Structural Characterization of the

Tn7 Target Selection Protein TnsE

Structural Characterization of the Tn7 Target Selection Protein TnsE

By Jeremy J. Caron, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Master of Science

MASTER OF SCIENCE (2016) McMaster University

(Biochemistry and Biomedical Sciences) Hamilton, Ontario

TITLE: Structural Characterization of the Tn7 Target Selection Protein TnsE

AUTHOR: Jeremy J. Caron, B.Sc. (University of British Columbia)

SUPERVISOR: Dr. Alba Guarné

NUMBER OF PAGES: 75

Abstract

Tn7 and Tn7-like transposons are complex elements found in disparate environments and are responsible for mobilizing a wide variety of genes and forming pathogenicity/fitness islands. They are novel in their ability to recognize both a single site in the chromosome and specifically target transposition into mobile plasmids via dedicated TnsD and TnsE targeting proteins. TnsE recognizes mobile plasmids through an association with the processivity clamp and a 3′ recessed DNA end during conjugal replication. However, the mechanism for the specific recognition of 3′ recessed DNA ends remains unclear. Structural analyses of the C-terminal domain of TnsE identified a novel protein fold including a central V-shaped loop that toggles between two distinct conformations. The structure of a robust TnsE gain-of-function variant has this loop locked in a single conformation, suggesting that conformational flexibility regulates TnsE activity. Structure-based analysis of a series of TnsE variants relates transposition to DNA binding stability. Follow up studies of full length TnsE bound to DNA are in progress.

Acknowledgements

Firstly, I would like to thank all my friends and family for their support over the course of my studies and for helping me grow into a person able to undertake this work. In particular, I would mention my mother, Cynthia Falksoki, who was always there to help me see the brighter side of hard situations.

All members of the Guarné Lab present during my studies made a great impact on my growth and development in this academic endeavor. Dr. Yu Seon Chung initiated this project before I began my studies and provided a unique example of making your work fit your character. I thank Tamiza Nanji, Dr. Monica Pillon, and Dr. Lindsay Matthews for being ideal role models in their passion for science in the beginning of my studies. I thank Ahmad Almawi and Meg Rashev for numerous discussions about our projects and enduring the trials of graduate school alongside me. Huasheng Wang in particular deserves acknowledgement for the time and effort he dedicated to this project during his term as an undergraduate student in the Guarné Lab. He provided an attentive ear as I learned about teaching and his work led to a structure that provided significant insight into the mechanism we sought to characterize.

I thank my committee members, Dr. Murray Junop and Dr. Lori Burrows, for their invaluable advice during our meetings. Most of all, I would like to thank my supervisor Dr. Alba Guarné, who showed me the value of a scrupulous approach to my work. She provided immeasurable guidance and insight into my research, and I am certain I could not have achieved as much under anyone else. I know the skills and values I learned from her will benefit me throughout the rest of my career.

Table of Contents

Chapter 3: Structural Characterization of TnsE

Chapter 4: Characterization of the TnsE+DNA complex

Chapter 5: Conclusion

List of Figures and Tables

List of abbreviations and symbols

bp	Base pairs
C	Celsius
CTD	C-terminal domain
º	Degree
DLS	Dynamic light scattering
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
DSF	Differential scanning fluorimetry
dsDNA	Double-stranded DNA
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assays
IPTG	Isopropyl-β-D-thiogalactopyranoside
IR	Inverted repeat
NTD	N-terminal domain
PIC	Pre-integration complex
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
ssDNA	Single-stranded DNA
w/v	Weight/volume

Declaration of Academic Achievement

I subcloned the C-terminal domains of wild type TnsE (TnsE-CTD) and the TnsE-A453V+D523N variant. I developed protein production and purification protocols for these two variants of TnsE-CTD. I crystallized and solved the structures of these two variants with assistance from an undergraduate student (Huasheng Wang) who helped me purify and set crystallization drops of the TnsEA453V+D523N-CTD variant. I collected data at the Canadian Light Source and solved and refined the two structures with the assistance of my supervisor. This work was included in a manuscript published last year (Conformational toggling controls target site choice for the heteromeric transposase element Tn7. Qiaojuan Shi*, Marco R. Straus*, **Jeremy J. Caron**, Huasheng Wang, Yu Seon Chung, Alba Guarné and Joseph E. Peters. *Nucleic Acids Research* 43(22):10734-45). I also worked devised a purification protocol for the N-terminal domain of TnsE and grew preliminary crystals. However, these crystals are not reproducible and are not of enough quality to determined the structure of this domain. Lastly, I used differential scanning fluorimetry to map the minimal DNA recessed substrates recognized by TnsE and completed preliminary crystallization screens of TnsE bound to DNA.

Chapter 1

Introduction

1.1 Transposable elements in bacteria

Transposons are mobile nucleic acid sequences that encode the protein machinery required to move themselves from one location to another within the genomes of host organisms without requiring homology with the target site. This transposition machinery generally performs functions like target site recognition, DNA structure manipulation, strand transfer catalysis, and self-regulation. While not all these functions are ubiquitous, strand transfer catalysis is absolutely required and is carried out by an enzyme usually called the transposase. The other functions are commonly also performed by the transposase, but may be carried out by additional proteins. Transposons are extremely prevalent in nature, where they and their inactive derivatives can account for more than 40% of eukaryote genomes and 1- 5% of prokaryote genomes, in both cases contributing to natural mutagenesis of their hosts and to the dissemination of genes (Curcio and Derbyshire, 2003). In bacteria, transposons and the related integration cassettes that move with plasmids during bacterial conjugation are responsible for the majority of horizontal transfer events that spread antibiotic resistance genes (Perry and Wright, 2013).

Transposons hosted by all kingdoms may be broadly classified into one of two groups: retrotransposons and DNA transposons. Retrotransposons are the most prevalent form of active transposons in higher eukaryotes and mobilize via an RNA intermediate, which

provides a template for a self-encoded reverse transcriptase in producing the cDNA that is inserted into a target site (Ewing et al., 2013). DNA transposons propagate without an RNA intermediate and are the most prevalent form in prokaryotes. They are mobilized through one of two mechanisms (Snyder et al., 2013). The first mechanism is replicative transposition, in which only one strand is nicked at either end of the element. Nucleophilic attack by the resulting 3′ hydroxyl group on the integration site covalently links the donor and target DNA molecules. Replicative extension of the strands duplicates the transposon and recombination within the element separates the donor and target DNA molecules. This results in two copies of the transposon, one remaining in the donor molecule and the other in the target site. The second mechanism is non-replicative and is referred to as cut and paste transposition. The transposon is excised from the donor DNA molecule by cleavage of both strands at the both ends, resulting in a double strand break in the donor DNA. Then, insertion by nucleophilic attack of the 3′ hydroxyl groups into the target site is mediated by the transposase+DNA complex, or transposasome. Unlike replicative transposition, this mechanism does not preserve a copy of the element in the donor DNA molecule.

The mobilization of many DNA transposons depends upon the assembly of a nucleoprotein complex in which inverted repeat (IR) sequences at both ends of the transposon are bound by a transposase protein, forming a paired end pre-integration complex (PIC). Subsequently, a target DNA molecule containing a suitable recognition sequence or structure is positioned for integration. The PIC bound to a target DNA as well as any other system-specific factors makes up the transposasome, which becomes transposition competent at various points for each type of transposon. A feature common to most transposasomes that assemble in this manner is a transposase subunit containing a DDE motif (Curcio and Derbyshire, 2003). This nomenclature refers to three residues in the motif that coordinate two

catalytic divalent metal cations (Yang et al., 2006). DDE transposase activity is common to many of the well-characterized DNA transposons, including that of bacteriophage Mu, the prokaryotic transposon families Tn3, Tn5/Tn10 (Snyder et al., 2013), Tn7 (Li et al., 2013), P Element of *Drosophila melanogaster* (Levin and Moran, 2011), and eukaryotic *mariner* transposons (Richardson et al., 2009). DDE transposases are involved in both replicative and non-replicative transposition mechanisms.

There is a fitness burden on transposons to self-regulate their activity (Nagy and Chandler, 2004). If transposition occurs too frequently or indiscriminately, insertion may occur at a site that is deleterious to its host. Transposition that occurs too infrequently presents a larger window of opportunity for mutations to accumulate and inactivate the transposon. So, successful transposons are highly self-regulated.

1.2 The transposon Tn7

The bacterial transposon Tn7 has been identified in a broad range of hosts adapted to a variety of environments (Parks and Peters, 2007). Originally identified in *E. coli* in both the chromosome and plasmids as a mobile element that carries trimethoprim and streptomycin resistance genes (Barth et al., 1976). Tn7-like elements carrying host fitness factors have since been found in a broad range of prokaryotes in diverse environments (Parks and Peters, 2009). Furthermore, Tn7-like elements have recently been identified in a variety of clinical pathogens, in which the transposition protein sequences are identical or nearly identical to the *E. coli* element (Shi et al., 2015).

The high fitness of Tn7 is seen both in the complexity of its regulation, and in the diversity of its host bacteria. Most bacterial transposons identify a target site in the genome and catalyze the strand transfer reaction using a single transposase protein. Tn7, however,

encodes five transposition proteins, each with a distinct function in the process (Li et al., 2013). These five transposition proteins, TnsA, B, C, D and E, allow for transposition through either of two pathways that target insertion to different types of sites. Each pathway is directed by either of Tn7's targeting proteins, TnsD or TnsE, which bind their respective target sites and recruit the transposition complex composed of TnsA, TnsB and TnsC. In the TnsABC+D pathway, TnsD recognizes a highly conserved sequence called *attTn7* at the 3′ end of the glucosamine-6-phosphate synthase (*glmS*) gene. TnsE targeting is less well understood, but has been shown to direct insertion to the lagging strand following replication forks in an orientation determined by the direction of DNA synthesis (Peters and Craig, 2001). This pathway displays a targeting bias for insertion into DNA structures that facilitate horizontal transfer, such as conjugal plasmids and bacteriophage genomes (Finn et al., 2007; Wolkow et al., 1996). Furthermore, unlike many other transposons, strand cleavage in both the donor and target DNA is not catalyzed until the complete transpososome has assembled, preventing the formation of double strand breaks in uncommitted reactions (Holder and Craig, 2010). The relative success of Tn7 compared to other prokaryotic transposons is attributed to strategic regulation of the TnsD- and TnsE-directed pathways, which provides the element mechanisms to alternately seek a safe haven in the host chromosome or seek a DNA structure that will provide dissemination.

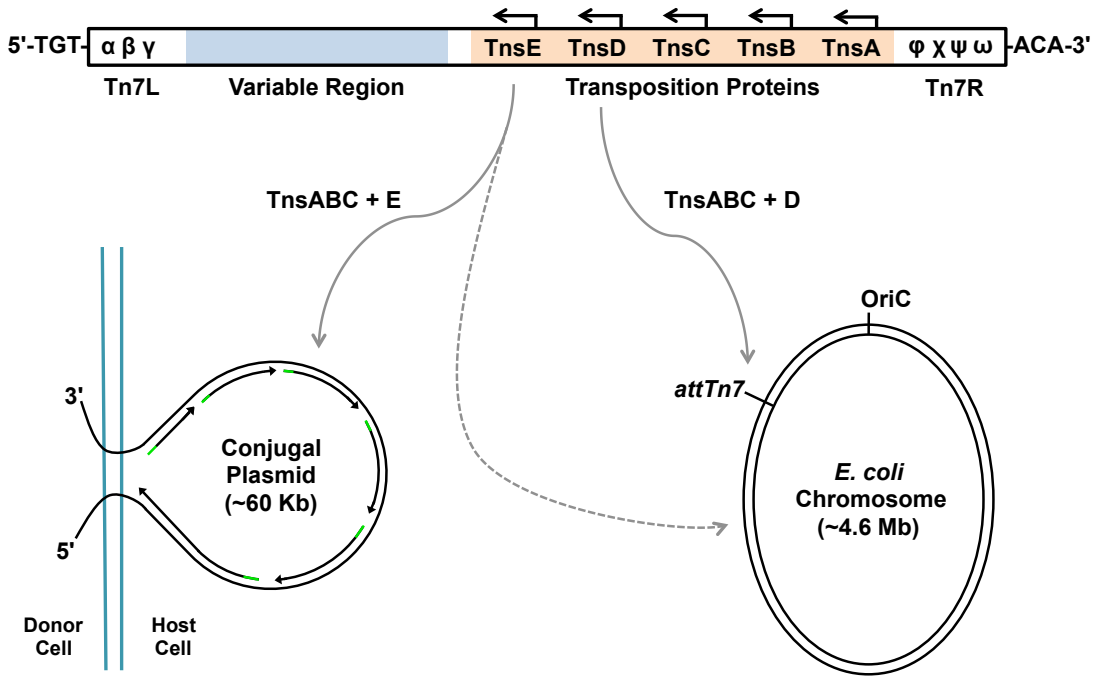

Figure 1.1: Schematic representation of Tn7's two targeting pathways. The diagram of Tn7's linear structure shows the relative arrangement of inverted repeats, transposition protein genes, variable region, and TnsB binding sites (indicated by greek symbols). Solid arrows indicate the preferred sites targeted by the TnsABC+D and TnsABC+E transposition pathways, and the dashed arrow indicates the less preferred target of the TnsABC+E pathway.

1.3 TnsABC: the core transposition machinery of Tn7

Tn7 and Tn7-like mobile elements are the only known transposons in bacteria that employ a transposase consisting of two separate proteins (Peters et al., 2014). In other mobile elements, a single polypeptide carries out all functions involved in catalysis of the strand transfer reaction. The Tn7 proteins TnsA and TnsB form the heteromeric transposase, in which each has a distinctly different role in Tn7's cut and paste mechanism. TnsA cleaves Tn7 at both 5′ ends, but in a fashion that depends on prior assembly of the transposition complex since it does not itself bind DNA (Bainton et al., 1993). The partial structure of TnsA resembles FokI endonuclease suggesting it acts through an endonuclease mechanism of DNA hydrolysis (Hickman et al., 2000). To complement this activity, TnsB cleaves at the 3′ ends, but first

binds recognition sequences in the Tn7 ends to form a pre-transposition paired-end complex (Skelding et al., 2002). TnsB cannot catalyze 3' end-joining to the target DNA in the absence of TnsA, however the presence of catalytically inactive TnsA results in a replicative transposition mechanism (Sarnovsky et al., 1996).

The core Tn7 tranposon machinery also includes TnsC, a regulator protein that passes on the target recognition signal from TnsD or TnsE to the TnsAB transposase (Stellwagen and Craig, 1998). While TnsA and TnsB function interdependently, TnsC regulates the activity of TnsAB in response to signals from other transposition proteins. If Tn7 is already in a potential target, TnsB binds the transposon ends and provides TnsC with a down-regulating target immunity signal (Stellwagen and Craig, 1998).

1.4 TnsD directs transposition into *attTn7*

TnsD and TnsE function in separate pathways for selecting Tn7 integration target sites (Bainton et al., 1993). TnsD directs integration to a specific, non-mutagenic site in the bacterial chromosome, whereas TnsE, preferentially targets sites on the lagging strand during replication of a conjugal plasmid (Wolkow et al., 1996). The alternate TnsD- and TnsE-mediated pathways allow Tn7 to persist and disseminate throughout a bacterial population without negatively influencing host survival (Li et al., 2013).

TnsD is a sequence-specific DNA binding protein that recognizes a single site in the bacterial chromosome and activates high-frequency insertion. The chromosome provides a safe haven for Tn7, thereby ensuring its vertical proliferation. Conversely, TnsE is a structure-specific DNA binding protein that recognizes features found on the lagging strand

during replication and strongly prefers conjugal plasmids undergoing lagging strand replication as they enter a cell, in turn ensuring horizontal spread of Tn7.

While both TnsD and TnsE have different targets, there are some similarities between them. In both, a 5-nucleotide duplication of host DNA occurs at either end of the integrated transposon, suggesting they use the same mechanism for the transposition reaction (Peters and Craig, 2001). Both pathways also introduce an orientation bias to the integrated transposon, allowing only one orientation of Tn7's left and right ends with respect to the target site. The TnsD-mediated pathway has been very well characterized. TnsD recognizes a highly conserved 33-nucleotide DNA sequence within the open reading frame of the housekeeping gene encoding glucosamine-fructose-6-phosphate aminotransferase, *glmS* (Li et al., 2013). TnsD's recognition sequence spans the termination codon and codons for the 10 C-terminal amino acids of GlmS. Insertion through this pathway occurs at high frequency, but is non-deleterious to the host since the strand transfer reaction is directed 22 bp downstream of the *glmS* termination codon (Mitra et al., 2010). The site encompassing the TnsD recognition sequence and the point of insertion is collectively referred to as the attachment site *attTn7*. Repeated transposition into this region is also not a problem since Tn7 contains a target immunity mechanism, which prevents insertion into target DNA already containing a copy of Tn7 (Arciszewska et al., 1989). Upon binding TnsD, the 5′ end of *attTn7* becomes distorted (Kuduvalli et al., 2001), which together with the surface of TnsD acts as an asymmetrical binding site for TnsC (Choi et al., 2014). This asymmetry is believed to impose the orientation bias observed in Tn7 integrated through TnsD. In this pathway, an interaction between TnsB and TnsC is essential for transposition, and is postulated to link the donor and target nucleoprotein complexes(Choi et al., 2014).

1.5 TnsE directs transposition into replicating DNA

The TnsE-mediated pathway is less understood, however it is of particular interest since integration into a conjugal plasmid is a determinant step in the horizontal transfer of Tn7. During cell-to-cell transfer of conjugal plasmids, one strand at the site *oriT* is nicked and then passed 5′-first into the recipient cell, undergoing lagging-strand replication in the new host (Cabezón et al., 2014). A suite of genes encoding conjugation machinery, called the *tra* operon also often occurs on conjugal plasmids. TnsE does not recognize any specific sequence or position on a transferred plasmid, but rather directs integration into any plasmid containing *oriT* if the *tra* operon is provided in trans (Wolkow et al., 1996). Therefore, it is the process of conjugation that is targeted by Tn7, not a primary feature of the plasmid.

To understand what TnsE recognizes in a conjugal plasmid, we first need to know what TnsE recognizes in any target site. TnsE-directed integration is critically dependent on TnsE binding features that are found on the lagging strand following replication forks: DNA structures with a 3′-recessed end (Peters and Craig, 2001) and the sliding clamp (Parks et al., 2009). A motif at the N-terminus is essential for the interaction with the sliding clamp and this interaction occurs in a species-dependent manner in the Tn7-like transposons. TnsE's activity is inhibited by the origin sequestration protein SeqA, which binds hemimethylated GATC sites following replication forks. The finding that TnsE interacts with SeqA led to a model in which one molecule of TnsE removes SeqA from the target, while another carries out target recognition and signaling (Li Z., and Peters J.E., unpublished). My thesis work aims to define features of the TnsE-DNA interaction that produce a 'target recognition' signal leading to the transposition reaction.

1.6 Distribution of gain-of-activity mutations in TnsE

In 2001, Drs. Nancy Craig and Joseph Peters published a body of work in which they identified TnsE variants that stimulate increased transposition frequency in vivo and increase its affinity for DNA target structures in vitro (Peters and Craig, 2001). This was done by chemically mutagenizing TnsE-encoding plasmids and subsequently transforming them into a cellular context that selects for transposition activity. Then, changes in transposition frequency were quantified in vivo using the 'lambda hop' or 'mating out' assays, which provide relative measures of frequency (McKown et al., 1988; Waddell and Craig, 1988). The mutations in the TnsE variants that allowed the highest levels of transposition occurred in pairs and were M37I+D523N, E516K+D523N, E522K+D523N, A453V+G483R, A453V+G515R and A453V+D523N.

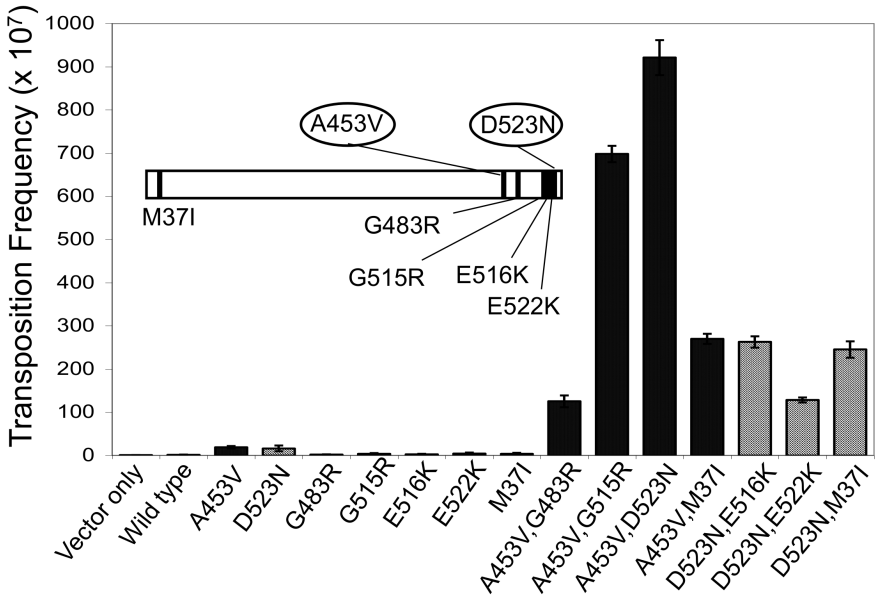

Figure 1.2: *In vivo* transposition activity of TnsE and TnsE variants. Transposition frequency was monitored in cells expressing TnsABC (pCW15) and mutant or wild-type TnsE, (pJP103) (n = 3). Error bars indicate the standard error of the mean (*extracted from Shi et al. (2015) NAR*).

Curiously, these variants did not influence the location of insertion sites in the host chromosome compared to wild type. Wild type TnsE was previously shown to direct transposon insertion into the lagging strand following replication forks in a single orientation determined by the direction of replication (Peters and Craig, 2000). The Peters group recently measured the transposition frequency conveyed by single amino acid variants of TnsE (Shi et al., 2015). As shown in figure 1.2, the individual variants A453V and D523N produced the greatest increases in transposition frequency, however the increase produced by the other mutations paired with one of A453V or D523N was far more than additive.

From a structural biology perspective, these results were interesting because all mutations occurred in the C-terminal region of TnsE, thereby suggesting that this region bears the primary function of regulating transposition frequency through the TnsABC+E pathway.

1.7 Thesis Objectives

Tn7 transposition into replicating DNA through the TnsE pathway is not well understood. To understand how TnsE preferably recognizes conjugal plasmids, we first need to understand how TnsE recognizes DNA and define the structural features that modulate transposition frequency. My thesis work has been aimed at: 1) determining the structure of TnsE, with the final goal of understanding how certain TnsE variants result in high transposition frequency; and 2) defining the minimal DNA substrate recognized by TnsE.

Chapter 2

Materials and Methods

2.1 Subcloning TnsE variants

PCR amplification of the plasmid encoding the wild type TnsE gene (pAG8623) was used to obtain two gene fragments corresponding to putative domains that were separated at a site estimated to mimic the major trypsin cleavage site. The N-terminal coding sequence was amplified with primers AG1881 and AG1882, while the C-terminal sequence was amplified with primers AG1883 and AG1884 (Table 2.1). Restriction sites at the 5′ end of each primer were included to allow for recombination of the amplicon with a linearized pET-22b expression vector. By incorporating a XhoI site at the 3′ end of the coding strand, a hexahistidine (His_6) tag followed by a termination codon was placed in frame immediately downstream of the insert. Successful recombination was verified by DNA sequencing (MOBIX Laboratories, McMaster University).

Table 2.1: Primers used for subcloning TnsE domains.

Name	Sequence
AG1881	5′CCTCTAGAGCTTAAACAGGAAGATGTTG3′
AG1882	5′AACTCGAGAGTTTGGCTACTTTTTCTTGC3′
AG1883	5′AACATATGTTAGAAAAAGAAGAAGCAACAACAAG3′
AG1884	5′TGCTCGAGATGCGTAAATTGCTCTC3′

2.2 Production and purification of TnsE and TnsE variants

All TnsE variants were produced in BL21 STAR pRARE pLysS cells transformed with IPTG-inducible pET-22b expression vectors. Cultures were grown to OD_{600} of 0.7 and protein production was induced by addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Cultures were then incubated at 37 ºC with orbital agitation for 3 hours (except for wild-type TnsE that was incubated at 25 ºC for 5 hours). Cell were harvested by centrifugation, washed with 20 mL phosphate buffer saline, and the cell pellets were stored at -80 °C.

Cell pellets were resuspended in buffer 'Ni' (20 mM Tris pH 8, 0.5 M NaCl, 1.4 mM 2-mercaptoethanol, and 5% glycerol) and lysed by sonication. The lysate was clarified by centrifugation at 39 000 g at 4 °C for 40 minutes before loading onto a HisTrap HP metal-chelating column (GE Healthcare) equilibrated with the 'Ni' buffer. The column-bound protein was washed extensively with Ni buffer containing 30 mM imidazole, followed by 60 mM imidazole, before elution with 240 mM imidazole. For full-length TnsE, a HiTrap HP Q Sepharose ion exchange column was equilibrated and attached to collect contaminating nucleic acids. The eluted protein was diluted to a NaCl concentration of 100 mM using 'Q' buffer (20 mM Tris pH 8, 1 mM EDTA, 5 mM DTT, 5% glycerol) and then was further purified by ion exchange chromatography on a Mono Q 5/50 GL anion exchange column (GE healthcare). TnsE-CTD and TnsE-CTD$^{A453V+D523N}$ were diluted to 50 mM NaCl to bind the MonoQ resin. After loading, the column-bound protein was extensively washed with loading buffer before protein elution with a linear salt gradient to 500 mM NaCl. The eluted proteins were then concentrated and the buffer was exchanged to the TnsE Storage Buffer (20 mM Tris pH 8, 150 mM NaCl, 0.1 mM EDTA, 5 mM DTT, 5% glycerol). Dynamic light scattering (DLS) was used to assess purified protein for aggregation. For this, 17 µl of

concentrated protein was measured at 4 °C under default settings for protein samples on a ZetaSizer Nano S instrument (Malvern).

Selenomethionine-labeled TnsE-CTD for phase determination, $CaCl_2$-competent B834 cells (methionine auxotroph) were transformed with pRARELysS and a pET-22b vector carrying the truncated gene for TnsE-CTD expression. Minimal-media culture preparation and protein expression was conducted as described previously with minor modifications (Hendrickson et al., 1990). Subsequent purifications were carried out following the same method as for native TnsE-CTD.

2.3 Crystallization and structure determination of TnsE-CTD

Initial high-throughput commercial crystallization screens were done for native TnsE-CTD (0.3 µl 0.1 mM TnsE-CTD + 0.3 µl well buffer over 50 µl well buffer) using a Crystal Phoenix liquid handling robot (Art Robbins Instruments). Plates were incubated at 4 °C with minimal exposure to light, and drop images were recorded periodically using a RockImager 180 (Formulatrix). Original crystals grew in 27% (w/v) PEG 3350, 0.2 M NaCl, 0.1 M Tris pH 8.0 and all were cryoprotected by the addition of 3-6% glycerol to the mother liquor before flash-freezing in liquid nitrogen.

Selenomethionine-labeled TnsE-CTD was crystallized by mixing equal volumes of 0.25 mM TnsE-CTD with arrays of crystallization buffers previously optimized for the wild type domain within the following composition range: 24-29% PEG 3350, 0.23-0.26 M NaCl, 0.1 M Tris pH 8.0 - pH 8.5. Hanging drop experiments were set over ice and equilibrated at 4 °C. Initial screening of cryo-cooled crystals was performed in house using a rotating copper anode generator and an R-AxisIV++ image plate detector (Rigaku). Full data sets of single crystals diffracting to the highest resolutions of those available were collected on a

Dectric Pilatus 6M detector at beam line X29 of the National Synchrotron Light Source (Brookhaven National Laboratory). To maximize the signal from anomalous scattering, X-rays of wavelength 0.979 Å were used, which is near the K-edge for selenium.
Data were indexed, processed and merged using HKL2000 (Otwinowski and Minor, 1997). None of the individual data sets yielded interpretable experimental phases using AutoSol (Phenix). However, a merged data set derived from the three crystals resulted in a good quality experimental map in the $P2_1$ space group, where the 12 molecules of TnsE contained in the asymmetric unit could be readily traced. Despite the high number of molecules in the asymmetric unit, refinement and model building were done using standard protocols in Phenix and COOT (Afonine et al., 2005; Emsley and Cowtan, 2004). Non-crystallographic averaging was not used for phasing in AutoSol, but non-crystallographic symmetry restraints were used during the refinement in Phenix.

Crystals of the TnsE-CTD$^{A453V+D523N}$ double mutant were grown in 25% (w/v) PEG 3350, 0.1 M Bis- Tris pH 5.5 and also belonged to the $P2_1$ space group. A complete data set was collected on an MX300HE CCD detector (Rayonix) at beamline 08B1-1 of the Canadian Light Source (CLS; Fodje et al., 2014) with X-rays of wavelength 0.979 Å. Diffraction data were processed by Autoprocess, a CLS-developed pipeline of XDS, POINTLESS, and SCALA; and the structure was solved by molecular replacement using Phaser (Phenix; McCoy et al., 2007) and Autobuild (Phenix; Terwilliger et al., 2008). For molecular replacement, the structure of wild type TnsE-CTD was used as the search model with residues 457 – 462 of the flexible arm of the V-loop deleted to prevent local model bias from arising in that region. Iterative building and fitting of the TnsE-CTD$^{A453V+D523N}$ model was done in COOT (Emsley and Cowtan, 2004) and refinement was performed using standard procedures in Refine (Phenix; Adams et al., 2010).

2.4 Crystallization and diffraction of TnsE-NTD

TnsE-NTD was produced and purified as described in section 2.2. To optimize the initial crystals obtained in previous in-situ proteolysis experiments, larger format experiments were assembled with well buffer conditions ranging in composition (19–29% PEG 3350, 0.1–0.3 M $(NH_4)_2SO_4$, 0.1 M Bis-Tris pH 6.5). For in-situ proteolysis, papain (Worthington Biochemical Co.) was added and mixed with a 200 μM protein stock to a proportion of 1/500, 1/1000, and 1/2000 the mass of TnsE-NTD. Drops of two protein:condition ratios were set by adding 1 μl of TnsE-NTD/papain stock to 1 and 2 μl of well buffer. The experiments were incubated at 4 °C protected from light.

2.5 DNA substrates

The oligonucleotides NLC 707 and NLC 710 (**Table 2.2**) were previously used in electrophoretic mobility shift assays to show that TnsE specifically binds DNA substrates with 3′-recessed ends (Peters and Craig, 2001). These two oligonucleotides served as the template to design shorter DNA substrates used to test for minimal substrate binding in differential scanning fluorimetry (DSF) experiments and crystallization trials. All DNA substrates preserved the sequence around the 3′-recessed end that forms in the NLC707+NLC710 duplex (**Table 2.2**), though some single-stranded 5′ ends were modified to prevent the formation of predicted alternate structures.

Table 2.2 Oligonucleotide pairs used in DNA-binding assays

Label	ID	Sequence
30+35		
	NLC710	5′GGCAAAGATGTCCTAGTTTAGTTGGCTGAT3′
	NLC707	3′CCGTTTCTACAGGATCAAATCAACCGACTAGCAACGTAACCACTGCAAAGGCCGGAACGATTACC5′
15+15	*JJC004*	5′GTTTAGTTGGCTGAT3′
	JJC013	3′CAAATCAACCGACTAGCAACGTAACCACGG5′
15+12	*JJC004*	5′GTTTAGTTGGCTGAT3′
	JJC027	3′CAAATCAACCGACTAGCAACGTAACCG5′
15+10	*JJC004*	5′GTTTAGTTGGCTGAT3′
	JJC014	3′CAAATCAACCGACTAGCAACGTAGG5′
15+5	*JJC004*	5′GTTTAGTTGGCTGAT3′
	JJC025	3′CAAATCAACCGACTAGCAAG5′
12+15	*JJC018*	5′TAGTTGGCTGAT3′
	JJC017	3′ATCAACCGACTAGCAACGTAACCACCG5′
10+15	*JJC006*	5′GTTGGCTGAT3′
	JJC010	3′CAACCGACTAGCAACGTAACCACAC5′
30+0	*JJC013*	5′GTTTAGTTGGCTGATCGTTGCATTGGTGCC3′
	JJC026	3′CAAATCAACCGACTAGCAACGTAACCACGG5′
0+30	*JJC026*	5′GTTTAGTTGGCTGATCGTTGCATTGGTGCC3′

Oligonucleotides were synthesized by Integrated DNA Technologies Inc. and were analyzed for melting temperature, hairpin formation and self-complementation using their online tool Oligo Analyzer 3.1. For experimental use, oligonucleotides were resuspended in sterile Millipore water before annealing by boiling sealed tubes in a 1L water bath for 5 minutes and cooling slowly to 4 °C over several hours. Nucleoprotein complexes were prepared by incubating an annealed DNA substrate with TnsE for 30 minutes at 25 °C. Label conventions for DNA substrates with 3′-recessed ends are such that one labeled 15+10 would have 15 nucleotides in the duplex region and 10 nucleotides in the single-stranded region (**Figure 2.1**).

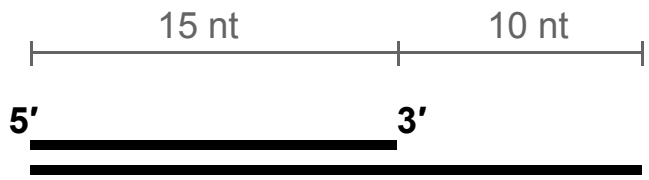

***Figure 2.1* DNA nomenclature in this study.** DNA substrates are labeled #+# such that the number before the plus sign indicates the length of the dsDNA region and the number following the plus sign indicates the length of the ssDNA region. This example shows the 15+10 substrate.

2.6 Electrophoretic mobility shift assays

Electrophoretic mobility shift assays (EMSA) were performed as described by Peters & Craig (2001). Before annealing, the shorter oligonucleotide of each pair was 5′-labeled with ^{32}P-ATP (Perkin Elmer). The concentration of this substrate was raised to 10 nM in the reactions to account for a smaller well capacity of the gels used. The plasmid pRM2 was linearized with HincII (New England Biolabs) to serve as double-stranded background DNA at a 50-fold mass excess over the labeled substrate.

The EMSA reaction mixtures were each prepared with 0.15 pmol ^{32}P-labeled DNA substrate and a stoichiometric increase in TnsE, TnsE-NTD, TnsE-CTD, or TnsE$^{A453V+D523N}$. Assembled reactions were incubated for 30 minutes at 30 °C before separation by gel electrophoresis at 80 V for 45 minutes on a 5% polyacrylamide gel at 25 °C in 1x TBE buffer. Autoradiographic gel images were collected on a phosphor screen (Molecular Dynamics Inc.), using a Typhoon Variable Mode Imager (GE Healthcare).

2.7 Size exclusion chromatography

Ten µL of 80 µM samples containing either TnsE$^{A453V+D523N}$, a 15+15 DNA substrate of oligonucleotides JJC003 and JJC004 (Table 2.2), or their nucleoprotein complex were loaded onto a Superdex 200 PC 3.2/30 gel filtration column (GE), which had been pre-equilibrated with de-gassed running buffer (20 mM Tris pH 8, 150 mM NaCl, 1 mM EDTA, 5 mM DTT,

5% glycerol). With the system at room temperature (~25 °C), the samples were fractionated at 0.04 ml/min in the same running buffer and the UV absorbance of the eluate was monitored at wavelengths of 260 nm and 280 nm. Marker proteins were also run in equivalent conditions to provide column calibration. These included ferritin (440 kDa), aldolase (158 kDa), albumin (67 kDa), ovalbumin (44 kDa) and ribonuclease A (14 kDa).

2.8 Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) was performed according to the protocol described by Boivin, Kozak, and Meijers (Boivin et al., 2013). 5 µl each of 5x stocks of TnsE+DNA complexes, buffers, and competitor DNA were added to wells of a 96-well RT-PCR plate, then brought up to a final volume of 23 µl with sterile Millipore water. 2 µl 63x Sypro Orange dye (Life Technologies) was added to each well immediately before beginning a 1 °C/min temperature gradient from 4 °C to 95 °C with plate scans at every degree increment. 5x complexes of $TnsE^{A453V+D523N}$ and 3′-recessed substrates were prepared by mixing 1.5 fold excess of annealed DNA substrate with purified protein. These, as well as controls, were incubated in TnsE storage buffer at 25 °C for 30 minutes before returning to 4 °C. Final concentrations were as follows: 5 µM $TnsE^{A453V+D523N}$, 7.5 µM 3′-recessed DNA substrates, 15 µM competitor DNA (30+0), 150 mM NaCl, 20 mM Tris pH 8.0, 0.1 mM EDTA, 1.4 mM 2-mercaptoethanol, and 5% (v/v) glycerol. The experiment was performed in a C1000 thermal cycler outfitted with a CFX96 optical module (Bio-Rad) and fluorescence data were collected on the SYBR Green channel.

Relative melting temperature (t_m) values were determined by fitting logistic functions to the DSF datasets, a relatively simple procedure for $TnsE^{A453V+D523N}$ since it produced a single fluorescence peak while melting. Due to thermal degradation of the Sypro Orange fluorophore, a logistic function with an asymmetry parameter (d) was applied following the form of equation 2-1. The log-scaled form of this function provides one of the most accurate contemporary fits for qPCR data (Spiess et al., 2008).

$$Fluorescence(x, a, b, g, d, t) = \frac{a}{\left(1-e^{-b(x-g)}\right)^{d}} + t \qquad (2\text{-}1)$$

The fit of each curve was optimized between the minimum and maximum values of the peak region by varying the parameters *a, b, g, d,* and *t* with the Generalized Reduced Gradient algorithm in Excel 2011 (Microsoft) while targeting a maximum value of the non-linear coefficient of determination (R^2). The t_m for each set was determined from the inflection point of the optimized curve, which was calculated from the root of the second derivative of equation 2-1 as determined by the Wolfram|Alpha math engine (Wolfram|Alpha, 2016). This root is shown in equation 2-2.

$$Fluorescence''(x, a, b, g, d, t) = 0, \ when \ x = \frac{bg + \log d}{b} \qquad (2\text{-}2)$$

2.9 Crystallization screens of TnsE$^{A453V+D523N}$+DNA complexes

TnsE$^{A453V+D523N}$ was produced and purified as described in section 2.2. Three-prime-recessed end DNA duplexes preserving the junction sequence of NLC707+NLC710 were annealed with 13-16 nucleotides in the ssDNA and dsDNA arms (Table 2.3). TnsE-DNA complexes (200 μM) were prepared by incubating a 1.2-fold excess of DNA substrates with TnsE$^{A453V+D523N}$ at 25 °C. To test a variety of crystallization conditions for TnsE$^{A453V+D523N}$+DNA complexes, screens were prepared using the commercially prepared condition sets Classics II,(QIAGEN), Top96 (Microlytic), PACT-premier, and JCSG-plus (Molecular Dimensions). The Phoenix liquid handling robot (Art Robbins Instruments) was used to set sitting drops consisting of 0.4 μl of the well buffer and 0.4 μl of TnsE$^{A453V+D523N}$+DNA complex, which were then sealed and then incubated at 4 °C with minimal exposure to light. Images were collected using a RockImager 180 (Formulatrix).

Table 2.3: Oligonucleotides used in TnsE$^{A453V+D523N}$+DNA crystallization screens.

Label	ID	Sequence
16+16	*JJC030*	5′AGTTTAGTTGGCTGAT 3′
	JJC031	3′TCAAATCAACCGACTAGCAACGTAACCACTGC 5′
16+15	*JJC030*	5′AGTTTAGTTGGCTGAT 3′
	JJC032	3′TCAAATCAACCGACTAGCAACGTAACCACTG 5′
16+14	*JJC030*	5′AGTTTAGTTGGCTGAT 3′
	JJC033	3′TCAAATCAACCGACTAGCAACGTAACCACT 5′
16+13	*JJC030*	5′AGTTTAGTTGGCTGAT 3′
	JJC034	3′TCAAATCAACCGACTAGCAACGTAACCAC 5′
15+16	*JJC004*	5′ GTTTAGTTGGCTGAT 3′
	JJC027	3′ CAAATCAACCGACTAGCAACGTAACCACTGC 5′
15+15	*JJC004*	5′ GTTTAGTTGGCTGAT 3′
	JJC003	3′ CAAATCAACCGACTAGCAACGTAACCACTG 5′
15+14	*JJC004*	5′ GTTTAGTTGGCTGAT 3′
	JJC028	3′ CAAATCAACCGACTAGCAACGTAACCACT 5′
15+13	*JJC004*	5′ GTTTAGTTGGCTGAT 3′
	JJC029	3′ CAAATCAACCGACTAGCAACGTAACCAC 5′
14+16	*JJC035*	5′ TTTAGTTGGCTGAT 3′
	JJC036	3′ AAATCAACCGACTAGCAACGTAACCACTGC 5′
14+15	*JJC035*	5′ TTTAGTTGGCTGAT 3′
	JJC037	3′ AAATCAACCGACTAGCAACGTAACCACTG 5′
14+14	*JJC035*	5′ TTTAGTTGGCTGAT 3′
	JJC038	3′ AAATCAACCGACTAGCAACGTAACCACT 5′
14+13	*JJC035*	5′ TTTAGTTGGCTGAT 3′
	JJC039	3′ AAATCAACCGACTAGCAACGTAACCAC 5′
13+16	*JJC040*	5′ TTAGTTGGCTGAT 3′
	JJC041	3′ AATCAACCGACTAGCAACGTAACCACTGC 5′
13+15	*JJC040*	5′ TTAGTTGGCTGAT 3′
	JJC042	3′ AATCAACCGACTAGCAACGTAACCACTG 5′
13+14	*JJC040*	5′ TTAGTTGGCTGAT 3′
	JJC043	3′ AATCAACCGACTAGCAACGTAACCACT 5′
13+13	*JJC040*	5′ TTAGTTGGCTGAT 3′
	JJC044	3′ AATCAACCGACTAGCAACGTAACCAC 5′

Chapter 3

Structural Characterization of TnsE

3.1 Initial deconstruction of TnsE's domain architecture

TnsE displays no amino acid sequence homology with other known proteins outside its homologs in the Tn7-like family of transposable elements. Yu Seon Chung, a former member of our research group used secondary structure prediction as well as limited trypsin proteolysis to show that TnsE has two domains connected by a flexible linker, information which proved to be crucial for continuing the structural study of this targeting protein. PsiPred (Buchan et al., 2013) predicted two regions populated uniformly with secondary structure elements separated by an extended region of coil. This suggests the initial model of two domains connected by a linker, which was further tested by limited trypsin proteolysis. Purified TnsE could be digested to completion in low concentrations of trypsin, forming two fragments (~25 kDa and ~48 kDa as analyzed by SDS-PAGE), which were then able to resist further degradation up to ~16-fold greater trypsin concentrations (Fig. 3.1A). The TnsE used in this experiment had a C-terminal

hexahistidine (His_6) tag allowing the initial degradation products to be identified by elution from a metal affinity column. The 48 kDa fragment eluted under a step gradient from 0 mM to 30 mM imidazole, while the 25 kDa fragment did not elute until the imidazole concentration was stepped from 60 mM to 300 mM, suggesting that the former is the N-terminal fragment and the latter is the tagged C-terminus (Fig. 3.1B).

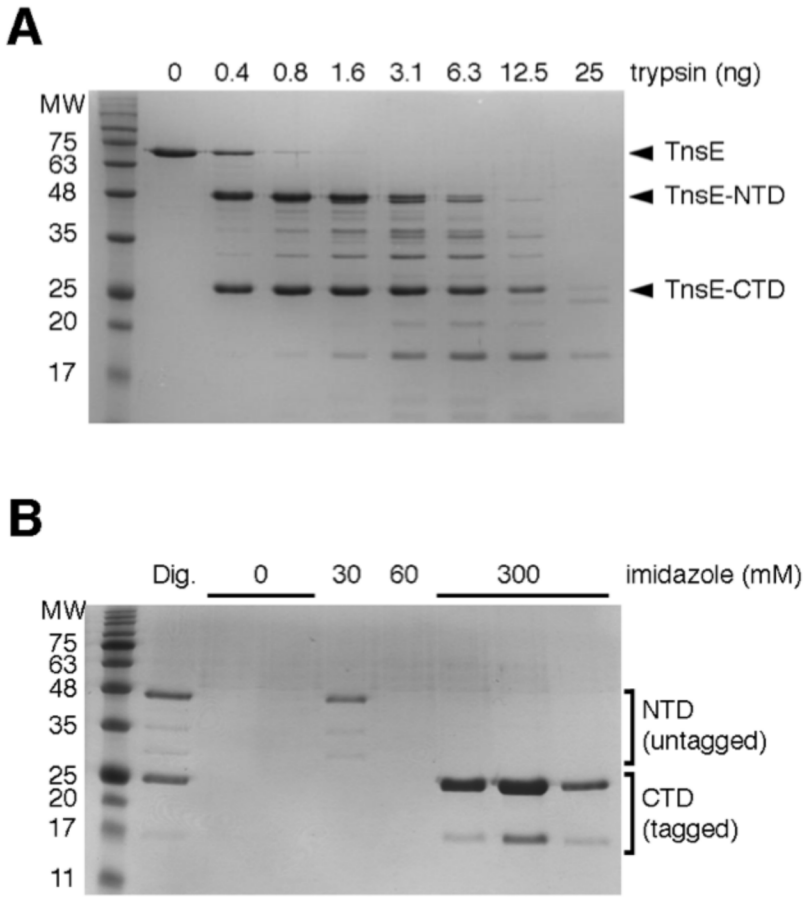

Figure 3.1: Trypsin proteolysis identifies two domains in TnsE. (A) SDS-PAGE of limited trypsin proteolysis products indicates TnsE is very susceptible to digestion into two primary structures that are more resistant to further degradation. (B) Elution of the two major digestion products from a metal affinity column sequentially resolves them since the C-terminal product retains the His_6 tag. (*Figure generated by Dr. Y.S. Chung*)

3.2 Production of the independent TnsE domains

Based on the predicted secondary structure of TnsE and the molecular weights of the TnsE fragments identified by Dr. Chung, I estimated the positions of several potential trypsin cleavage sites, which were grouped together in the region spanning residues Arg332 to Lys344. To study functional separation between the putative domains and facilitate independent structural analysis of each, the *tnsE* gene was divided by subcloning between the codons for residues Thr340 and Ile341. The putative N-terminal domain (TnsE-NTD) includes the TnsE N-terminus through residue Thr340 as well as a C-terminal His_6 tag. The putative C-terminal domain (TnsE-CTD) introduces the residue variant I341M to enable production in culture and extends to the C-terminus, which is likewise appended with a His_6 tag. This separation site was chosen because it is within the basic region accessible to trypsin and the I341M mutation should minimize any effects on folding while introducing an initiation codon.

Following independent expression of the two putative domains, both were produced in high quantity and had excellent solubility in lysis buffer, which indicated neither was dependent on the other for native folding. A standard purification protocol was developed for each domain (See Methods and Materials), which involved an initial step separating the His_6-tagged domains from other soluble cellular lysate on a metal-affinity column. This was followed by anion exchange chromatography, which eluted the domains as single peaks over a linear gradient of sodium chloride, separating them from most residual contaminants. In both cases, the peak fraction corresponding to the tagged protein was of a purity (as determined by SDS-PAGE) and yield (2-8 mg purified protein

per litre of production culture) suitable for biochemical assays and crystallography (**Figure 3.2**).

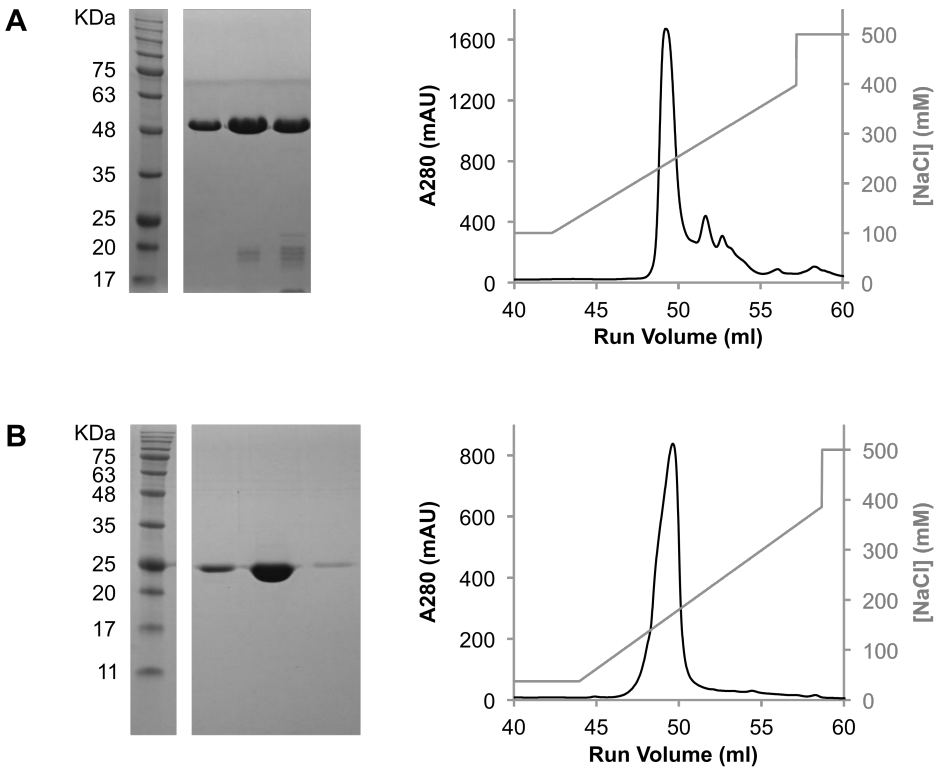

Figure 3.2: Purification of TnsE domains. **(A)** SDS-PAGE of fractions collected from an anion exchange column loaded with TnsE-NTD indicates the major peak fractions were an adequate purity for crystallography and biochemical assays. **(B)** SDS-PAGE gel of mid-peak and peak-edge fractions of eluate from an anion exchange column loaded with TnsE-CTD shows no visually distinguishable contamination. This purity was sufficient to proceed with crystallization trials and biochemical assays.

3.3 The two domains of TnsE are required for DNA binding

Mutagenesis experiments localized regulation of TnsE-directed transposition to the C-terminus of TnsE, which correlated increases in transposition frequency to increases in DNA binding affinity (Peters and Craig, 2001). To test the fragment of TnsE that binds DNA, electrophoretic mobility shift assays (EMSA) were performed on a 3′-recessed DNA substrate (30+35) with full-length TnsE, each of its domains separately, and both domains together (**Figure 3.3**). Full-length TnsE shifted the labeled substrate at a concentration similar to previous studies (Peters and Craig, 2001), but neither domain alone shifted DNA at comparably low concentrations. Contrary to expectations, TnsE-NTD produced a shift more readily than TnsE-CTD, though only at very high concentrations where non-specific binding becomes more likely to drive the shift. A partial rescue effect was observed when both domains are present, though it still occurred at a much higher protein concentration than for full-length TnsE. Together, this supports a domain cooperation model for the DNA-binding mechanism and indicates the linker tethering the two domains is critical, likely because it increases their local concentration with respect to each other. This lack of autonomous functional control over DNA binding in TnsE-CTD, despite mutagenic data, prompted structural analysis of the domain to elucidate the mechanism.

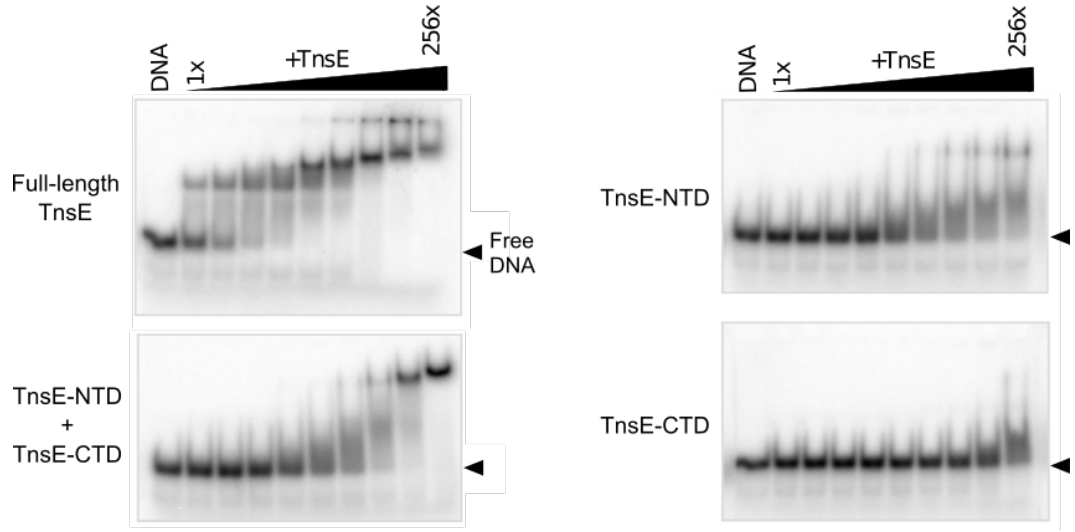

Figure 3.3: EMSA of TnsE and its domains using 3′-recessed DNA substrates. Full-length TnsE produced discrete shifts at low concentrations, which were not reproduced by any assays using independently purified domains. When both domains were independently purified and then combined, the shifted bands became more discrete, but did not occur at the same low protein concentration as in the full-length TnsE assay.

3.4 Determination of the TnsE-CTD crystal structure

With no homologous structures available, selenomethionyl derivative crystals of TnsE-CTD were produced and the structure was solved by single-wavelength anomalous diffraction. Individual datasets were sufficient to determine the space group ($P2_1$) and unit cell parameters (See Table 3.1), but a solution was initially not obtained due to low anomalous signal-to-noise ratios ($\Delta F/\sigma$). However, merging datasets from multiple crystals can improve $\Delta F/\sigma$, a strategy that was applied to produce a solution using data from three of the selenomethionyl crystals diffracted (Liu et al., 2011).

Table 3.1: Data collection and refinement statistics for TnsE-CTD

Data Collection	
Wavelength	0.979
Cell: a, b, c (Å) / α, β, γ (°)	110.3, 87.1, 141.6 / 90, 95, 90
No. mol asu	12
Resolution (Å)	50-2.85 (2.9-2.85)
R_{meas}	0.12 (1.00)
I/σ(I)	20.5 (1.9)
Completeness (%)	100 (100)
Redundancy	10.6 (10.1)
Refinement	
Resolution (Å)	43.2 – 2.85
Completeness (%)	99.9
No. Reflections	122,621
R_{work} / R_{free} (%)	21.1 / 24.2
Atoms refined	15,020
Solvent atoms	139
Solvent content (%)	49.6
Rmsd in bonds (Å)	0.005
Rmsd in angles (°)	0.97
Mean B values (Å)	84.6

Data derived from three merged crystals.
Data in the highest resolution shell is shown in parentheses.

The asymmetric unit of these crystals contains 12 copies of TnsE-CTD arranged into 2 distinct pseudohexamers, each of which is organized as a trimer of dimers (**Figure 3.4**). Since this domain has not been observed to multimerize in vitro, the structure was submitted to the PISA server, which found that the surface buried by each order of multimerization was below the cutoff value for specific protein-protein interactions (Krissinel and Henrick, 2007). Therefore, the formation of these pseudo-multimers is most likely driven by crystal packing.

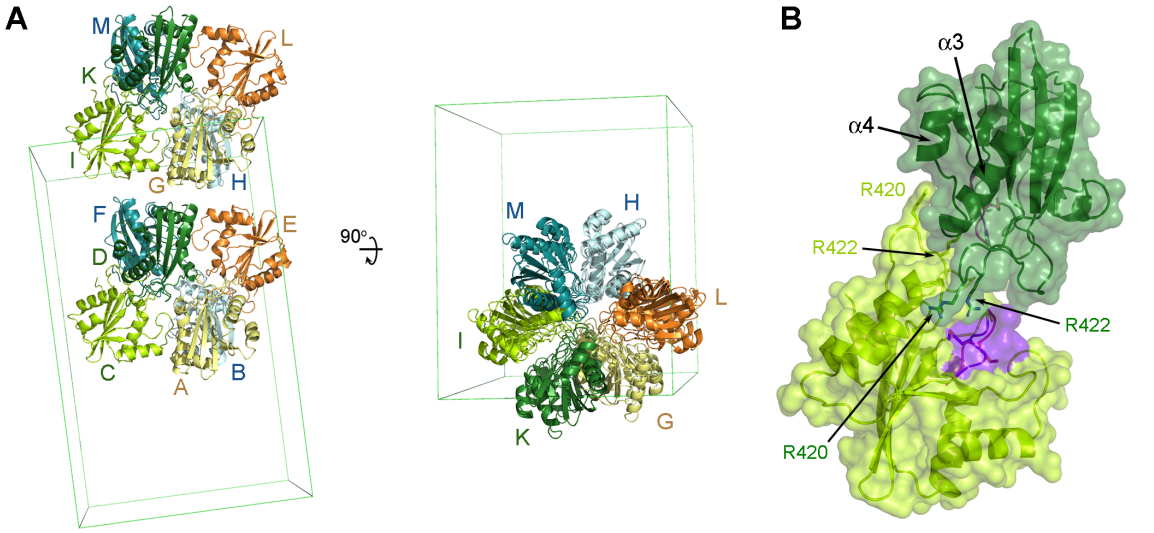

Figure 3.4: Crystal packing of TnsE-CTD. (A) The 12 TnsE-CTD molecules in the ASU are organized into pseudodimers (similar hues, ex: I-K G-L H-M), three of which form wind around each other to form one of two pseudohexamers in the ASU. The unit cell is represented as the outine of the edges of the prism. **(B)** The interface between pseudodimer components is the most extensive and is largely mediated by hydrogen bonding between the guanidinium groups of residues R420 and R422 from one partner with the loop connecting helices α3 and α4 of the other. The V-shaped loop described in section 3.6 participates minimally in this interface.

The ordered region in the TnsE-CTD structure spans residues Asp377 to Arg533, with the exception of several flexible surface-exposed sidechains. The domain core is defined by a 5-strand antiparallel β-sheet covered by 3 α-helices (α1,α2, α3) on one face and 2 α-helices (α3, α4) on the other (**Figure 3.5**).

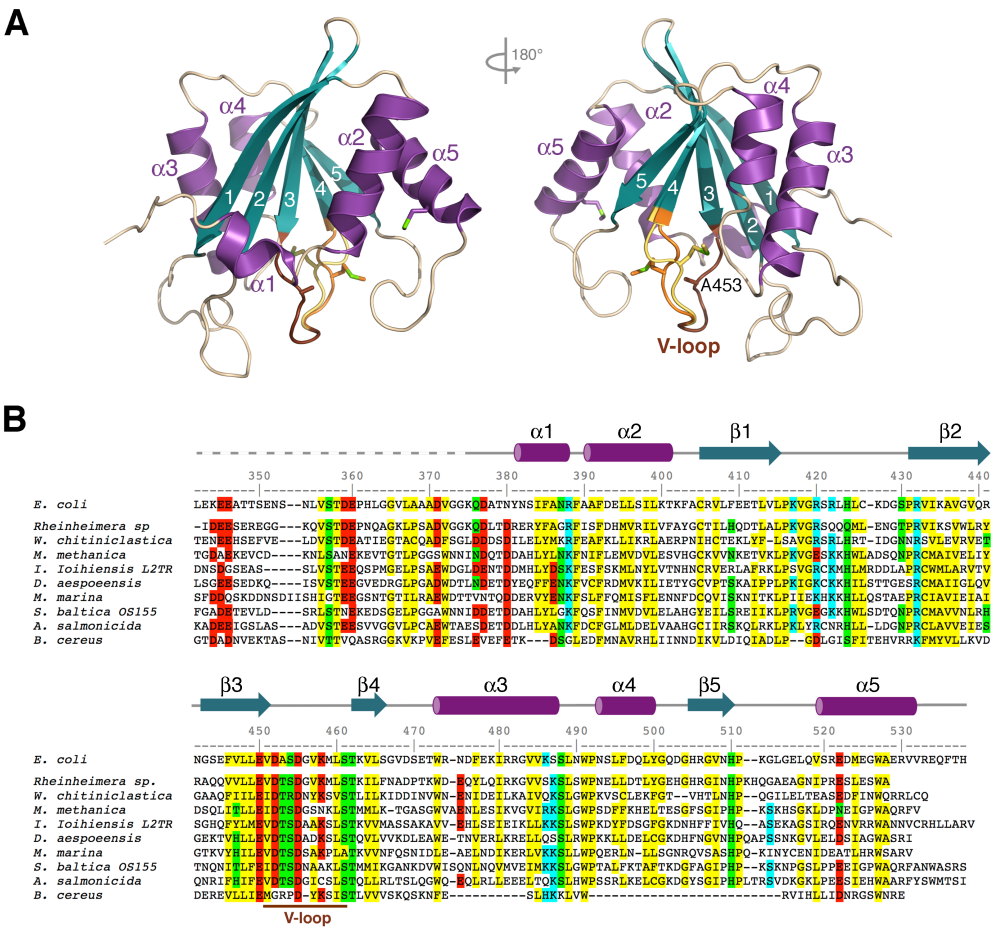

Figure 3.5: Crystal structure and sequence conservation of TnsE-CTD. (A) Opposite views of TnsE-CTD. Two conformations of the extended loop between β3 and β4 are shown. **(B)** Sequence alignment shows conservation in secondary structure regions and in the lengths of extended loops between them.

The loops connecting secondary structure elements at one end of the β-sheet (ex: β2 to β3) are minimal in size, while those at the other end (ex: β1 to β2) are comparatively extended and define an entire face of the domain (**Figure 3.5**). A network of hydrogen bonds allows these extended loops to remain ordered. The importance of the integrity of this side of the domain is made apparent by conservation of the loop lengths and sequences in TnsE homologs (**Figure 3.5**). This conservation occurs across a diverse

range of host species and also extends well into other ordered regions, with moderate exception in *Bacillus cereus*.

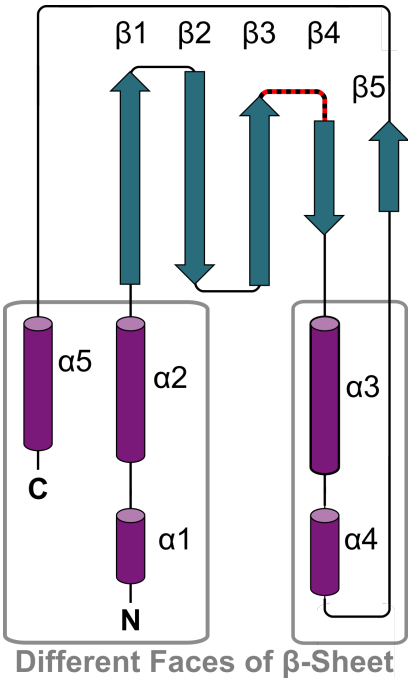

Figure 3.6: TnsE-CTD topology diagram. Two sets of α-helices are grouped on either of the two faces of the core β-sheet. Section 3.6 describes a V-shaped loop, indicated here by red dashes.

3.5 TnsE-CTD contains a novel fold

No sequence homology has been determined between TnsE and any other protein outside the Tn7-like transposon family, however this does not preclude it adopting the same fold as a previously characterized protein. To determine whether any functional information about TnsE could be extrapolated from structurally related proteins, searches were performed to identify homologous folds. The model coordinates of TnsE-CTD were compared pair-wise with every other entry in the Protein Data Bank using the DALI Server (Holm and Rosenstrom, 2010), returning no significant matches. The most similar

structure returned, the FP domain of the human proteasome inhibitor PI31, has 8% sequence identity with TnsE-CTD and a partially similar topology. The topological similarity includes only TnsE-CTD's central β-sheet and helices α2 and α5, and the interactions stabilizing the domain cores are very different (Fig. 3.5.1). These two domains share no common functionality to rationalize similar elements of their folds. The exposed surface of the PI31 FP domain's β-sheet serves as an interface for a protein-protein interaction (Kirk et al., 2008), while the equivalent face of TnsE-CTD's β-sheet constitutes much of the hydrophobic core. Furthermore, the PI31 FP domain lacks a region with multiple extended loops. The functional significance that accompanies these structural differences allowed us to determine with certainty, that TnsE-CTD contains a novel fold.

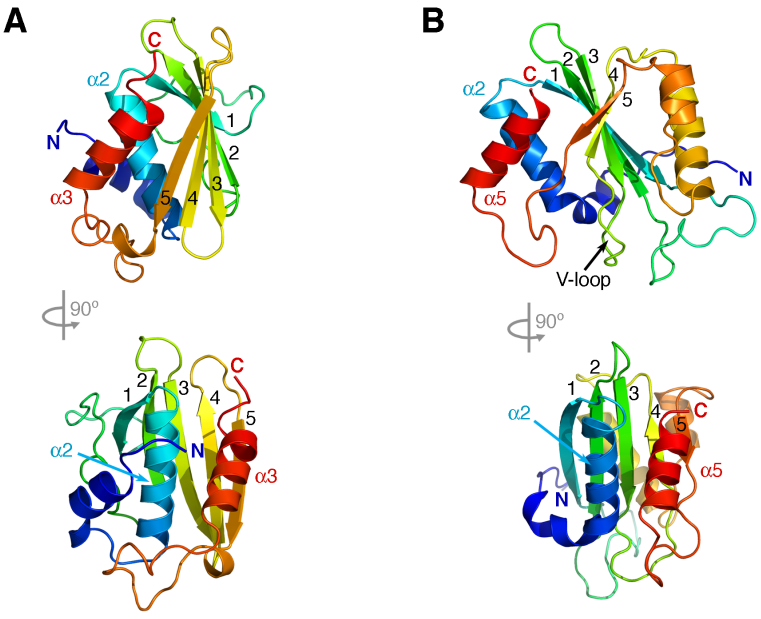

Figure 3.7: TnsE-CTD has a novel fold. The FP domain of human PI31 (A) positions some se- condary structure ele-ments similarly to those in TnsE-CTD (B), however the topology of connec-tions between these ele-ments differs and equi-valent faces of the β sheet are solvent-exposed in PI31 and buried by helices in TnsE-CTD. Functional separation of these structural elements further distinguishes the two domains.

3.6 The V-shaped loop connecting β3-β4 adopts two conformations

The extended loop connecting strands β3 and β4 (V-loop) has the shape of a V and can be defined by assigning residues Val451-Asp455 to the first arm (arm 1) of the V and residues Val457-Ser461 to the second arm (arm 2). In the crystal structure, arm 1 is buried in the hydrophobic core of the domain and is stabilized by numerous hydrogen bonds, and is therefore well defined in the electron density. This is not the case for arm 2, for which the electron density does not uniformly resemble an ordered polypeptide. However, this is reconciled by the presence of two anomalous peaks at the position of residue M459, one on either side of the polypeptide backbone.

Although the construct of TnsE-CTD expressed contains 3 selenomethionine (Mse) residues, we expected to see 2 anomalous peaks in each copy since the Mse341 of this construct falls within TnsE's presumably flexible linker region and the N-terminal methionine of recombinant proteins is often removed by an endogenous methionine aminopeptidase (Liao et al., 2004). We observed 3 anomalous peaks for each of the 12 copies of TnsE-CTD in the asymmetric unit (ASU), although as expected, residue Mse341 was not ordered if present at all. The 12 strongest peaks in the asymmetric unit mapped to the position of residue Mse524, while the remaining 24 mapped in pairs of comparable occupancy on either side of Mse459.

To account for the ambiguous electron density and the paired anomalous peaks in arm 2 of the V-loop, two alternate conformations were modeled (**Figure 3.8**). Along with changes to the configuration of the polypeptide backbone, the most radical differences between the alternate conformations are the orientations of the side chains of residues

Lys458, Mse459, and Leu460. In conformation I, the sidechain of Lys458 points toward a turn in the loop between sheet β5 and helix α5. While several electronegative atoms in this turn are within range for hydrogen bonding, the structure of the Lys458 sidechain becomes disordered beyond Cγ. In conformation II, this sidechain points directly out into the solvent and is disordered beyond Cγ. The sidechains of residues Mse459 and Leu460 alternately occupy a small hydrophobic pocket (Mse459 in conformation I, Leu460 in conformation II) created by residues Leu489, Trp491, and Val451. In conformation II, when Mse459 does not occupy the pocket, its sidechain is instead flipped to the opposite side of the backbone, whereas in conformation I, when Leu460 does not occupy the pocket, its sidechain is undergoes a lesser rotation to extend directly out into the solvent. Although the complete exposure of the Leu460 sidechain to solvent should be disfavoured, the similar occupancies of the two anomalous peaks for Mse459 suggests the energetic favourability of the alternate conformations is comparable.

The V-shaped loop is adjacent to the pseudodimer interface, which occurs on the same surface of both participating monomers. This ensures a consistent packing environment in all 12 copies of arm 2. Furthermore, its proximity to this interface sequesters it from external crystallographic contacts, ensuring the conformational heterogeneity of arm 2 is not the result of irregularities in its environment throughout the crystal lattice.

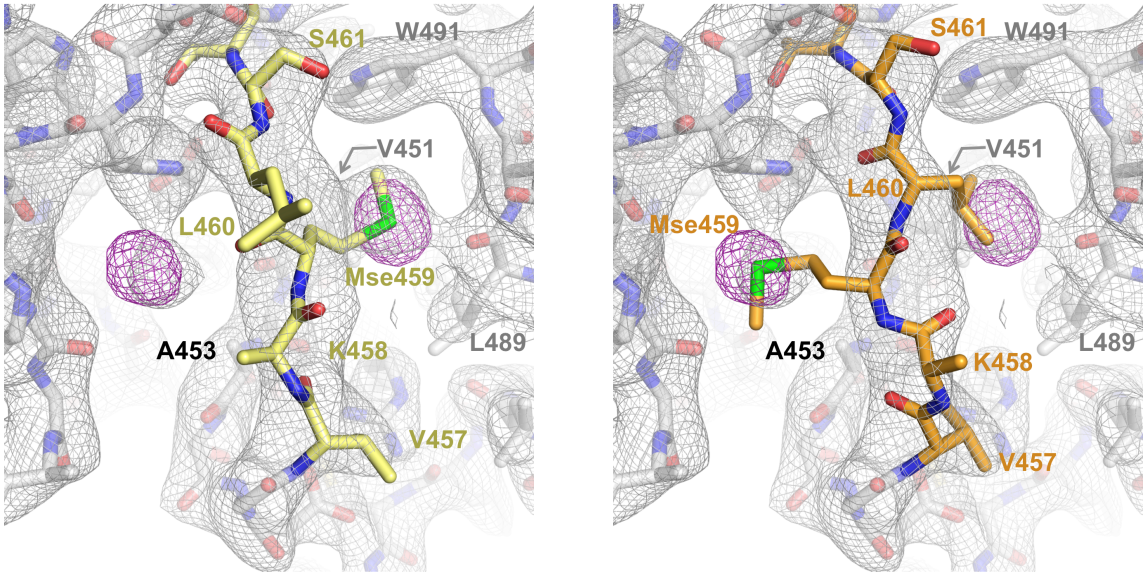

Figure 3.8: Arm 2 of the V-loop adopts two conformations. Paired anomalous peaks (purple mesh, σ = 6.0) indicate the positions of residue Mse459 in the alternate conformations of arm 2 (orange and yellow). Electron density for other sidechains in this arm was not defined. The sidechain of V451 lies behind the right anomalous peak in this view.

In vivo, the amino acid variants A453V and D523N have a particularly significant role in upregulating transposition frequency. Residue Ala453 is positioned in the centre of the highly supported arm 1 and its sidechain Cβ comes very close to the backbone of arm 2. This suggests that the wild type structure of arm 1 may permit flexibility in arm 2, while the added sidechain bulk in A453V variants may not. If this were the case, it would indicate that the conformational state of the V-loop has a key function in regulating transposition through the TnsABC+E pathway. Bacterial transposons and insertion sequences possess a number of other regulatory mechanisms to limit transposition to a non-deletorious frequency (Nagy and Chandler, 2004). Here, it is suggested that the conformational flexibility of TnsE's V-loop demonstrates a new class of regulation in the targeting of Tn7-like transposons. I hypothesized that the mechanism underlying this

regulation relates the targeting avidity of TnsE to the conformational state of arm 2, and that the freedom to toggle between conformations prevents excessive target recognition. To test this, we investigated the structure of the most proficient non-lethal variant of the domain: TnsE-CTD$^{A453V+D523N}$.

3.7 The side chain of Ala453 determines the toggling of the V-loop

TnsE-CTD$^{, A453V+D523N}$ also crystallized in the $P2_1$ space group, although unit cell dimensions were smaller with only 2 copies of the domain in the asymmetric unit. These crystals diffracted to a much higher resolution than the wild type crystals, largely due to the lower solvent content (Heras and Martin, 2005) (**Table 3.2**). The TnsE-CTD$^{,A453V+D523N}$ structure was solved by molecular replacement, using a monomer from the wild type structure as a search model. To prevent the search model from generating local bias, the flexible arm of the V-loop (arm 2) was removed.

The structure of the A453V+D523N variant is nearly identical to wild type TnsE-CTD, with differences seen only in residues Gly419-Leu423 of the loop between strands β1 and β2, and arm 2 of the V-loop. The difference in residues Gly419-Leu423 only occurred in one of the two copies of TnsE-CTD$^{A453V+D523N}$ and is attributed to the different packing environment for this loop. In wild type crystals, this loop constitutes nearly half the interface between pseudodimer subunits, which occurs for all 12 ASU copies. In one monomer in the variant crystals, the segment is exposed to a solvent channel and the ordered atoms align to match the conformation of this loop in the wild type structure. In the other monomer, the sidechains of residues Arg420 and Arg422 are involved in crystal

contacts with two other monomers, which appear to influence the position of the local main chain.

Table 3.2: Data collection and refinement statistics for TnsE-CTD$^{A453V+D523N}$

Data Collection	
Wavelength	0.979
Cell: a, b, c (Å) / α, β, γ (°)	47.0, 71.3, 50.4 / 90, 97.3, 90
No. mol asu	2
Resolution (Å)	32.5-1.76 (1.82-1.76)
R_{meas}	0.038 (1.13)
I/σ(I)	20.5 (1.4)
Completeness (%)	99.8 (99.9)
Redundancy	3.8 (3.8)
Refinement	
Resolution (Å)	29.3 – 1.76
Completeness (%)	99.7
No. Reflections	32,706
R_{work} / R_{free} (%)	19.6 / 22.3
Atoms refined	2,650
Solvent atoms	202
Solvent content (%)	32.2
Rmsd in bonds (Å)	0.004
Rmsd in angles (°)	0.76
Mean B values (Å)	47.7

Data derived from a single crystal.
Data in the highest resolution shell is shown in parentheses.

The difference in arm 2 of the V-loop is that it only adopts conformation II in the TnsE-CTD$^{A453V+D523N}$ crystals. While the packing environment for each monomer is different in the variant crystals, the V-loop participates in no crystal contacts with other monomers, creating a local environment that matches wild type crystals. To confirm this,

the wild type structure was aligned with both monomers in the variant asymmetric unit and no clashes with symmetry mates appeared possible with conformation I of the V-loop. So, the ubiquity of conformation II in the TnsE-CTD$^{A453V+D523N}$ structure is not the result of crystal packing, but rather steric influence from the bulkier sidechain of variant residue Val453 (**Figure 3.9**).

Given the apparent role of V-loop arm 1 (residues Val451-Asp455) in regulating the conformational state of arm 2, an evolutionary precedent is set upon it to maintain its structural integrity. This requirement is partly accounted for by the burial of this arm in the hydrophobic core of the domain, but also by stabilization from a network of hydrogen bonds and a salt bridge. Most notably, the sidechain of Asp452 forms hydrogen bonds with Arg388 and Ser454, while Asp455 and Arg432 form a salt bridge (**Figure 3.9**).

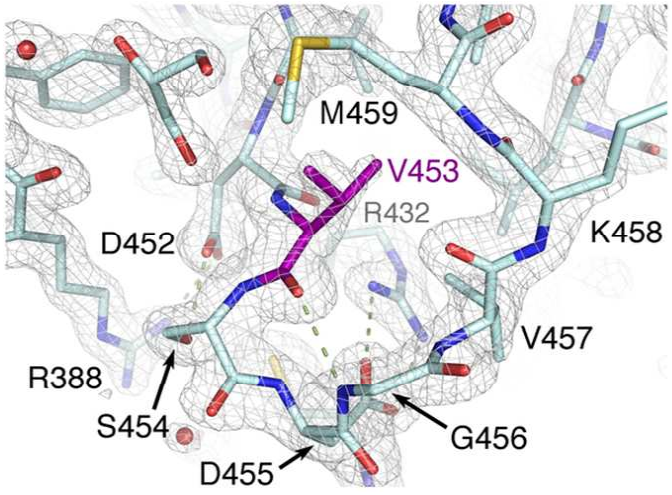

Figure 3.9: Conserved interactions stabilize arm 1 of the V-loop. Val453 (purple) in the TnsE-CTD A453V+D523N variant alters the influence of arm 1 over arm 2, locking arm 2 in conformation II. The perspective of this diagram has rotated ~60° to the left relative to the view in Figure 3.8.

Conservation of all these residues in TnsE homologs provides evidence for the importance of maintaining structural integrity in arm 1. Furthermore, the Peters group found that mutations disrupting the ability of residues Asp452 and Asp455 to form supporting interactions decreased transposition frequency in vivo (Shi et al., 2015). The D452A mutation greatly decreased transposition frequency while D452N did not, indicating that side chains with similar capacity for hydrogen bonding in that position also provide adequate structural support to the V-loop. Arg388, with which Asp452 hydrogen bonds, is often a lysine in TnsE homologs, showing that the primary requirement for structural support by the residue pair is the ability to form at least one hydrogen bond. Conversely, both the D455A and D455N variants, which introduce non-ionizable sidechains, disrupt the formation of the salt bridge with Arg432 and decrease transposition frequency equally. Both Asp455 and Arg432 are heavily conserved, even more so than the arm 1 residues forming hydrogen bonds, and the salt bridge itself is held in place by three hydrogen bonds with the main chain of residues Leu423, Cys426, and Lys427 in the extended loop joining strand β1 to β2.

It is clear from both the structural information and mutational analysis that the motifs determining the behaviour of the flexible arm 2 of the V-loop can have a significant impact on transposition frequency. This makes it initially surprising to observe that most TnsE homologs have a threonine at position 453, not an alanine as in the Tn7 element hosted by *E. coli*. If a threonine generates the same effect as the A453V mutation, it would seemingly be deleterious to the fitness of the transposable element, allowing it to insert too frequently and increase the chance of harming the host. Without

seeking structural information for a TnsE-CTD homolog, we had two alternate ways to explain the prevalence of the threonine allele. Firstly, by modeling a threonine residue into the position of Ala453, we saw that its side chain may stabilize conformation I of arm 2 by hydrogen bonding with its polypedtide backbone, as the non-polar valine variant could not. Secondly, the Peters group generated variants of *E. coli* TnsE containing the A453T amino acid change and tested their effect on transposition frequency in vivo (**Figure 3.10**).

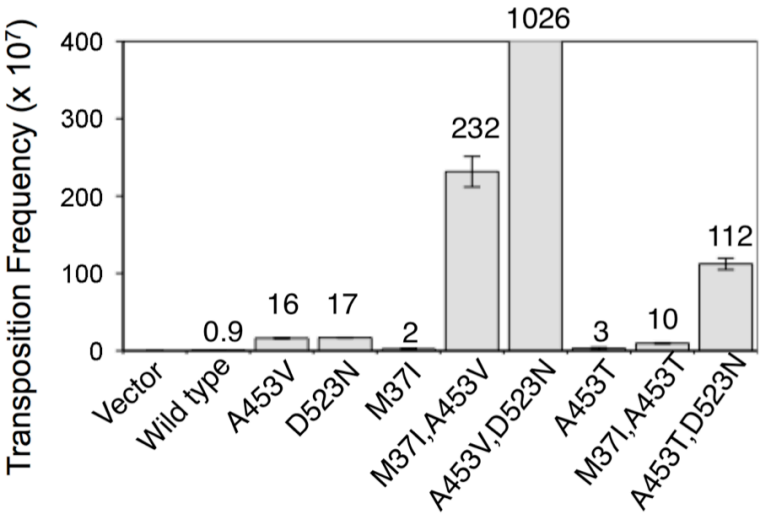

Figure 3.10: Transposition frequency associated with *E. coli* TnsE A453V and A453T double variants. Data were produced by the Peters Lab using the lambda hop assay (Shi et al., 2015). The mean transposition frequency is displayed above each plot and error bars indicate the standard error of the mean (n=3) (*extracted from Shi et al. (2015) NAR*).

The A453T mutation alone did not increase transposition frequency as much as the A453V mutation, but it was also able to produce larger increases when paired with secondary mutations expected to increase affinity fro target DNA (**Figure 3.10**). However, these increases are diminished in comparison to the tremendous upregulation of transposition frequency when secondary mutations are paired with the A453V amino acid change. This likely means that while TnsE homologs with the Thr453 allele may have similar activity to *E. coli* TnsE, they may also be more prone to changes in transposition

frequency by secondary mutations. The amino acid sequence of arm 2 is not as highly conserved as that of arm 1 (**Figure 3.5**), which could mean that TnsE homologs are better adapted to maintaining arm 1 mobility with a threonine at position 453 and therefore may not be so sensitive to secondary mutations.

3.8 A dual mechanism to regulate Tn7 transposition

Apart from Ala453, the other gain-of-activity amino acid variants identified in TnsE-CTD (G483R, G515R, E516K, E522K, D523N) occur in surface residues along one face of the domain (**Figure 3.11**). These changes all either add a positive charge, remove a negative charge, or both, suggesting that this face facilitates the interaction with target DNA. A defining characteristic of the face is a shallow, positively charged groove, which passes across the rotational axis of the V-loop's flexible arm. The conservation of wild type positively charged residues along this groove (Lys417, Arg420, Arg422, Lys486, Lys512, Arg530, and Arg533) further supports a DNA binding role for this region.

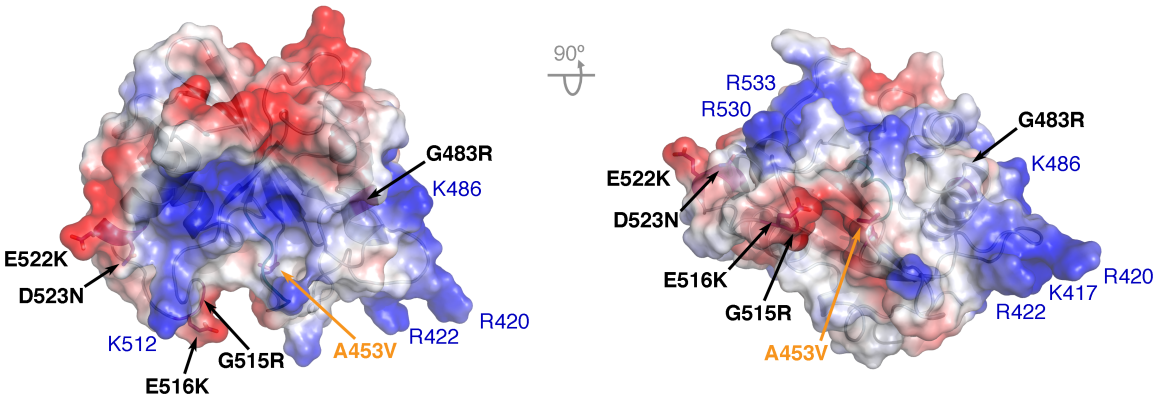

Figure 3.11: Electrostatic surface potential map of TnsE-CTD shows many gain-of-function residue variants clustering around a positively-charged surface (blue), which transverses the V-loop. Electronegative sidechains (red) surround this region. Black residue labels indicate gain-of-function residue variants. Blue labels indicate conserved electropositive residues.

All the surface residue changes on the putative DNA binding face, except for G483R, occur on one side of the V-loop, in the loop connecting strand β5 to helix α5 or within α5. From the transposition frequency increases of paired mutants (**Figure 1.2**), a trend exists in which mutations of residues on this side of the proposed DNA-interacting surface have a much more transposition-stimulating effect when paired with A453V. Consider that there is already a concentrated patch of positive charge around the position of G483R, while the surface around the sites of the other mutations is a mixture of positive and negative charges. This could lead G483R to produce a smaller change to the distribution of electrostatic potential on the DNA-interacting surface, leading to a lesser impact on affinity for target DNA. An alternative explanation, if the surface surrounding G483 were more significantly involved in the DNA binding function, is that mutations in this area tend to be lethal to the host and were therefore not discovered.

The difference observed between the two TnsE-CTD structures as well as the effect of A453V on other gain-of-activity point mutations suggests a mechanism in which excessive transposition is prevented by at least two aspects of the TnsE-CTD surface in the TnsABC+E pathway.

3.9 Crystallization of TnsE-NTD

While the structure of TnsE-CTD has produced considerable insight into how the target recognition activity of TnsE is regulated, it does not provides a complete picture in light of the finding that DNA binding depends on both domains. While the most complete mechanistic information would be obtained from a structure of full-length TnsE in

complex with target DNA, the structure of TnsE-NTD could provide some initial clues. Since the production capacity of TnsE-NTD was already at a scale suitable for crystallography, screening trials and subsequent condition optimizations were performed.

Initial crystals of TnsE-NTD were obtained in commercial high-throughput screens, but could not be reproduced on a larger scale in optimization experiments using freshly made crystallization solutions. A fungus was found to have contaminated the commercial solution that produced the original crystal hit and may have secreted proteases, which could have degraded flexible loops or tails preventing native TnsE-NTD from forming ordered contacts. Re-screening of crystallization conditions including trace amounts of papain for controlled in-situ proteolysis uncovered a new condition [0.2 M $(NH_4)_2SO_4$, 25% PEG-3350 (w/v), 0.1 M Bis-Tris pH 6.5], which yielded crystals that could be reproduced unreliably in scaled-up optimization experiments. In the optimizations, TnsE-NTD initially formed a light precipitate, producing crystals after three weeks at 4 °C. The productive optimization condition varied somewhat [0.1 M $(NH_4)_2SO_4$, 25% PEG-3350 (w/v), 0.1 M Bis-Tris pH 6.5], and only generated crystals when the protein storage buffer was altered [0.1 M NaCl, 0.05 M $MgCl_2$, 0.1 mM EDTA, 5 mM DTT, 5% Glycerol (v/v), 0.02 mM Tris pH 8]. The optimization crystals also formed as clusters of plates and were fairly small (<50 μm). As a control, papain was omitted from a set of optimization experiments, which yielded no crystals. While papain itself is apt to crystallize, the concentration used was far lower than one that could produce crystals ($1/1000^{th}$ the amount of TnsE-NTD by mass in screens, $1/500^{th}$ in productive optimizations).

Dozens of these crystals were frozen for diffraction analysis, however most diffracted very weakly (>15 Å) or contained multiple lattices and still diffracted somewhat weakly (>6 Å). This could have been the result of damage caused by the handling required to separate individual crystals from clusters. One of the crystals separated from a cluster diffracted very well in comparison to the others (**Table 3.3**). Under the microscope, it had greater depth than any of the others and it was the only one screened that had the diffraction power to produce meaningful data. A full dataset was collected from this crystal at the Canadian Light Source on beamline 08B1 (Fodje et al., 2014).

Since this was the only exploitable sample we had for this domain, a wavelength of 1.6102 Å was used for the experiment in an attempt to measure anomalous scattering from native sulfur atoms to solve the structure. The data collected extends to an appropriate resolution and completeness for anomalous phasing, however several other important factors were lacking in quality. The redundancy and overall I/σI, while sufficient for molecular replacement, are too low to provide a reliable measure of anomalous differences, especially with a native sample. After scaling the data, it was analyzed with PHENIX.XTRIAGE (Adams et al., 2010), which indicated insufficient anomalous signal/noise in all resolution shells. In disregard of statistical expectations, sulfur-SAD phasing was attempted with PHENIX.AUTOSOL (Adams et al., 2010), but an interpretable solution could not be obtained.

Table 3.3 Data collection and processing statistics for TnsE-NTD.

Data Collection	
Wavelength (Å)	1.610
Space Group	C2
a, b, c (Å) / α, β, γ (°)	99.2, 44.9, 147.3 / 90, 106.7, 90
Resolution (Å)	47.52 – 1.95 (2.02 – 1.95)
R_{meas}	0.237 (1.745)
I/σ(I)	7.7 (1.0)
Completeness (%)	99.7 (99.6)
Redundancy	3.6 (3.5)
Total Reflections	166266
Unique Reflections	45636
Mosaicity	0.126

Data derived from a single crystal.
Data in the highest resolution shell is shown in parentheses.

Chapter 4

Characterization of the TnsE+DNA complex

4.1 Production of full-length TnsE$^{A453V+D523N}$

A trend has been established in which gain-of-activity variants of TnsE display a greater affinity for target DNA and a greater ability to form higher order protein-DNA complexes than wild type TnsE (Peters and Craig, 2001). While some insight into the regulation of TnsE may be drawn from structural investigation of the individual domains, a more complete regulatory model could be constructed from the structure of full-length TnsE in complex with a target DNA. The double mutant TnsE$^{A453V+D523N}$ forms a more stable interaction with DNA than wild type TnsE (Shi et al., 2015), and therefore has greater potential to provide homogeneous complexes in the search for initial crystals.

Full-length TnsE and the TnsE$^{A453V+D523N}$ variant had previously been prone to precipitation during purifications. Following elution from a metal affinity chromatography (MAC) column, the eluate was prepared for anion exchange chromatography by diluting it to a lower the ionic strength, at which point the majority of the sample was often lost to precipitation. In the interest of improving purification yields and product stability, the purification protocol was optimized around this problem. Initial

work focused on bypassing anion exchange chromatography to avoid the dilution step, since protein of apparently acceptable purity was obtained following MAC alone. The

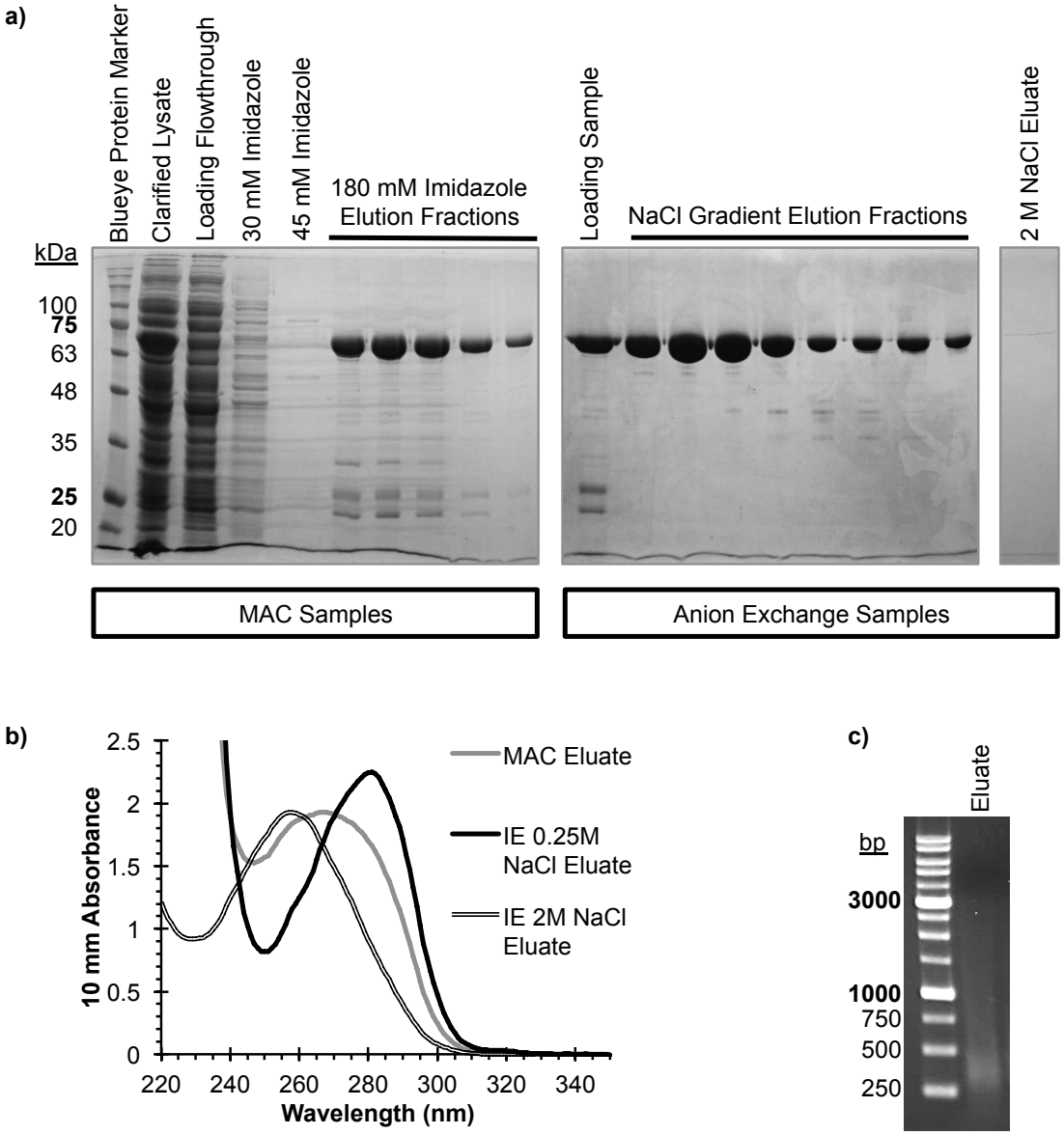

Figure 4.1: Purification of full-length TnsE. **(A)** SDS-PAGE shows that the TnsE$^{A453V+D523N}$ sample obtained from the original purification protocol is very pure (anion exchange eluate), but also that the contaminant peak eluted from anion exchange resin in 2M NaCl does not contain protein. **(B)** UV absorbance spectra of purification samples show a shift in maximum absorbance between peaks eluted from anion exchange resin in 0.25 M NaCl (280 nm) and in 2 M NaCl (260 nm). **(C)** Agarose gel electrophoresis followed by staining with ethidium bromide confirms the presence of nucleic acid in a size range appropriate for sonicated samples.

MAC eluate was exchanged to a buffer lacking imidazole, but of equivalent ionic strength (0.5 M NaCl) on a desalting column. This produced a temporarily stable sample that aggregated and precipitated while concentrating.

Even with the precipitation following MAC, the remaining soluble TnsE that could be purified by anion exchange remained stable in a lower ionic strength environment (150 mM NaCl) even at high concentrations. Following protein elution from the anion exchange column, a large absorbance peak (280 nm) eluted when the column was washed with a 2 M NaCl solution. While this fraction was presumed to be a protein contaminant in earlier purifications, UV spectrophotometry and gel electrophoresis suggested it is primarily nucleic acid (**Figure 4.1**). Since the contaminant remained bound to the anion exchange resin in buffer with the ionic strength used for MAC (0.5 M NaCl), a test was performed to determine whether it could be removed immediately following elution from the MAC column. An anion exchange column was equilibrated with MAC buffer and then connected to follow the MAC column at the time of elution. The resulting eluate did not precipitate when diluted to lower ionic strength for subsequent ion exchange chromatography. This step was integrated into the routine purification protocol for full-length TnsE, and eliminated precipitation following MAC in all following purifications.

4.2 Identification of a minimal substrate for TnsE$^{A453V+D523N}$

To crystallize TnsE$^{A453V+D523N}$ with DNA, it is important to determine the characteristics of a DNA substrate suitable for complete binding and for crystal packing. Some of the

specific DNA characteristics required for crystallization of a complex may only be determined experimentally in crystallization trials, however an initial DNA substrate search was performed to narrow the permutations required in subsequent crystallization experiments. This principle tenet guiding this search was as follows: the ideal DNA substrate should be minimal in length to prevent flexible protrusions from interfering with crystal packing, but must also present a complete interaction surface to bind TnsE with enough occupancy to produce a homogeneous population of complexes. Here, the nomenclature for DNA substrates with a 3′-recessed end follows the convention such that one labeled 15+10 would have 15 nucleotides in the duplex region and 10 nucleotides in the single-stranded region.

Initially, EMSAs indicated that TnsE shifts a 15+15 DNA duplex at lower concentration than duplexes with shorter single-stranded regions (**Figure 4.2**). In these conditions, the $TnsE^{A453V+D523N}$ variant did not shift a greater proportion of DNA than $TnsE^{WT}$ at equivalent protein concentrations, but did form more discrete shifted bands consistent with two stable complex states. In the gel for 15+15 DNA substrate, a secondary DNA band was resolved from the annealed duplex and was believed to be a single-stranded portion of the labeled 15-mer oligonucleotide used. This portion was likely also present with the shorter duplexes, though its mobility in the gel could have better matched the longer, but more charge-dense duplex DNA. The secondary single-stranded DNA was shifted at high TnsE concentrations (64x) across all experiments, but only after the DNA duplex had been completely shifted. This occurred in concert with the accumulation of labeled DNA remaining in the loading well. This indicates that TnsE

has the ability to bind non-specifically to alternate DNA structures in a manner that affects its stability, however the 3′-recessed duplex was still shifted preferentially by both $TnsE^{WT}$ and $TnsE^{A453V+D523N}$. Testing shorter double-stranded regions with this method was not viable due to the melting temperature of such DNA duplexes falling below the temperature of the experiment.

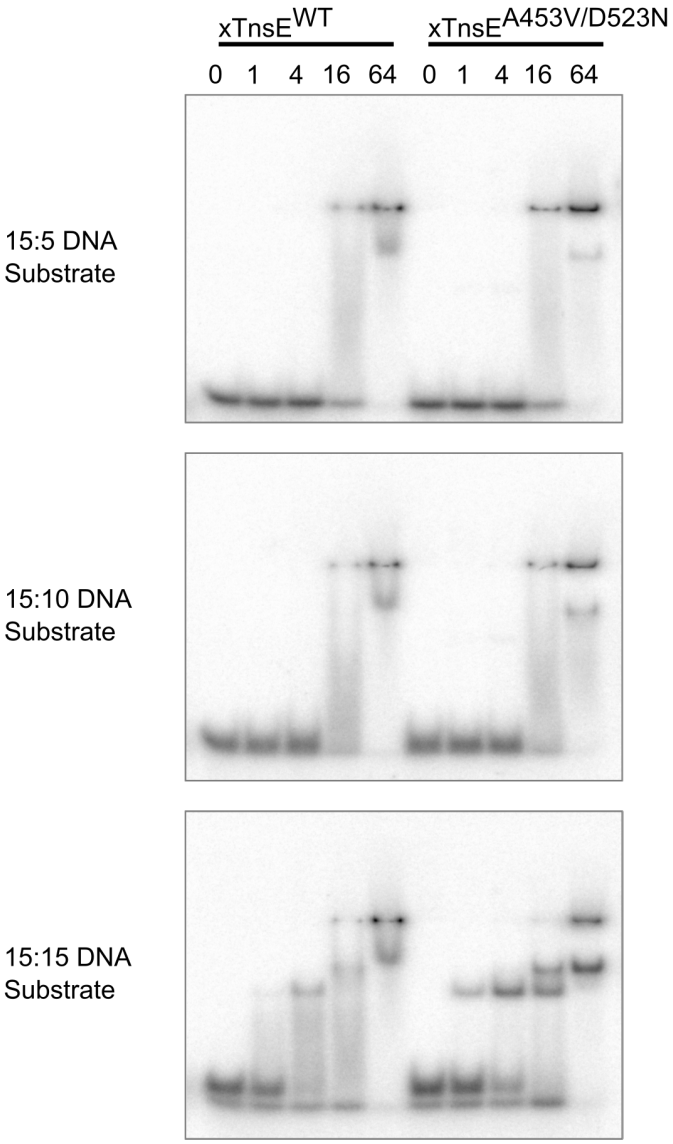

Figure 4.2: EMSA to characterize a minimal DNA substrate for TnsE. EMSA experiments identify a limit between 15 and 10 nucleotides in the length of the single-stranded region in 3′-recessed DNA duplexes that can be effectively shifted by TnsE and $TnsE^{A453V+D523N}$. For the most effective 15+15 DNA substrate, $TnsE^{A453V+D523N}$ formed more discrete shifted complexes than TnsE. The most-shifted bands correspond to labeled substrate retained in the loading well.

The observations that TnsE shifted a secondary labeled DNA substrate and formed an immobile nucleoprotein structure in the EMSA experiments raised concerns about the uniformity and stability of the nucleoprotein complex at higher concentrations that would be required for crystallization. An 80 μM complex of $TnsE^{A453V+D523N}$ with the current shortest effective 3′-recessed DNA duplex was prepared and evaluated for these characteristics with size exclusion chromatography (SEC). The complex eluted as a single peak at a run volume indicating a greater size than either component alone along with smaller secondary peaks consistent with unbound $TnsE^{A453V+D523N}$ and DNA (**Figure 4.3**).

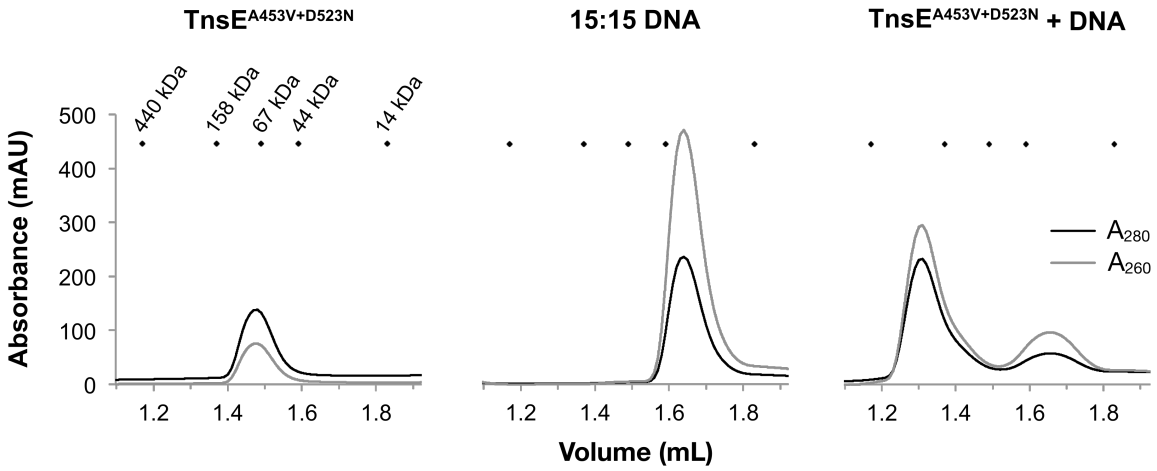

Figure 4.3: SEC of $TnsE^{A453V+D523N}$ with 3′-recessed DNA indicates the formation of a uniform and stable nucleoprotein complex at a viable crystallographic concentration. A peak shift to a lower elution volume for the mixed sample indicates complex formation while the smaller secondary peaks are consistent with a remainder of unbound components. No significant absorbance concurred with elution of the void volume (0.80 mL, not shown), precluding the formation of a persistent aggregate in these conditions.

No significant source of UV absorbance eluted with the void volume of the column at 0.80 ml (not shown). Together, this indicated the predominance of a stable and uniform complex not prone to aggregation as would typically be required for

crystallization. The SEC apparatus was calibrated with globular proteins of known molecular weight and so the molecular weight of the predominant species in each of the samples was estimated (**Table 4.1**).

Table 4.1 Analysis of the SEC data for TnsE$^{A453V+D523N}$ with 3′-recessed DNA

Sample	Elution Volume (mL)	Estimated MW (kDa)	Expected MW (kDa)
TnsE$^{A453V+D523N}$	1.48	88	62.3
DNA	1.64	38	13.8
TnsE$^{A453V+D523N}$ + DNA	1.31	215	76.0

The measured molecular weights of TnsE alone and the DNA substrate alone are somewhat greater than expected, however, the elution volume depends both on the size and shape of the sample and neither TnsE nor the DNA duplex have the shape of a single globular domain. The estimated molecular weight of the complex was several-fold greater than expected, suggesting either that the complex is extended in shape or that a higher order species predominated.

A preliminary differential scanning fluorimetry (DSF) experiment showed that the presence of a 15+15 DNA duplex increases the melting temperature of TnsE$^{A453V+D523N}$ by more than 7 °C (not shown) while maintaining a unimodal melting curve. To establish a range of minimal effective lengths of DNA substrates that form a thermally stable complex with TnsE$^{A453V+D523N}$, triplicate DSF experiments were used to screen more finely for minimal 3′-recessed DNA substrates with each permutation of double-stranded and single-stranded regions having 15, 12, or 10 nucleotides. It was previously shown with DNA shift assays that wild type TnsE as well as A453V+G515R and

A453V+G483N variants interact preferentially with dsDNA over ssDNA (Peters and Craig, 2001). So, for each DNA substrate a parallel set of replicates was included with an excess of 30+0 dsDNA to test the thermal stability of complexes formed in the presence of a non-specific binding partner. Sets with only 30+0 (dsDNA) and 0+30 (ssDNA) were also included to establish a baseline for stabilization by DNA structures to which TnsE is not known to bind specifically.

Of note in the controls, $TnsE^{A453V+D523N}$ was stabilized more by the ssDNA than by dsDNA control substrate (**Figure 4.4**). While the oligonucleotide used for ssDNA was not predicted to have potential for hairpin formation and self-complementation, it is possible that a transient structure could have been stabilized upon binding $TnsE^{A453V+D523N}$. Alternatively, this result indicates either a different DNA structure preference in the A453V+D523N variant or an inherent difference between the meaning of DSF and shift assay results in evaluating DNA binding preferences.

Of the DNA substrates tested, only 15+15 stabilized $TnsE^{A453V+D523N}$ significantly relative to an excess of the double-stranded DNA. Shortening either the double-stranded or single-stranded arm of the DNA from 15 to 12 or 10 nucleotides resulted in a loss of specific stabilization. A much larger 30+35 DNA duplex was included as a positive control, and was found to have a greater stabilizing effect than 15+15. This indicates that $TnsE^{A453V+D523N}$ affinity for DNA substrates within a certain size range decreases with the lengths of the single-stranded and double-stranded arms until a minimal threshold is reached. In the context of substrates bound more tightly than background double-stranded DNA, 15+15 remains the approximate minimal DNA substrate. Further

refinement of a candidate minimal DNA substrate through these methods was not warranted since the unpredictable nature of crystallization would have demanded screening an array of similar substrates regardless.

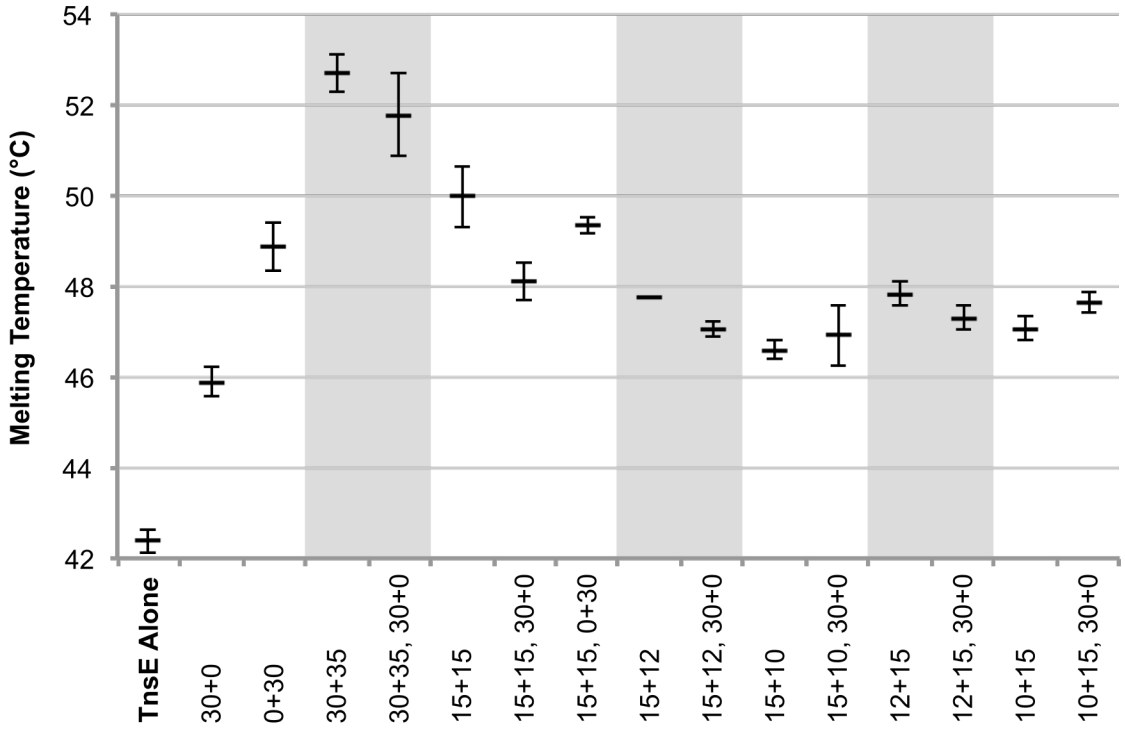

Figure 4.4: DSF refines the properties of a minimal DNA substrate for TnsE$^{A453V+D523N}$, requiring at least 15+15 nucleotides for stable binding. DSF melting temperatures were measured for 5 µM TnsE$^{A453V+D523N}$ in the presence of 7.5 µM candidate minimal DNA substrates derived from the DNA duplex formed by oligonucleotides NLC707 and NLC710. 30+0 and 0+30 are respectively 15 µM dsDNA and ssDNA controls. Error bars indicate 2 standard deviations from the mean. For TnsE$^{A453V+D523N}$ alone, n = 6, and for TnsE$^{A453V+D523N}$ with DNA substrates, n = 3.

4.3 Crystallization trials of TnsE$^{A453V+D523N}$ bound to DNA

Early attempts to crystallize TnsE$^{A453V+D523N}$ in complex with the 15+15 DNA substrate yielded no crystals in high-throughput screens, prompting the consideration of different DNA substrates. Since shift assays, SEC, and DSF identified the 15+15 DNA duplex as the minimal resolved DNA substrate and both the 15+12 and 12+15 substrates as sub-minimal, crystallization trials were designed with the objective of identifying a 3′-recessed DNA substrate with ideal properties for crystal nucleation and growth. Under the premise that an ideal substrate would be as short as possible to prevent flexible protruding DNA from occupying sites of crystal contacts, but of sufficient length to form a homogeneous population of stable complexes, all 16 permutations of DNA substrates with both the double-stranded and single-stranded regions ranging from 13 to 16 nucleotides in length were designed to be tested. Nucleoprotein complexes with about one third of the shortest substrates in this pool were each screened for crystallization across 384 conditions from contemporary commercial kits, but none had yielded crystals within several months. A new graduate student in the Guarné laboratory is currently pursuing this project.

Chapter 5

Conclusion

5.1 Future directions – Target recognition

Using DSF to assay candidate 3′-recessed DNA substrates with and without a background of double stranded DNA (dsDNA), we have a method to determine whether stabilization of TnsE is due to a specific interaction with the substrate. Even background dsDNA alone with TnsE generates a temperature shift. Binding drops off when either arm is shortened from 15 to 12 nucleotides, and even with 15 nucleotides in both DNA arms, stabilization is not as pronounced as with the much larger 35+30 DNA. This suggests that the true minimal substrate may lie with both arm lengths in the intermediate range. Since the difference persists even in the presence of excess background dsDNA, it is attributed to a stronger specific interaction with the larger 3′-recessed DNA, indicating that even the 15+15 substrate may be too small to capture the full interaction. To identify the truly best substrate size for crystallization, high-throughput crystallization experiments with 16 substrates with both arms ranging in size from 13 to 16 nucleotides are underway. If this approach does not yield crystals, it may be worth considering other ways to modify the complex for crystallization.

A large region in the middle of TnsE's primary sequence is thought to form a flexible linker separating the two domains. While some success toward the crystallization of TnsE-NTD has been attained using in-situ proteolysis, this may be a good technique to remove flexible regions in the complex to promote crystallization. Before beginning crystallization experiments, it would be important to determine whether the quaternary structure of the complex tolerates proteolysis. This could be tested in a controlled setting by performing limited proteolysis experiments to determine the extent of digestion within TnsE. Analytical gel filtration chromatography will test whether the complex remains intact after proteolysis. For this, papain is a good candidate protease since it has a broad range in cut sites and has provided some success in the crystallization of TnsE-NTD. Alternatively, portions of the flexible linker region between TnsE's domains may be removed by site-directed mutagenesis to limit space it occupies. Before starting crystallization trials, deletion mutants would be assessed for DNA-binding activity by EMSA and for in vivo functionality using a papillation assay, which provides a colony-based measure of transposition frequency (Stellwagen and Craig, 1997).

5.2 Future Directions – Transposase activation

TnsE directs Tn7 integration to sites of discontinuous replication by communicating with the TnsA/B transposase via the ATPase regulator TnsC, however the mechanism of this communication is not known (Peters, 2015). In the TnsABC+D pathway, TnsC displays an ATP-dependent interaction with target DNA distorted by TnsD, but also binds TnsD, resulting in a two-factor authentication mechanism for activation of the transposition

reaction (Choi et al., 2014). Does TnsE stimulate transposition in a similar manner to TnsD, by providing a binding site for TnsC on target DNA and verifying the location through a protein-protein interaction?

To test whether TnsE induces a similar distortion in target DNA to attract TnsC, one may apply footprinting assays similar to those that identified the distortion induced by TnsD (Kuduvalli et al., 2001) if TnsE distorts the double-stranded region of its recognition site. The dsDNA region protected by TnsE would be identified by hydroxyl radical footprinting, missing nucleoside analysis, and dimethyl sulfate footprinting, while distorted dsDNA would be identified by hypercleavage analysis of TnsE-bound DNA using potassium permanganate and 5-phenyl-1,10-phenanthroline copper. TnsE binds target DNA in a semi-stable manner, however TnsE variants ($TnsE^{A453V+D523N}$ and $TnsE^{E516K+D523N}$) have greatly increased binding stability and still direct insertion to TnsE's natural target type (Shi et al., 2015). For these reasons, such variants would provide the clearest footprinting and hypercleavage analyses, while maintaining biological meaning.

Does TnsE mimic TnsD's direct interaction with TnsC? A TnsC variant has been identified ($TnsC^{A225V}$) with decreased ATPase activity and the ability to activate Tn7 transposition into random targets at low frequency in the absence of both TnsD and TnsE targeting proteins (Stellwagen and Craig, 1997). TnsE stimulates transposition controlled by $TnsC^{A225V}$, but does not influence target selection, suggesting it can license TnsC to activate transposition even without recognizing a target site (Shi et al., 2015). These findings could be explained by a direct interaction between TnsC and TnsE, however this

interaction has not yet been identified. We would test for this interaction biochemically using a yeast two-hybrid assays and chromatography co-elution assays with permutations of full-length and soluble fragments of both TnsC and TnsE. If a direct interaction is not identified, we would attempt to gather structural information by co-crystallization of soluble protein fragments with statistical analysis of contact surfaces and by cryo-electron microscopy of full-length proteins bound to a DNA substrate.

5.3 Impact of this work

Before this work, interested groups had constructed an extensive body of information while investigating TnsE's mechanism in targeting Tn7 transposition, however no structural information was available to rationalize their observations. By solving the crystal structure of the C-terminal domain of TnsE, where gain-of-function mutations tend to appear, we were able to construct a model explaining the phenotypes observed in variants. In addition to explaining the biological mechanism of TnsE, TnsE-CTD was also adopted a novel protein fold, expanding the scientific knowledge base of possible tertiary structures and structures that can be used for binding DNA.

Gain-of-function variants that increase transposition frequency tend to also increase target DNA binding stability and lead to the greatest increases in transposition frequency when the residue variant A453V is paired with other single residue variants. A flexible loop adjacent to residue 453 was identified connecting TnsE-CTD β strands 3 and 4, and was observed to be rigid in an A453V variant. This finding led to a model of regulation in TnsE where this loop can alternate between an 'active' and 'inactive' state,

likely giving TnsE the ability to browse many potential targets while limiting its ability to remain at unfavourable sites. Other gain-of-function variants either introduced a positive charge or removed a negative charge from the periphery of a positively charged surface surrounding the flexible loop. The increased DNA-binding stability of these variants as well as the nature of the surface suggests it mediates the interaction with target DNA. However, we also observed in electrophoretic mobility shift assays that neither TnsE-CTD nor the N-terminus region alone is sufficient to bind DNA, with stable interactions occurring only with full-length TnsE. We are left wondering how the two TnsE domains work together to bind a DNA target and how TnsE communicates with the transposase assembly through TnsC. In summary, while this work provided key insight into TnsE's mechanism of target selection, it also opened the door to the next layer of questions.

References

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Afonine, P.V., Ralf W. Grosse-Kunstleve, and Adams, P.D. The Phenix refinement framework. CCP4 Newsl. 1–7.

Arciszewska, L.K., Drake, D., and Craig, N.L. (1989). Transposon Tn7: cis-acting sequences in transposition and transpostion immunity. J. Mol. Biol. *207*, 35–52.

Bainton, R.J., Kubo, K.M., Feng, J.N., and Craig, N.L. (1993). Tn7 transposition: target DNA recognition is mediated by multiple Tn7-encoded proteins in a purified in vitro system. Cell *72*, 931–943.

Boivin, S., Kozak, S., and Meijers, R. (2013). Optimization of protein purification and characterization using Thermofluor screens. Protein Expr. Purif. *91*, 192–206.

Buchan, D.W.A., Minneci, F., Nugent, T.C.O., Bryson, K., and Jones, D.T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res. *41*, W349–W357.

Cabezón, E., Ripoll-Rozada, J., Peña, A., de la Cruz, F., and Arechaga, I. (2014). Towards an integrated model of bacterial conjugation. FEMS Microbiol. Rev.

Choi, K.Y., Spencer, J.M., and Craig, N.L. (2014). The Tn7 transposition regulator TnsC interacts with the transposase subunit TnsB and target selector TnsD. Proc. Natl. Acad. Sci.

Curcio, M.J., and Derbyshire, K.M. (2003). The outs and ins of transposition: from mu to kangaroo. Nat. Rev. Mol. Cell Biol. *4*, 865–877.

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

Ewing, A.D., Ballinger, T.J., Earl, D., Broad Institute Genome Sequencing and Analysis Program and Platform, Harris, C.C., Ding, L., Wilson, R.K., and Haussler, D. (2013). Retrotransposition of gene transcripts leads to structural variation in mammalian genomes. Genome Biol. *14*, R22.

Finn, J.A., Parks, A.R., and Peters, J.E. (2007). Transposon Tn7 Directs Transposition into the Genome of Filamentous Bacteriophage M13 Using the Element-Encoded TnsE Protein. J. Bacteriol. *189*, 9122–9125.

Fodje, M., Grochulski, P., Janzen, K., Labiuk, S., Gorin, J., and Berg, R. (2014). 08B1-1: an automated beamline for macromolecular crystallography experiments at the Canadian Light Source. J. Synchrotron Radiat. *21*, 633–637.

Hendrickson, W.A., Horton, J.R., and LeMaster, D.M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO J. *9*, 1665–1672.

Heras, B., and Martin, J.L. (2005). Post-crystallization treatments for improving diffraction quality of protein crystals. Acta Crystallogr. D Biol. Crystallogr. *61*, 1173–1180.

Hickman, A.B., Li, Y., Mathew, S.V., May, E.W., Craig, N.L., and Dyda, F. (2000). Unexpected structural diversity in DNA recombination: the restriction endonuclease connection. Mol. Cell *5*, 1025–1034.

Holder, J.W., and Craig, N.L. (2010). Architecture of the Tn7 posttransposition complex: an elaborate nucleoprotein structure. J. Mol. Biol. *401*, 167–181.

Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Res. *38*, W545–W549.

Kirk, R., Laman, H., Knowles, P.P., Murray-Rust, J., Lomonosov, M., Meziane, E.K., and McDonald, N.Q. (2008). Structure of a Conserved Dimerization Domain within the F-box Protein Fbxo7 and the PI31 Proteasome Inhibitor. J. Biol. Chem. *283*, 22325–22335.

Krissinel, E., and Henrick, K. (2007). Inference of Macromolecular Assemblies from Crystalline State. J. Mol. Biol. *372*, 774–797.

Kuduvalli, P.N., Rao, J.E., and Craig, N.L. (2001). Target DNA structure plays a critical role in Tn7 transposition. EMBO J. *20*, 924–932.

Levin, H.L., and Moran, J.V. (2011). Dynamic interactions between transposable elements and their hosts. Nat. Rev. Genet. *12*, 615–627.

Li, Z., Craig, N., and J E Peters (2013). Chapter 1: Transposon Tn7. In Bacterial Integrative Mobile Genetic Elements, (Landes Bioscience),.

Liao, Y.-D., Jeng, J.-C., Wang, C.-F., Wang, S.-C., and Chang, S.-T. (2004). Removal of N-terminal methionine from recombinant proteins by engineered E. coli methionine aminopeptidase. Protein Sci. Publ. Protein Soc. *13*, 1802–1810.

Liu, Q., Zhang, Z., and Hendrickson, W.A. (2011). Multi-crystal anomalous diffraction for low-resolution macromolecular phasing. Acta Crystallogr. D Biol. Crystallogr. *67*, 45–59.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

McKown, R.L., Orle, K.A., Chen, T., and Craig, N.L. (1988). Sequence requirements of Escherichia coli attTn7, a specific site of transposon Tn7 insertion. J. Bacteriol. *170*, 352–358.

Mitra, R., McKenzie, G.J., Yi, L., Lee, C.A., and Craig, N.L. (2010). Characterization of the TnsD-attTn7 complex that promotes site-specific insertion of Tn7. Mob. DNA *1*, 18.

Nagy, Z., and Chandler, M. (2004). Regulation of transposition in bacteria. Res. Microbiol. *155*, 387–398.

Parks, A.R., and Peters, J.E. (2007). Transposon Tn7 Is Widespread in Diverse Bacteria and Forms Genomic Islands. J. Bacteriol. *189*, 2170–2173.

Parks, A.R., and Peters, J.E. (2009). Tn7 elements: Engendering diversity from chromosomes to episomes. Plasmid *61*, 1–14.

Parks, A.R., Li, Z., Shi, Q., Owens, R.M., Jin, M.M., and Peters, J.E. (2009). Transposition into replicating DNA occurs through interaction with the processivity factor. Cell *138*, 685–695.

Perry, J.A., and Wright, G.D. (2013). The antibiotic resistance "mobilome": searching for the link between environment and clinic. Front. Microbiol. *4*.

Peters, J.E. (2015). Tn7. In Mobile DNA III, A.M. Lambowitz, M. Gellert, M. Chandler, N.L. Craig, S.B. Sandmeyer, and P.A. Rice, eds. (American Society of Microbiology), pp. 647–667.

Peters, J.E., and Craig, N.L. (2000). Tn7 transposes proximal to DNA double-strand breaks and into regions where chromosomal DNA replication terminates. Mol. Cell *6*, 573–582.

Peters, J.E., and Craig, N.L. (2001). Tn7 recognizes transposition target structures associated with DNA replication using the DNA-binding protein TnsE. Genes Dev. *15*, 737–747.

Peters, J.E., Fricker, A.D., Kapili, B.J., and Petassi, M.T. (2014). Heteromeric transposase elements: generators of genomic islands across diverse bacteria. Mol. Microbiol.

Richardson, J.M., Colloms, S.D., Finnegan, D.J., and Walkinshaw, M.D. (2009). Molecular architecture of the Mos1 paired-end complex: the structural basis of DNA transposition in a eukaryote. Cell *138*, 1096–1108.

Sarnovsky, R.J., May, E.W., and Craig, N.L. (1996). The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. EMBO J. *15*, 6348–6361.

Shi, Q., Straus, M.R., Caron, J.J., Wang, H., Chung, Y.S., Guarné, A., and Peters, J.E. (2015). Conformational toggling controls target site choice for the heteromeric transposase element Tn7. Nucleic Acids Res.

Skelding, Z., Sarnovsky, R., and Craig, N.L. (2002). Formation of a nucleoprotein complex containing Tn7 and its target DNA regulates transposition initiation. EMBO J. *21*, 3494–3504.

Snyder, L., Henkin, T.M., Peters, J.E., and Champness, W. (2013). Molecular Genetics of Bacteria, 4th Edition (American Society of Microbiology).

Spiess, A.-N., Feig, C., and Ritz, C. (2008). Highly accurate sigmoidal fitting of real-time PCR data by introducing a parameter for asymmetry. BMC Bioinformatics *9*, 221.

Stellwagen, A.E., and Craig, N.L. (1997). Gain-of-function mutations in TnsC, an ATP-dependent transposition protein that activates the bacterial transposon Tn7. Genetics *145*, 573–585.

Stellwagen, A.E., and Craig, N.L. (1998). Mobile DNA elements: controlling transposition with ATP-dependent molecular switches. Trends Biochem. Sci. *23*, 486–490.

Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Zwart, P.H., Hung, L.-W., Read, R.J., and Adams, P.D. (2008). Iterative model building, structure refinement and density modification with the *PHENIX AutoBuild* wizard. Acta Crystallogr. D Biol. Crystallogr. *64*, 61–69.

Waddell, C.S., and Craig, N.L. (1988). Tn7 transposition: two transposition pathways directed by five Tn7-encoded genes. Genes Dev. *2*, 137–149.

Wolfram|Alpha. 2016. Wolfram Alpha LLC.
https://www.wolframalpha.com/input/?i=(a+%2F+((1+-+e%5E(-b(x-g)))%5Ed))%2Bt
(accessed February 11, 2016)

Wolkow, C.A., DeBoy, R.T., and Craig, N.L. (1996). Conjugating plasmids are preferred targets for Tn7. Genes Dev. *10*, 2145–2157.

Yang, W., Lee, J.Y., and Nowotny, M. (2006). Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity. Mol. Cell *22*, 5–13.

Zbigniew Otwinowski, and Wladek Minor (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. In Methods in Enzymology, (Academic Press (New York)), pp. 307–326.