Transposon Tn7 Protein TnsD Binding to *Escherichia coli* attTn7 DNA and Its Eukaryotic Orthologs†

Atis Chakrabarti, Priyanka Desai,‡ and Eric Wickstrom*

Department of Biochemistry and Molecular Pharmacology, Department of Microbiology and Immunology, Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, Pennsylvania 19107

Received August 27, 2003; Revised Manuscript Received January 9, 2004

ABSTRACT: Transposon Tn7 inserts itself into the attTn7 target DNA sequence at the 3′ end of the *Escherichia coli glmS* gene with high specificity and efficiency. This site in the *E. coli* genome displays amino acid conservation and nucleotide similarity with orthologous sequences in Archaebacteria and eukaryotes. On the basis of the high degree of nucleotide similarity, particularly with eukaryotes, we examined the interactions of a set of 20-bp duplex DNA sequences with the Tn7 protein TnsD. The protein was overexpressed in the IPTG-inducible vector pET14b-TnsD in *E. coli* BL21(DE3)-RIL-Codon-Plus, and purified by nickel chelation and ion exchange chromatography. Changes in the conformation of DNA duplexes upon interaction with TnsD were monitored by circular dichroism (CD) spectroscopy. TnsD binding to and dissociation from immobilized DNA duplexes were monitored by total internal reflectance (TIR). CD and TIR results were analyzed to calculate k_{on}, k_{off}, and K_D values. The 20-bp DNA duplex corresponding to attTn7 at the 3′ end of *E. coli glmS* displayed strong affinity for TnsD protein, with $K_D \approx 20$ nM. Eukaryotic orthologs of attTn7 from yeast and mammalian *GFPT1* displayed lower affinity, with $K_D \approx 500$ nM. Mutant DNA sequences with a single central mismatch did not display any detectable interaction with TnsD. The physical studies validate our biological observation of Tn7 transposition into a plasmid containing the mammalian attTn7 ortholog sequence [Cleaver, S. H., and Wickstrom, E. (2000) *Gene 254*, 37−44], and suggest that 1−2 amino acid substitutions in TnsD might be sufficient to permit binding to mammalian orthologs that is as strong as wild-type TnsD binding to attTn7.

Transposons are mobile genetic elements that have the ability to translocate to a variety of sites on both chromosomal and extrachromosomal DNA. Transposon Tn7 is unique among transposons due to its high efficiency and high specificity of insertion in *Escherichia coli* (*1*). TnsA, TnsB, and TnsC proteins compose the core Tn7 transposase responsible for the transposition event, independently of target site selected. TnsD provides recognition of a specific target sequence in *E. coli* called attTn7, while TnsE protein, when present, facilitates nonspecific insertion (*1*).

The 63-bp attTn7 recognition−insertion domain at the 3′ end of *E. coli* glutamine synthase *glmS* gene (*2*) that is recognized by TnsD protein displays significant similarity to the corresponding 3′ termini of the glutamine:fructose-6-phosphate transaminase (*GFPT*) genes of the thermophilic archaebacterium *Pyrococcus horikoshii* (*3*), yeast *Saccharomyces cerevisiae*, on chromosome XI (*4*), of mice, on chromosome 6 (*5*), and of humans (*6*), located at 2p13 (*7*) (Figure 1). The 63 bp of attTn7 at the 3′ terminus of *glmS* and the corresponding 63 bp of *Homo sapiens GFPT1* share 56% identity (35/63), compared with 49% identity (31/63) with the same 63 bp of *P. horikoshii GFPT*. In a narrower range, the 41-bp TnsD binding zone of attTn7 displays 71% identity (29/41) with its orthologous 41 bp of human *GFPT1* (*8*), and 76% identity (31/41) with its orthologous 41 bp of *P. horikoshii GFPT*. In the zone of greatest homology, comparison of the apparent 19-bp TnsD binding site in *glmS* (*1, 8*) with that in human *GFPT1* and *P. horikoshii GFPT* revealed 74% identity (14/19). In fact, the human and *P. horikoshii* sequences are more similar to each other than they are to *E. coli* attTn7.

The similarities are not coincidental, but exist because the 3′ ends of all of the genes homologous to glucosamine: fructose-6-phosphate aminotransferase in all the organismic kingdoms code for the identical amino acid sequence "PRNLAKSVTVE". These alignment results establish that these 3′-terminal sites were present in the common ancestors of bacteria and eukaryotes, and then survived for the intervening three billion years with only minor differences, as a result of the selection pressure to maintain enzymatic activity. Nucleotide sequence variations were limited to the wobble position and alternate codon selection.

Intrigued by the similarity of the *E. coli* and human sequences, we found that Tn7 could insert itself precisely into a donor plasmid containing the orthologous *GFPT1* domain at precisely the base pair corresponding to the insertion site in attTn7, at the same frequency as into a plasmid containing attTn7 (*8*). Since the transposition event

† The work was supported by NIH Grant CA76290 to E.W.

* To whom correspondence should be addressed. Department of Biochemistry and Molecular Pharmacology, Thomas Jefferson University, Philadelphia, PA 19107. Phone: 215.955.4578. Fax: 215.955.4580. E-mail: eric@tesla.jci.tju.edu.

‡ Current address: VaxInnate Corporation, New Haven, CT 06511.

10.1021/bi035535u CCC: $27.50
Published on Web 02/18/2004

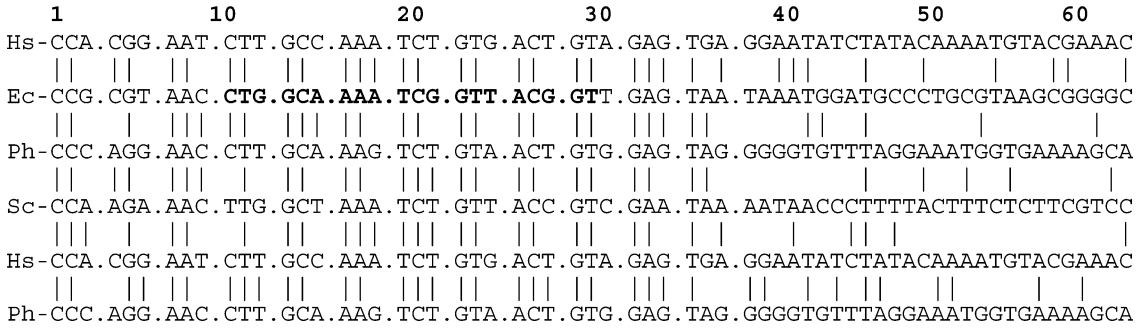

FIGURE 1: Comparison of the 63-bp attTn7 recognition−insertion domain at the 3′ end of *E. coli* (Ec) glutamine synthase *glmS* gene (*2*) with the homologous glutamine:fructose-6-phosphate transaminase (*GFPT*) genes of *P. horikoshii* (Ph), *S. cerevisiae* (Sc), and *H. sapiens* (Hs). The putative 19-bp TnsD binding sequence is shown in bold. The Hs and Ph sequences are shown twice to facilitate comparison.

depends on the binding of TnsD protein to attTn7, or to its orthologs in other species, the determination of binding rate constants, dissociation rate constants, and equilibrium constants could elucidate the specificity and strength of TnsD binding. We report here a study of interaction between protein TnsD with a set of 20-bp duplex DNA substrates using circular dichroism (CD)[1] and total internal reflectance (TIR). CD was used to follow changes in the conformation of DNA sequences upon interaction to TnsD. Kinetic studies of these interactions were carried out in a total internal reflectance (TIR) instrument and the interaction profiles were analyzed to get the k_{on}, k_{off}, and K_D values. DNA duplexes corresponding to *E. coli glmS* attTn7 and its eukaryotic *GFPT1* orthologs in yeast and mammals displayed high binding affinity, while a single point mutation in the middle eliminated detectable binding to TnsD.

MATERIALS AND METHODS

Purification of TnsD. Unless otherwise indicated, all steps were carried out at room temperature, 22 °C, and reagent grade chemicals were obtained from Sigma Chemical (St. Louis MO) or Fisher Chemical (Pittsburgh PA). The TnsD gene was subcloned into pET14b-TnsD vector (Novagen, Madison WI), an overexpression system containing a bacteriophage T7 promoter followed by six codons of histidine, six codons of thrombin cleavage site, and the TnsD gene (*9*). The thrombin cleavage site is useful to separate TnsD protein from the 6× His-tag when desired. The resulting pET14b-TnsD plasmid was transformed into *E. coli* BL21 (DE3)-Codon Plus-RIL (Stratagene, San Diego CA) for overexpression to overcome the codon bias limitations of TnsD. This strain has additional copies of tRNA for the most underutilized codons in *E. coli* the AGG/AGA codon (*argU*, arginine), AUA codon (*ileY*, isoleucine) and the CUA codon (*leuW*, lysine). The BL21 (DE3)-Codon Plus RIL: pET14b-TnsD clone was inoculated into 30 mL LB containing 100 μg/mL carbenicillin (ampicillin substitute) and 34 μg/mL chloramphenicol, then incubated overnight at 37 °C, 175 rpm. This saturated culture was used as a preinoculum for 1.5 L of LB in a 5-L flask and was incubated at 37 °C at 175 rpm to an $OD_{600\ nm}$ of 0.55 followed by 3 h induction with 1 mM IPTG.

Cells were harvested at 4000 rpm at 4 °C for 10 min and lysed with 80% B-Per (Pierce, Rockford IL), 20% glycerol, plus high-salt buffer (300 mM NaCl, 10 mM Tris-HCl, pH 8.0, 20% glycerol, 1 mM DTT, and 1 mM PMSF) and incubated for 20 min at 22 °C. Hen egg white lysozyme (Roche, Nutley NJ) was added to a final concentration of 1 mg/mL and incubated at 4 °C for 30 min. Benzonase nuclease (Novagen, Madison WI) was added to a final concentration of 25 U/mL and incubated at 22 °C for 20 min to degrade the nucleic acids, reducing the viscosity of the lysate. This mixture was sedimented 15 min at 1680*g*, at 4 °C. The pellet was discarded and the lysate supernatant was applied to a Ni-NTA superflow column (Qiagen, Valencia, CA). The nonspecifically bound proteins were washed out of the column with 10 column volumes of 20 mM imidazole in high-salt buffer (10 mM Tris-HCl, pH 8.0, 20% glycerol, 300 mM NaCl, 1 mM DTT, 1 mM PMSF). The TnsD protein was eluted with 250 mM imidazole in high-salt buffer. The TnsD fraction was diluted 10-fold with 25 mM NaCl in Tris buffer (10 mM Tris-HCl, pH 8.0, 20% glycerol, 1 mM DTT, 1 mM PMSF) to bring down the NaCl concentration before its application to a 5 mL Hi-Trap Q-Sepharose HP anion-exchange column (Amersham Pharmacia Biotech, Piscataway NJ). The column was then eluted with a linear 25−500 mM NaCl gradient in Tris buffer. The TnsD protein peak was centered at 225−250 mM NaCl.

Computational analysis using the program ProtParam (Swiss Institute of Bioinformatics, Lausanne, Switzerland) showed that the theoretical isoelectric point of TnsD protein is 9.42, which is unusually high. Thus, we predicted that cation exchange chromatography also be effective for TnsD. In this case, induced cells were lysed with 25 mM NaCl in Tris buffer. The lysate supernatant was diluted with 10 volumes of 25 mM NaCl in Tris buffer and filtered through a 0.45 μm filtration unit. A 5-mL Hi-trap SP Sepharose HP column (Amersham Pharmacia Biotech, Piscataway NJ) was preequilibrated with 10 column volumes of 25 mM NaCl in Tris buffer. The diluted, filtered lysate supernatant was then applied to the column, followed by steps of 25, 50, 100, 150, 200, 225, 250, 275, 300, 325, and 500 mM NaCl in Tris buffer. Each gradient step consisted of 4 column volumes, except that at the 100 mM step, 10 column volumes were used. The TnsD protein eluted at 225−250 mM NaCl.

The lysates, pellets, flow-throughs, washes, and fractions were electrophoresed in the presence of 0.1% $NaDodSO_4$ on a precast 12% polyacrylamide gel containing 25 mM Tris-HCl, 192 mM glycine, pH 8.3 (Invitrogen, Carlsbad CA),

[1] Abbreviations: CD, circular dichroism; DTT, dithiothreitol; $NaDodSO_4$, sodium dodecyl sulfate; EDTA, ethylenediaminetetraacetic acid; PMSF, phenylmethanesulfonyl fluoride; TIR, total internal reflectance.

Table 1: Sequences of 20-bp DNA Duplex Analogues of attTn7[a]

name	sequence
E. coli attTn7	Fl-5′-TGG CAA AAT CGG TTA CGG TT-3′
	3′-ACC GTT TTA GCC AAT GCC AA-5′-Biotin
E. coli attTn7 mutant	Fl-5′-TGG CAA AAT AGG TTA CGG TT-3′
	3′-ACC GTT TTA TCC AAT GCC AA-5′-Biotin
yeast *GFAT*	Fl-5′-TTG GCT AAA TCT GTT ACC GT-3′
	3′-AAC CGA TTT AGA CAA TGG CA-5′-Biotin
mammalian *GFPT1* A	Fl-5′-TTG CCA AAT CTG TGA CTG TA-3′
	3′-AAC GGT TTA GAC ACT GAC AT-5′-Biotin
mammalian *GFPT1* A mutant	Fl-5′-TTG CCA AAT ATG TGA CTG TA-3′
	3′-AAC GGT TTA TAC ACT GAC AT-5′-Biotin
mammalian *GFPT1* B	Fl-5′-TCT TGC CAA ATC TGT GAC TG-3′
	3′-AGA ACG GTT TAG ACA CTG AC-5′-Biotin

[a] The yeast sequence is shifted 1 bp upstream relative to *E. coli* and mammalian A. The mammalian B sequence is shifted 2 bp upstream relative to mammalian A, thereby lacking the 3′ terminal T of the putative 19-bp TnsD interaction sequence, but gaining one more identical nucleotide.

followed by Coomassie blue staining (Gel Code Blue Stain reagent, Pierce). Stained gels were scanned on a Kodak 440CF imager. Following purification, single bands of TnsD were excised from denaturing gels for mass spectroscopic analysis of in-gel tryptic digests (*10*) by Surface Enhanced Laser Desorption/Ionization (SELDI-TOF MS) Time-of-Flight on a SELDI PBS mass spectrometer (Ciphergen Biosystems, Inc. Fremont, CA 94555), and N-terminal sequencing on an Applied Biosystems 470A peptide sequencer. The concentration of TnsD was estimated with the BCA Protein Assay reagent kit (Pierce), standardized against bovine serum albumin (BSA). The purified yield of TnsD protein was 20−24 mg/L of bacterial culture. The protein was stored at 4 °C.

Preparation and Purification of DNA Duplexes. We hypothesized that the conserved 19-bp TnsD recognition sequence should be statistically unique (1 out of 4^{19} or 2.7×10^{11}) within the 3×10^9 bp of the human genome. On the basis of BLAST analysis of the human genome on Genbank that supported the hypothesis, double-stranded DNA duplexes (Table 1) were formed from a set of 20-bp single-strand oligonucleotides (TriLink Biotechnologies, San Diego CA) corresponding to the apparent TnsD binding site of *E. coli glmS* attTn7, and the orthologous sequences from yeast and mammalian *GFPT1* (*8*). Sequences from *P. horikoshii* were not studied because we are not considering site-specific gene insertion in that species.

Each duplex included a fluorescein tag on the 5′ end of one strand and biotin on the 5′ end of the complementary strand. Equimolar amounts of complementary oligonucleotide strands were annealed in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) at 95 °C for 3 min on a thermal block. The samples were left inside the thermal block and allowed to cool gradually over 1.5 h to room temperature. The duplex DNA strands with slower mobility were separated from free single strands by electrophoresis on native 20% polyacrylamide gels. The duplex bands were collected as small gel pieces in Eppendorf tubes, homogenized in 1 mL of TE buffer, then extracted from the homogenized gel pieces by diffusion overnight with tumbling at 4 °C. Duplex DNAs eluted into the supernatant were purified by ethanol precipitation (*11*), then reanalyzed on native 20% polyacrylamide gels and quantified on a Shimadzu UV160 spectrophotometer.

Circular Dichroism Studies. CD spectra were obtained by a Jasco 810 spectropolarimeter thermostated at 25 °C using a quartz cell of 1-cm path length in CD buffer (10 mM Tris-HCl, pH.7.4, 110 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 10% glycerol, 1 mM EDTA). The CD spectra were recorded over a wavelength range (250−320 nm) where the protein TnsD had virtually no signal, relative to the DNA. The CD spectrum of each DNA duplex alone was recorded, followed by the spectrum upon the addition of the protein. The titration profiles for the interaction of the *E. coli* sequence and its eukaryotic orthologs were monitored at 270 nm, the wavelength of maximum difference between the peak of the DNA spectrum and that of the TnsD−DNA complex. Titrations were recorded at least twice. The data were fit to the ligand binding equation for single-site saturation, $y = (B_{max}x)/(K_D + x)$, where $B_{max} = \max(y)$ (*12*).

Total Internal Reflectance Studies. The kinetic analysis of the interaction of the different synthetic DNA duplexes with TnsD was carried out on an IAsys biotinylated sensor chip in an IAsys Auto Plus total internal reflectance (TIR) spectrometer (Affinity Sensors, Franklin MA) at a temperature of 25 °C. Each new biotinylated chip was preconditioned before use by treating with two pulses of flow buffer (10 mM Tris-HCl, pH 7.4, 110 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM EDTA, 1.8 mM DTT, 0.8 mM ATP, 10% glycerol, 0.05% Tween). After equilibrating the sensor chip surface with flow buffer for 10 min, 80 μL of streptavidin (100 μg/mL in flow buffer) was injected across the surface to saturate the biotinylated layer. After 10 min of binding, the sensor surface was washed with 80 μL of flow buffer to remove excess streptavidin. A total of 80 μL of 10 μg/mL biotinylated DNA duplex in flow buffer were then immobilized on the streptavidin-coated chips. Noncovalently bound DNA duplex was removed from the surface two injections of flow buffer. The interaction was monitored at several concentrations of protein TnsD. The association and dissociation profiles were recorded in real time as a function of increasing concentration of TnsD. A control reaction was carried out with the addition of TnsD to the streptavidin to monitor and measure the change in the bulk refraction index that is independent of any specific interaction. Each series of sensorgrams was recorded at least twice.

The binding and dissociation data were modeled and analyzed using IASys Grafit v.4 (Affinity Sensors) to calculate $k_{association}$, $k_{dissociation}$, and the ratio $K_D = k_{dissociation}/k_{association}$. The instrument response R (measured in arc seconds) is proportional to the mass of bound ligate, according to $R_t = (R_{eq} - R_0)[1 - e^{-k_{on}t})] + R_0$, where R_t is the response at time t, R_0 is the initial response, R_{eq} is the maximal response, and k_{on} is the measured forward rate constant. The association rate constant $k_{association}$ was obtained as the slope of the measured k_{on} values plotted against the concentrations of the protein used in the titration using the equation $k_{on} = k_{dissociation} + k_{association}[L]$, where L is the ligate concentration, i.e., TnsD. Theoretically, the intercept of this plot is $k_{dissociation}$. But the error was always very high in this case, so we obtained $k_{dissociation}$ values directly from the actual individual dissociation profiles of each titration using the equation $R_t = R_0 + \Delta R(1 - e^{-k_{diss}t})$. Although the $k_{dissociation}$ values determined by each means should be the same, the

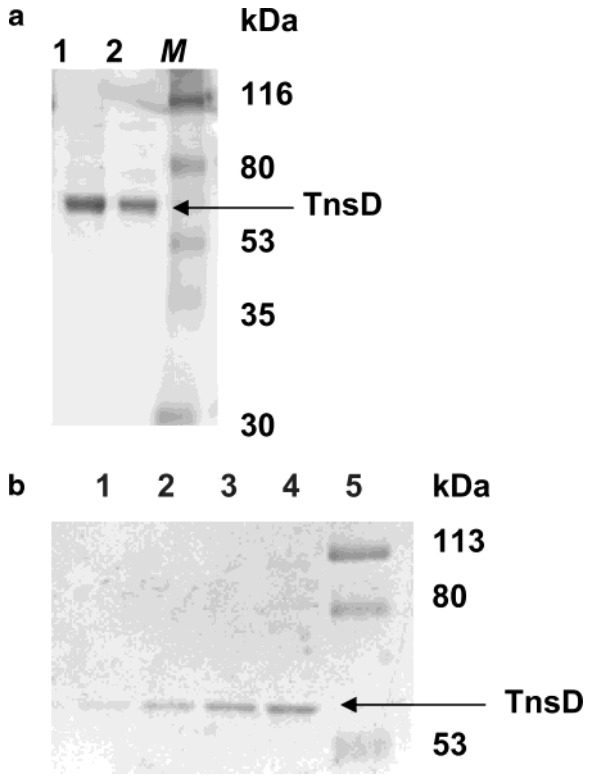

FIGURE 2: (a) Purification of TnsD protein by Ni-NTA affinity and anion exchange chromatography. Lanes 1 and 2 contain 20 μL of the TnsD fractions obtained from anion exchange chromatographic separation at 225 and 250 mM NaCl, respectively. Lane M is prestained low range protein standards (Bio-Rad). The arrow denotes the TnsD band. (b) Purification of TnsD protein by cation exchange chromatography. Lanes 1−4 contain 20 μL of the fractions obtained from cation exchange chromatographic steps at 150, 200, 250, and 275 mM NaCl, respectively. Lane 5 is prestained low range protein standards (Bio-Rad). The arrow denotes the TnsD band. The gel in (b) was electrophoresed longer than the gel in (a).

intercept results showed more variability, so we did not use the intercept in our calculations of K_D.

RESULTS

Protein Purification. TnsD purification by Ni-NTA chromatography and anion-exchange chromatography yielded a single band upon denaturing gel electrophoresis (Figure 2a). Purification by cation-exchange chromatography, based on the calculated theoretical pI of 9.42, yielded a single TnsD peak centered at 280 mM NaCl, which yielded a single band upon denaturing gel electrophoresis (Figure 2b). Surprisingly, this route yielded purer protein than His-tag affinity chromatography. That might indicate that the His_6 tags are not fully exposed or accessible for strong binding at the N-terminus of native TnsD protein. The yield of the protein per liter of culture following purification ranged from 20 to 24 mg over multiple batches.

SELDI-TOF MS analysis of in-gel tryptic digests (*10*) was consistent with the known amino acid sequence of TnsD protein, with a probability of 1.00 relative to a probability of 1.4×10^{-15} for the next most likely *E. coli* ATP binding protein of 62.98 kDa. Analysis of the N-terminal amino acid sequence of an excised gel band showed agreement for the six His residues and the first 8 N-terminal residues. The result is consistent with the band also showing fluorescence when stained with GelCode 6× His protein tag staining kit (Pierce).

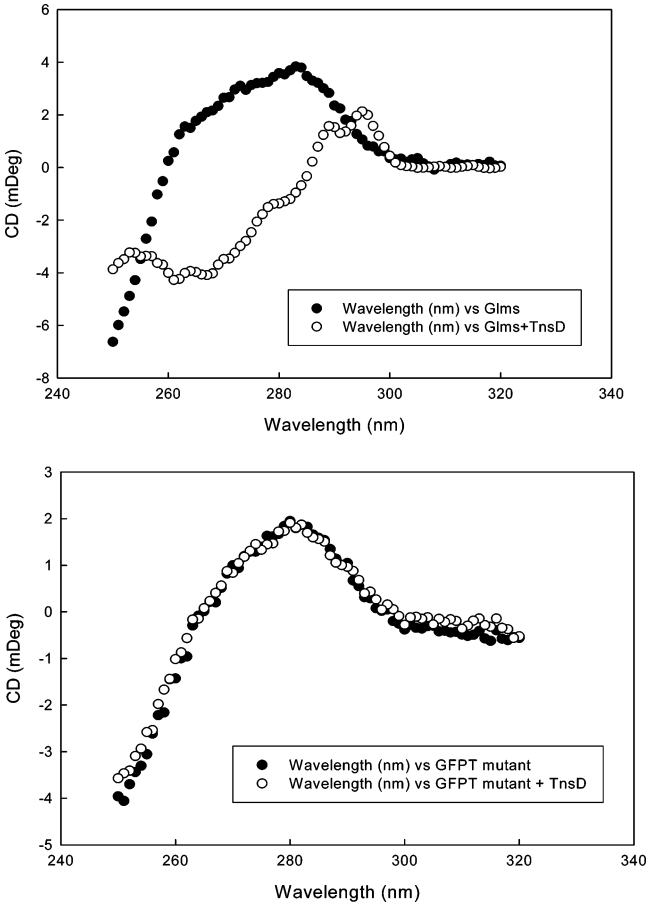

FIGURE 3: CD spectra of 4.1 μM *E. coli* attTn7 (top) and *E. coli* attTn7 mutant (bottom) sequences upon interaction with 11.2 μM TnsD protein at 25 °C in CD buffer, as described in Materials and Methods.

Searches through publicly accessible databases did not reveal the existence of any similar sequences or structures. In particular, no homology could be found with any other DNA binding protein.

Circular Dichroism Spectra. Interaction of excess TnsD with the *E. coli* attTn7 20-bp duplex sequence (Table 1) strongly decreased the magnitude and changed the shape of the DNA peak (Figure 3a). Comparable interactions were also observed between TnsD and the attTn7 orthologs from yeast and mammals (not shown). Remarkably, replacement of a single central C:G basepair of the *E. coli* sequence by an A:T basepair eliminated any perturbation of the CD spectrum (Figure 3b). Similarly, TnsD had no effect on the mutant mammalian DNA duplex (not shown).

Titration of DNA CD at the wavelength of maximum reduction by addition of TnsD allowed estimation of K_D. Spectra measured during titrations did not show any peak flattening or red shift, nor was turbidity observed in the cell, implying that no aggregation occurred.

Figure 4 shows a representative titration profile for TnsD interacting with the mammalian A sequence at 270 nm. The titration displayed a hyperbolic decay curve. The solid line is the best hyperbolic fit of the experimental data representing single site binding; a normalized titration is shown in the inset. Best-fit single-site K_D values are shown in Table 2. The three eukaryotic ortholog sequences showed about 10-fold lower affinities to TnsD than did the *E. coli* 20-bp TnsD binding sequence.

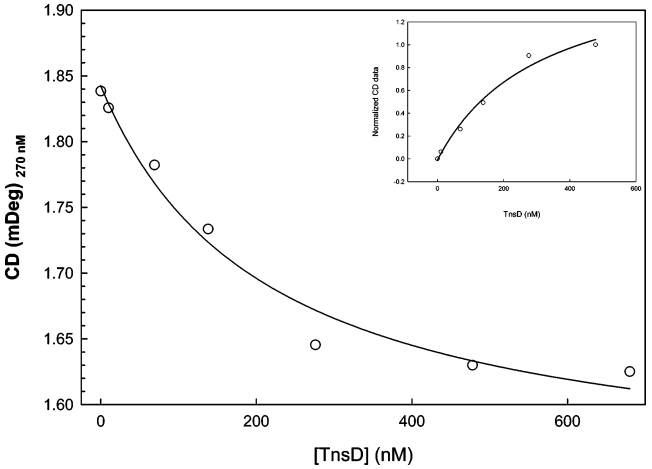

FIGURE 4: CD titration profile for the interaction of TnsD protein with mammalian *GFPT1* A duplex monitored at 270 nm. Open circles are the raw data, and the solid line represents the fit as described in Materials and Methods. The inset shows the normalized data of the same titration profile, with 1.0 representing the maximum change and 0 representing no protein added.

Total Internal Reflectance. Figure 5 displays representative TIR profiles for the association and dissociation of TnsD with the immobilized *E. coli* 20-bp TnsD binding sequence. Both sensorgrams fitted well to the equation for single-phase interaction, consistent with a single-site binding model. The association and dissociation profiles for TnsD interactions with the orthologous eukaryotic DNA duplexes (Table 1) were measured in the same way. The association rate constants in Table 2 are of the order of 10^4/M·s, several orders of magnitude lower than the theoretical diffusion limit in the range of 10^6 to 10^9/M·s, suggesting that sequence-specific association reactions are controlled by a rare event or by a rate-limiting rearrangement step (*13*, *14*). This phenomenon is consistent with the lack of detectable binding in the association step by the single central mismatch sequences. The three eukaryotic ortholog sequences showed about 50-fold lower affinities (Table 2).

Figure 6 is a representative plot of k_{on} values vs TnsD concentrations interacting with the *E. coli* sequence. As with the CD measurements, TnsD had no detectable effect on the TIR signal of the central mismatch mutant sequences. Data fitting (Table 2) revealed that the *E. coli* sequence displayed the strongest affinity for the TnsD protein, with $K_D = 9 \pm 3$ nM. The three eukaryotic ortholog sequences showed about 10-fold lower affinities (Table 2).

DISCUSSION

Transposon Tn7 protein TnsD exhibits the noteworthy ability to recognize a 19-bp sequence target in *E. coli*, to direct TnsA, TnsB, and TnsC to insert the Tn7 transposon 33 bp downstream, while sharing no discernible homology to any other characterized protein. Our previous observation

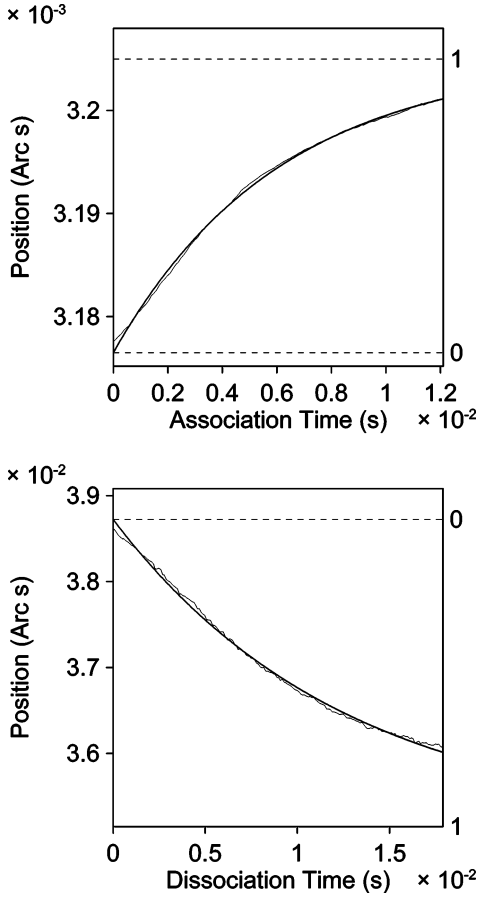

FIGURE 5: Total internal reflectance kinetics of TnsD protein binding to *E. coli* attTn7 20-bp DNA duplex. Association (top) and dissociation (bottom) profiles of immobilized *E. coli* DNA duplex interacting with 7.25 μM TnsD were fit by nonlinear regression. Solid continuous line represents single-phase fitted curve in both cases. Calculated k_{on} was 0.01649 ± 0.00024/M·s and k_{off} was 0.00794 ± 0.00018/s.

that Tn7 could insert itself precisely into a donor plasmid containing the mammalian attTn7 ortholog at precisely the base pair corresponding to the insertion site in attTn7, at the same frequency as into a plasmid containing attTn7 (*8*), despite five mismatches between the bacterial and mammalian 19-bp sequences, invited further study of the recognition mechanism.

Biophysical analysis of protein−DNA interactions requires appreciable quantities of purified components. With 24 low-usage codons, beginning immediately after the initiation codon, the *tnsD* mRNA provides intrinsically low expression (*9*). The use of BL21-codon plus (DE3)-RIL containing extra copies of the *argU*, *ileY*, and *leuW* tRNA genes encoding tRNAs that recognize the arginine codons AGA and AGG, the isoleucine codon AUA, and the leucine codon CUA, respectively, as the host strain for expression permitted compensation for the 24 low-expression codons in *tnsD*. The

Table 2: Calculated Dissociation Constants, K_D, of Transposon Tn7 Protein TnsD Binding to attTn7 DNA and Its Eukaryotic Orthologs

sequence name	K_D (CD) (nM)	$k_{association}$ (TIR) ($\times 10^3$/M·s)	$k_{dissociation}$ (TIR)	K_D (TIR) (nM)
E. coli attTn7	35 ± 16	10.0 ± 0.4	9.0 ± 0.1 × 10^{-5}/s	9 ± 3
yeast *GFAT*	458 ± 64	11.4 ± 2.1	8.9 ± 0.3 × 10^{-3}/s	781 ± 170
mammalian *GFPT1* A	284 ± 43	5.6 ± 0.4	1.7 ± 0.8 × 10^{-3}/s	304 ± 164
mammalian *GFPT1* B	130 ± 11	68.8 ± 1.1	13.3 ± 0.1 × 10^{-3}/s	193 ± 4

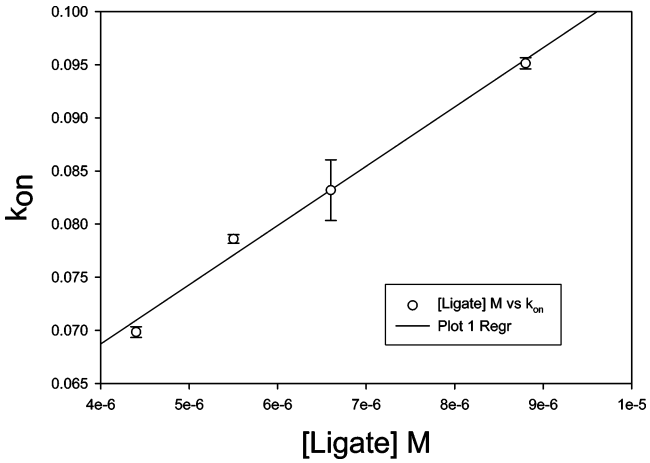

FIGURE 6: Determination of $k_{association}$ for the interaction of free TnsD ligate with immobilized mammalian *GFPT1* A duplex DNA ligand from k_{on} values as a function of TnsD concentration.

TnsD protein was almost entirely purified in a single step by cation exchange chromatography, which was inspired by the calculated pI of 9.42. That protocol allowed sufficient TnsD expression, 20−24 mg/L, for physical measurements. This is a remarkable improvement compared to the earlier reported yield of 0.16 mg of purified TnsD from a 200-L culture (*15*).

Preliminary agarose gel shift experiments with crude cellular lysates suggested that TnsD could bind to 20-bp duplexes corresponding to the apparent TnsD recognition domain in attTn7 and its eukaryotic orthologs, but not to single central mismatch sequences (not shown). Precise measurements of TnsD association to those 20-bp duplexes were obtained at equilibrium in a homogeneous solution by CD titration, and dynamically in a solution phase TnsD/solid-phase DNA system by TIR kinetic measurements.

CD spectra immediately revealed that TnsD strongly reduces the DNA peak at 270 nm, suggesting that TnsD destabilizes its duplex target sequence upon binding, consistent with the expectation that chain scission and splicing in of a transposon would require some unwinding of the double helix. The lack of DNA perturbation by TnsD in the case of single central mismatches illustrated the strong sequence specificity of the interaction. Overlap of protein troughs with DNA troughs precluded assignment of meaning to changes in spectra below 250 nm.

CD and TIR estimates of TnsD dissociation constants with attTn7 and its eukaryotic orthologs (Table 2) roughly agreed with each other, with $K_D \approx 20$ nM for the attTn7 sequence, and $K_D \approx 500$ nM for the eukaryotic orthologs. The differences in the ratios of the K_D(CD) and K_D(TIR) values of *E. coli* attTn7, yeast *GFAT*, mammalian *GFPT1A* and *GFPT1B* may be due to the inherent discrepancies in the homogeneous and heterogeneous systems used in CD and TIR, respectively. Mutant DNA sequences with a single central mismatch yielded no detectable phenomena upon TnsD addition, even during the kinetic association step, supporting the hypothesis that TnsD could be used to locate a unique insertion site in the human genome at 2p13. The lack of detectable binding to the mismatch sequences may signal the formation of uniquely kinked structures as a prerequisite for precise protein binding (*16*, *17*) in the attTn7 orthologs, which remain in the normal B form in the mutant sequences.

The 10−50-fold increase in K_D for the eukaryotic orthologs is much less of a weakening than might be expected from so many mismatches, but correlates with pattern recognition analysis of which residues are important for recognition (*8*), which predicted only one significant mismatch among six mismatches with the mammalian sequences, and no significant mismatches among three mismatches with the yeast sequence. Thus, the weaker K_D, roughly 8 kJ/mol less favorable, might be due to the loss of only two hydrogen bonding or hydrophobic interactions between TnsD and DNA. If one knew the crystal structure of TnsD bound to attTn7 and its mammalian counterpart, one might be able to predict an amino acid mutation that could make up the difference in affinity.

To understand all the selection rules for sequence recognition in Tn7 transposition as a tool for gene therapy, future studies will examine other single point mutations in attTn7 orthologs, and will also examine the DNA binding interactions of Tn7 proteins TnsA, TnsB, and TnsC.

ACKNOWLEDGMENT

We thank Dr. Yuri Sykulev and Dr. Tatiana Lebedeva for their advice and suggestions regarding total internal reflectance studies.

REFERENCES

1. Craig, N. L. (1996) *Curr. Top. Microbiol. Immunol 204*, 27−48.
2. Lichtenstein, C., and Brenner, S. (1982) *Nature 297*, 601−3.
3. Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y., Sakai, M., Ogura, K., Otsuka, R., Nakazawa, H., Takamiya, M., Ohfuku, Y., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., and Kikuchi, H. (1998) *DNA Res. 5*, 55−76.
4. Cheret, G., Pallier, C., Valens, M., Diagnan-Fornier, B., Fukuhara, H., Bolotin-Fukuhara, M., and Sor, F. (1993) *Yeast 9*, 1259−65.
5. Sayeski, P. P., Paterson, A. J., and Kudlow, J. E. (1994) *Gene 140*, 289−90.
6. McKnight, G. L., Mudri, S. L., Mathewes, S. L., Traxinger, R. R., Marshall, S., Sheppard, P. O., and O'Hara, P. J. (1992) *J. Biol. Chem. 267*, 25208−12.
7. Whitmore, T. E., Mudri, S. L., and McKnight, G. L. (1995) *Genomics 26*, 422−3.
8. Cleaver, S. H., and Wickstrom, E. (2000) *Gene 254*, 37−44.
9. Cleaver, S. H. (2001) *Transposons and Retroviruses as a Paradigm for Site-Specific Gene Insertion.* Dissertation in Molecular Pharmacology and Structural Biology, Thomas Jefferson University, Philadelphia.
10. Merchant, M., and Weinberger, S. R. (2000) *Electrophoresis 21*, 1164−77.
11. Maniatis, T., Fritsch, E., and Sambrook, J. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
12. Wickstrom, E., and Laing, L. G. (1988) *Methods Enzymol. 164*, 238−58.
13. Eigen, M., and Hammes, G. G. (1963) *Adv. Enzymol. Relat. Areas Mol. Biol. 25*, 1−38.
14. Kim, J. G., Takeda, Y., Matthews, B. W., and Anderson, W. F. (1987) *J. Mol. Biol. 196*, 149−58.
15. Bainton, R. J., Kubo, K. M., Feng, J. N., and Craig, N. L. (1993) *Cell 72*, 931−43.
16. Wang, T., Piefer, A. J., and Jonsson, C. B. (2001) *J. Biol. Chem. 276*, 14710−7.
17. Drotschmann, K., Yang, W., Brownewell, F. E., Kool, E. T., and Kunkel, T. A. (2001) *J. Biol. Chem. 276*, 46225−9.

BI035535U