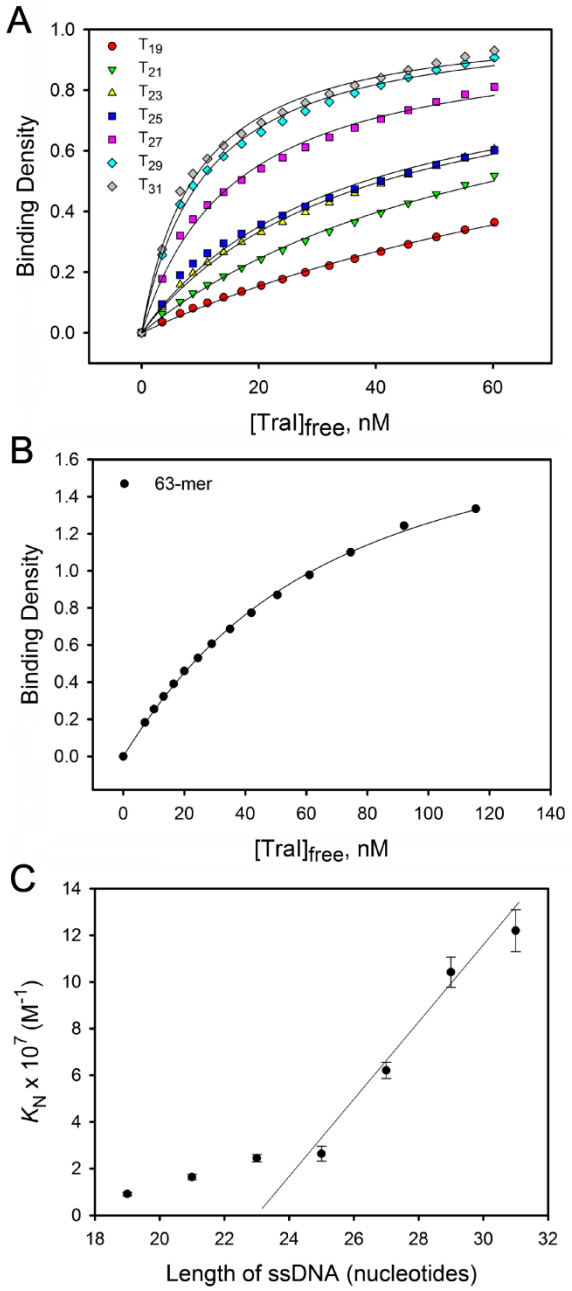


Figure 1.9 Determination of stoichiometry and binding parameters for TraI binding to dsDNA

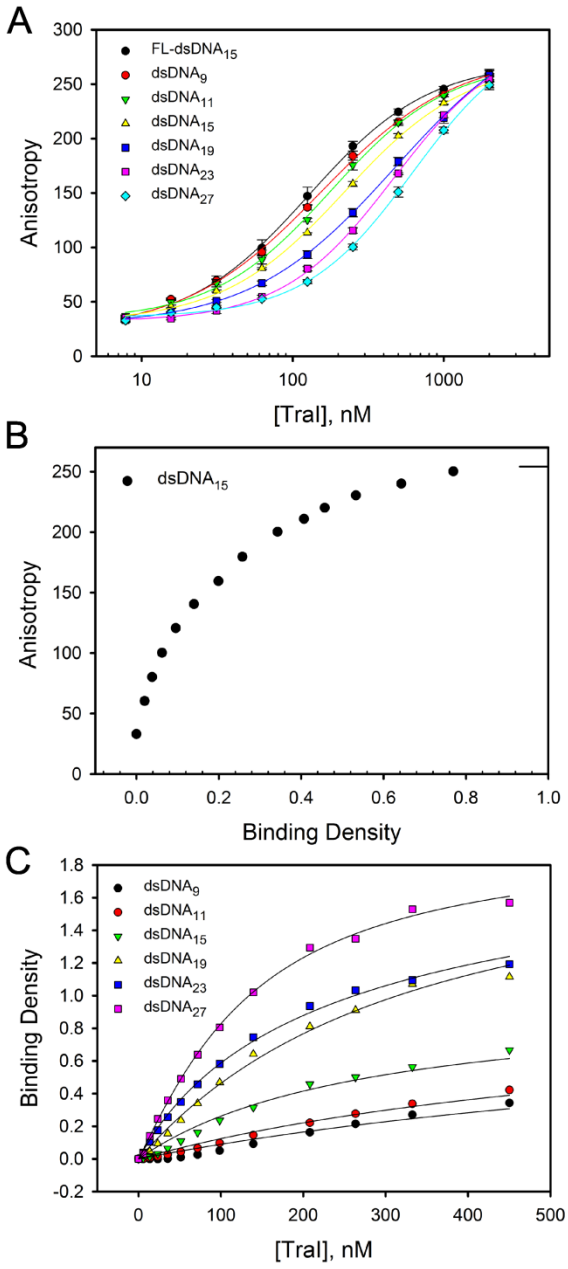


Figure 1.10 Identification of the location of the dsDNA binding site

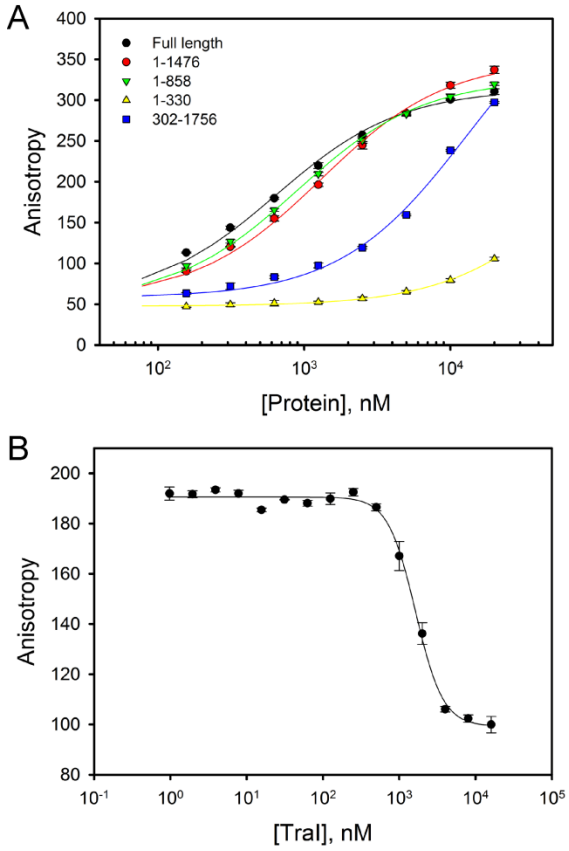


Figure 1.11 ADP and AMPPNP increased TraI affinity for ssDNA

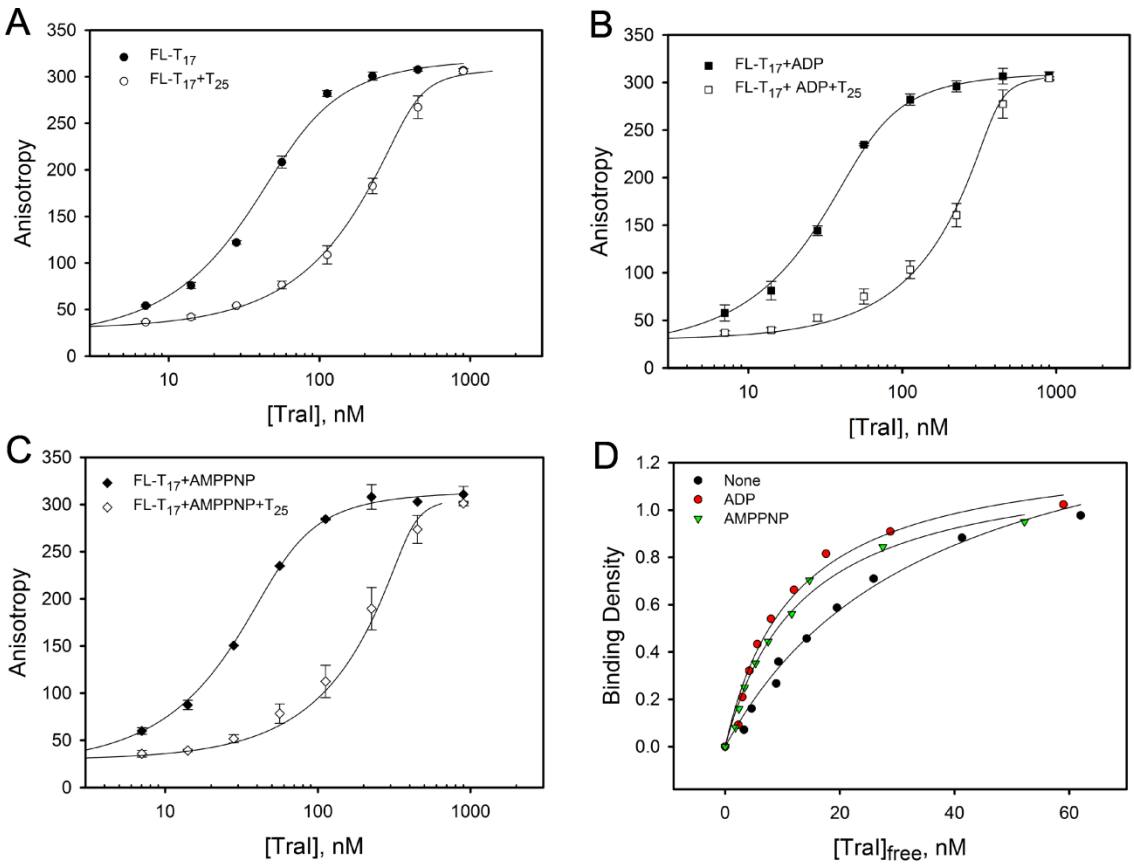


Figure 1.12 The role of ionic strength and base specificity in TraI DNA binding

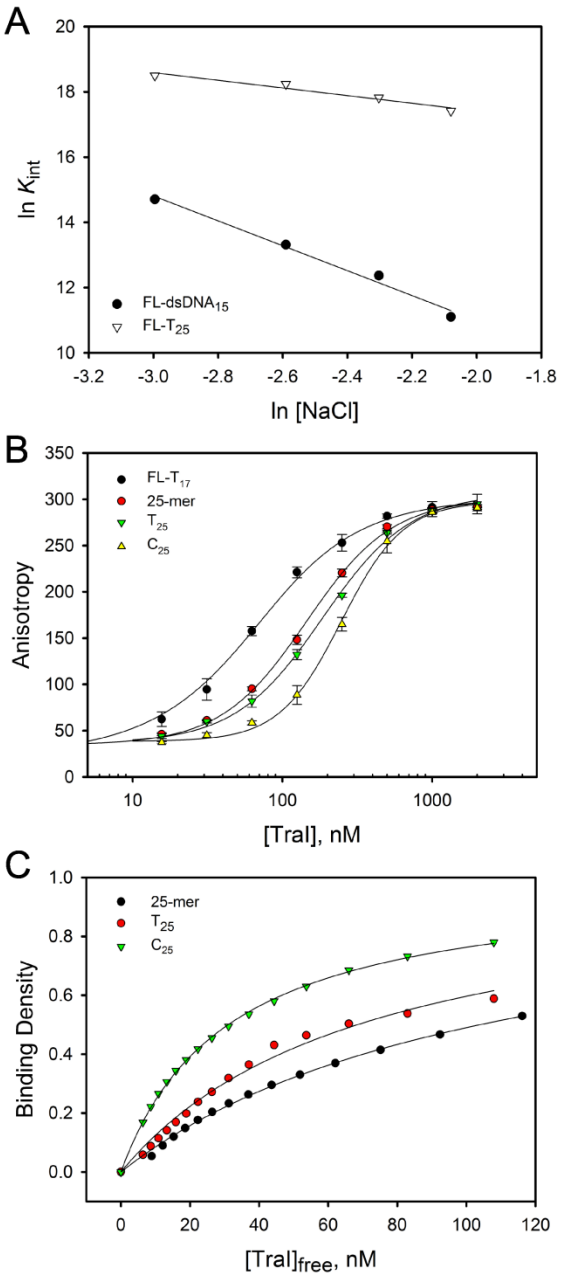


Figure 1.13 Solution structure of TraI

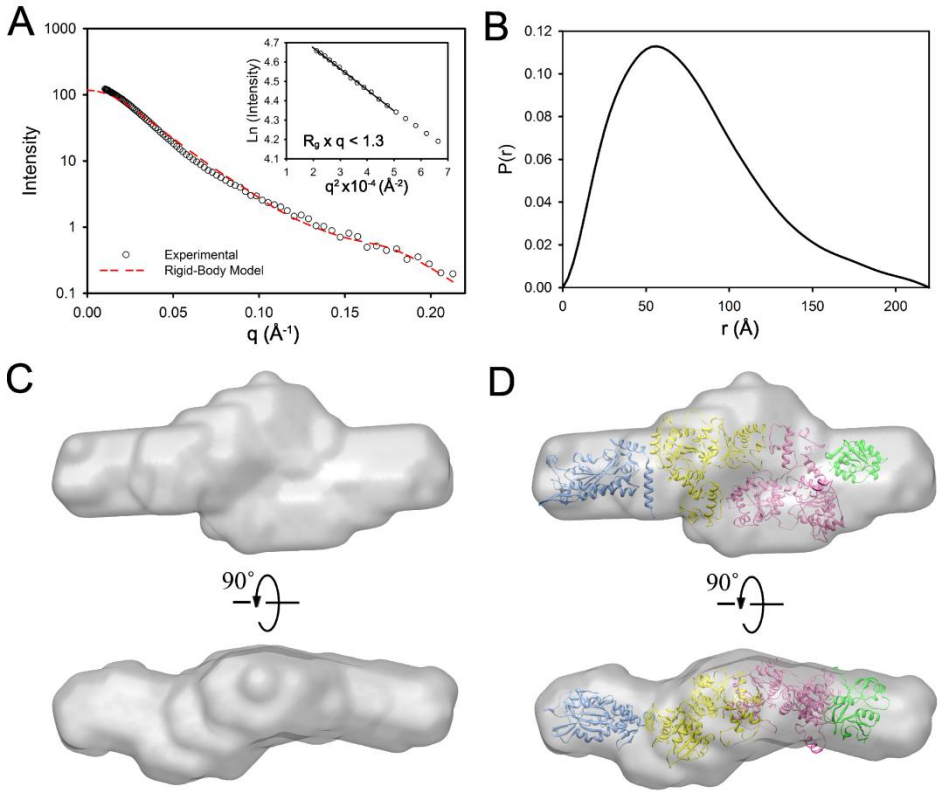
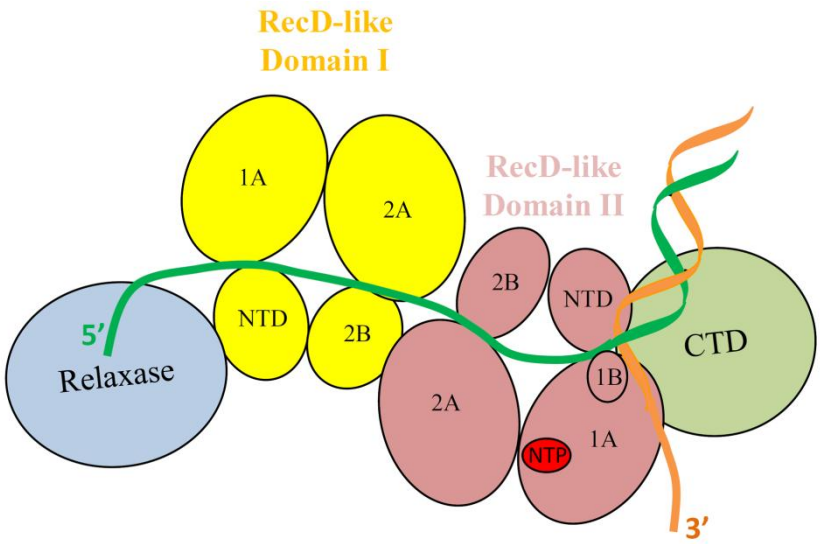


Figure 1.14 Model of TraI association with a dsDNA-ssDNA junction



Chapter 2. Structural and functional studies of *Arabidopsis thaliana* protein arginine methyltransferase 10 (PRMT10)

(This chapter has been submitted to *EMBO J*)

2.1 Introduction

Protein arginine methyltransferases (PRMT) are a family of enzymes that catalyze the transfer of methyl groups from S-adenosylmethionine (SAM) to arginine residues of target proteins (*56*). As a ubiquitous post-translational modification in eukaryotes, arginine methylation plays essential roles in many biological processes, such as signal transduction, chromatin remodeling, RNA processing, gene transcription, DNA repair and cellular transport (*56-60*). The dysfunction of PRMTs have been correlated with the development of cancer as well as autoimmune, cardiovascular, pulmonary and neuro-developmental diseases (*61-67*). The PRMTs are mainly classified as type I (PRMT1, 3, 4, 6, 8) or type II enzymes (PRMT5, 7 and FBXO11) (*56*) . Type I and type II PRMTs both catalyze the production of ω-N^{G}-monomethylarginine, but they differ in the production of dimethyl derivatives. Type I enzymes specifically produce asymmetric ω-N^{G},N'^{G}-dimethylarginine, while type II enzymes only produce symmetric dimethylarginine (*68*).

While the PRMTs have a relatively conserved catalytic core, they exhibit remarkable variations in the sequences N-terminal to the catalytic core. Indeed, the most N-terminal regions of the various enzymes (the "N-terminal additions") are divergent in sequence and been demonstrated to be important for the substrate specificity. For

example, the zinc-finger domain within the N-terminal addition of PRMT3 is essential for its recognition of RNA-associated targets (*69*). Structural studies reveal that the catalytic core of PRMTs has a three-domain architecture: an N-terminal SAM binding domain, a central dimerization domain and a C-terminal β-barrel domain (*70*). The main substrate binding site is located in the cleft formed between the SAM binding domain and the β-barrel domain (*71, 72*). As a conserved feature of PRMTs, dimerization has been established to be essential for the methyltransferase activities of various PRMTs (*71, 72*). Disruption of dimer formation can cause the loss of SAM binding (*72*), although the underlying mechanism of this impact on enzyme function remains unclear.

PRMT methyltransferase activity is regulated by several features of the target protein. The local sequence of the methylation site is one of the major determinants (*73, 74*). PRMT-catalyzed methylation typically occurs within glycine and arginine rich motifs (or GAR motifs), such as "RG", "RGG" and "RXR" (*75*), but there are exceptions to these rules (*74*). The activity of PRMTs can also be affected by the sequences distal to the methylation site (*76*) and by protein binding partners (*77, 78*). Circumstantial evidence has suggested that PRMTs often form complexes with other proteins *in vivo*, and that these proteins impact subcellular location and substrate recognition (*79, 80*).

PRMT10 is a plant-specific type I PRMT that plays an essential role in the regulation of flowering-time in *Arabidopsis thaliana* (*81*). Genetic disruption of PRMT10 causes late flowering by up-regulating the transcription of *FLOWERING LOCUS C* (*FLC*). Biochemical studies showed that PRMT10 can specifically methylate arginine-3 of histone H4 (H4R3) and arginine-3 of histone H2A (H2AR3) *in vitro*, and preferentially produces asymmetrical dimethylarginines (*81*). However, the physiological substrates of

PRMT10 have remained unknown. Besides PRMT10, eight other PRMTs have been identified in the *Arabidopsis thaliana* genome, including PRMT1a, PRMT1b, PRMT3, PRMT4a, PRMT4b, PRMT5, PRMT6 and PRMT7. These PRMT paralogs likely have diverse properties in cellular location, substrate specificity and protein-protein interaction (*56*).

In this work, we report the first crystal structure of PRMT10 in complex with a product of its enzymatic reaction, S-Adenosylhomocysteine (SAH). This structure provides insights into how PRMT10 interacts with peptides, and reveals numerous structural features likely to confer unique substrate specificity to PRMT10, including the role of the PRMT10 N-terminal addition in the enzyme function. Our studies also show that PRMT10 exists predominantly in a dimeric state in solution, and disruption of dimerization cause the loss of activity. Finally, we characterize the impact PRMT10 dimerization has on enzyme motion using molecular dynamics (MD) simulations. Our results highlight distinct differences between PRMT10 and other structually-characterized PRMTs, but also suggest that motions may be a conserved element of PRMT function.

2.2 Materials and methods

Cloning, expression and purification of PRMT10

The expression plasmids encoding wide-type *Arabidopsis thaliana* PRMT10 (1-383) and its various mutants and related constructed were created using the standard ligation-independent cloning techniques, as described by Stols *et al.* (*20*). All expression plasmids used in this study were sequence verified. The N-terminally 6×His-MBP-tagged PRMT10 was overexpressed in *Escherichia coli* BL21-CondonPlus (DE3) RIPL (Stratagene). The cells were grown at 37 ^{o}C to an OD_{600} of 0.6 in Luria-Bertani medium containing 50 μg/mL chloramphenicol and 50 μg/mL ampicilin. Protein expression was induced by the addition of isopropyl β-D-1 thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM and the culture was grown for another 16 h at 18 ^{o}C. The harvested cells were resuspended in buffer A (50 mM Na phosphate pH 7.4, 50 mM NaCl and 20 mM imidazole) supplemented with 0.5 mM EDTA, 0.1% Triton X-100, 1 mM phenylmethylsulphonyl fluoride (PMSF), one tablet of a protease inhibitor cocktail (Roche), and 1 mg/mL lysozyme. After 45 min of incubation on ice, the resuspended cells were sonicated on ice for 3 min and the lysate was centrifuged at 50,000 × g for 60 min at 4 ^{o}C. The supernatant was passed through a 0.2 μm filter (Millipore) and then loaded onto a 5 mL high performance HisTrap^{TM} column (GE Life Sciences), equilibrated with buffer A. The column was washed with 100 mL buffer A to remove nonspecifically bound proteins; the bound protein was then eluted with buffer B (50 mM Na phosphate pH 7.4, 50 mM NaCl and 250 mM imidazole). The elutant was loaded onto a HiPrep^{TM} 26/10 desalting column (GE Healthcare Life Sciences) equilibrated with

buffer C (20 mM Tris-HCl pH 8.0 and 150 mM NaCl), and the protein fractions were collected. To remove the His-MBP tag, TEV protease was added into the pooled protein fractions with a ratio of 1:100 (w/w) TEV to PRMT10. After 12 hr of incubation at 4 ^{o}C, the mixture was reloaded onto 5 mL high performance HisTrap^{TM} column (GE Life Sciences), equilibrated with buffer A. The flow-through fractions were collected and concentrated in a Centricon YM10 (Amicon) concentrator. Concentrated protein was loaded on a HiLoad^{TM} 16/60 Superdex 200 column (GE, Life Sciences) equilibrated with sizing buffer (20 mM Tris-HCl pH 7.5, 250 mM NaCl, 5% Glycerol). PRMT10 containing fractions were concentrated, flash frozen in liquid nitrogen and stored at -80 ^{o}C. Purified protein was >95% pure by SDS-PAGE.

Crystallization, data collection, structure determination, and refinement

Diffraction-quality crystals of PRMT10 (residues 11-383)-SAH complex were obtained by the hanging-drop vapour-diffusion method at 22 ^{o}C, with the mother solution containing 0.1 M Tris-HCl pH 7.6, 2.3 M Na_2HPO_4 and 100 mM arginine. Crystals grew to the size of 250 × 200 × 50 μm in approximately 10 days. Since flash-frozen crystals diffracted poorly and could not be used for structural determination, diffraction data were collected from warm-mounted crystals to 2.6 Å resolution using a Rigaku X-ray generator MicroMax-007HF. Data from four different crystals were reduced and merged using the program HKL2000 (*82*) (Table 2.1). Data quality was examined using the program PHENIX (*83*). The structure was determined in space group $P2_1$ by molecular replacement using the program PHENIX (*83*). The crystal structure of rat PRMT3 (PDB entry, 1F3L), processed using the program chainsaw of the CCP4 package (*84*), was used

as the template for molecular replacement. Due to the salient difference between PRMT10 and PRMT3 in the sequence of the dimerization arm, the dimerization arm of PRMT3 (residues 370-399) was not incorporated into the template. Since psudomerohedral twinning (approximately 50%) was detected with the crystals used for the structure determination, least square twin refinement was performed using the program PHENIX. The structural model was further built manually using the program Coot (*85*), and refined using the program PHENIX.

Methyltransferase assay

In vitro methyltransferase assays were performed as described by Cao *et al*. (*81*). The reaction mixture contained 20 mM Tris-HCl pH 8.0, 4 mM EDTA, 1 mM DTT, 0.5 mM PMSF, 4 μM *S*-[Methyl-^{3}H] Adenosyl-*L*-methionine (Perkin Elmer [NET155]) and indicated concentrations of PRMTs and substrates. After being incubated for 5 hr at 30 ^{o}C in a total volume of 20 μl, the reaction mixtures were separated on a 15% SDS-PAGE and stained with Commassie blue. The gel was then treated with Amplifier (Amersham Biosciences), dried and exposed to Kodak Biomax MS film at -80 ^{o}C.

Dynamic light scattering (DLS)

The hydrodynamic radii of various PRMT10 constructs were measured by a DynoPro instrument (Wyatt Technology Corporation). All samples and buffers (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.5 mM EDTA) were filtered through 0.2 μM filters (Millipore) or centrifuged at 17,000 × g at 4 ^{o}C for 30 min before measurement. Three replicates were performed for each sample. The hydrodynamic radii and molecular

weights of PRMT10 samples were estimated using the assumption of globular protein shape.

Molecular dynamics (MD) simulation

MD simulations of PRMT10 were performed using the AMBER 2003 force field (*86*) as described previously (*87*). All production runs were generated using the PMEMD module of Amber 9.0 (*88*) with a 2 fs time step. The topology and parameter files were created using the LEaP program within AMBER (*88*). To maintain charge neutrality, the protein molecule was surrounded by a truncated octahedron of water and sodium ions in the simulation system. Electrostatic interactions were calculated using the particle-mesh Ewald algorithm (*89*) with a cutoff of 10 Å applied to Lennard-Jones interactions. Energy minimization was conducted using the SANDER package within AMBER (*88*). Equilibration consisted of 20 ps of constant volume conditions with heating from 100 to 300 K and subsequent 100 ps of constant temperature conditions.

Simulation results were analyzed by the PTRAJ package in AMBER (*88*). The pair-wise correlation coefficient, C_{ij}, was calculated between the α-carbons of two residues as described by Sharma *et al*. (*90*). When the two residues i and j move in a correlated fashion (the angle between the motion of i and j is less than 90 º), $0<C_{ij}\leq 1$; when they move in an anti-correlated way (the angle between the motion of i and j is more than 90 ºbut less than 180 º), $-1\leq C_{ij}<0$; finally, when they move in a non-correlated manner (randomly), $C_{ij}=0$. The more positive the value of C_{ij} is, the smaller the angle between the motion of the two residues is. Single-linkage clustering analysis was

performed to identify groups of residues that move in a correlated or anti-correlated fashion, as described by Leese *et al*. (*91*).

2.3 Results

Crystal Structure of the PRMT10-SAH Complex

The structure of *Arabidopsis thaliana* PRMT10 (residues 11-383) in complex with SAH was determined in space group $P2_1$ by molecular replacement and subsequently refined to 2.6 Å resolution (Table 2.1). Four PRMT10-SAH complexes were observed per asymmetric unit. The crystals used to solve the structure contain nearly 50% pseudomerohedral twinning as indicated by the L-test and Britton plot carried out by the program PHENIX (*83*). Notably, the β angle (89.98^o) of the unit cell is very close to 90^o. Consequently, the diffraction data could also be reduced into the orthorhombic space group P222 and its derivatives. Serious violations of systematic absences were observed, however, in space groups $P2_12_12_1$, $P2_12_12$, $P222_1$. Consequently, molecular replacement was only performed in the space group P222. As predicted by the Matthew's coefficient, two monomers were identified in each asymmetric unit. All solutions in space group P222, however, were finally rejected owing to the presence of significant main-chain clashes during crystal packing. Taken together, these results indicated that $P2_1$ was the correct space group and pesudomerohedral twinning was present. The data were detwinned using the twining refinement function of the program PHENIX (*83*). No violations in systematic absences for $P2_1$ were observed, and the structure was determined and refined with confidence in this space group (Table 2.1).

The N-terminal twenty residues of this PRMT10 construct (residues 11-30) were missing from the structure due to lack of electron density. PRMT10 exhibits three

sequentially folded domains, an N-terminal SAM binding domain (residue 31-173), a central dimerization arm (residues 187-235), and a C-terminal β-barrel domain (residues 174-185 and residues 238-383) (Figure 2.1A & 2.1B). The SAM binding domain is composed of two N-terminal helices (αX & αY, residues 31-50) followed by a classical Rossman fold (residues 51-173) consisting of five α helices (αZ, αZ', αA, αB, αD) and five β strands (β1 to β5). The consensus Rossman fold has been observed in other known SAM-dependent methyltransferases (*92, 93*), while the two N-terminal helices (αX & αY) are unique to PRMTs (*71*). The β-barrel domain, forming close contacts with the SAM-binding domain at one end of its barrel, harbors ten β-strands (β6 to β15) and two short α-helices (αH and αI). The dimerization arm, exhibiting a helix-turn-helix fold, is inserted in between β6 and β7 of the β-barrel domain and protrudes from the main body of the protein. Sequence analysis reveals four PRMT signature motifs in PRMT10 (Figure 2.2). Motif I (YFxxY) and Motif II (DVGxGxG) are directly involved in the binding of cofactor SAM. Motif III (SExMGxxLxxExM), harbors two critical catalytic residues E143 and E152. Mutation of any of these two residues completely disrupted the methyltransferase activity of PRMT10 (data not shown). Motif IV (or the THW motif) is the most highly conserved sequence among PRMTs and directly involved in the formation of the active site. Disruption of motif IV is accompanied with complete loss of the methyltransferase activity of PRMT10 (data not shown).

The structure of PRMT10 (residues 31-383) exhibits a similar overall fold when compared to other known PRMT structures, sharing, for example, 1.8 Å root-mean-square deviation over 245 Cα positions with PRMT1 (residues 41-354). However, a strikingly unique feature of PRMT10 is its dimerization arm, consisting of two straight

anti-parallel α-helices, which is significantly longer (41 Å) than that of other PRMTs of known structure (*e.g.*, PRMT1, 22 Å; PRMT3, 22 Å; CARM1, 34 Å; REFs) (Figure 2.1C). PRMT10 also differs from other PRMTs in two loop regions of the β-barrel domain (Figure 2.1C). Sequence alignment indicates that these loops are relatively conserved among PRMT10 orthologs (Figure 2.8), but highly divergent among PRMT paralogs (Figure 2.2). Loop I is located close to a conserved substrate binding site of PRMTs (see below). Acidic residues in Loop II have been shown to be important for the interaction of PRMT1 with its substrates (*94*).

PRMT10 Active Site

In the PRMT10-SAH complex, SAH binds to a deep pocket formed by the three N-terminal α-helices (αX, αY and αZ) and the carboxyl ends of the parallel β-strands (β1 to β5) (Figure 2.1D). Most of the residues involved in SAH binding are highly conserved among type I PRMTs (Figure 2.2), indicating that members of the type I PRMT family likely share similar mechanisms in cofactor binding and catalysis. Hydrogen bonding plays a major role in the interaction of PRMT10 with SAH, with six hydrogen bonds formed between PRMT10 and all three moieties of SAH (adenine, ribose and methioine). R54 of the helix αZ forms bifurcated hydrogen bonds with the terminal carboxylate group of the methionine moiety. For the ribose moiety, hydrogen bonds are observed between the two hydroxyl groups and the side chains of E100 of strand β2 and Q45 of helix αY. The adenine group is recognized by the E129 from the loop between β2 and β4. In addition to hydrogen bonding, the main-chains of the glycine rich loop (Motif II) and the side-chains of seven other residues (A101, V128, F36, M154, S157, Y35 and Y39)

contact SAH through van der Waals forces. Given the small difference between the chemical structure of cofactor SAM and its derivative SAH, it is expected that SAM binds to the active site in a manner similar to that observed here for SAH.

PRMT10 Dimer

In the crystal structure, PRMT10 forms a ring-like homodimer through the interaction between the dimerization arm (αE-loop-αG) of one monomer and the outer surface (αY, αZ, αA & αD) of the SAM binding domain of the other monomer (Figures 2.3A & 2.3B). The active sites are located at the periphery of this central cavity. As observed in other PRMTs, hydrophobic interactions play a major role in the formation of PRMT10 dimers. Significant conservation was maintained in the residues on the surface of the SAM binding domain that form the dimer interface, but not in the residues that form the dimerization arm (Figure 2.4). Hydrogen bonding is also observed on the dimer interface, with the side-chains of Q90 and N115 forming hydrogen bonds with the main-chains of G215 and D217 respectively (Figure 2.3C). The hydrogen bonding between N115 and D217 are highly conserved among PRMTs (Figure 2.2), highlighting its importance for dimer formation. Another conserved residue on the dimer interface is G215, whose small side-chain is apparently favorable for the formation of the sharp turn at the tip of the dimerization arm.

Notably, due to the longer dimerization arm, the central cavity of PRMT10 is significantly larger than those of other PRMTs of known structure (Figure 2.5). PRMT10 creates a cavity 15x13 Å, while those of PRMT1, PRMT3 and CARM1 exhibit cavities of 8x12, 8x13 and 8x11 Å, respectively (Figure 2.5). The longer "vertical" distance as

depicted in Figure 2.5 is generated by the longer PRMT10 dimerization arm. Consistent with the observation in the crystal structure, our results from dynamic light scattering and gel filtration experiments confirmed that PRMT10 exists predominately in a dimeric state in solution (Table 2.2). In addition, unlike PRMT1, which dissociates into a smaller oligomer upon SAH binding (*72*), the oligomeric state of PRMT10 is independent of the SAH binding (Table 2.2).

To test the role of the dimer interface observed in the crystal structure in PRMT10 function, we designed two mutants, Δ214-218 and Δ202-225, in which different sections of the dimerization arm were replaced with stretches of glycine residues. Both arm mutants displayed wild-type level expression and stability in *E.coli*, suggesting that they are well folded. The oligomeric states of these mutants were examined using dynamic light scattering and gel filtration experiments (Table 2.2). Deletion of the dimerization arm that forms the dimer interface (Δ202-225) in the crystal structure disrupted dimer formation. Notably, substitution of the short loop at the tip of the dimerization arm (residues 214-218, YGVDM) with a stretch of five glycine residues also caused the complete loss of dimer formation, highlighting the important role of this short loop in dimer formation. The role of dimerization in the methyltransferase activity of PRMT10 was examined by measuring the activities of two arm mutants, Δ214-218 and Δ187-235, that are monomers in solution. Both arm mutants displayed no observable activities toward H2A and H4 (Figure 2.6A & 2.6B), indicating that dimerization is essential for the methyltransferase activity of PRMT10.

PRMT10 Surface Electrostatics

Surface charge distribution appears to be an important functional property of PRMTs. For example, published data has suggested that surface charges are crucial for the interaction with substrates and other proteins (*71, 72*). Figure 2.7 illustrates the surface charge distribution of PRMT10. As seen in other PRMTs, the surface of PRMT10 contains numerous acidic patches, especially around the active site. However, there are notable differences in surface charge distribution due to the low conservation in the surface residues among different PRMTs (Figure 2.7). The uniquely long dimerization arm of PRMT10 contains ten acidic residues, which are much more than what is observed in other PRMTs. Another major difference is located at one end of the β-barrel domain, where PRMT10 has a large acidic patch formed by residues E374, E367, E336, D339 and E281. The acidic residue in this region has been shown to be important for the substrate interaction of PRMT1 (*94*). Due to the low sequence conservation, the surface charge properties of this region are highly variable among PRMTs. Such differences are likely to impact PRMT substrate recognition. Structural studies of PRMT1 have indicated the location of the substrate binding groove of this enzyme (*72*). Based on the location of acidic patches and the shape of the PRMT10 surface in light of other PRMTs of known structure, we have identified four putative substrate binding grooves on the surface of PRMT10 (Figure 2.7). Binding grooves I and II are located in the cleft formed between the SAM binding domain and the β-barrel domain and are directly connected to the active site. Binding grove III and IV lie on the surface of the β-barrel domain. Substrates can also enter the active site through binding groove III. A high degree of conservation is maintained in the residues that form binding grooves I and II (Figure 2.4 and 2.7), suggesting the conserved role of these two binding grooves in substrate interaction. In

contrast, little conservation is observed in the residues that form binding grooves III and IV (Figure 2.4 and 2.7). It is possible that the diversity in binding grooves III and IV may confer unique substrate specificities to different PRMTs.

Increased Active Site Accessibility in PRMT10

As already mentioned, members of the PRMT family share a consensus three-domain architecture and form dimers using a similar mechanism. However, the relative orientation of the two monomers in a functional dimer significantly varies between different PRMTs, due to the diversity in the length and structure of their dimerization arms. Consequently, the dimeric forms of different PRMTs do not superimpose well. When we align different PRMTs based on one of their two monomers (the "bottom monomers" in Figure 2.8A, left panel), the other monomers (the "top monomers") end up in distinct locations. In Figure 2.8A (left panel), the top monomers of PRMT1 (cyan) and PRMT3 (yellow) sit right above their bottom monomers, while the top monomers of PRMT10 (magenta) and CARM1 (blue) swing away from the vertical direction by 35^{o} and 20^{o}, respectively. The top monomers of PRMT10 and CARM1 are observed to be translated leftward 21 Å and 13 Å, respectively, relative to those of PRMT1 and PRMT3 (Figure 2.8A, middle panel). Furthermore, the angles between the two monomers in a PRMT dimer vary significantly between enzyme orthologues, ranging from 30^{o} in PRMT3 to 52^{o} in PRMT10 (shown schematically in Figure 2.8A, right panel).

The differences in the relative orientation of the two monomers in PRMT dimers, together with the differences in the size of the central enzyme cavities, results in the variations in active site accessibility across the enzymes of known structure. To provide a

quantitative measure of active sites accessibility for different PRMTs, we defined an accessibility angle, or A-angle, for each enzyme. For PRMT10, the vertex of the A-angle is located at the side-chain O atom of Y147 and the two rays are made by connecting the side-chain O atom of Y147 and the Cα atoms of H318 from monomer I and monomer II respectively (Figure 2.8B). Y147 is located closed to the middle of the substrate binding groove that runs through the active site and H318 lies at the outer edge of the monomer. The A-angles of other PRMTs were defined by connecting three atoms with homologous positions with those of PRMT10 (Figure 2.8C-E). Based on these A-angles, the active sites of PRMTs have diverse accessibility, with an order of PRMT10>CARM1>PRMT1>PRMT3.

PRMT10 Motion

Although dimerization has been widely shown to be essential for the methyltransferase activity of PRMTs, the underlying structural mechanism has remained unclear. Here, we have chosen to examine the impact dimerization has on the motion of PRMT10 using molecular dynamics (MD) simulations. Monomeric and dimeric forms of PRMT10 were analyzed over a 30 ns timescale. As shown in Figure 2.9, the total energy of each system, calculated as the sum of the kinetic and potential energy at each time point, was relatively constant after the first 5 ns, particularly in the last 10 ns. Therefore, the averages of the MD trajectories in the last 10 ns were used for the following analysis. The effects of dimerization on the degree of motion of PRMT10 were analyzed by computing the atomic position fluctuations (APFs) of Cα atoms. PRMT10 exhibits similar APFs in monomeric and dimeric states, except for two regions (αY-loop-αZ,

residues 40-68; the dimerization arm, residues 187-235,) where dimeric PRMT10 displays significantly lower APFs than monomeric PRMT10 (Figures 2.10A & 2.10B). The reduced fluctuations within these two regions likely result from their direct involvement in the formation of dimer interface (Figure 2.3C). Notably, the region αY-loop-αZ (residues 40-68) is directly involved in the binding of SAH and the formation of substrate binding groove I. Therefore, stabilization of this region by dimerization likely positively impacts the binding of SAH and substrate proteins.

We computed normalized covariance matrices to classify the relationships between the motions of all residue pairs in the protein (Figures 2.10C & 2.10D). Normalized covariance matrices generate the residue-residue correlation coefficients (C_{ij}s), which inform the relative motion between a residual pair. Based on the value of C_{ij}s, the motions of all residue pairs can be classified into three groups, correlated motion (two residues moving toward the same direction) as indicated by C_{ij} approaching 1, anti-correlated motion (two residues moving toward the opposite direction) as indicated by C_{ij} approaching -1, and uncorrelated motion (two residues moving with the lack of a dynamic relationship) with C_{ij} values near zero. The SAM binding domain of dimeric PRMT10 exhibits considerably greater residue-residue correlations relative to that of monomeric PRMT10 (Figure 2.10C). Increased residue-residue correlations are also observed in several discrete regions of the β-barrel domain.

To better understand the biological significance of residue-residue correlations, single-linkage clustering analysis was then conducted to identify groups of residues that move together (see Experimental Procedures for details). Clustering of dimeric PRMT10 at a correlation coefficient above 0.7 resulted in five clusters, while clustering of

monomeric PRMT10 under the same criterion only resulted in three clusters (Figures 2.10E & 2.10F). One notable difference between monomeric PRMT10 and dimeric PRMT10 lies in the SAM binding domain. Most parts of the SAM binding domain, except the two N-terminal helices (αX and αY) and two loop regions (L1 and L2), are clustered in dimeric PRMT10 (Figure 2.10G), while only helix B is self-clustered in monomeric PRMT10. In addition, one end of the β-barrel domain is also clustered in dimeric PRMT10, but not in monomer PRMT10. Our results clearly show that the SAM binding domain and one end of the β-barrel domain to move as a cohesive unit in dimeric PRMT10, but not in monomeric PRMT10. These observations provide a structural mechanism that explains the essential nature of PRMT dimerization for enzyme function.

PRMT10 N-terminus in Enzyme Function

Finally, we examined the role of the N-terminal addition (residues 1-30) in the oligomeric state and methyltransferase activity of PRMT10. We created three N-terminal deletion mutants, including ΔN10 (residues 11-383), ΔN20 (residues 21-383) and ΔN30 (residues 31-383), and compared them to wild-type PRMT10. The oligomeric states of these mutants were investigated using dynamic light scattering (DLS) and gel filtration experiments (Table 2.2). As observed in the wide-type enzyme, all N-terminus deletion mutants remain as dimers in solution. Moreover, the oligomeric states of these mutants are SAH-independent. Together, these data suggest that the N-terminal addition does not impact PRMT10 dimerization.

The methyltransferase activities of wide-type PRMT10 and three N-terminal deletion mutants were measured as previously described (see experimental procedures)

(2.5A & 2.5B). Purified calf thymus core histones, which are a mixture of histones H2A, H2B, H3 and H4, were chosen as the substrate. Interestingly, ΔN10 had approximately 3-fold greater activities toward H2A relative to the wide-type enzyme. Further deletions of the N-terminus (ΔN20 and ΔN30) do not enhance the activity over H2A. When H4 was used as the substrate, however, all three N-terminus mutants displayed wide-type level activities. In the crystal structure, helix αX (residues 32-40) plays an important role in cofactor bining by covering the opening of the SAM binding pocket. As expected, the deletion of the helix αX of PRMT10 (ΔN40) caused a dramatic drop in its activity over both H2A and H4. Taken together, our results indicate that first ten residues of PRMT10 impact enzyme methyltransferase activity in a substrate-dependent manner.

Previous studies of PRMT1 have shown that the substrate sequence distal to the methylation site can affect its activity toward H4. Thus, we examined if the substrate sequence outside of the methylation site also impacts the activity of PRMT10 using purified full-length histone H4 and H4N1-20, a peptide covering the N-terminal twenty residues of histone H4. First, we found that the activity of PRMT10 on the full-length H4 substrate was markedly higher than that on the H4N1-20 substrate (Figure 2.6C), in spite of the fact that 10-fold more H4N1-20 peptide was employed in these assay. Thus, it appeared that the sequence outside the N-terminal 20 residues of histone H4 was important for the methyltransferase activity of PRMT10.

To probe further the impact of protein regions outside the methylation site, we examined a glutathione-S-transferease-tagged H4 substrate. The arginine-3 methylation site of PRMT10 in both histone H2A and histone H4 is located at the far N-terminus of these proteins. To examine if a bulkly protein fused to the N-terminus of H4 would

impact PRMT10 activity on H4R3, we compared the methylation of histone H4 and N-terminally GST-tagged histone H4 (GST-H4) by PRMT10 (Figure 2.6C). Our results show that the presence of a N-terminal GST tag reducing the activity of PRMT10 by ~2-fold relative to untagged H4. These data indicate that PRMT10 can methylate H4R3 even when it is not located at the far N-terminus of this histone protein.

2.4 Discussion

We present the first structure of a plant protein arginine methyltransferase, that of PRMT10, and highlight unique features of this enzyme, including a long dimerization arm and a distinctly open conformation in the catalytic dimer. However, we also establish for the first time that the family of PRMTs exhibit conserved domain motions, particularly within the enzyme region that binds the SAM substrate that donates the methyl group to arginines on target proteins. Together, these data advance our understanding of features shared by the PRMT enzymes, which function as both epigenetic and non-epigenetic factors, as well as unique aspects of particular family members that may impact substrate preference.

In a functional PRMT dimer, the enzyme active sites are located at the periphery of a central cavity. This configuration likely impacts access of substrate proteins to the PRMT catalytic site. Indeed, most known methylation sites are located in disordered regions of substrates, and the structural flexibility around the methylation site has been shown to be essential for PRMT function. Comparing the PRMT dimer to that of other PRMTs of known structure led to the appreciation that PRMT paralogs exhibit diverse central cavity sizes, with an order of PRMT10>CARM1>PRMT1>PRMT3 (Figure 2.8). This variation mainly results from the difference in the relative orientation of the two monomers in a functional PRMT dimer and the difference in the size of the dimerization arm. Previous studies have shown that the activity and substrate specificity of PRMTs are directly correlated with active site accessibility (*95*). Thus, the more accessible PRMT10 active site may allow this enzyme to methylate arginine residues that do not serve as substrates for other PRMT enzymes.

We show that the *Arabidopsis thaliana* PRMT10 enzyme functions only as a dimer. Although this now appears to be a feature common to all PRMTs, the precise mechanism of how PRMT dimerization impacts enzyme function has remained unclear. Our data from MD simulations on both the monomeric and dimeric forms of PRMT10, as well as PRMT1 and PRMT3, show that dimer formation produces coherent motions in key catalytic domains. PRMT dimers exhibit reduced fluctuations in the N-terminal αY-loop-αZ region, which not only forms direct contacts with the SAM methyl donor, but also forms a portion of substrate binding groove I that is conserved among PRMTs. Furthermore, dimerization results in more correlated motions throughout the SAM binding domain. Previous studies have shown that oligomerization can facilitate protein-ligand interaction by increasing the correlation in the motion of the structural elements involved in ligand binding. (*87*). Importantly, our results show that the effects of dimerization on the motion of PRMT10 can be generalized into other members of the PRMT family. Thus, we conclude that dimerization facilitates the methyltransferase activity of the PRMTs by producing coherent protein motions involving the SAM cofactor.

Members of the PRMT family have a relatively conserved catalytic core, but exhibit remarkable diversity in the length and sequence of their N-terminal regions. Multiple lines of evidence suggest that the variations in the N-terminus diversify the functions of the PRMT family by modulating the substrate specificities. PRMT10 has a 30-residue N-terminal addition, which is one of the shortest among known PRMTs. Secondary structure analysis predicts that the N-terminal addition of PRMT10 remains in a disordered state. In support of this prediction, the PRMT10 N-terminal addition is prone

to proteolysis (data not shown), and is not ordered in our crystal structure. Although PRMT1 also has a short N-terminal region (~31 residues), its length varies more among different PRMT1 isoforms and these variations have been shown to alter the substrate specificity of PRMT1 (*96*).

In this work, we have established that the N-terminal addition (residues 1-10) can also affect the substrate specificity of PRMT10. The deletion of the N-terminal addition enhances the activity of PRMT10 over H2A, but does not change its activity over H4 significantly. This variation may result from the difference in the way that H4 and H2A interact with PRMT10. Based on the crystal structure of dimeric PRMT10, the 30-residue N-terminal addition is likely located at one side of the ring and affects substrate binding grooves III and IV, having less or no effect on substrate binding grooves I and II. Our results suggest that H2A may approaches the active site through substrate binding groove III or IV, but H4 employs substrate binding groove I. The local sequence of the methylation site in H2A (SG***R***$_3$GKGG) is identical to that of H4 (SG***R***$_3$GKGG), indicating that the sequence outside the methylation site is also important for the interaction of PRMT with its substrates. In support of this notion, our results clearly show that deletion of the sequence C-terminal to the residue 20 of H4 dramatically reduced the methylation at arginine 3 by PRMT10. Considering that fact that PRMT10 displays similar activity over H4N1-20 and H4 when the concentration of H4N1-20 is 500 μM (data not shown), we hypothesize that the substrate sequence outside of the methylation site likely facilitates the methylation by enhancing the binding affinity between PRMT10 and its substrate.

PRMT10 displayed comparable activities over histone H4 and N-terminally GST-tagged histone H4 (GST-H4). In histone H4, the methylation site arginine-3 is located proximal to the N-terminus. Therefore, the substrate can bind to the substrate binding grooves in a linear structure as shown in the crystal structure of PRMT1-substrate complex. However, due to the blockage of the N-terminus by GST in GST-H4, a loop structure has to form in order for the argnine-3 to enter the active site of PRMT10. Our results suggest that the proximal of the methylation site to the N-terminus is not essential for the methylation catalyzed by PRMT10 and the methylation site located far away from the ends can also reach the active site of PRMT10 by forming a loop structure. Taken together, these data expand our view of PRMT structure and function. They indicate that while the family of enzymes share some key traits (e.g., a functional dimer and coherent SAM-binding domain motion), unique features of particular PRMTs, like the larger central cavity of the PRMT10 dimer, may lead to unique methylation patterns and even target substrate proteins.

2.5 Figure Legend

Figure 2.1. **Crystal structure of PRMT10**. (A) Domain architecture of PRMT10 from *Arabidopsis Thailian*. The SAM binding domain (residues 31-174) is shown in red; the β-barrel domain (residues 174-186, 236-383) in blue and the dimerization domain (or arm domain, residues 186-236) in yellow. (B) A ribbon show of the crystal structure of PMT10-SAH complex with labeling of secondary structure elements (helices X through I and β-strands 1 through 15). The bound SAH is shown as sticks and spheres. The structure is colored as displayed in Figure 2.1A. The start and end residues of PRMT10 are indicated. (C) Two views of the superimposition of PRMT10 (residues 31-383, magenta) with rat PRMT1 (residue 41-353, cyan, PDB entry 1ORI). Structure differences between PRMT10 and PRMT1 (located in the dimerization domain and two loops in the β-barrel domain) are indicated by arrows. (D) A Stereo-view representation of SAH binding. The Simulated annealing omit map of SAH contoured at 1σ (colored in blue) is shown. Hydrogen bonds are indicated by red dashed lines.

Figure 2.2. **Structure-based sequence alignment of PRMT10 and other PRMT paralogs,** including rat PRMT1 (PDB, 1ORI), rat PRMT3 (PDB, 1F3L), mouse CARM1 (PDB, 3B3F) and yeast RMT1 (PDB, 1G6Q). The secondary-structure elements are shown across the top of the sequences. The residue numberings are shown on the right side. Invariant and similar residues are highlighted in black and gray, respectively. The SAM binding domain is colored in red; the dimerization domain in yellow; and the β-barrel domain in blue. The four PRMT signature motifs are labeled. Residues involved in SAM binding and dimerization are highlighted by red and black stars respectively.

Figure 2.3. **Dimer formation of PRMT10**. (A) A ribbon show of the crystal structure of a PRMT10 dimer. The structure is colored and labeled as shown in Figure 2.1B. The location of dimer interfaces are highlighted by arrows. The dimer is formed by the interaction between the dimerization arm of one monomer and the outer surface of the other monomer. (B) Two views of a PRMT10 dimer in surface representation, with wwo monomers are colored in gray and pink respectively. (C) An expanded stereo-view of the dimer interface. Two monomers are colored as shown in Figure 2.2B. Residues involved in dimer formation are shown as sticks and labeled.

Figure 2.4. Conserved residues between PRMT10 and other PRMT paralogs, including rat PRMT1, rat PRMT3, yeast RMT1 and mouse CARM1/PRMT4, were mapped onto the structure of PRMT10. All three views are shown in surface representation where magenta represents residues with 100% conservation and yellow denotes residues with high similarity (>80% conservation). The top left view has the same orientation as shown in Figure 2.1B. The location of dimer interface and active sites are labeled.

Figure 2.5. **Surface representation of various PRMT paralogs,** including rat PRMT1 (PDB: 1ORI) (A), rat PRMT3 (PDB: 1F3L) (B), mouse CARM1 (PDB: 3B3F) (C) and Arabidopsis PRMT10 (D). For consistency, the sequences N-terminal to helix X (including helix X) are deleted from the structure. The sizes of the central cavity of different PRMT paralogs are labeled.

Figure 2.6. Methyltransferase activities of different PRMT10 constructs *in vitro*. (A) Indicated PRMT10 constructs (5 μg) were incubated with the reaction buffer containing 10 mM Tris pH 8.0, 4 mM EDTA, 1 mM DTT and 0.5 mM PMSF, 10 μM calf thymus

core histones and 4 μM methyl donor *S*-adenosyl-*L*-[methyl-^{3}H] at 30 ℃ for 5 hours (total reaction volume 20 μl). The reaction mixture was separated on a 15% SDS-PAGE and the autoradiograph of the gel is recorded. The experiments were performed in triplicate and a typical result is shown here. (B) Quantification of the results from Figure 2.7(A). The relative activities presented here were calculated by considering the activity of wide-type PRMT10 over H2A as one.

Figure 2.7. **PRMT10 exhibits a uniquely open conformation**. (A) Superimposition of PRMT10 (magenta) with rat PRMT1 (cyan, PDB: 1ORI), rat PRMT3 (yellow, PDB: 1F3L) and mouse CARM1 (blue, PDB: 3B3F). The bottom monomers are used for the alignment. The left view rotates about 90^{o} along the vertical axis with respect to the view shown in Figure 2.2A. For clarity, the bottom monomers are not shown in the middle panel. The left edges of four monomers are indicated by vertical dashed lines with corresponding colors. The distance of the leftward translation of PRMT10 and CARM relative to PRMT3 are shown. The right panel is a schematic representation of the upper right view, with PRMT monomers shown as rectangles. The superimposed bottom monomers are shown and colored in white. (B)-(E) Views of the dimeric rat PRMT1, rat PRMT3, mouse CARM1 and PRMT10. The accessibility of the active sites was represented by an angle (the A-angle) whose vertex is located at the center of the central cavity and the two rays are made by connecting the vertex and two surface residues.

Figure 2.8. **Surface electrostatics of PRMT10**. Acidic surfaces are represented in red; basic surfaces in blue and neutral surfaces in gray. Acidic surface residues are labeled and colored based on their conservation among five PRMT paralogs as indicted by the sequence alignment in Figure 2.2 (black, conservation≥80% ; green,

40%<conservation<80%; blue, 20%<conservation≤40%; unique for PRMT10, red). The location of putative substrate binding grooves, the active site, the dimer interface and helix αX are highlighted and labeled. The top left view has the same orientation as shown in Figure 2.1B.

Figure 2.9. **Conservation of total energy during PRMT10 simulations**. Total energy, an indicator of the overall simulation stability, remains relatively constants during the course of MD simulations of monomeric PRMT10 (A) and dimeric PRMT10 (B). The average of the total energy is shown in a blue lines (A) and a green line (B), respectively. The final 10 ns of each simulation, highlighted by dashed red square, was used for analysis.

Figure 2.10. **Effects of dimerization on the motion of PRMT10**. (A) Local fluctuation (B factors) of residues in dimeric and monomeric PRMT10. Major differences between monomeric and dimeric PRMT10 are highlighted by arrows. (B) The structure of a dimeric PRMT10. The two monomers are colored in blue and green respectively. The N-terminal region αY-loop-αZ , which displayed dramatically reduced local fluctuations in dimeric PMT10, is colored in magenta. Covariance analysis of dimeric PRMT10 (C) and monomeric PRMT10 (D). The values of residue-residue correlation coefficients range from blue (anticorrelated, -0.62) to red (correlated, +1.00), with non-correlated residue pairs colored in yellow. Schematic representations of the secondary structure corresponding to the residues on x-axis and y-axis are presented from left to right and bottom to top. (E), (F) Clustering of correlated residues in dimeric PRMT10 (E) and monomeric PRMT10 (F). Clusters with correlation coefficients higher than 0.7 are shown in different colors (other than gray) and all other regions are colored in gray. SAH is

shown as sticks and spheres to highlight the location of the SAM binding pocket, although it is removed from the crystal structures before MD simulations. (G) An expanded stereo-view of the SAM binding domain of Figure 2.10E, with the secondary structures labeled.

Figure 2.11. **Sequence alignment of PRMT10 orthologs from various plants**. The aligned sequences include PRMT10 from *A.thaliana, G.max, V. vinifera, P.trichocarpa, Z.mays, S.bicolor* and *O.sativa*. The alignment was conducted using Clustal W, and then manually adjusted. The secondary-structure elements and residue numbering of PRMT10 are shown across the top of the sequences, colored and labeled as shown in Figure 2.1B. Residues involved in SAH binding and dimer formation are highlighted by red and black stars, respectively. Four PRMT signature motifs are lableled.

2.6 Figures and Tables

Table 2.1 Crystallographic data and refinement statistics for PRMT10

Resolution range (Å)	33.0-2.61 (2.67-2.61)
Space group	$P2_1$
Unit-cell parameters (Å,°)	a=80.55 b=86.69 c= 114.74 β=89.98
Twinning fraction	0.5
Number of total reflections	48469
Number of unique reflections	44374
R_{sym} [a] (%) (highest shell)	12.0 (58.0)
Completeness (%) (highest shell)	96.6 (91.0)
Mean I/σ (highest shell)	13.0 (2.30)
Average redundancy	2.90 (2.60)
R_{cryst} [b] (%) (highest shell)	18.4 (36.9)
R_{free} [c] (%) (highest shell)	21.0 (40.0)
RMSD	
Bond length (Å)	0.008
Bond angels (°)	1.24
Dihedral angles (°)	16.8
Number of atoms per AU	
Protein	10745
Ligand	104
Solvent	166

[a] $R_{sym} = \Sigma|I - \langle I \rangle| / \Sigma I$ where I is the observed intensity and $\langle I \rangle$ is the average intensity of multiple symmetry-related observation of the reflection.

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

[c] $R_{free} = \Sigma||F_{obs}| - |F_{cal}|| / \Sigma|F_{obs}|$ for 10% of the data not used at any stage of the refinement.

Table 2.2 Oligomeric states of PRMT10 mutants

	Monomer Size (kDa)	- AdoHcy		+ AdoHcy		GF size (-AdoHcy) (kDa)	GF size/ Monomer size
		DLS Size (kDa)	DLS Size/ Monomer Size	DLS Size (kDa)	DLS Size/ Monomer Size		
wt	43.1	93	2.2	92	2.1	79.5	1.8
ΔN10	42.1	89	2.1	87	2.1	N.D.	N.D.
ΔN20	41.4	83	2.0	83	2.0	N.D.	N.D.
ΔN30	40.4	74	1.8	80	2.0	68	1.7
Δ214-218	42.5	111	1.3	114	1.3	37.9	0.9
Δ202-225	40.8	66	1.6	61	1.5	34.4	0.9

Figure 2.1 Crystal structure of PRMT10

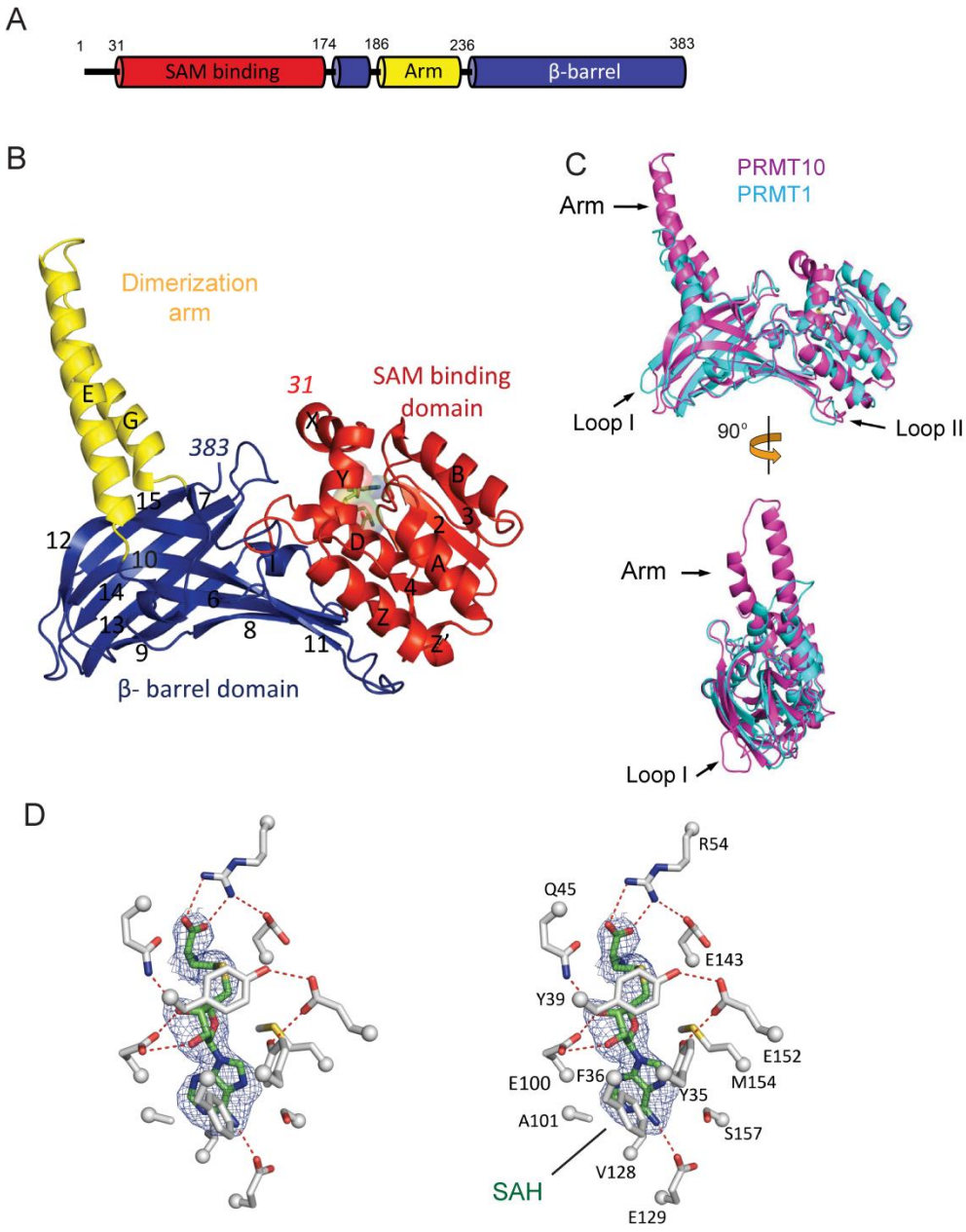


Figure 2.2 Structure-based sequence alignment of PRMT10 and other PRMT paralogs

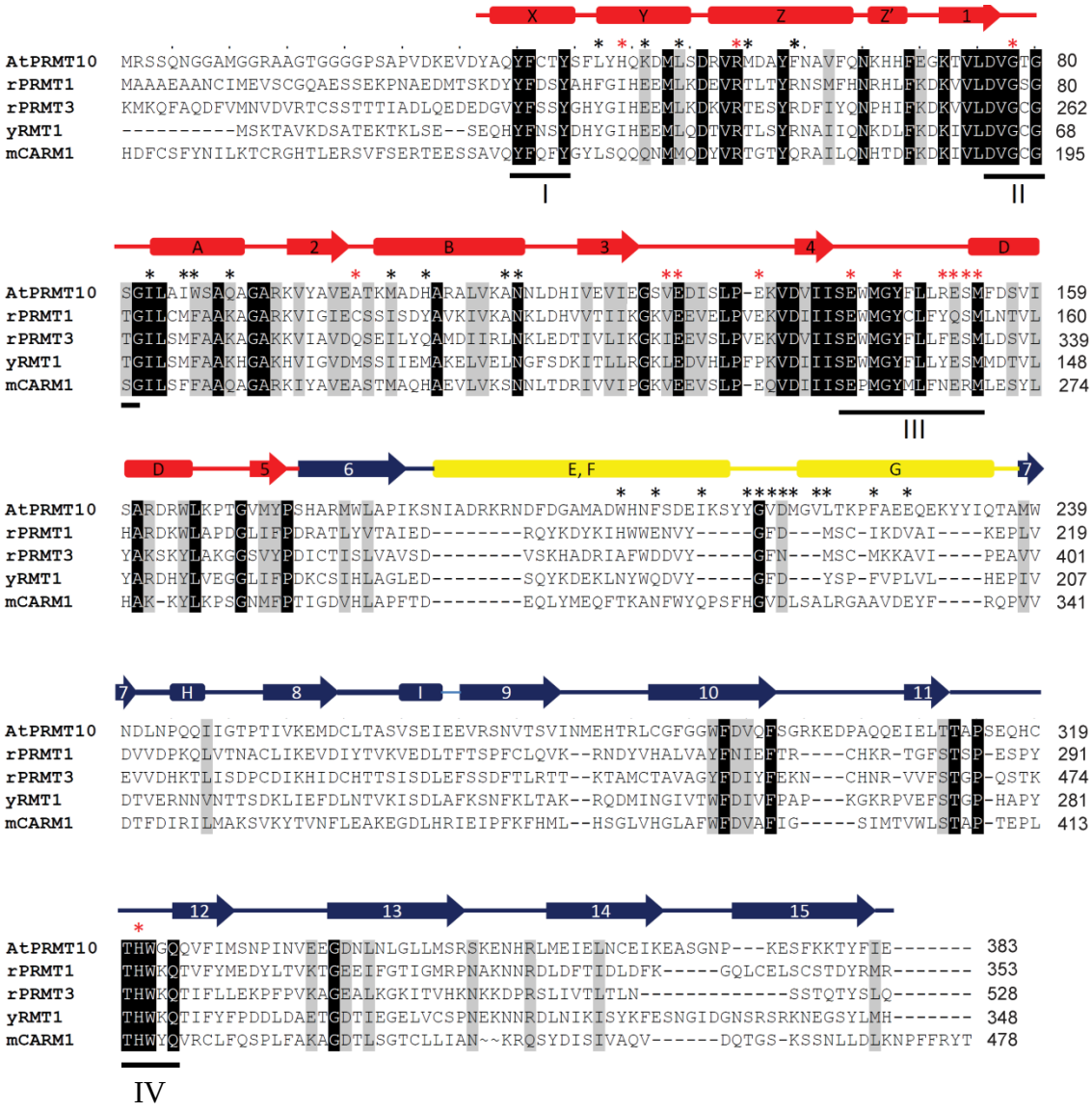


Figure 2.3 Dimer formation of PRMT10

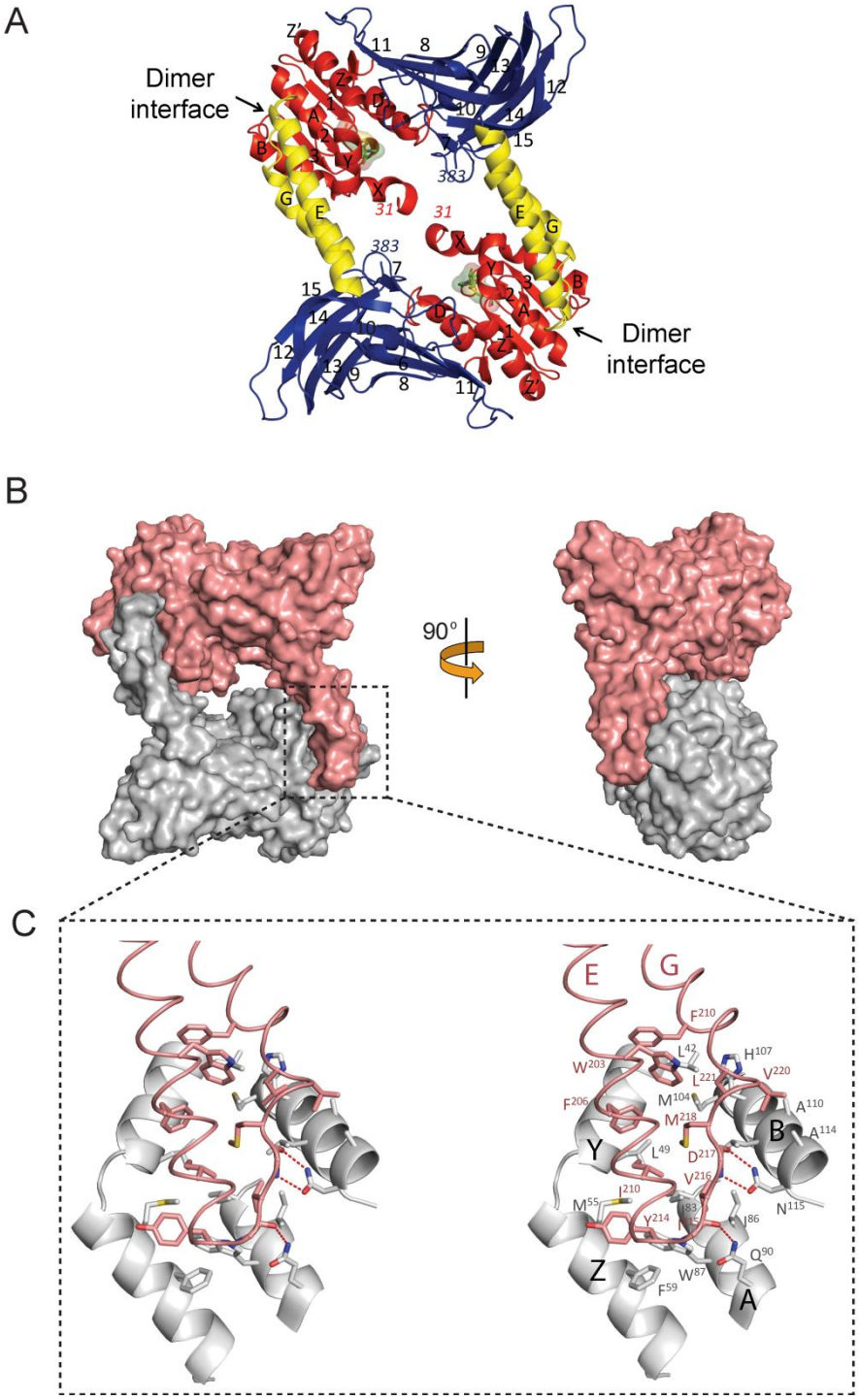


Figure 2.4 Conserved residues between PRMT10 and other PRMT paralogs

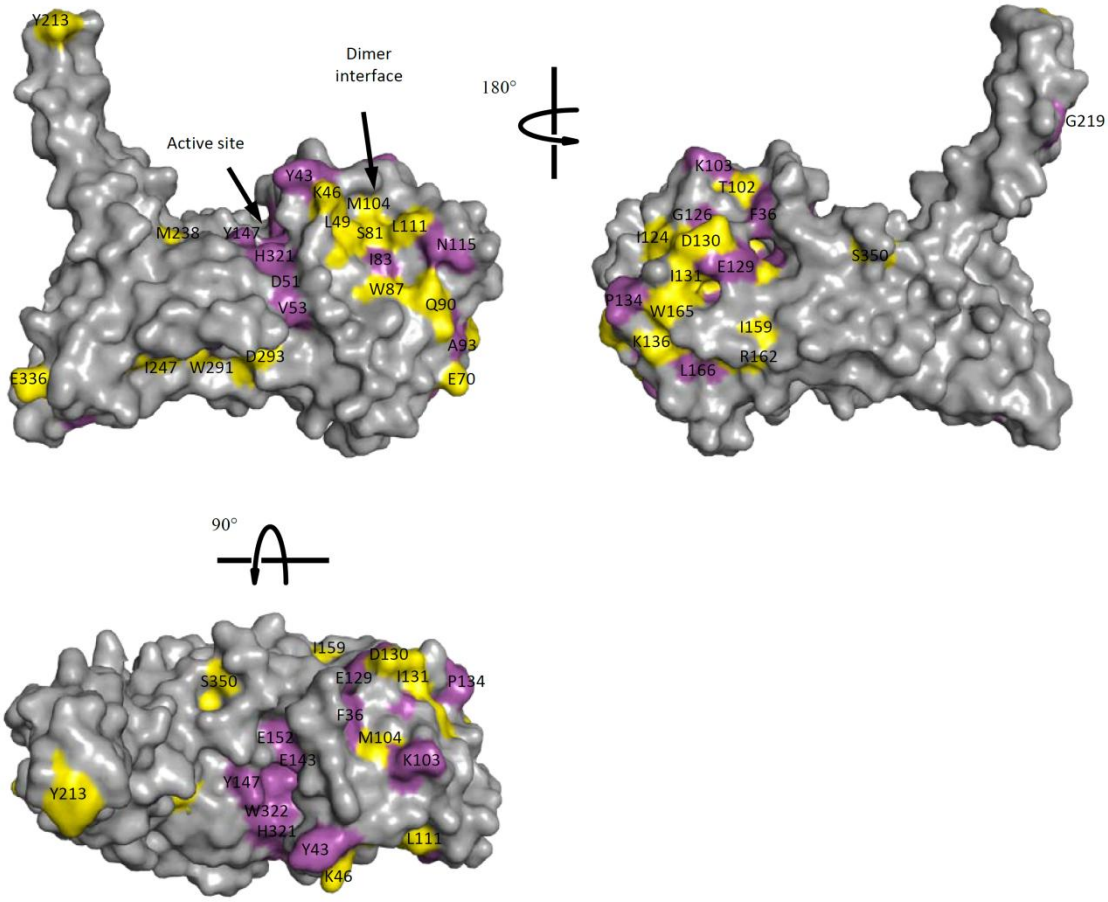


Figure 2.5 Surface representation of various PRMT paralogs

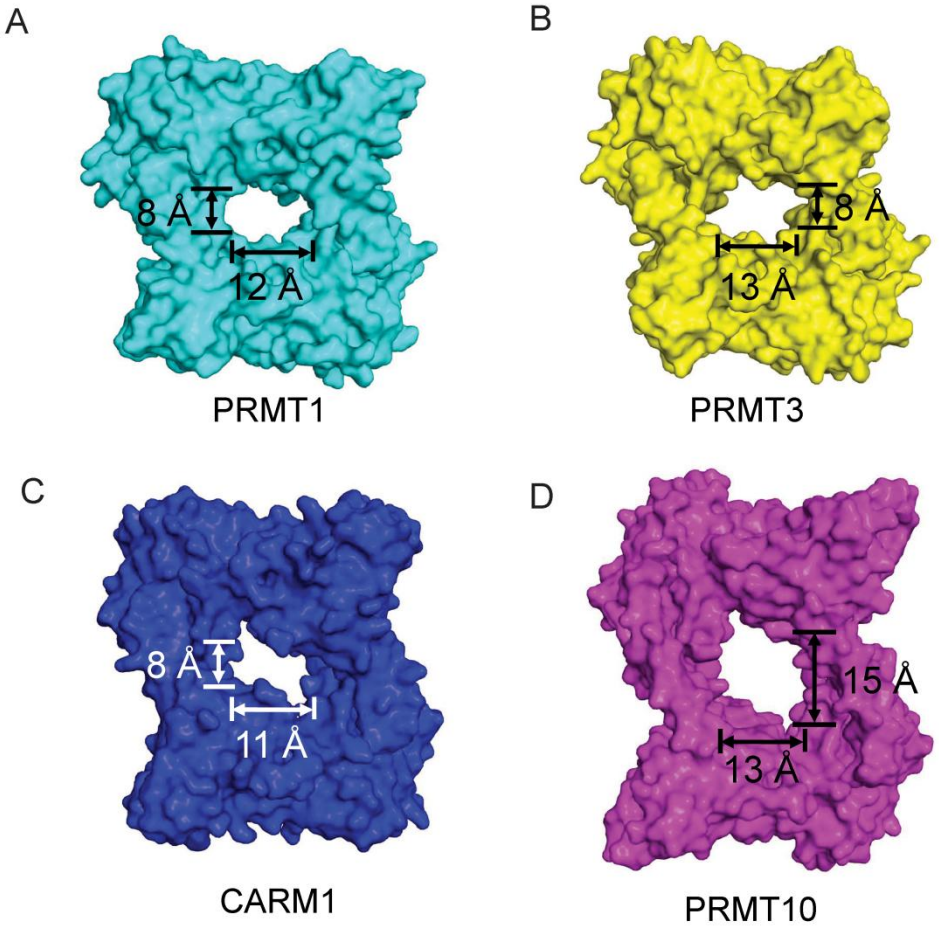


Figure 2.6 Methyltransferase activities of different PRMT10 constructs

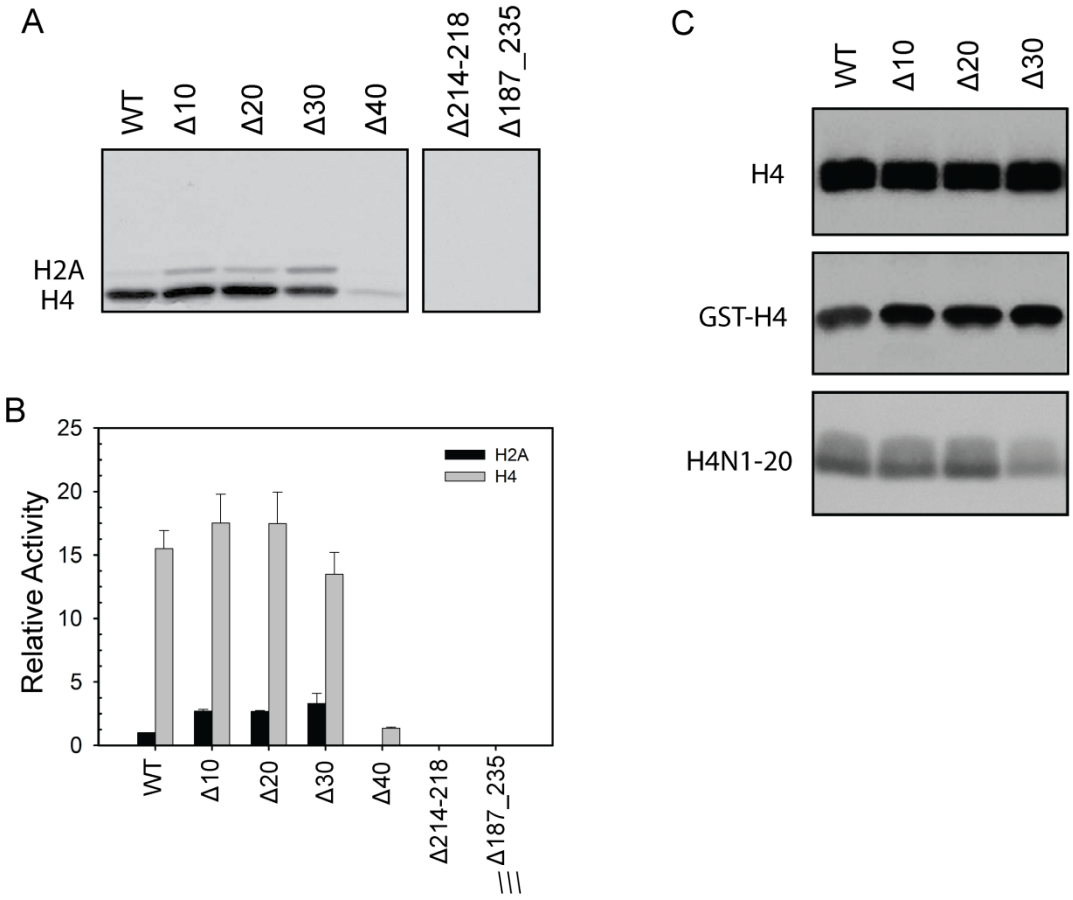


Figure 2.7 Surface electrostatic potential of PRMT10

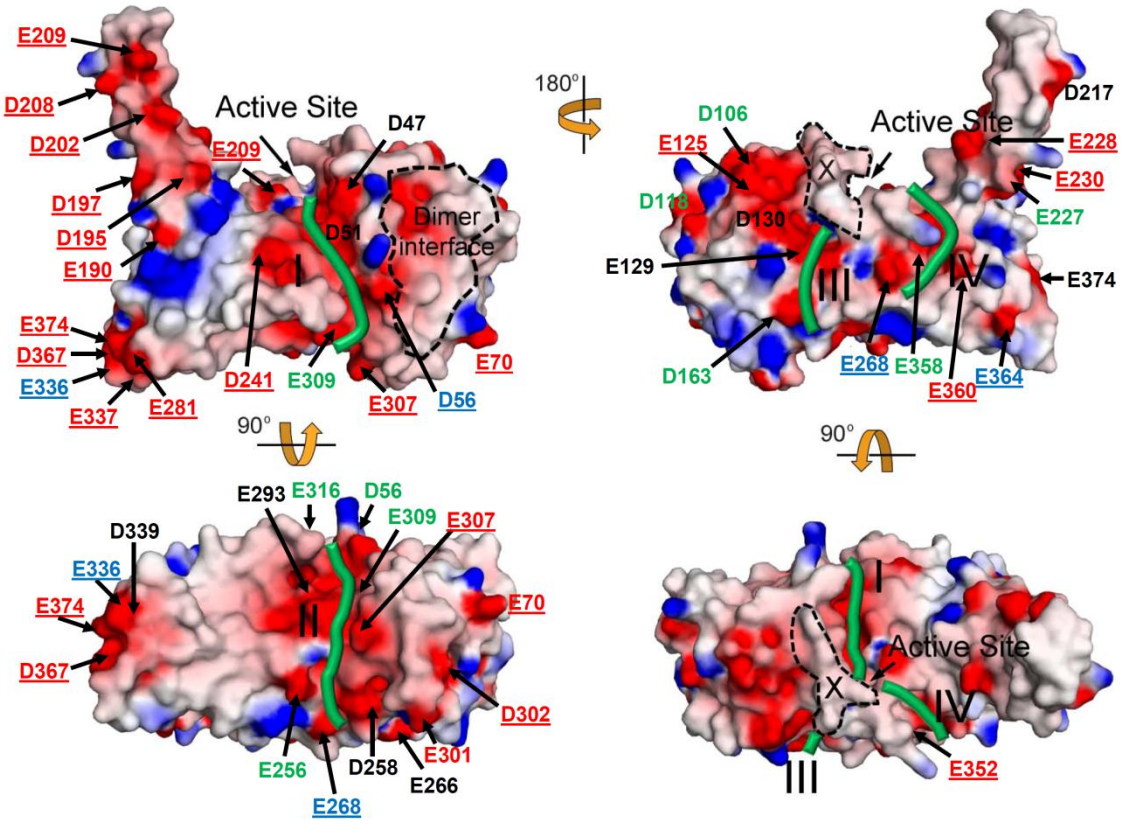


Figure 2.8 PRMT10 exhibits a uniquely open conformation

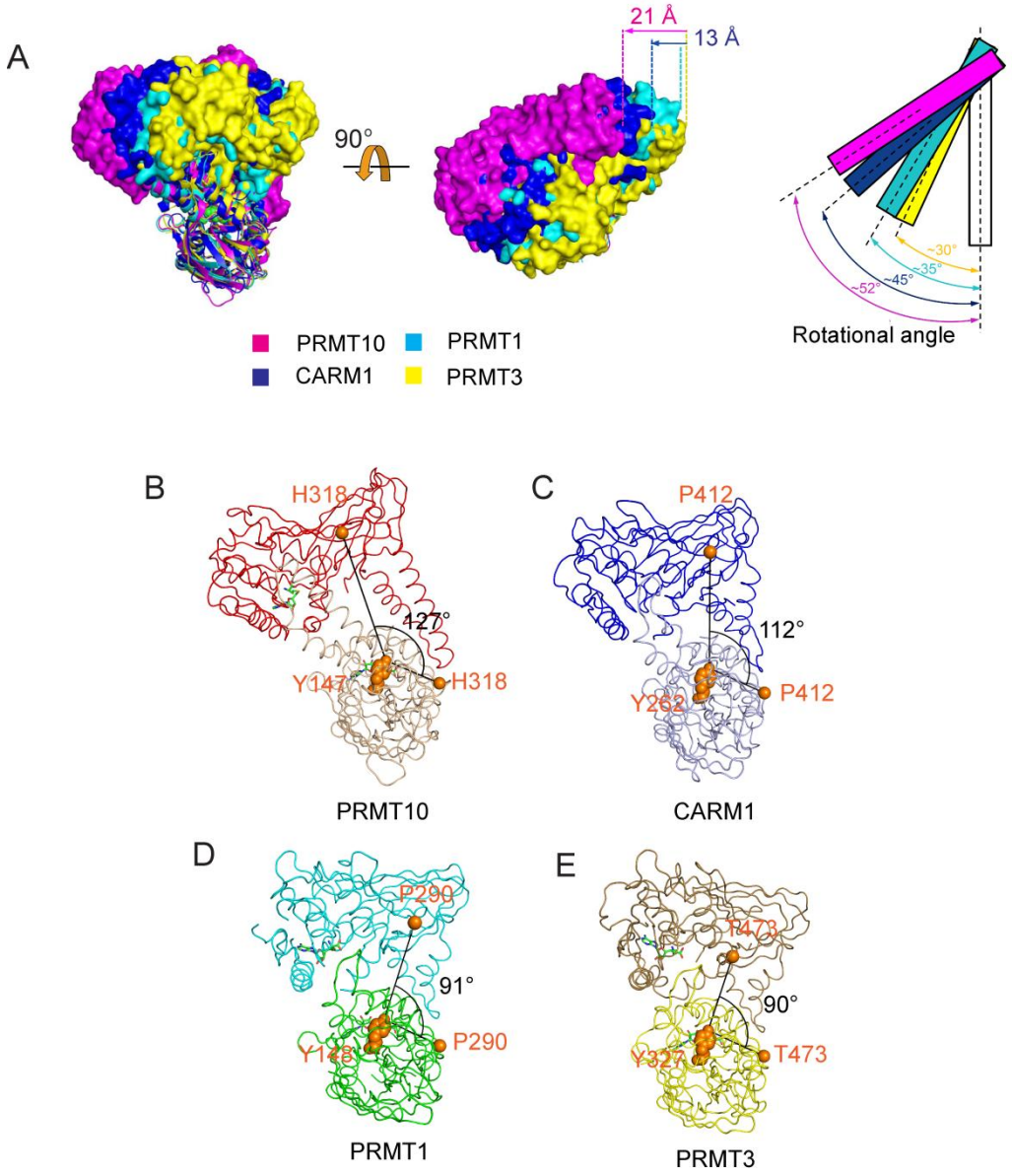


Figure 2.9 Conservation of total energy during PRMT10 simulations

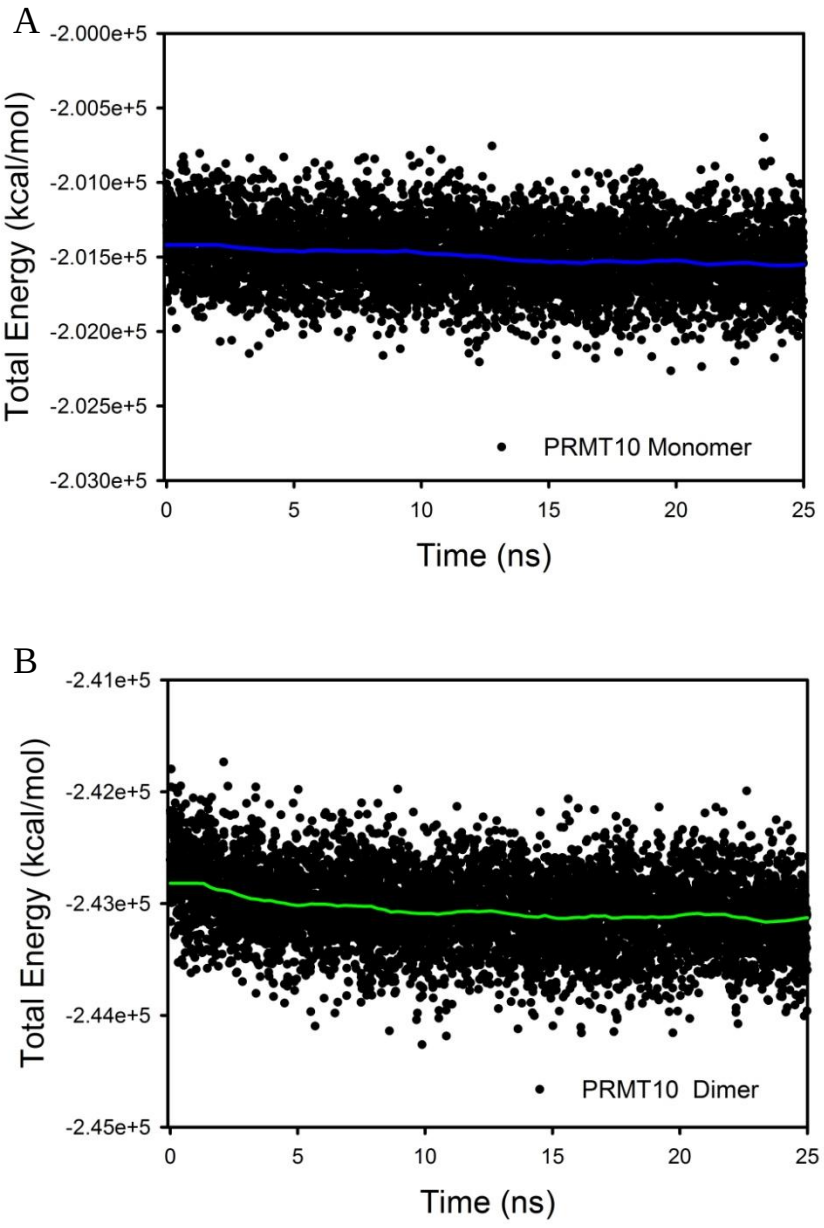


Figure 2.10 Effects of dimerization on the motion of PRMT10

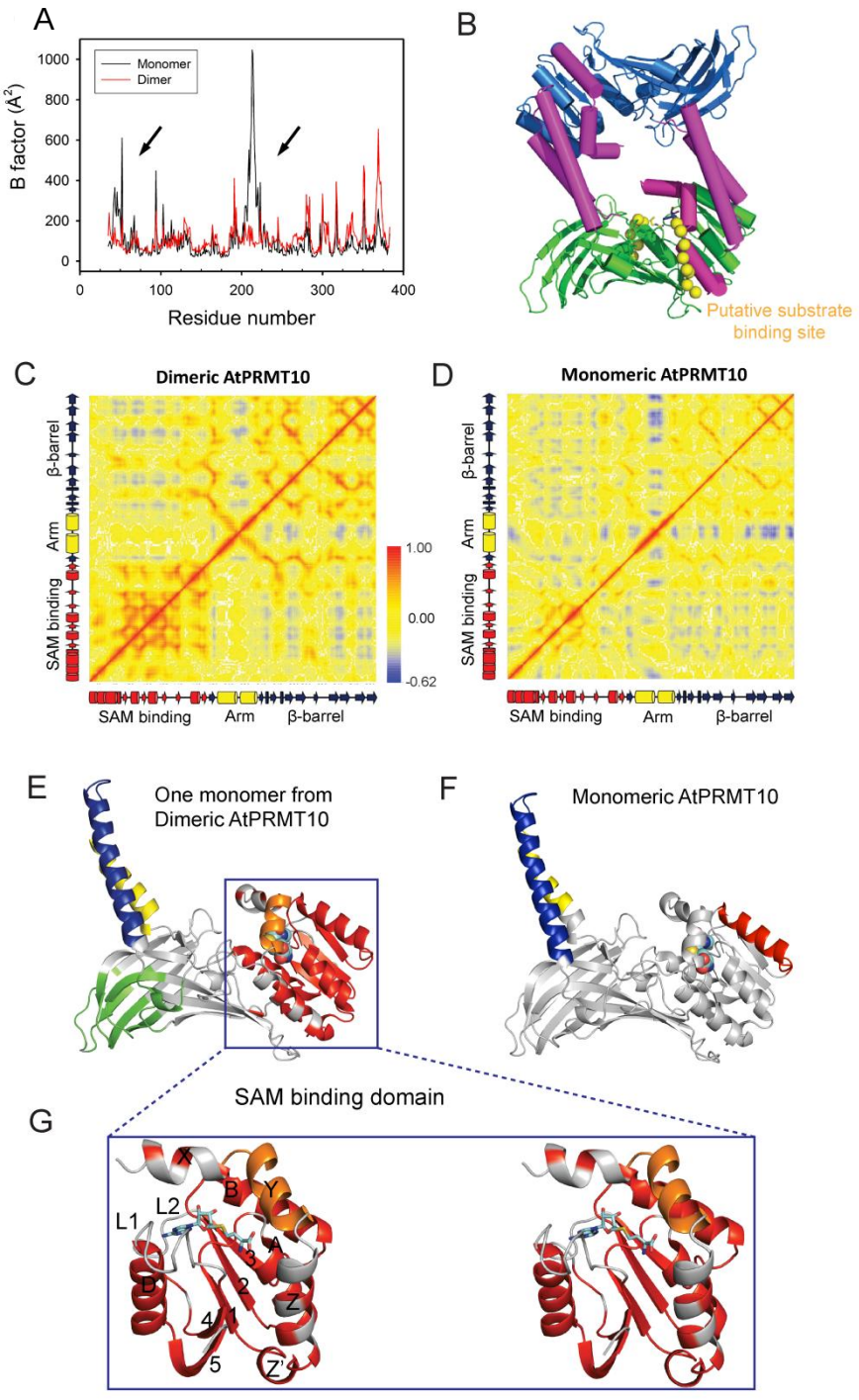
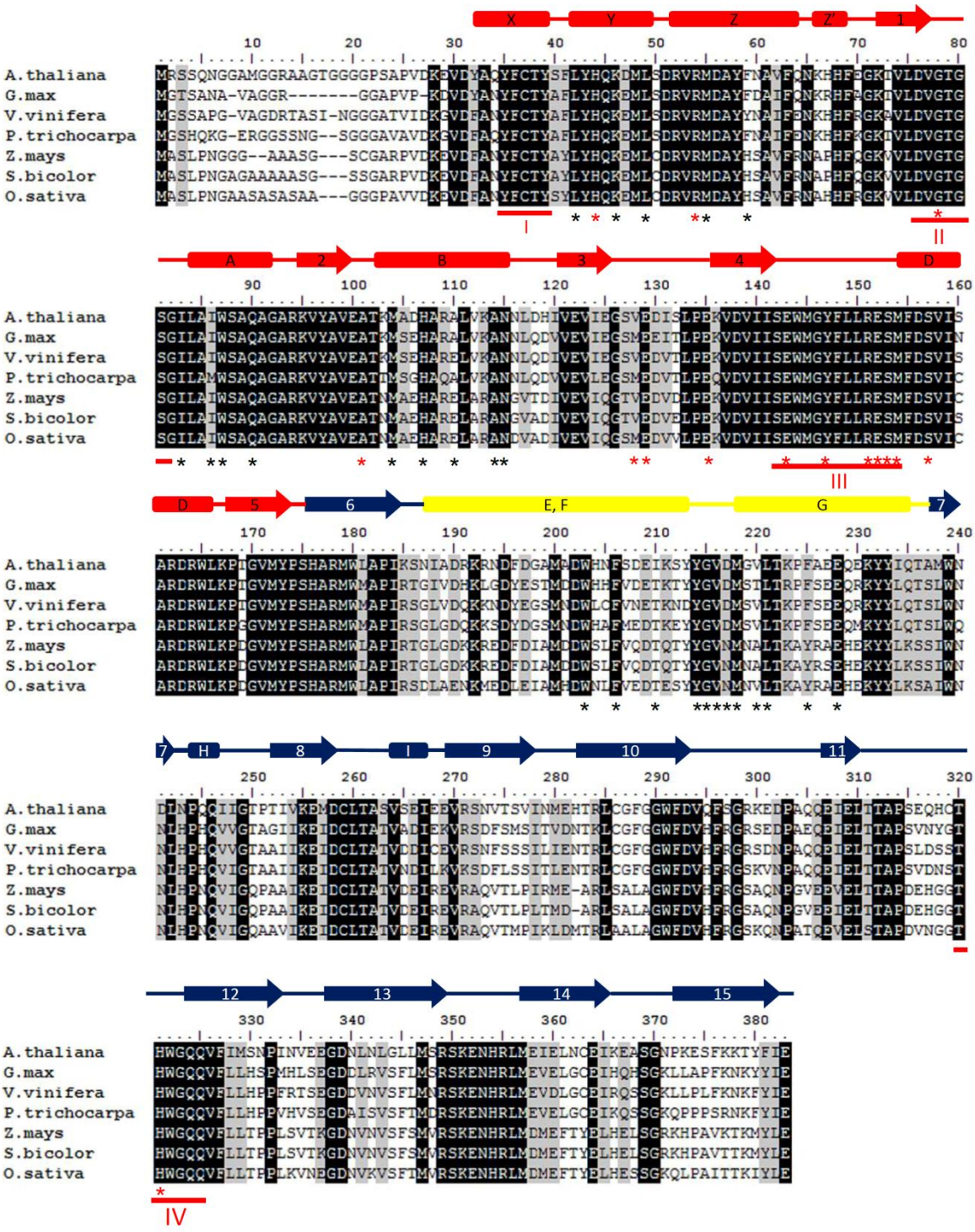


Figure 2.11 Sequence alignment of PRMT10 orthologs in various plants



Chapter 3. Activation of the human nuclear xenobiotic receptor PXR by the reverse transcriptase-targeted anti-HIV drug PNU-142721

(This chapter has been submitted to *Protein Science*)

3.1 Introduction

Nucleoside (or nucleotide) reverse transcriptase inhibitors (NRTIs) and non-nucleoside (or nucleotide) reverse transcriptase inhibitors (NNRTIs) are the two popular types of anti-HIV drugs that target RT (*97*). NRTIs directly bind to the catalytic site of RT and competitively inhibit enzyme activity, while NNRTIs bind to an allosteric lipophilic site adjacent to the catalytic site of the RT and function in a non-competitive manner (*98*). Due to their distinct mechanisms, NRIs and NNRTIs are commonly used together in HIV treatment regimen to achieve optimal efficacy (*99, 100*). Like other antiviral drugs, the usefulness of anti-HIV drugs is often compromised by mutations in the drug binding sites of HIV RT, which arise quickly due to the high mutation rate of the virus (*101*). Over the past two decades, the design of new drugs that can efficiently treat drug-resistant HIV has become a significant challenge (*102*). Indeed, the combinational use of multiple anti-HIV drugs with distinct mechanisms of action, such as the triple-cocktail treatment, has thus-far been proven to be the most effective anti-HIV regimen available (*103*).

A fundamental problem associated with coadministering multiple drugs is the risk of drug-drug interactions. Cytochrome P450 (CYP) enzymes have been shown to be centrally involved in such interactions due to the principal role these proteins play in eliminating potentially toxic xenobiotics (*104*). The nuclear receptor pregnane X receptor (PXR) is a master regulator of the expression of CYP enzymes involved in drug metabolism. PXR is the

primary sensor of xenobiotic stress in the liver and other first-pass tissues and can be activated by a structurally diverse collection of molecules including many drugs (*105, 106*). Like other nuclear receptors, PXR contains a DNA binding domain (DBD), a ligand-binding domain (LBD), and a flexible linker domain (*106*). The PXR-LBD consists of a three-layered -helical sandwich, common to nuclear receptors and a five-stranded anti-parallel-sheet, unique to PXR (*107-112*). The PXR-DBD binds specifically to the DNA response elements in the regulatory regions of CYP genes as a heterodimer with the retinoid X receptor (RXR) (*105*). Upon activation by ligand binding to the LBD, PXR recruits coactivators that facilitate the formation of the transcriptional initiation complex and the expression of CYP gene products (*105*). Coactivator proteins bind to the activation function 2 (AF-2) region of the PXR-LBD using a surface formed, in part, by the C-terminal AF helix in the receptor (*113*).

PNU-142721 was first identified as a broad-spectrum second-generation NNRTI that demonstrated potent efficacy against various type I HIV mutants resistant to first-generation antivirals like delavirdine and zidovudine (Figure 3.1) (*114*). Despite its promising *in vitro* performance, studies showed that PNU-142721 causes strong drug-drug interactions. Here, we test and validate the hypothesis that PNU-142721 causes drug-drug interactions by activating PXR. We also elucidate the structure basis for PNU-142721 mediated PXR activation by solving the crystal structure of PXR-LBD in complex with PNU-142721.

3.2 Materials and methods

Protein Expression and Purification The human PXR-LBD (residues 130-434) was coexpressed with an 88-amino acid fragment of human steroid receptor coactivator 1 (SRC-1) (623-710) in *E. coli* and purified as described previously (*111*). Purified hPXR-LBD was concentrated to 3 mg/mL in buffer containing 20 mM Tris-HCl (pH 7.8), 250 mM NaCl, 5 mM DTT, 2.5 mM EDTA and 5% Glycerol (v/v) using a Centri-prep 30K unit (Amicon) in the presence of 20-fold molar excess PNU-142721. Purified protein was flash-frozen using liquid nitrogen and stored at -80 ^{o}C.

Crystallization, Data Collection and Structural Determination. The hPXR-PNU142721 complex was crystallized by hanging-drop vapor diffusion at room temperature against 50 mM imidazole (pH 7.8) and 16% 2-propanol (v/v). Crystals were handled for cryoprotection as described previously (*111*). Data collection was conducted at SER-CAT at the Advanced Photon Source in Argonne National Labs (Beamline 22-ID). Diffraction data were indexed, integrated and scaled using HKL2000 (*115*). The structure of hPXR-PNU142721 complex was determined by molecular replacement with the MolRep module of the CCP4 suite (*116, 117*) using the crystal structure of the apo PXR as the search model. Clear rotational and translational solutions were obtained in space group $P4_32_12$. The structure was manually adjusted using WinCoot 3.1 (*118*), and refined using the Refmac module of CCP4 (*119*), with 10% of the reflections set aside for Rfree validation prior to any structural refinement. Simulated annealing omit map was generated using the program CNS (*120*). Graphica; figures were created using Pymol (*121*)

3.3 Results

PXR-PNU142721 Complex Crystal Structure

To understand the structural basis of receptor activation by this potential therapeutic lead, we determined the crystal structure of the human PXR-LBD in complex with PNU-142721. One protein-ligand complex was observed in each asymmetric unit. The structure was determined by molecular replacement in space group $P4_32_12$ and refined to 2.8 Å resolution (Figure 2A and Table 3.1). The PXR-LBD in this complex consists of a three-layered α-helical sandwich (α1/ α 3, α 4/ α 5 /α8, and α7/ α10) common to nuclear receptors and a five-stranded anti-parallel β-sheet unique to PXR (Figure 3.2A). Three regions located adjacent to the ligand-binding pocket (the loops 176-198 and 309-315, as well as residue 217) are disordered and not built in the final model (Figure 3.3). The region between residues 176 and around 198 have been reported to be disordered in other PXR-LBD structures resolved to date (*107-113*). The PXR-PNU142721 complex exhibits the same overall fold observed in previous PXR structures, sharing, for example, 0.22 Å root-mean-square deviation over 254 Cα positions with the structure of the unliganded PXR-LBD (*111*).

PNU-142721 Binding to PXR-LBD

A 2.8 Å simulated annealing omit map contoured at 2.5 σ was used to identify the electron density for PNU-142721, which binds to the PXR ligand-binding pocket in a single orientation (Figure 3.2B). PNU-142721 forms van der Waals contacts with the side chains of fourteen residues that line the ligand-binding pocket (Figure 3.2C). In addition, the side-chain amine group of Gln-285 forms a 3.3 Å weak hydrogen bond with the nitrogen atom of the furopyridine ring of PNU-142721 (Figure 3.2C). As observed in other complexes, van der Waals contacts play a principal role in stabilizing PNU-142721 within PXR's ligand-binding

pocket. Phe-288 forms a parallel aromatic stacking interaction with the furopyridine ring of PNU-142721, while Trp-299 and Tyr-306 form edge-to-face aromatic stacking contacts with the same ligand ring. The pyrimidine ring of PNU-142721 forms van der Waals interactions with the side chain of eleven residues, including a 3.6 Å interaction with Met-425 and a 5.0 Å interaction with Phe-429, which are located on AF of the PXR AF-2 surface. These interactions likely help to stabilize the active conformation of the AF-2 surface for coactivator binding.

PNU142721 Metabolite Binding to PXR-LBD

PNU-173575, a thiomethyl metabolite of PNU-142721, is a more potent activator of PXR than the unmodified compound (unpublished data from Dr. Evan Smith) (Figure 3.1B). As shown in Figure 3.1, the only difference between these two compounds is that the chloro group on the pyrimidine ring of PNU-142721 is replaced by a thiomethyl group in PNU-173575. To understand why this small change improves PNU-173575 PXR activation, we modeled PNU-173575 into PXR's ligand-binding pocket using the structure of PNU142721 as a template (Figure 3.4). To accommodate PNU-173575, the side chain of Leu-411 and His-407 were shifted by favored rotamers by alterations of 100.3^{o} and 88.5^{o}, respectively (Figure 3.4). Although the rotamer change in His-407 is likely to cause the loss of a 4.8 Å van der Waals contact between the ligand and PXR, two new van der Waals contacts are observed between the thiomethyl group and the side chain of residues that are located on the AF of the PXR AF-2 surface, including a 3.9 Å contact with Met-425, and a 4.1 Å contact with Phe-429. Thus, these new van der Waals contacts likely explain the improved PXR activation of PNU-173575, as they are better able to stabilize the active conformation of receptor's AF-2 surface.

Modeling of the structure of rat PXR-LBD in complex with PNU-142721

As shown in Figure 3.3, PNU-142721 can induce PXR-mediated CYP3A expression in rats. Due to the relatively high sequence similarity (76% identity) shared between the rat and human PXRs, the rodent receptor is expected to have a structure highly analogous to human PXR. To understand rat PXR activation by PNU-142721, we modeled the structure of the ligand-binding pocket of rat PXR based on the structure of human PXR. The following five residues that interact with PNU-142721 in human PXR were altered to the corresponding residue (listed second) in the rat sequence: Leu209Met, Met243Leu, Met246Val, Gln285Ile and His407Gln (Figure 3.5). Because they are largely conservative with respect to amino acid type, no significant change in the interaction between PNU-142721 and the PXR-LBD was observed in this modeled binding pocket. Indeed, PNU-142721 still forms fourteen van der Waals contacts with residues that line the ligand binding pocket of rat PXR. Thus, for this particular lead scaffold, rat PXR served as an excellent model for potential activation of human PXR.

3.4 Discussion

As the use of anti-HIV drug cocktails becomes increasingly necessary, dangerous drug-drug interactions may emerge as a major challenge in the antiretroviral field. Studies showed that potent HIV-1 RT inhibitor PNU-142721 produced drug-drug interactions, although the underlying mechanisms remained elusive. It has been shown that PNU-142721 activates the nuclear xenobiotic receptor PXR in a concentration-dependent manner and PNU-173575, the metabolite of PNU-142721, is a better PXR-agonist than PNU-142721 (unpublished data from Dr. Evan Smith). Altogether, these results established that PNU-142721 is a efficacious PXR agonist and it likely causes drug-drug interactions by activating PXR.

We elucidated the crystal structure of the PXR-LBD in complex with PNU-142721. The mechanism of ligand-induced PXR activation has been well studied by x-ray crystallography and other techniques. To activate PXR, a ligand needs to binds to the ligand-binding pocket of PXR and form direct interactions with the αAF helix, thereby stabilizing the AF-2 surface in the active conformation (*106*). Due to the large size and flexibility of the PXR's ligand binding pocket, however, it has been difficult to predict the exact binding mode of a novel ligand in PXR's ligand binding pocket by known PXR structures, although recent studies have demonstrated progress in this regard (*122*). Our PXR-PNU-142721 complex crystal structure reveals that van der Waals interactions play a major role in stabilizing the interaction between drug and receptor (Figure 3.2C). PNU-142721 forms van der Waals contacts with fourteen residues in PXR's ligand-binding pocket, which covers all of five ligand-binding hot spots recently identified by Ngan et.al, indicating a strong affinity between PNU-142721 and PXR. As in previous PXR-ligand structures, Gln-285 is found to

form a hydrogen bond with the ligand, indicating its conserved role in ligand interaction. PNU-142721 forms direct van der Waals contacts with Met-425 and Phe-429, which are located on the αAF helix. The contact with Met-425 has been commonly observed in previous published PXR-ligand structures, while the contact with Phe-429 only appears in the PXR-estradiol structure to date.

The crystal structure of the PXR-PNU142721 complex was then used to examine two recent observations: 1) the thiomethyl metabolite PNU-173575 exhibits improved PXR activation relative to PNU-142721; and 2) PNU-142721 is also an agonist for rat PXR (unpublished data from Dr. Evan Smith). First, we found that the modeled structure of PNU-173575 in PXR's ligand binding pocket indicates that 173575 can bind to PXR in a similar way as PNU-142721. By means of the longer and more hydrophobic thiomethyl group, PNU-173575 appears capable of forming improved interactions with the αAF helix, and thereby stabilize the AF-2 surface in an active conformation. Second, the ligand-binding pocket of rat PXR can be modeled simply by mutating five residues in human PXR to their counterparts in rat PXR. Despite several changes in relevant residues, key ligand interactions are conserved between human PXR and rat PXR, indicating that rat PXR can bind PNU-142721 in a similar way to human PXR. Our results reinforce the hypothesis that, in this particular case, the rat could be used as a good model for the prediction of potential drug-drug interaction in human.

If one wanted to modify PNU-142721 to eliminate its activation of PXR, the crystal structure of the complex could be used as a guide. The van der Waals interaction between the chloro-group on the pyrimidine ring of PNU-142721 and the αAF helix of PXR appear key to the stabilization of AF-2 surface in the active conformation. Therefore, disrupting the interaction between PNU-142721 and the AF-2 surface could be achieved by removing the

chloro-group from the pyrimidine ring. However, previous studies have shown that the pyrimidine thioether moiety is the major anti-HIV pharmacophore of PNU-142721 (*114*). The furopyridine moiety, however, was not as important (*114*), and thus, it might be safer to disrupt the interaction between PNU-142721 and PXR by modifying this ring. As observed in the complex structure, the furopyridine ring of PNU-142721 interacts with five residues in the ligand-binding pocket including Phe-288, Trp-299, Tyr 306, His-327 and Gln-285. Among these contacts, the π-π interaction between the side-chain of Phe-288 and the furopyridine ring provide the major stabilizing forces, keeping the furopyridine ring in the current orientation. Disruption of this interaction by introducing a bulky group on the C2 atom of the ligand (see Figure 3.1), which is spatially closest to Trp-299 (see Figure 3.5), may effectively destabilize the interaction of PNU-142721 with PXR. Either PNU-142721 or Trp-299 would have to move to accommodate this bulky group. Trp299 is located on four-stranded β-sheet, which is known for its rigidity and does not change its conformation even upon the binding of large PXR ligands like rifampcin. Therefore, it is not likely that Trp-299 would adjust its position to accommodate the bulky group. Also, there is no observable space around Trp-299 for its side-chain to take alterative rotamers to accommodate the bulky group. Therefore, it would be expected that PNU-142721 might move away from Trp-299 toward the other side of the ligand-binding pocket. This movement will break three important interactions between PXR and the furopyridine ring of PNU-142721, including the π-π interaction with Phe-288, the van der Waals interaction with Tyr-306, and the hydrogen bonding with the side chain of Gln-285. Although several new van der Waals interactions might form meanwhile, they are not likely to compensate for the losses mentioned above.

Taken together, the proposed modification here might mitigate PXR agonist activity by disrupt the binding of PNU-142721 to PXR.

In summary, the data presented here reveal the detailed mechanism of PNU-142721-mediated PXR activation. Using the structural model presented here, we explained why the metabolite PNU-173575 is a better PXR agonist and why PNU-142721 can also activate rat PXR. Based on the structural data presented here, several routes toward the redesign of PNU-142721 were proposed that may limit PXR activation while keeping intact the HIV-1 RT inhibitory activity of this compound.

3.5 Figure Legends

Figure 3.1 Chemical Structures of PNU-142721 and its metabolite PNU-173575

Figure 3.2 Crystal structure of the human PXR-LBD complex with PNU-142721. (A) The overall structure of PXR-LBD in complex with PNU-142721. The α-helices, β-strands and loops are rendered in cyan, magenta and grey respectively. PNU-142721 is colored by elements, with green, red, blue and yellow representing carbon, oxygen, nitrogen and sulfur atoms respectively**.** Note the proximity of the PNU-142721 to αAF of the AF-2 surface of PXR. (B) The 2.8 Å simulated annealing omit map contoured at 2.5 σ is shown in blue mesh. PNU-142721 is shown and colored by elements. (C) Stereoview of the ligand-binding pocket of the PXR-PNU142721compelx. Residues that line the ligand-binding pocket of PXR are shown in grey and labeled. The hydrogen bond is represented by a red dash line.

Figure 3.3 Analysis of the ligand binding pocket of PXR-LBD-PNU complex. Stereoview of a superimposition between the apo-PXR (blue), PXR-PNU142721 (green) and PXR-rifampicin (red) complexes. Major secondary structures around the ligand-binding pocket are labeled. The hydrogen bond is represented by a red dash line.

Figure 3.4 Modeled structure of human PXR-LBD in complex with PNU-173575. Residues that line PXR-LBD are shown in grey, except for His-407 and Leu-411 which undergo rotamer changes in the presence of PNU-173575. The side-chain position of His-407 and Leu-411 in the presence of PNU-142721 and PNU-173575 are represented in blue and magenta respectively.

Figure 3.5 Modeled structure of rat PXR-LBD in complex with PNU-142721. Residues that line rat PXR-LBD and conserved between human PXR and rat PXR are labeled and colored in grey, while residues that are different between human PXR and rat PXR are highlighted in yellow (rat PXR) and magenta (human PXR) respectively

3.6 Figures and Tables

Table 3.1 Crystallographic statistics for the PXR-PNU142721 Complex

Resolution (Å) (highest shell)	46-2.8 (2.84-2.80)
Space group	$P4_32_12$
Unit Cell dimension (Å)	a=b= 92.3, c=84.4
No. of total reflections	9166
No. of unique reflections	8944
R_{sym} [a] (%) (highest shell)	9.4 (53)
Completeness (%) (highest shell)	99.5 (96.9)
Mean I/σ (highest shell)	42 (5.1)
R_{cryst} [b] (%) (highest shell)	23.8 (28)
R_{free} [c] (%) (highest shell)	28.9 (43)
RMSD	
Bond lengths (Å)	0.044
Bond angles (^{0})	3.324
Dihedral angles (^{0})	23.6
Improper angles (^{0})	3.9
Number of atoms	
Protein	2157
Solvent	16
Ligand	20

[a] $R_{sym} = \sum \left| I - \langle I \rangle \right| / \sum I$ where I is the observed intensity and $\langle I \rangle$ is the average intensity of multiple symmetry-related observation of the reflection.

[b] $R_{cryst} = \sum \left\| F_{obs} \right| - \left| F_{cal} \right\| / \sum \left| F_{obs} \right|$ where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.

[c] $R_{free} = \sum \left\| F_{obs} \right| - \left| F_{cal} \right\| / \sum \left| F_{obs} \right|$ for 10% of the data not used at any stage of the refinement.

Figure 3.1 Chemical structures of PNU-142721 and its metabolite PNU-173575

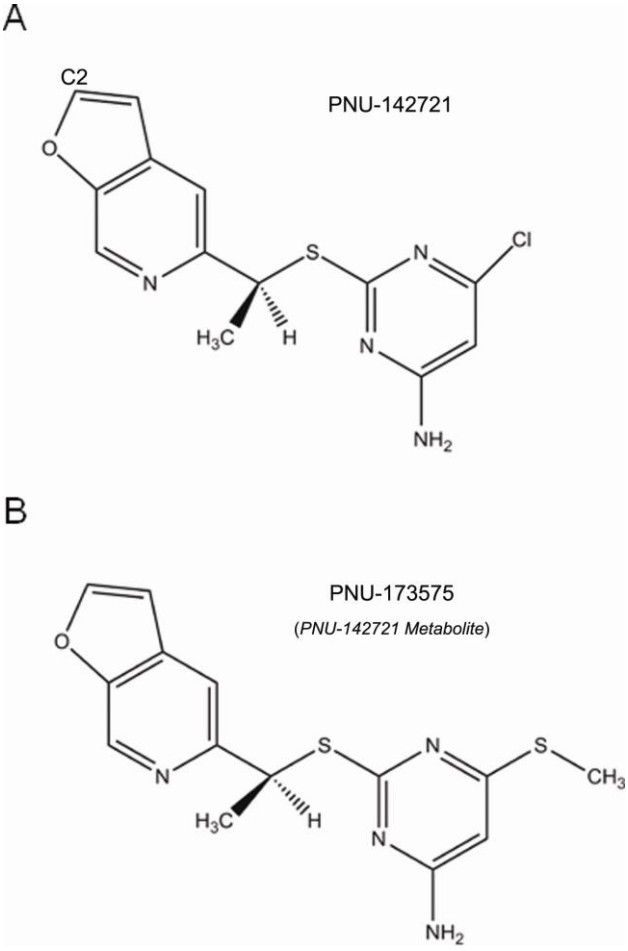


Figure 3.2 Crystal structure of the human PXR-LBD complex with PNU-142721

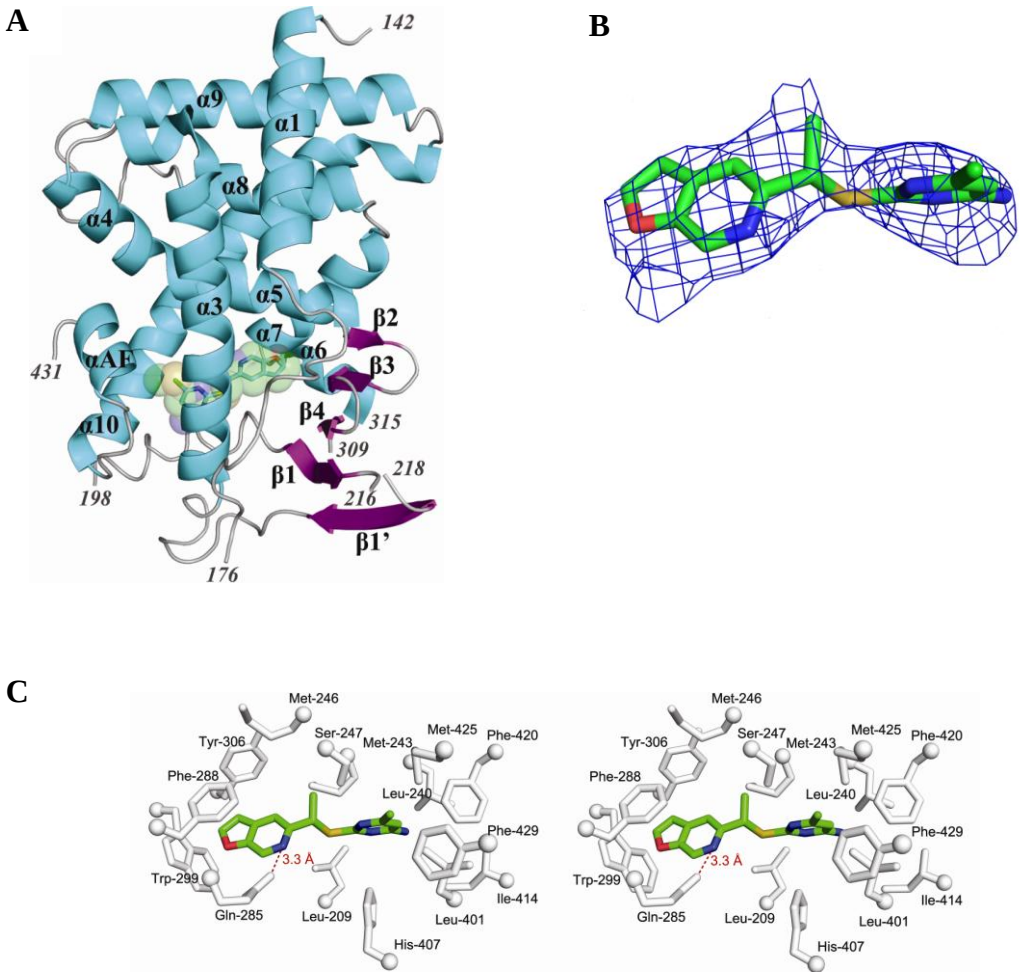


Figure 3.3 Analysis of the ligand binding pocket of PXR-LBD-PNU complex

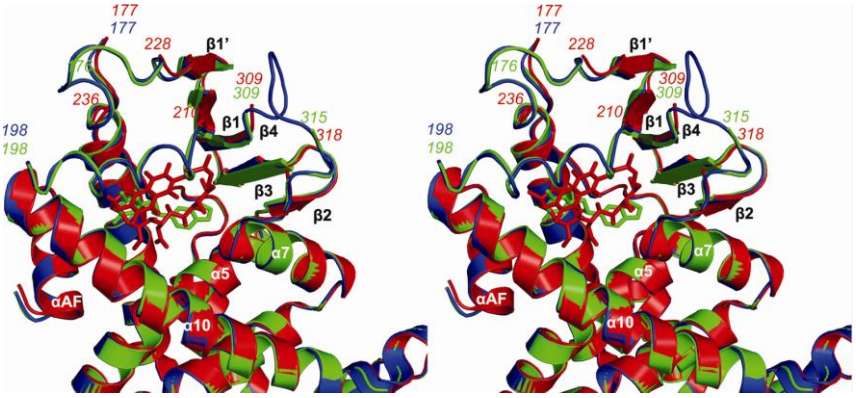


Figure 3.4 Modeled structure of human PXR-LBD in complex with PNU-173575

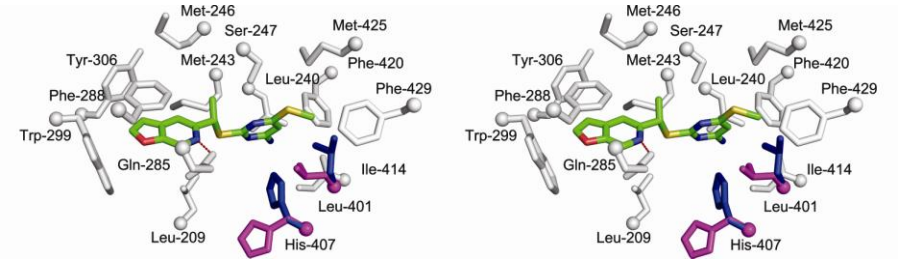
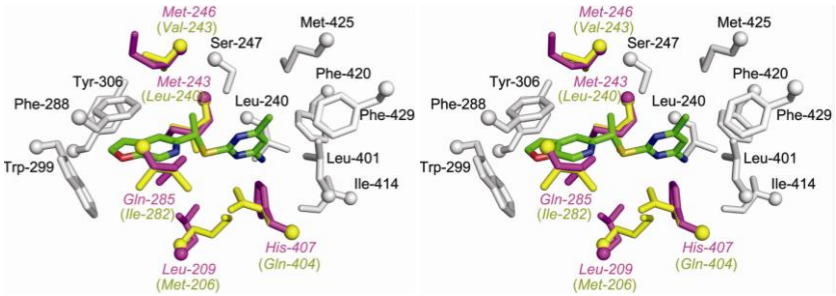


Figure 3.5 Modeled structure of rat PXR-LBD in complex with PNU-142721



Appendix: Preliminary studies of the adhesin PilC1 from *Neisseria gonorrhoeae*

Abstract

PilC1 is a protein essential for pilus-mediated adherence of *Neisseria gonorrhoeae* to human cells. In this work, we developed a new method for recombinant expression of PilC1 in *E.coli*. We identified a calcium binding site in the C-terminal region of PilC1 (residues 519-1037). Our experiments showed that calcium binding increased the thermostability of PilC1, but did not have a significant impact on the secondary structure of PilC1. We also examined the role calcium binding plays in the function of PilC1 *in vivo*. Our results show that calcium binding facilitates the surface localization of PilC1. Moreover, calcium binding to PilC1 is essential for pilus-mediated adherence of *N. gonorrhoeae* to human ME180 cells, as well as the aggregation of *N. gonorrhoeae*. Taken together, our results suggest the essential role of calcium binding in the normal function of PilC1.

Introduction

N. gonorrhoeae is a species of Gram-negative bacteria which cause a variety of human diseases including gonorrhea, conjunctivitis, pharyngitis, proctitis and urethritis. *N. gonorrhoeae* infection starts from the adherence of the bacteria to human epithelial cells, which is mediated by the hair-like pili extended from the bacterial surface (*123*). Following the initial attachment, the bacteria use other surface machineries to bind more tightly to and invade the epithelial cells (*124*).

The pili of *N. gonorrhoeae*, or the type IV pili, are polymeric structures composed of the major pilin (PilE) and a variety of minor pilins, such as PilC, PilV and PilX (*125*). Type IV pili are also found in many other Gram-negative bacteria such as *Pseudomonas aeruginosa, Dichelobacter nodosus* and *Vibrio cholera* (*126*). In addition to colonizing host cells, type IV pili are essential for many other bacterial activities, such as twitching motility, DNA uptake and biofilm formation (*127, 128*). Disruption of type IV pili is often associated with dramatically reduced virulence (*129*).

The 110 kDa protein PilC was first identified as a type IV pilus associated adhesin (*130*). Later studies showed that PilC can also be associated with the outer-membrane of the bacteria (*131*). The N-terminal region of PilC mediates the interaction with human cell receptors (*132*). Purified PilC can block the adherence of *N. gonorrhoeae* to human cells (*133*). PilC is involved in pilus biogenesis and the regulation of pilus retraction (*134*). Two copies of *PilC* genes, *PilC1* and *PilC2*, are present in the genome of *N. gonorrhoeae*. The protein PilC1 and PilC2 are very similar in sequence and functionally interchangeable (*132*). Several lines of evidence have suggested that the initial adherence of *N. gonorrhoeae* to human cells is mediated by the human cell receptor CD46 (*135, 136*). However, some studies

indicated that CD46 may not be essential for the initial attachment of *N. gonorrhoeae* and PilC to human cells (*137*).

The yield of PilC from current methods is too low to meet the requirement of extensive *in vitro* studies. In this work, we have developed a new expression and purification method that is significantly more efficient than previous methods. In addition, we identified a calcium binding site in the C-terminal region of PilC1 and examined the role of calcium binding in the structure and function of PilC1.

Materials and methods

Protein constructs, expression and purification

Standard ligation independent cloning techniques, as described by Stols *et al.* (*20*), were employed in the construction of expression plasmids encoding the segments of PilC1 used in this study. The amplified DNA fragments were treated and cloned into empty pMCSG7-Lic-His expression vector. All expression plasmids used in this study were sequence verified.

The expression plasmids were transformed into *E.coli* BL21(DE3)Origami 2 (Stratagene). Bacteria were grown in LB medium supplemented with ampicillin, streptomycin and tetracycline at 37 ^{o}C with shaking. After the OD_{600} reached 0.6, isopropyl β-D-thiogalactoside was added to a final concentration of 0.2 mM and bacteria were grown for another 12 hours at 16 ^{o}C with shaking. Bacteria were harvested and resuspended in loading buffer (50 mM sodium phosphate pH 7.6, 500 mM NaCl, 25 mM imidazole) supplemented with 0.5 mM EDTA, 0.1% Triton X-100, 1 mM PMSF, one tablet of a protease inhibitor cocktail (Roche) and 1 mg/ml lysozyme. After 1 hour of incubation on ice, the resuspended cells were sonicated on ice for 2 min and the lysate was centrifuged at 45,000 × g for 90 min at 4 ^{o}C. The supernatant was passed through a 0.2 μm filter (Millipore) and then loaded onto a 5 ml high performance HisTrap^{TM} column (GE Life Sciences), equilibrated with loading buffer. The column was washed with 100 ml loading buffer before bound protein was eluted with elution buffer (50 mM sodium phosphate pH 7.6, 500 mM NaCl, 500 mM imidazole). The eluted protein was concentrated and loaded on a HiLoad^{TM} 16/60 Superdex 200 column (GE Life Sciences) equilibrated with sizing buffer (20 mM Tris-HCl pH 7.5, 250 mM NaCl, 2 mM DTT, 5% Glycerol). Protein containing fractions were

concentrated, flash frozen in liquid nitrogen and stored at -80 ^{o}C. Purified protein was >95% pure by SDS-PAGE.

Calcium binding assay

Terbium ion was used as a substitute of calcium ion due to their similar properties (ionic radius and coordination properties). More importantly, terbium can be excited by tryptophan emission and produce an emission at a certain wavelength. Terbium binding to protein, then, can be quantified by measuring its specific emission. FRET experiment was carried out at 25 ^{o}C using a SPEX Fluorolog-3 Research T-format Spectrofluorometer (Horiba Jobin Yvon, Edison, NJ). Terbium chloride was diluted in protein buffer (20 mM Tris-HCl, ph7.5, 250 mM NaCl) to a working concentration of 10 mM. Protein (400 μl at 0.2 mg/ml) was added to the cuvette, and terbium chloride was added in 0.2 μl increments. The tryptophan residues in protein were excited at 283 nm, and terbium fluorescence emission was recorded at 543 nm. The error bars represent the standard deviation of three replicates. The data was fitted to a one-site binding equation and the dissociation constant is determined to be 0.9 ±0.2 μM.

Circular Dichroism spectroscopy

Far-UV CD spectra of PilC1 were measured with a JASCS J715 CD spectropolarimeter for the 205-260 nm using a 1 mm path-length cell region at 22 ^{o}C. Thermal denaturation of PilC1 was analyzed by monitoring ellipticity changes at the wavelength of 222 nn while the sample was heated at a constant rate of 2 min/ ^{o}C. The basic protein buffer contained 20 mM Tris-HCl, ph7.5 and 250 mM NaCl.

Cell lines and growth conditions

The ME180 cells were grown in McCoy's 5A medium supplemented with 10% fetal bovine serum (FBS) at 37 ^{o}C, 5% CO_2. *N. gonorrhoeae* strains were grown on GCB-agar plates, containing Kellogg's supplement (*138*) at 37 ^{o}C in 5% CO_2.

Cell adherence assay

The experiment was performed as previously described (*137*) but employed the human epithelial ME180 cell line. ME180 cells were incubated with *N. gonorrhoeae* strain FA1090. After one hour, ME180 cells were washed three times with medium and lysed with 1% saponin. The colony forming units were quantified by plating serial dilutions of cell lyses. The assay was performed in triplicate. The error bars represent the standard error of the normalized fraction of GC adherence. To examine the impact of recombinant PilC1 on the adherence of GC to human cells, ME180 cells were pre-incubated with different concentrations of PilC1 or BAS for 30 min before the incubation with *N. gonorrhoeae* strain FA1090. The adherence of GC to ME180 cells was inhibited in a dose-dependent manner by full-length PilC1 protein (residues 23-1037), but not by PilC1-CTD (residues 519-1037) and BSA.

Cell surface exposure assay

Purified PilC1 specific monoclonal antibodies, 4B5.10 and 1D5, are biotinylated with NHS-PEO4-Biotin. They were incubated with *Neisseria gonorrhoeae* strain FA1090 for 30 min and the antibodies bound to the GC surface were detected by binding of HRP-streptavidin. HRP was then detected by production of light in the presence of chemiluminescent substrate. All experiments were done in triplicate.

Aggregation of *N. gonorrhoeae*

N. gonorrhoeae strains were grown at 37 ^{o}C overnight on GCB-agar plate. Bacteria were scraped off and resuspended in PBS buffer to OD_{600} ≈1. The suspension was kept at room temperature and the OD_{600} of the supernatant was measured at different time-points.

Results and Discussion

To identify the PilC1 constructs that can be expressed as soluble proteins in *E.coli*, we test-expressed a series of PilC1 constructs in *E.coli* cell lines (Table 1). Our results show that only full-length (FL) PilC1 (residues 23-1037) and the C-terminal domain (CTD) of PilC1 (residues 519-1037) can be expressed as soluble proteins in *E.coli* BL21 (DE3) Origami 2. As indicated by SDS-PAGE, the purity of purified FL-PilC1 and PilC-CTD was over 95% (Figure 1). The expression level of PilC1-CTD is considerably higher than that of FL-PilC1. Accordingly, these two PilC1 constructs were used in the following *in vitro* studies of PilC1.

Previous studies showed that purified FL-PilC can block the adherence of *N. gonorrhoeae* to human cells by competing with pilus-associated native PilC (*133*). Here, we examined the impact of the recombinant FL-PilC1 obtained in this study on the adherence of *N. gonorrhoeae* strain FA1090 to human ME180 cells. As shown in Figure 2, FL-PilC1 inhibited the adherence in a dose-dependent manner with an IC_{50} of about 10 nM. As expected, bovine serum albumin that does not specifically interact with human cells did not have a significant effect on the adherence. Taken together, our results suggest that the recombinant FL-PilC1 obtained in this study is likely well folded. Our results also showed that PilC1-CTD did not affect the adherence of GC to human ME180 cells, which is consistent with the previous observation that the N-terminal region of PilC is responsible for the interaction with human receptors (*132*).

Recent studies from our lab identified a calcium binding site in the PilY1 protein from *Pseudomonas aeruginosa* that shares high homology with the PilC protein from *N. gonorrhoeae (139)*. The result from sequence alignment suggested that PilC1 may also have a calcium binding site in its C-terminal domain (Figure 3). Calcium binding experiments showed that PilC1-CTD can bind calcium in a manner that can be fitted with a one-site binding model (K_d = 0.9 ± 0.2 μM) (Figure 4). Moreover, the aspartic residues (D731A, D733A and D739A) in the calcium binding site are essential for calcium binding as suggested by the fact that mutation of any of them could abolish the binding of calcium to PilC1 (Figure 4).

The impact of calcium binding on the secondary structure of PilC1-CTD was examined by Circular Dichroism (CD) spectroscopy. Our results clearly showed that the presence of calcium did not cause significant change in the CD profile of PilC1-CTD in the far-UV region, suggesting that calcium binding does not affect the secondary structure of PilC1-CTD (Figure 5). We also examined the impact of calcium binding on the thermostability of PilC1-CTD. The presence of calcium shifted the thermodenaturation curve to higher temperature, increasing the melting temperature (T_m) for about 5 ^oC. This observation indicates that calcium binding increases the thermostability of PilC1-CTD.

To examine the role calcium binding plays in the function of PilC1 *in vivo*, we engineered several *N. gonorrhoeae* strains, in which the wide-type *PilC1* gene was substituted with *PilC1* genes that carry different calcium binding site mutations. We first examined the expression of PilC1 mutants in engineered *N. gonorrhoeae* strains. Western-blot analysis showed that PilC1-D731A had wide-type level expression, while PilC1-D733A

and PilC1-D739A had significantly reduced expression (Figure 7). As a result, only PilC1-D731A was used in the following analysis.

We next investigated the exposure of PilC1 mutants on the surface of *N. gonorrhoeae* (see materials and methods for details). The monoclonal antibody 4B5.10 can recognize pilus-associated and outer-membrane associated PilC1, while 1D5 can only recognize pilus-associated PilC1. The non-piliated *N. gonorrhoeae* strain (*PilC*1^{+}/*PilC2*$^{-}$/*PilE*$^{-}$) was used as a negative control here. Our results showed that the amount of PilC1-D731A on type IV pili or the bacterial surface as a whole was significantly less than that of wide-type PilC1 (Figure 8), suggesting that calcium binding promotes the surface exposure of PilC1. In addition, the *N. gonorrhoeae* strain expressing PilC1-D731A (*PilC1-D731A*$^{+}$*/PilC2*$^{-}$) exhibited dramatically reduced adherence to human ME180 cells compared to the *N. gonorrhoeae* strain expressing wide-type PilC1 (*PilC1-D731A*$^{+}$*/PilC2*$^{-}$) (Figure 8).

Finally, we examined the aggregation of the *N. gonorrhoeae* that is mainly mediated by type IV pili. As expected, the hyperpiliated *N. gonorrhoeae* strain (*PilC1*$^{+}$*PilC2*$^{-}$/*PilT*$^{-}$) displayed higher levels of aggregation than the piliated strain (*PilC*1^{+}/*PilC2*$^{-}$), and the non-piliated strain (*PilC*1^{+}/*PilC2*$^{-}$/*PilE*$^{-}$) displayed the lowest level of aggregation (Figure 9). Interestingly, the GC strain expressing PilC1-D731A (*PilC1-D731A*$^{+}$/*PilC2*$^{-}$) displayed similar levels of aggregation with the non-piliated GC strain, suggesting that the *N. gonorrhoeae* strain expressing PilC1-D731A may have some defects in pilus biogenesis.

Figure lengends

Figure 1. SDS-PAGE of purified PilC1 constructs. (A) The C-terminal domain (CTD) of PilC1 (residues 519-1037). (B) Full-length (FL) PilC1 (residues 23-1037). M represents the protein marker. 12% SDS-PAGE were used here.

Figure 2. Purified full-length PilC1 protein inhibits in a dose-dependent manner the adherence of *Neisseria gonorrhoeae* to human epithelial cells. ME180 cells were preincubated with increasing concentrations of PilC1 proteins or Bovine Serum Albumin (BSA) for 30 minutes before being washed and incubated with GC strain FA1090 in our cell adherence assay. The assay was performed in triplicate on three days. The error bars represent the standard error of the normalized fraction of GC adherence.

Figure 3. Partial sequence alignment of *Pseudomonas aeruginosa* PilY1 and PilC from various *Neisseria* species. The residues involved in calcium ion binding in PilY1 are denoted by asterisks and labeled. Sequence conservation is highlighted in red.

Figure 4. Binding of Tb^{3+} to different PilC1 constructs. Different PilC1-CTD constructs were titrated with increasing concentrations of $TbCl_3$ as described in our materials and methods. The solid line is computer fit of the data using one-site binding model.

Figure 5. Impact of calcium binding on the secondary structure of PilC1. The far-UV CD spectra of PilC1-CTD in the absence or presence of $CaCl_2$ were recorded as described in the materials and methods.

Figure 6. Impact of calcium binding on the thermostability of PilC1. The change of ellipticity at 222 nm with the increase of temperature was recorded as described in the

materials and methods. The melting temperature (Tm) indicates the temperature that the protein loses half of its ellipticity at 222 nm.

Figure 7. Expression of PilC1 calcium binding mutants in *Neisseria gonorrhoeae*. Western-blot of the whole cell lysates of GC strains expressing different PilC constructs are shown here. N represents the GC strain that does not express PilC.

Figure 8. Impact of calcium binding on the surface exposure of PilC1 and the adherence of *Neisseria gonorrhoeae* to human ME180 cells. The surface exposure of PilC1 proteins and the adherence of different GC strains to human ME180 cells were performed as described in the materials and methods. The experiments were performed in triplicate on three days. The error bars represent the standard error of the normalized fraction of surface exposure or GC adherence.

Figure 9. Impact of calcium binding on aggregation of *Neisseria gonorrhea.* The OD_{600} of the supernatant of the solutions that contain different GC strains were measured at different time-points.

Table 1. Expression of pilC1 constructs in various *E.coli* cell lines. The sign "–"insoluble expression. The sign "+/-" indicates that the protein is soluble in the presence of the maltose binding protein (MBP) tag, but is insoluble after the removal of the MBP tag. The expression level is qualitatively reflected by the number of "+".

Protein Constructs	BL21 (DE3) Gold & pLysS & RIPL		BL21(DE3) Origami 2	
	Lic-his	Lic-MBP-his	Lic-his	Lic-MBP-his
Full-length				
23-1037	-	-	+	-
C-terminal domain				
519-974	-	+/	-	+/
519-982	-	+/	-	+/
519-984	-	+/	-	+/
519-986	-	+/	-	+/
519-1014	-	+/	-	+/
519-1037	-	+/	++++	+/
492-1037	-	+/	-	+/
513-1037	-	+/	-	+/
572-1037	-	+/	-	+/
592-1037	-	+/	-	+/
602-1037	-	+/	-	+/
N-terminal domain				
23-485	-	-	-	-
119-485	-	-	-	-
180-485	-	-	-	-
224-485	-	-	-	-

Figure 1. SDS-PAGE of purified PilC1 constructs

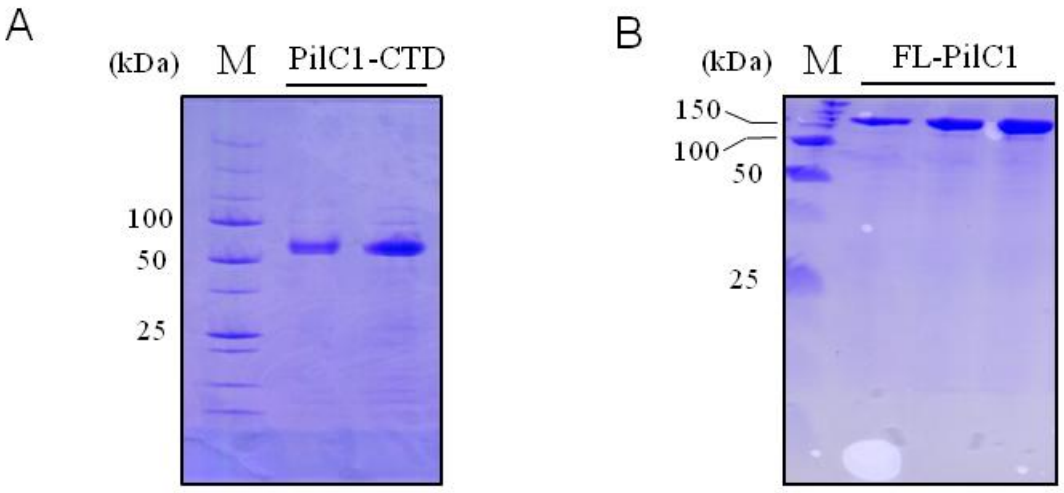


Figure 2. Purified full-length PilC1 protein inhibits in a dose-dependent manner the adherence of *Neisseria gonorrhoeae* to human epithelial cells

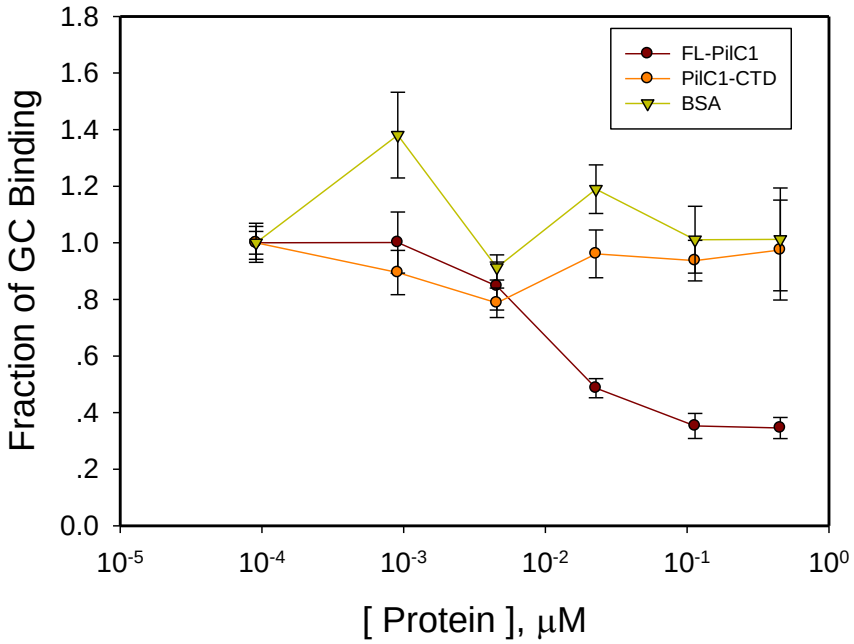


Figure 3. Partial sequence alignment of *Pseudomonas aeruginosa* PilY1 and PilC from various *Neisseria* species

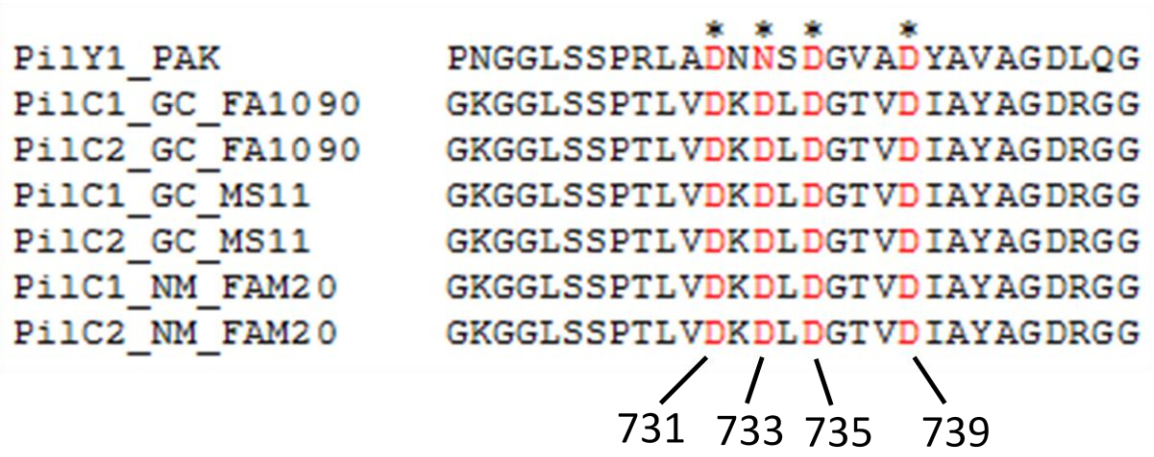


Figure 4. Binding of Tb^{3+} to different PilC1 constructs

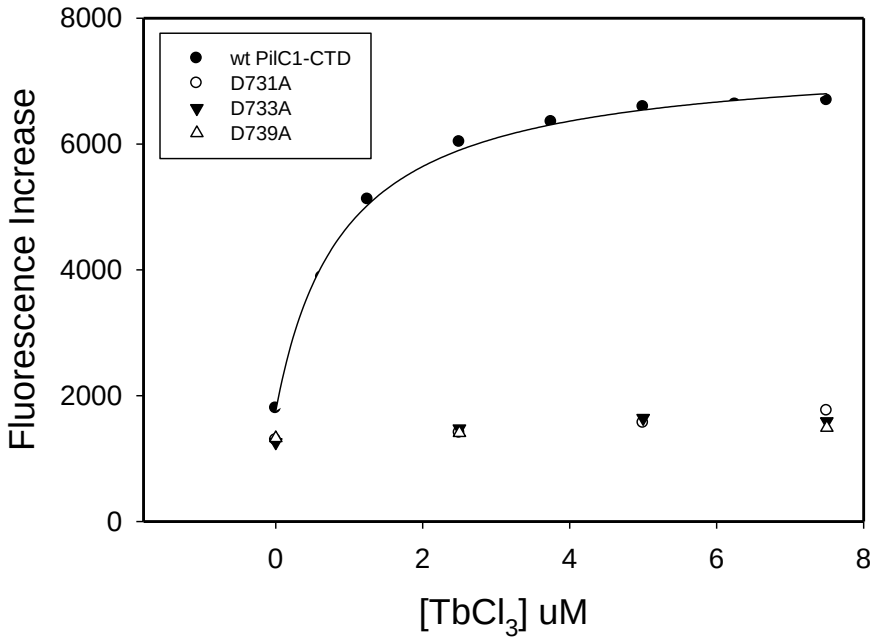


Figure 5. Impact of calcium binding on the secondary structure of PilC1

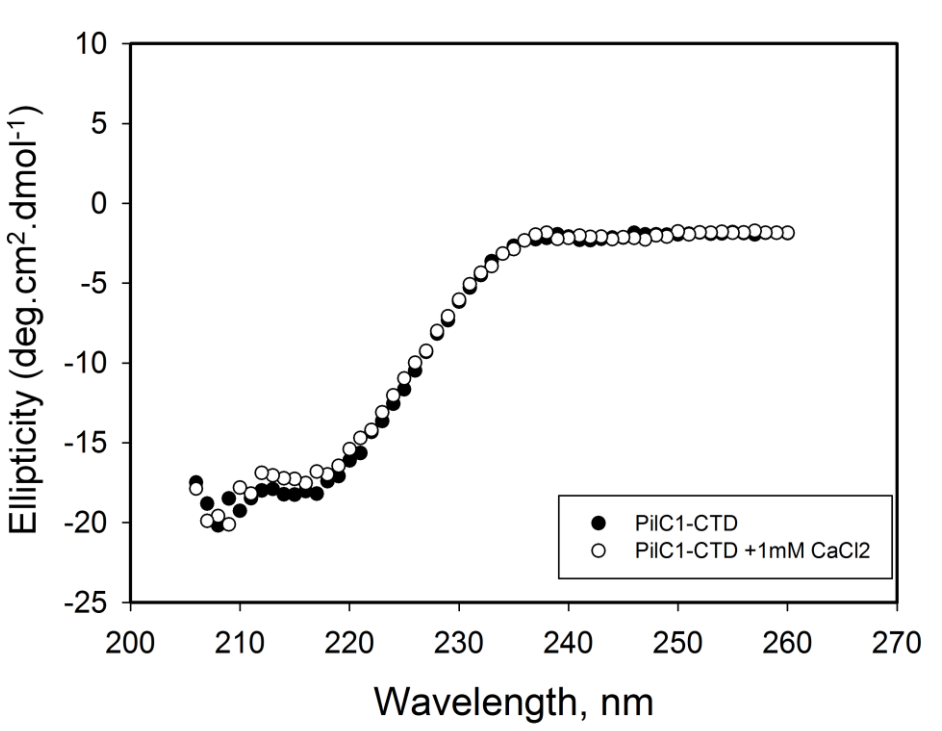


Figure 6. Impact of calcium binding on the thermostability of PilC1

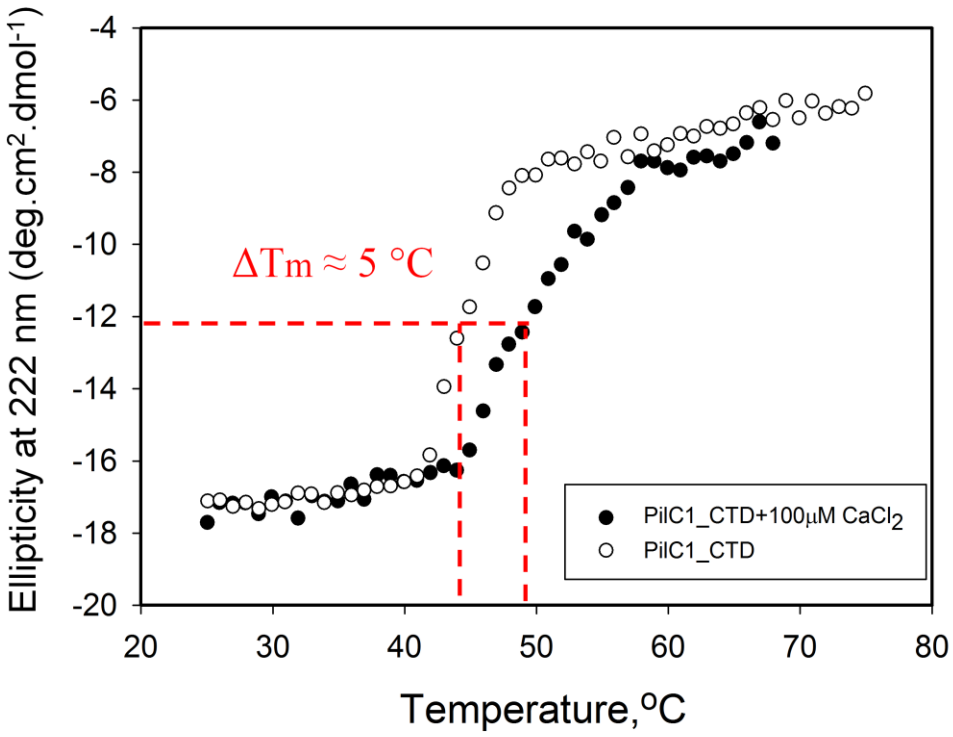


Figure 7. Expression of PilC1 calcium binding mutants in *Neisseria gonorrhoeae*

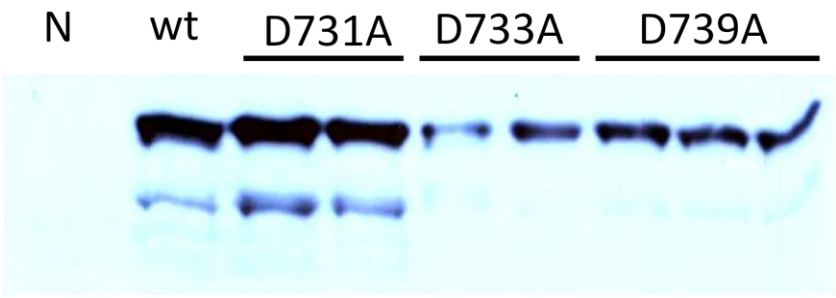


Figure 8. Impact of calcium binding on the surface exposure of PilC1 and the adherence of *Neisseria gonorrhoeae* to human ME180 cells

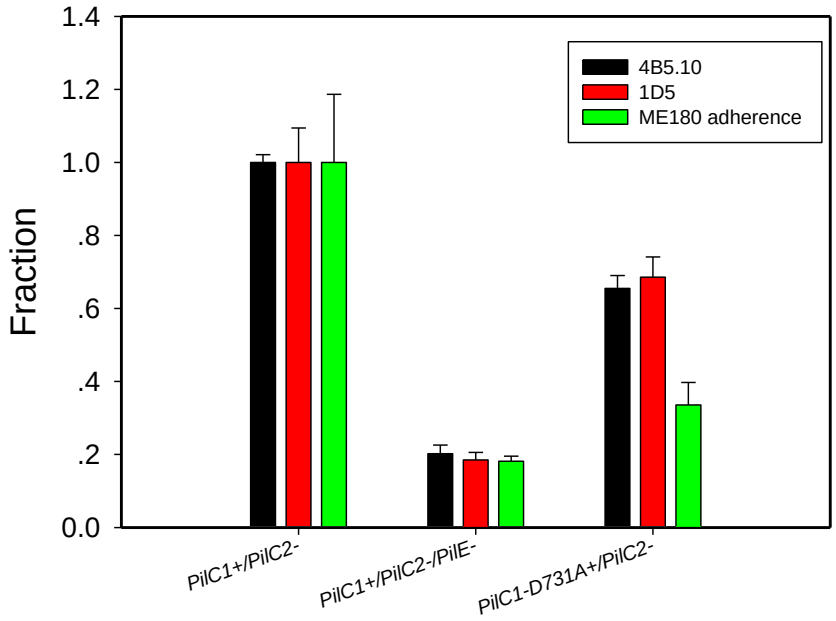
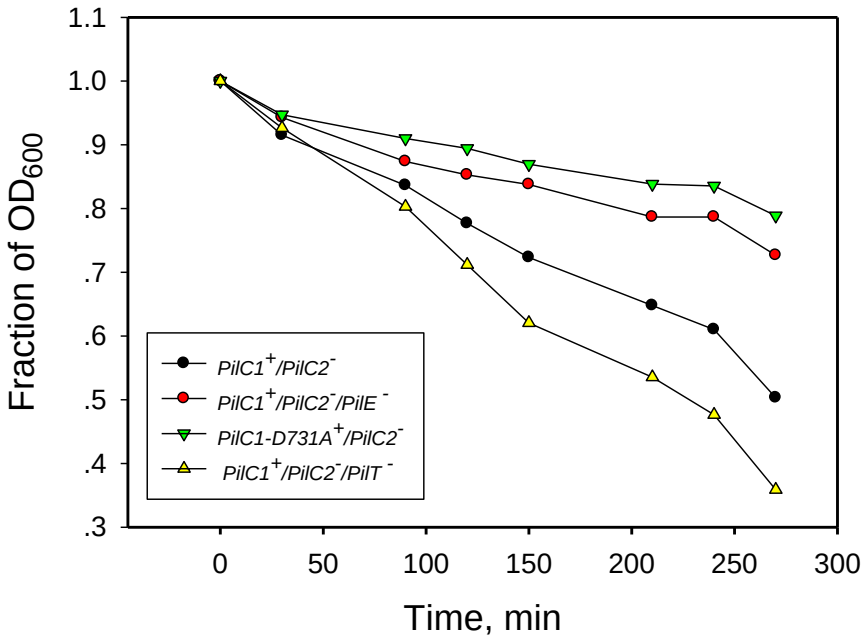


Figure 9. Impact of calcium binding on aggregation of *Neisseria gonorrhoeae*



References

1. Brochet, M., Rusniok, C., Couve, E., Dramsi, S., Poyart, C., Trieu-Cuot, P., Kunst, F., and Glaser, P. (2008) Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of *Streptococcus agalactiae*, *Proc Natl Acad Sci USA 105*, 15961-15966.

2. Frost, L. S., and Koraimann, G. (2010) Regulation of bacterial conjugation: balancing opportunity with adversity, *Future Microbiol 5*, 1057-1071.

3. Frost, L. S., Leplae, R., Summers, A. O., and Toussaint, A. (2005) Mobile genetic elements: the agents of open source evolution, *Nat Rev Microbiol 3*, 722-732.

4. Mazel, D., and Davies, J. (1999) Antibiotic resistance in microbes, *Cell Mol Life Sci 56*, 742-754.

5. de la Cruz, F., Frost, L. S., Meyer, R. J., and Zechner, E. L. (2010) Conjugative DNA metabolism in Gram-negative bacteria, *FEMS Microbiol Rev 34*, 18-40.

6. Gomis-Ruth, F. X., and Coll, M. (2006) Cut and move: protein machinery for DNA processing in bacterial conjugation, *Curr Opin Struct Biol 16*, 744-752.

7. Lanka, E., and Wilkins, B. M. (1995) DNA processing reactions in bacterial conjugation, *Annu Rev Biochem 64*, 141-169.

8. Abdel-Monem, M., Taucher-Scholz, G., and Klinkert, M. Q. (1983) Identification of *Escherichia coli* DNA helicase I as the traI gene product of the F sex factor, *Proc Natl Acad Sci U SA 80*, 4659-4663.

9. Traxler, B. A., and Minkley, E. G., Jr. (1988) Evidence that DNA helicase I and *oriT* site-specific nicking are both functions of the F TraI protein, *J Mol Biol 204*, 205-209.

10. Byrd, D. R., Sampson, J. K., Ragonese, H. M., and Matson, S. W. (2002) Structure-function analysis of Escherichia coli DNA helicase I reveals non-overlapping transesterase and helicase domains, *J Biol Chem 277*, 42645-42653.

11. Matson, S. W., and Ragonese, H. (2005) The F-plasmid TraI protein contains three functional domains required for conjugative DNA strand transfer, *J Bacteriol 187*, 697-706.

12. Ragonese, H., Haisch, D., Villareal, E., Choi, J. H., and Matson, S. W. (2007) The F plasmid-encoded TraM protein stimulates relaxosome-mediated cleavage at oriT through an interaction with TraI, *Mol Microbiol 63*, 1173-1184.

13. Street, L. M., Harley, M. J., Stern, J. C., Larkin, C., Williams, S. L., Miller, D. L., Dohm, J. A., Rodgers, M. E., and Schildbach, J. F. (2003) Subdomain organization

and catalytic residues of the F factor TraI relaxase domain, *Biochim Biophys Acta 1646*, 86-99.

14. Dostal, L., and Schildbach, J. F. (2010) Single-stranded DNA binding by F TraI relaxase and helicase domains is coordinately regulated, *J Bacteriol 192*, 3620-3628.

15. Matson, S. W., Sampson, J. K., and Byrd, D. R. (2001) F plasmid conjugative DNA transfer: the TraI helicase activity is essential for DNA strand transfer, *J Biol Chem 276*, 2372-2379.

16. Abdel-Monem, M., and Hoffmann-Berling, H. (1976) Enzymic unwinding of DNA. 1. Purification and characterization of a DNA-dependent ATPase from Escherichia coli, *Eur J Biochem 65*, 431-440.

17. Sikora, B., Eoff, R. L., Matson, S. W., and Raney, K. D. (2006) DNA unwinding by *Escherichia coli* DNA helicase I (TraI) provides evidence for a processive monomeric molecular motor, *J Biol Chem 281*, 36110-36116.

18. Kuhn, B., Abdel-Monem, M., Krell, H., and Hoffmann-Berling, H. (1979) Evidence for two mechanisms for DNA unwinding catalyzed by DNA helicases, *J Biol Chem 254*, 11343-11350.

19. Lahue, E. E., and Matson, S. W. (1988) *Escherichia coli* DNA helicase I catalyzes a unidirectional and highly processive unwinding reaction, *J Biol Chem 263*, 3208-3215.

20. Stols, L., Gu, M., Dieckman, L., Raffen, R., Collart, F. R., and Donnelly, M. I. (2002) A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site, *Protein Expr Purif 25*, 8-15.

21. Weber, J., and Senior, A. E. (1997) Binding of TNP-ATP and TNP-ADP to the non-catalytic sites of *Escherichia coli* F1-ATPase, *FEBS Lett 412*, 169-172.

22. Kiianitsa, K., Solinger, J. A., and Heyer, W. D. (2003) NADH-coupled microplate photometric assay for kinetic studies of ATP-hydrolyzing enzymes with low and high specific activities, *Anal Biochem 321*, 266-271.

23. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria, *J Appl Crystallogr 25*, 495-503.

24. Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing, *Biophys J 76*, 2879-2886.

25. VolKov, V. S., DI. (2003) Uniqueness of ab initio shape determination in small-angle scattering, *J Appl Crystallogr 36*, 860-864.

26. Wriggers, W., and Chacón, P. (2001) Using Situs for the registration of protein structures with low-resolution bead models from X-ray solution scattering., *J Appl Crystallogr 34*, 773-776.

27. Kelley, L. A., and Sternberg, M. J. (2009) Protein structure prediction on the Web: a case study using the Phyre server, *Nat Protoc 4*, 363-371.

28. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *J Comput Chem 25*, 1605-1612.

29. Svergun, D. I., Barberato, C & Koch, MHJ (1995) CRYSOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr 28*, 733-768.

30. Anderson, B. J., Larkin, C., Guja, K., and Schildbach, J. F. (2008) Using fluorophore-labeled oligonucleotides to measure affinities of protein-DNA interactions, *Methods Enzymol 450*, 253-272.

31. Jezewska, M. J., and Bujalowski, W. (1996) A general method of analysis of ligand binding to competing macromolecules using the spectroscopic signal originating from a reference macromolecule. Application to *Escherichia coli* replicative helicase DnaB protein nucleic acid interactions, *Biochemistry 35*, 2117-2128.

32. Epstein, I. R. (1978) Cooperative and non-cooperative binding of large ligands to a finite one-dimensional lattice. A model for ligand-oligonucleotide interactions, *Biophys Chem 8*, 327-339.

33. Dostal, L., and Schildbach, J. F. (2010) Single-stranded DNA binding by F TraI relaxase and helicase domains is coordinately regulated, *J Bacteriol 192*, 3620-3628.

34. Bujalowski, W., and Jezewska, M. J. (1995) Interactions of *Escherichia coli* primary replicative helicase DnaB protein with single-stranded DNA. The nucleic acid does not wrap around the protein hexamer, *Biochemistry 34*, 8513-8519.

35. Jezewska, M. J., Galletto, R., and Bujalowski, W. (2004) Interactions of the RepA helicase hexamer of plasmid RSF1010 with the ssDNA. Quantitative analysis of stoichiometries, intrinsic affinities, cooperativities, and heterogeneity of the total ssDNA-binding site, *J Mol Biol 343*, 115-136.

36. Jezewska, M. J., Rajendran, S., and Bujalowski, W. (2000) *Escherichia coli* replicative helicase PriA protein-single-stranded DNA complex. Stoichiometries, free energy of binding, and cooperativities, *J Biol Chem 275*, 27865-27873.

37. Bujalowski, W., Lohman, T. M., and Anderson, C. F. (1989) On the cooperative binding of large ligands to a one-dimensional homogeneous lattice: the generalized three-state lattice model, *Biopolymers 28*, 1637-1643.

38. McGhee, J. D., and von Hippel, P. H. (1974) Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice, *J Mol Biol 86*, 469-489.

39. Wang, J., and Julin, D. A. (2004) DNA helicase activity of the RecD protein from *Deinococcus radiodurans*, *J Biol Chem 279*, 52024-52032.

40. Maluf, N. K., Fischer, C. J., and Lohman, T. M. (2003) A Dimer of *Escherichia coli* UvrD is the active form of the helicase in vitro, *J Mol Biol 325*, 913-935.

41. Csitkovits, V. C., and Zechner, E. L. (2003) Extent of single-stranded DNA required for efficient TraI helicase activity in vitro, *J Biol Chem 278*, 48696-48703.

42. Bujalowski, W., and Lohman, T. M. (1987) Limited co-operativity in protein-nucleic acid interactions. A thermodynamic model for the interactions of *Escherichia coli* single strand binding protein with single-stranded nucleic acids in the "beaded", (SSB)65 mode, *J Mol Biol 195*, 897-907.

43. Kowalczykowski, S. C., Lonberg, N., Newport, J. W., and von Hippel, P. H. (1981) Interactions of bacteriophage T4-coded gene 32 protein with nucleic acids. I. Characterization of the binding interactions, *J Mol Biol 145*, 75-104.

44. Dou, S. X., Wang, P. Y., Xu, H. Q., and Xi, X. G. (2004) The DNA binding properties of the *Escherichia coli* RecQ helicase, *J Biol Chem 279*, 6354-6363.

45. Gourves, A. S., Tanguy Le Gac, N., Villani, G., Boehmer, P. E., and Johnson, N. P. (2000) Equilibrium binding of single-stranded DNA with herpes simplex virus type I-coded single-stranded DNA-binding protein, ICP8, *J Biol Chem 275*, 10864-10869.

46. Jezewska, M. J., Kim, U. S., and Bujalowski, W. (1996) Binding of *Escherichia coli* primary replicative helicase DnaB protein to single-stranded DNA. Long-range allosteric conformational changes within the protein hexamer, *Biochemistry 35*, 2129-2145.

47. Jezewska, M. J., and Bujalowski, W. (2000) Interactions of *Escherichia coli* replicative helicase PriA protein with single-stranded DNA, *Biochemistry 39*, 10454-10467.

48. Howard, M. T., Nelson, W. C., and Matson, S. W. (1995) Stepwise assembly of a relaxosome at the F plasmid origin of transfer, *J Biol Chem 270*, 28381-28386.

49. Matson, S. W., and Morton, B. S. (1991) *Escherichia coli* DNA helicase I catalyzes a site- and strand-specific nicking reaction at the F plasmid oriT, *J Biol Chem. 266*, 16232-16237.

50. Saikrishnan, K., Griffiths, S. P., Cook, N., Court, R., and Wigley, D. B. (2008) DNA binding to RecD: role of the 1B domain in SF1B helicase activity, *EMBO J 27*, 2222-2229.

51. Dostal, L., Shao, S., and Schildbach, J. F. (2010) Tracking F plasmid TraI relaxase processing reactions provides insight into F plasmid transfer, *Nucleic Acids Res*

52. Singleton, M. R., Dillingham, M. S., Gaudier, M., Kowalczykowski, S. C., and Wigley, D. B. (2004) Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks, *Nature 432*, 187-193.

53. Korangy, F., and Julin, D. A. (1993) Kinetics and processivity of ATP hydrolysis and DNA unwinding by the RecBC enzyme from Escherichia coli, *Biochemistry 32*, 4873-4880.

54. Myong, S., and Ha, T. (2010) Stepwise translocation of nucleic acid motors, *Curr Opin Struc Biol 20*, 121-127.

55. Wong, I., and Lohman, T. M. (1992) Allosteric effects of nucleotide cofactors on Escherichia coli Rep helicase-DNA binding, *Science 256*, 350-355.

56. Bedford, M. T., and Clarke, S. G. (2009) Protein arginine methylation in mammals: who, what, and why, *Mol Cell 33*, 1-13.

57. Stallcup, M. R. (2001) Role of protein methylation in chromatin remodeling and transcriptional regulation, *Oncogene 20*, 3014-3020.

58. Bedford, M. T. (2007) Arginine methylation at a glance, *J Cell Sci 120*, 4243-4246.

59. Lake, A. N., and Bedford, M. T. (2007) Protein methylation and DNA repair, *Mutat Res 618*, 91-101.

60. Smith, B. C., and Denu, J. M. (2009) Chemical mechanisms of histone lysine and arginine modifications, *Biochim Biophys Acta 1789*, 45-57.

61. Teyssier, C., Le Romancer, M., Sentis, S., Jalaguier, S., Corbo, L., and Cavailles, V. (2009) Protein arginine methylation in estrogen signaling and estrogen-related cancers, *Trends Endocrinol Metab*.

62. Aletta, J. M., and Hu, J. C. (2008) Protein arginine methylation in health and disease, *Biotechnol Annu Rev 14*, 203-224.

63. Boger, R. H., Cooke, J. P., and Vallance, P. (2005) ADMA: an emerging cardiovascular risk factor, *Vasc Med 10 Suppl 1*, S1-2.

64. Ueda, S., Yamagishi, S., Matsumoto, Y., Fukami, K., and Okuda, S. (2007) Asymmetric dimethylarginine (ADMA) is a novel emerging risk factor for cardiovascular disease and the development of renal injury in chronic kidney disease, *Clin Exp Nephrol 11*, 115-121.

65. Pullamsetti, S., Kiss, L., Ghofrani, H. A., Voswinckel, R., Haredza, P., Klepetko, W., Aigner, C., Fink, L., Muyal, J. P., Weissmann, N., Grimminger, F., Seeger, W., and Schermuly, R. T. (2005) Increased levels and reduced catabolism of asymmetric and symmetric dimethylarginines in pulmonary hypertension, *Faseb J 19*, 1175-1177.

66. Bedford, M. T., and Richard, S. (2005) Arginine methylation an emerging regulator of protein function, *Mol Cell 18*, 263-272.

67. Lin, C. H., Hsieh, M., Li, Y. C., Li, S. Y., Pearson, D. L., Pollard, K. M., and Li, C. (2000) Protein N-arginine methylation in subcellular fractions of lymphoblastoid cells, *J Biochem 128*, 493-498.

68. Wolf, S. S. (2009) The protein arginine methyltransferase family: an update about function, new perspectives and the physiological role in humans, *Cell Mol Life Sci 66*, 2109-2121.

69. Frankel, A., and Clarke, S. (2000) PRMT3 is a distinct member of the protein arginine N-methyltransferase family. Conferral of substrate specificity by a zinc-finger domain, *J Biol Chem 275*, 32974-32982.

70. Cheng, X., Collins, R. E., and Zhang, X. (2005) Structural and sequence motifs of protein (histone) methylation enzymes, *Annu Rev Biophys Biomol Struct 34*, 267-294.

71. Zhang, X., Zhou, L., and Cheng, X. (2000) Crystal structure of the conserved core of protein arginine methyltransferase PRMT3, *Embo J 19*, 3509-3519.

72. Zhang, X., and Cheng, X. (2003) Structure of the predominant protein arginine methyltransferase PRMT1 and analysis of its binding to substrate peptides, *Structure 11*, 509-520.

73. Kim, S., Merrill, B. M., Rajpurohit, R., Kumar, A., Stone, K. L., Papov, V. V., Schneiders, J. M., Szer, W., Wilson, S. H., Paik, W. K., and Williams, K. R. (1997) Identification of N(G)-methylarginine residues in human heterogeneous RNP protein A1: Phe/Gly-Gly-Gly-Arg-Gly-Gly-Gly/Phe is a preferred recognition motif, *Biochemistry 36*, 5185-5192.

74. Wooderchak, W. L., Zang, T., Zhou, Z. S., Acuna, M., Tahara, S. M., and Hevel, J. M. (2008) Substrate profiling of PRMT1 reveals amino acid sequences that extend beyond the "RGG" paradigm, *Biochemistry 47*, 9456-9466.

75. Gary, J. D., and Clarke, S. (1998) RNA and protein interactions modulated by protein arginine methylation, *Prog Nucleic Acid Res Mol Biol 61*, 65-131.

76. Osborne, T. C., Obianyo, O., Zhang, X., Cheng, X., and Thompson, P. R. (2007) Protein arginine methyltransferase 1: positively charged residues in substrate peptides distal to the site of methylation are important for substrate binding and catalysis, *Biochemistry 46*, 13370-13381.

77. Lin, W. J., Gary, J. D., Yang, M. C., Clarke, S., and Herschman, H. R. (1996) The mammalian immediate-early TIS21 protein and the leukemia-associated BTG1 protein interact with a protein-arginine N-methyltransferase, *J Biol Chem 271*, 15034-15044.

78. Singh, V., Miranda, T. B., Jiang, W., Frankel, A., Roemer, M. E., Robb, V. A., Gutmann, D. H., Herschman, H. R., Clarke, S., and Newsham, I. F. (2004) DAL-1/4.1B tumor suppressor interacts with protein arginine N-methyltransferase 3 (PRMT3) and inhibits its ability to methylate substrates in vitro and in vivo, *Oncogene 23*, 7761-7771.

79. Pal, S., Vishwanath, S. N., Erdjument-Bromage, H., Tempst, P., and Sif, S. (2004) Human SWI/SNF-associated PRMT5 methylates histone H3 arginine 8 and negatively regulates expression of ST7 and NM23 tumor suppressor genes, *Mol Cell Biol 24*, 9630-9645.

80. Xu, W., Cho, H., Kadam, S., Banayo, E. M., Anderson, S., Yates, J. R., 3rd, Emerson, B. M., and Evans, R. M. (2004) A methylation-mediator complex in hormone signaling, *Genes Dev 18*, 144-156.

81. Niu, L., Lu, F., Pei, Y., Liu, C., and Cao, X. (2007) Regulation of flowering time by the protein arginine methyltransferase AtPRMT10, *EMBO Rep 8*, 1190-1195.

82. Otwinowski, Z., and Minor, W. (1997) *"Processing of X-ray Diffraction Data Collected in Oscillation Mode", Methods in Enzymology, Volume 276: Macromolecular Crystallography, part A, p. 307-326, C.W.Carter, Jr & R.M.Sweet, Eds.* , Academic Press (New York).

83. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr D Biol Crystallogr 66*, 213-221.

84. (1994) The CCP4 suite: programs for protein crystallography, *Acta Crystallogr D Biol Crystallogr 50*, 760-763.

85. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. Features and development of Coot, *Acta Crystallogr D Biol Crystallogr 66*, 486-501.

86. Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J., and Kollman, P. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, *J Comput Chem 24*, 1999-2012.

87. Teotico, D. G., Frazier, M. L., Ding, F., Dokholyan, N. V., Temple, B. R., and Redinbo, M. R. (2008) Active nuclear receptors exhibit highly correlated AF-2 domain motions, *PLoS Comput Biol 4*, e1000111.

88. Case, D. A., Darden, T. A., Cheatham, T. E. I., Simmerling, C. L., and Wang, J. (2006) *AMBER 9*, University of California: San Francisco.

89. Essman, U., Perera, L., Berkowita, M. L., Darden, T. A., and Lee, H. (1995) A smooth particle mesh Ewald method, *J Chem Phys 103*, 8577-8593.

90. Sharma, S., Ding, F., and Dokholyan, N. V. (2007) Multiscale modeling of nucleosome dynamics, *Biophys J 92*, 1457-1470.

91. Everitt, B. S., Landau, S., and Leese, M. (2001) *Cluster analysis*, Oxford University Press: Oxford.

92. Fauman, E. B., Blumenthal, R. M., and Cheng, X. (1999) Structure and evolution of Adomet-dependent methyltransferase. In S-Adenosylmethionine-Depedent Methyltransferase: Structures and Functions, X.Cheng and R.M. Blumenthal, eds, pp 1-38, River Edge, NJ: World Scientific.

93. Cheng, X., and Roberts, R. J. (2001) AdoMet-dependent methylation, DNA methyltransferases and base flipping, *Nucleic Acids Res 29*, 3784-3795.

94. Lee, D. Y., Ianculescu, I., Purcell, D., Zhang, X., Cheng, X., and Stallcup, M. R. (2007) Surface-scanning mutational analysis of protein arginine methyltransferase 1: roles of specific amino acids in methyltransferase substrate specificity, oligomerization, and coactivator function, *Mol Endocrinol 21*, 1381-1393.

95. Smith, J. J., Rucknagel, K. P., Schierhorn, A., Tang, J., Nemeth, A., Linder, M., Herschman, H. R., and Wahle, E. (1999) Unusual sites of arginine methylation in Poly(A)-binding protein II and in vitro methylation by protein arginine methyltransferases PRMT1 and PRMT3, *J Biol Chem 274*, 13229-13234.

96. Pawlak, M. R., Banik-Maiti, S., Pietenpol, J. A., and Ruley, H. E. (2002) Protein arginine methyltransferase I: substrate specificity and role in hnRNP assembly, *J Cell Biochem 87*, 394-407.

97. Jochmans, D. (2008) Novel HIV-1 reverse transcriptase inhibitors, *Virus Res 134*, 171-185.

98. Ren, J., and Stammers, D. K. (2005) HIV reverse transcriptase structures: designing new inhibitors and understanding mechanisms of drug resistance, *Trends Pharmacol Sci 26*, 4-7.

99. Dieterich, D. T. (2003) Long-term complications of nucleoside reverse transcriptase inhibitor therapy, *AIDS Read 13*, 176-184, 187.

100. Nolan, D., and Mallal, S. (2004) Complications associated with NRTI therapy: update on clinical features and possible pathogenic mechanisms, *Antivir Ther 9*, 849-863.

101. Shafer, R. W., Rhee, S. Y., Pillay, D., Miller, V., Sandstrom, P., Schapiro, J. M., Kuritzkes, D. R., and Bennett, D. (2007) HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance, *Aids 21*, 215-223.

102. Geretti, A. M. (2006) Clinical implications of HIV drug resistance to nucleoside and nucleotide reverse transcriptase inhibitors, *AIDS Rev 8*, 210-220.

103. Bartlett, J. A., Fath, M. J., Demasi, R., Hermes, A., Quinn, J., Mondou, E., and Rousseau, F. (2006) An updated systematic overview of triple combination therapy in antiretroviral-naive HIV-infected adults, *Aids 20*, 2051-2064.

104. Pal, D., and Mitra, A. K. (2006) MDR- and CYP3A4-mediated drug-drug interactions, *J Neuroimmune Pharmacol 1*, 323-339.

105. Kliewer, S. A., Goodwin, B., and Willson, T. M. (2002) The nuclear pregnane X receptor: a key regulator of xenobiotic metabolism, *Endocr Rev 23*, 687-702.

106. Carnahan, V. E., and Redinbo, M. R. (2005) Structure and function of the human nuclear xenobiotic receptor PXR, *Curr Drug Metab 6*, 357-367.

107. Teotico, D. G., Bischof, J. J., Peng, L., Kliewer, S. A., and Redinbo, M. R. (2008) Structural basis of human pregnane X receptor activation by the hops constituent colupulone, *Mol Pharmacol 74*, 1512-1520.

108. Xue, Y., Moore, L. B., Orans, J., Peng, L., Bencharit, S., Kliewer, S. A., and Redinbo, M. R. (2007) Crystal structure of the pregnane X receptor-estradiol complex provides insights into endobiotic recognition, *Mol Endocrinol 21*, 1028-1038.

109. Xue, Y., Chao, E., Zuercher, W. J., Willson, T. M., Collins, J. L., and Redinbo, M. R. (2007) Crystal structure of the PXR-T1317 complex provides a scaffold to examine the potential for receptor antagonism, *Bioorg Med Chem 15*, 2156-2166.

110. 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., and Redinbo, M. R. (2003) 2.1 A crystal structure of human PXR in complex with the St. John's wort compound hyperforin, *Biochemistry 42*, 1430-1438.

111. Watkins, R. E., Wisely, G. B., Moore, L. B., Collins, J. L., Lambert, M. H., Williams, S. P., Willson, T. M., Kliewer, S. A., and Redinbo, M. R. (2001) The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity, *Science 292*, 2329-2333.

112. Chrencik, J. E., Orans, J., Moore, L. B., Xue, Y., Peng, L., Collins, J. L., Wisely, G. B., Lambert, M. H., Kliewer, S. A., and Redinbo, M. R. (2005) Structural disorder in

the complex of human pregnane X receptor and the macrolide antibiotic rifampicin, *Mol Endocrinol 19*, 1125-1134.

113. Watkins, R. E., Davis-Searles, P. R., Lambert, M. H., and Redinbo, M. R. (2003) Coactivator binding promotes the specific interaction between ligand and the pregnane X receptor, *J Mol Biol 331*, 815-828.

114. Campiani, G., Ramunno, A., Maga, G., Nacci, V., Fattorusso, C., Catalanotti, B., Morelli, E., and Novellino, E. (2002) Non-nucleoside HIV-1 reverse transcriptase (RT) inhibitors: past, present, and future perspectives, *Curr Pharm Des 8*, 615-657.

115. Otwinowski, Z., and Minor, W. (1997) *Processing of X-ray Diffraction Data Collected in Oscillation Mode*, Vol. 276, Academic Press, New York.

116. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr D Biol Crystallogr 53*, 240-255.

117. 4, C. C. P. N. (1994) The CCP4 suite: programs for protein crystallography, in *Acta Crystallogr D Biol Crystallogr*, pp 760-763.

118. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr D Biol Crystallogr 60*, 2126-2132.

119. Winn, M. D., Murshudov, G. N., and Papiz, M. Z. (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions, *Methods Enzymol 374*, 300-321.

120. 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., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr D Biol Crystallogr 54*, 905-921.

121. Delano, W. L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific LLC, San Carlos, CA, USA.

122. Ekins, S., Kortagere, S., Iyer, M., Reschly, E. J., Lill, M. A., Redinbo, M. R., and Krasowski, M. D. (2009) Challenges predicting ligand-receptor interactions of promiscuous proteins: the nuclear receptor PXR, *PLoS Comput Biol 5*, e1000594.

123. Swanson, J. (1973) Studies on gonococcus infection. IV. Pili: their role in attachment of gonococci to tissue culture cells, *J Exp Med 137*, 571-589.

124. McGee, Z. A., Stephens, D. S., Hoffman, L. H., Schlech, W. F., 3rd, and Horn, R. G. (1983) Mechanisms of mucosal invasion by pathogenic Neisseria, *Rev Infect Dis 5 Suppl 4*, S708-714.

125. Craig, L., Pique, M. E., and Tainer, J. A. (2004) Type IV pilus structure and bacterial pathogenicity, *Nat Rev Microbiol 2*, 363-378.

126. Wall, D., and Kaiser, D. (1999) Type IV pili and cell motility, *Mol Microbiol 32*, 1-10.

127. Strom, M. S., Nunn, D. N., and Lory, S. (1994) Posttranslational processing of type IV prepilin and homologs by PilD of Pseudomonas aeruginosa, *Methods Enzymol 235*, 527-540.

128. Mattick, J. S. (2002) Type IV pili and twitching motility, *Annu Rev Microbiol 56*, 289-314.

129. Alm, R. A., and Mattick, J. S. (1997) Genes involved in the biogenesis and function of type-4 fimbriae in Pseudomonas aeruginosa, *Gene 192*, 89-98.

130. Rudel, T., Scheurerpflug, I., and Meyer, T. F. (1995) *Neisseria* PilC protein identified as type-4 pilus tip-located adhesin, *Nature 373*, 357-359.

131. Rahman, M., Kallstrom, H., Normark, S., and Jonsson, A. B. (1997) PilC of pathogenic *Neisseria* is associated with the bacterial cell surface, *Mol Microbiol 25*, 11-25.

132. Morand, P. C., Tattevin, P., Eugene, E., Beretti, J. L., and Nassif, X. (2001) The adhesive property of the type IV pilus-associated component PilC1 of pathogenic *Neisseria* is supported by the conformational structure of the N-terminal part of the molecule, *Mol Microbiol 40*, 846-856.

133. Scheuerpflug, I., Rudel, T., Ryll, R., Pandit, J., and Meyer, T. F. (1999) Roles of PilC and PilE proteins in pilus-mediated adherence of *Neisseria gonorrhoeae* and *Neisseria meningitidis* to human erythrocytes and endothelial and epithelial cells, *Infect Immun 67*, 834-843.

134. Morand, P. C., Bille, E., Morelle, S., Eugene, E., Beretti, J. L., Wolfgang, M., Meyer, T. F., Koomey, M., and Nassif, X. (2004) Type IV pilus retraction in pathogenic *Neisseria* is regulated by the PilC proteins, *Embo J 23*, 2009-2017.

135. Kallstrom, H., Liszewski, M. K., Atkinson, J. P., and Jonsson, A. B. (1997) Membrane cofactor protein (MCP or CD46) is a cellular pilus receptor for pathogenic *Neisseria*, *Mol Microbiol 25*, 639-647.

136. Kallstrom, H., Blackmer Gill, D., Albiger, B., Liszewski, M. K., Atkinson, J. P., and Jonsson, A. B. (2001) Attachment of *Neisseria gonorrhoeae* to the cellular pilus receptor CD46: identification of domains important for bacterial adherence, *Cell Microbiol 3*, 133-143.

137. Kirchner, M., Heuer, D., and Meyer, T. F. (2005) CD46-independent binding of neisserial type IV pili and the major pilus adhesin, PilC, to human epithelial cells, *Infect Immun 73*, 3072-3082.

138. Kellogg, D. S., Jr., Cohen, I. R., Norins, L. C., Schroeter, A. L., and Reising, G. (1968) *Neisseria gonorrhoeae*. II. Colonial variation and pathogenicity during 35 months in vitro, *J Bacteriol 96*, 596-605.

139. Orans, J., Johnson, M. D., Coggan, K. A., Sperlazza, J. R., Heiniger, R. W., Wolfgang, M. C., and Redinbo, M. R. Crystal structure analysis reveals *Pseudomonas* PilY1 as an essential calcium-dependent regulator of bacterial surface motility, *Proc Natl Acad Sci U S A 107*, 1065-1070.