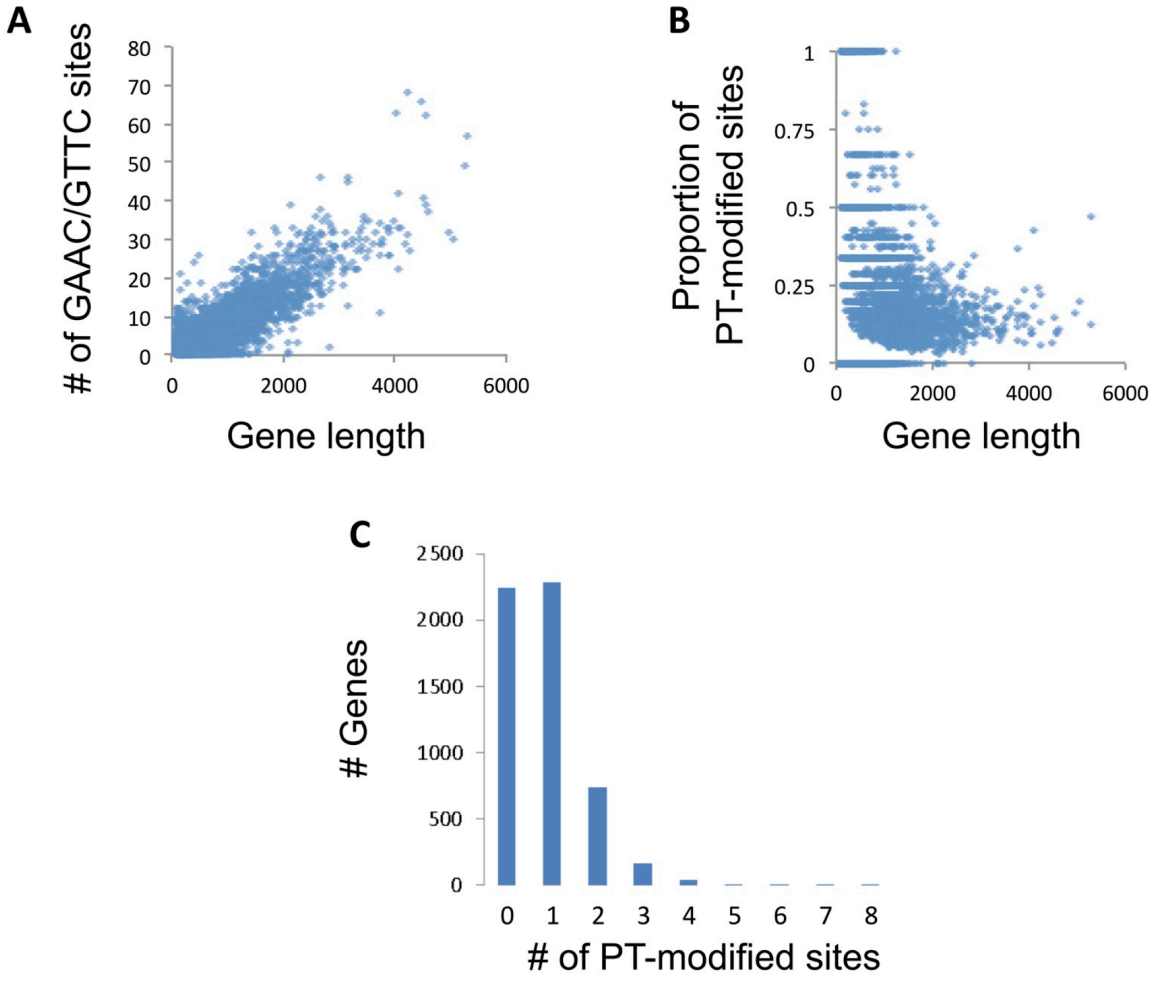
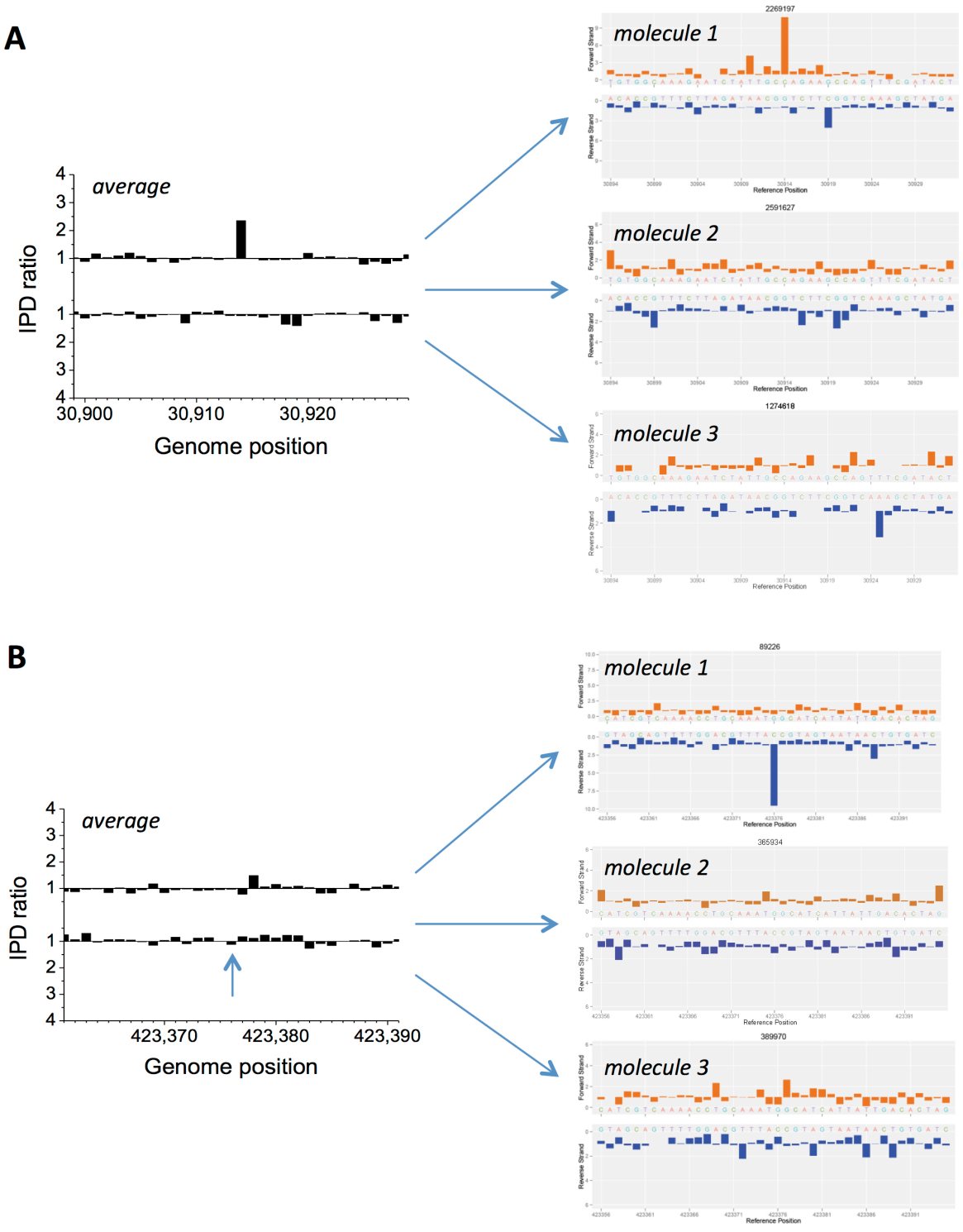
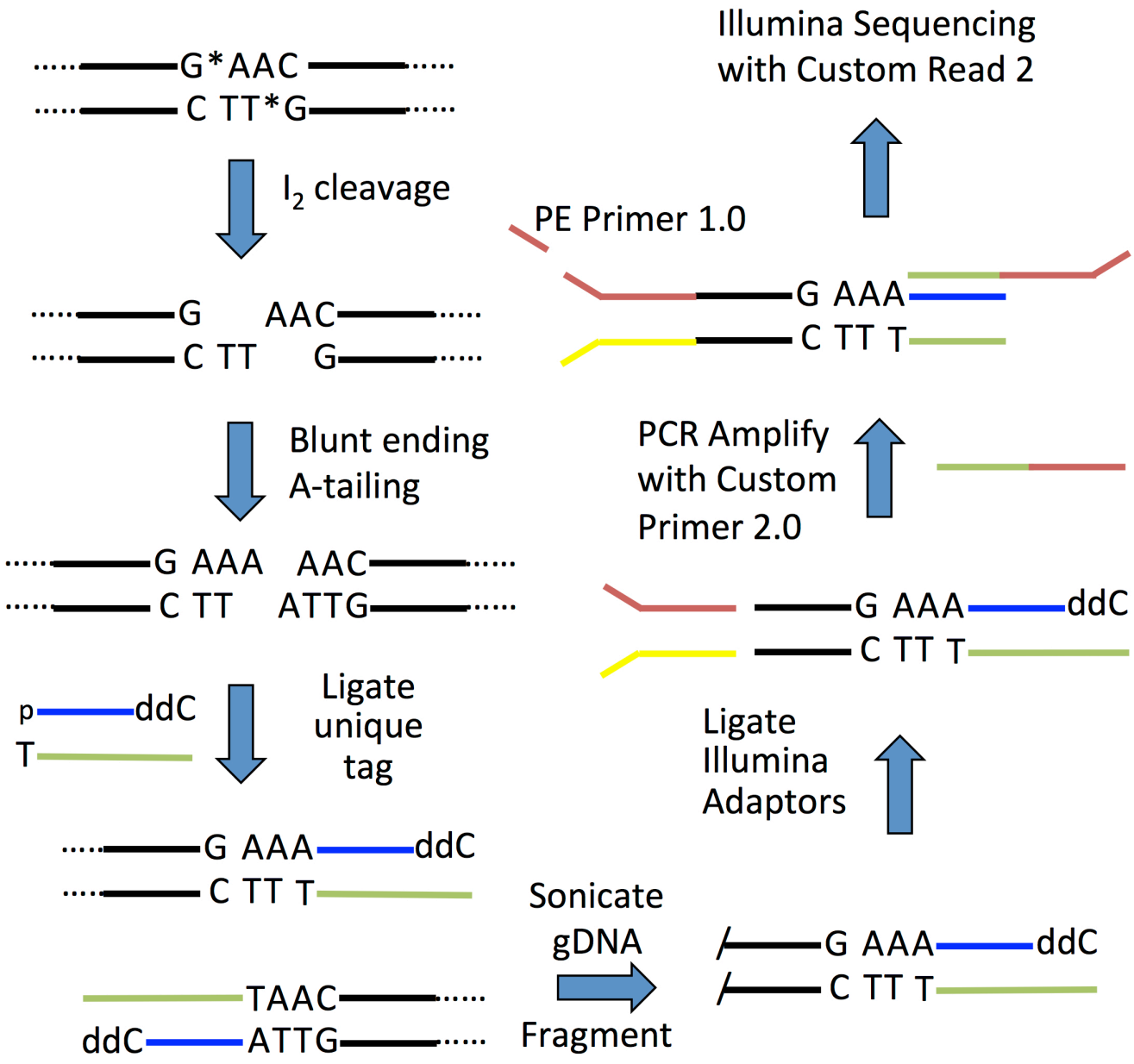
Supplementary Information for "Genomic mapping of phosphorothioates in bacteria reveals partial modification of short consensus sequences" by Cao *et al.*



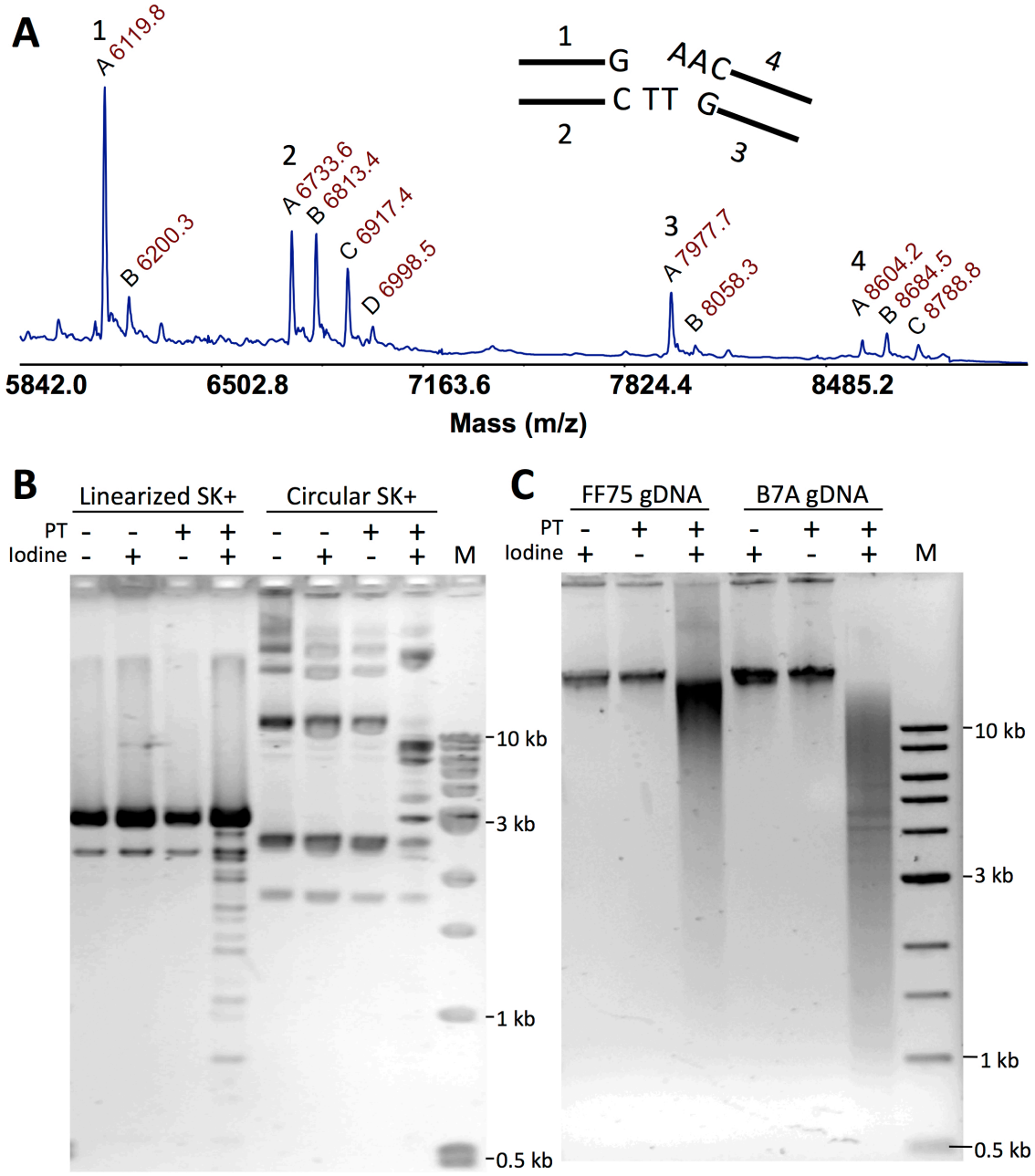
Supplementary Figure 1 – Proportion of PT-modified GAAC/GTTC sites as a function of gene length in B7A. Data from SMRT mapping of PT modifications in E. coli B7A were analyzed in terms of the frequency of modification of GAAC/GTTC sites and gene length. (**A**) Total number of GAAC/GTTC sites as a function of gene length in nucleotides. (**B**) The proportion of GAAC/GTTC sites modified with PT as a function of gene length in nucleotides. (**C**) Frequency distribution of PT-modified sites in genes across the B7A genome.



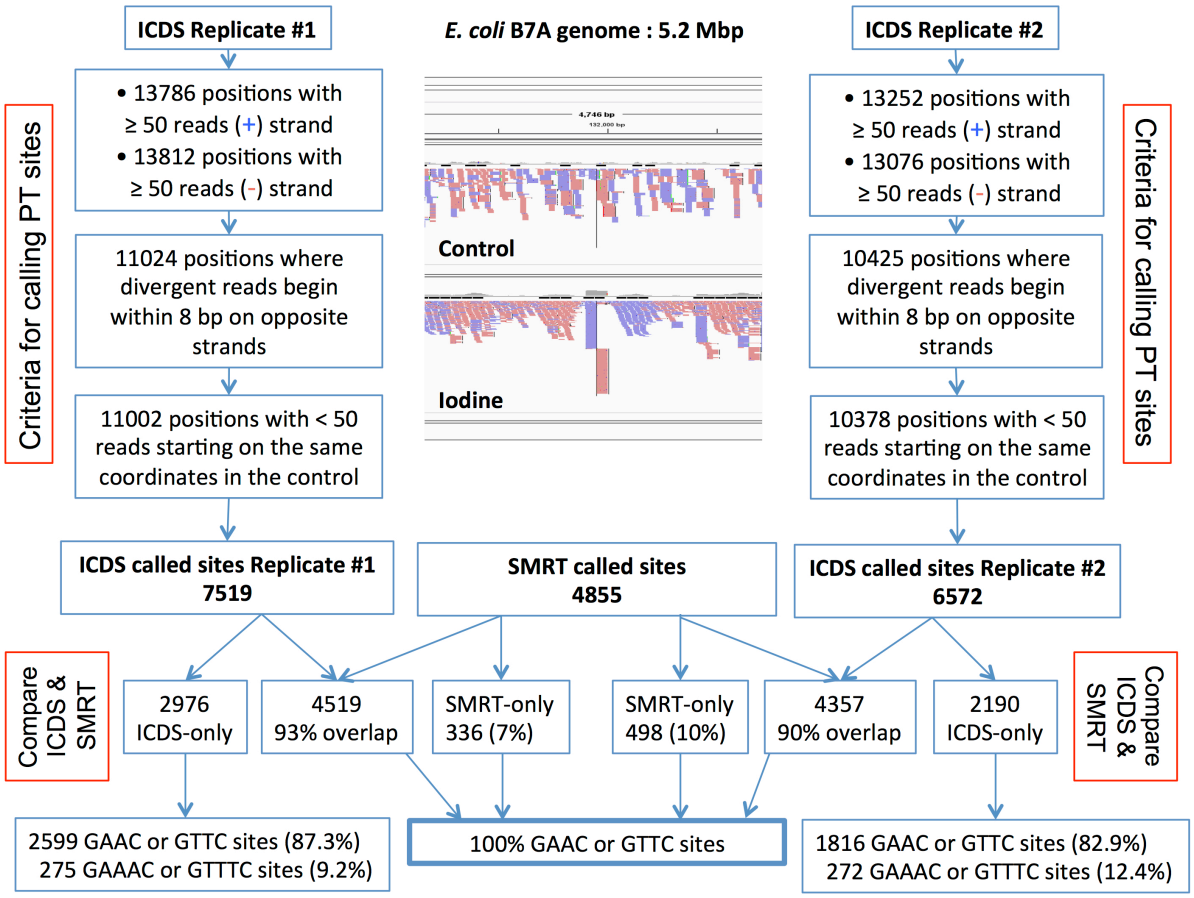
Supplementary Figure 2: Single-molecule analysis suggests partial or incomplete modification of specific sites with PT in *V. cyclitrohicus* FF75. (**A**) The kinetic signal average over all molecules at a selected genomic position showing a strong PT kinetic signal is shown on the left for FF75, with IPD ratio plot examples from single molecules underlying this average, which show between none and full PT modification signals, are shown on the right. (**B**) The kinetic signal average over all molecules at a selected genomic position for which there was no detectable PT kinetic signal is shown on the left for FF75, with IPD ratio plot examples from single molecules underlying this average, which show the presence of some molecules with a PT, are shown on the right.



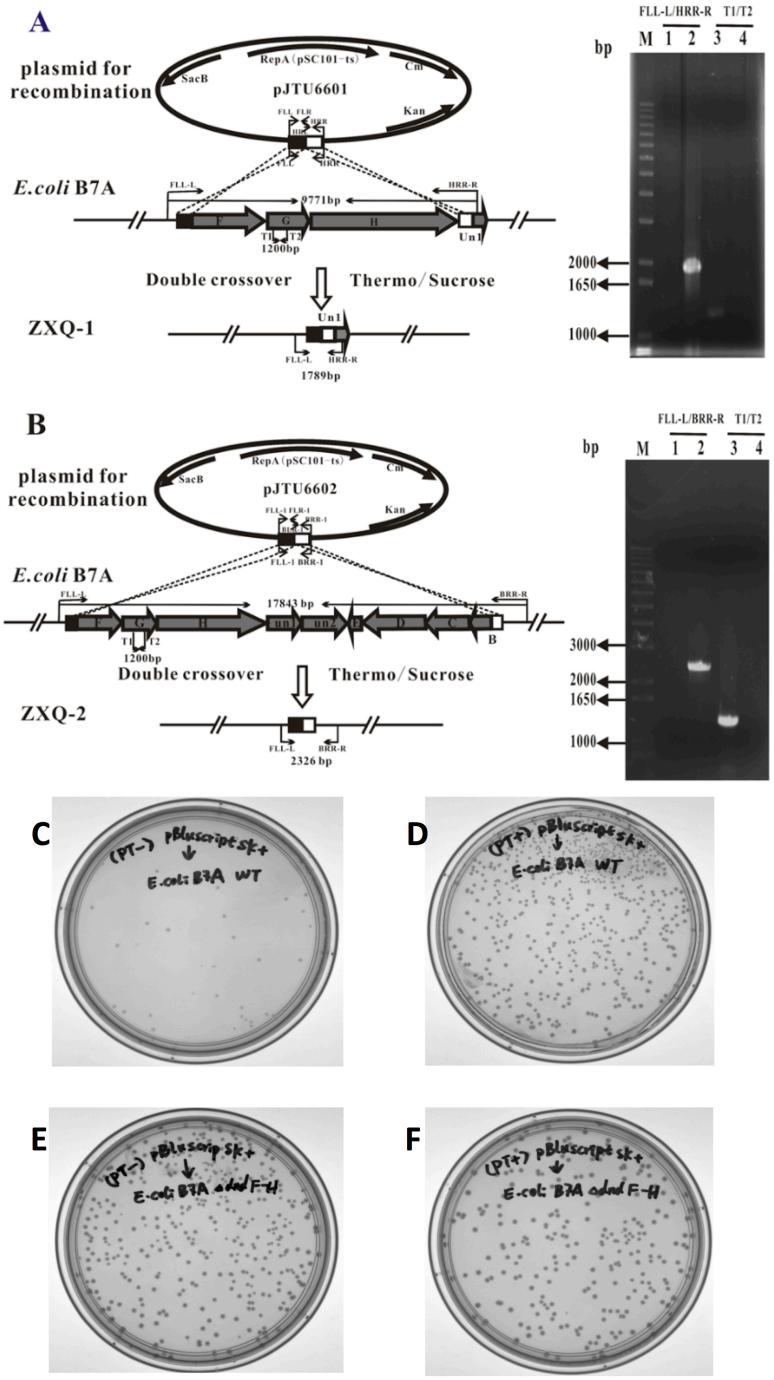
Supplementary Figure 3: Detailed schematic of the iodine cleavage, deep sequencing (ICDS) method. Oxidation of phosphorothioate (PT) modifications by I_2 results in a strand break at the site of the modification, with bistranded modifications producing a double-strand break. For mapping of PT modifications, the genomic DNA containing bistranded PT modifications is treated with I_2 and the resulting double-strand break is then processed and ligated to a unique double-stranded oligodeoxynucleotide tag, followed by fragmentation of the ligated genomic DNA. Following addition of deep sequencing adaptors (Illumina shown here), the DNA is subjected to PCR amplification and sequencing.



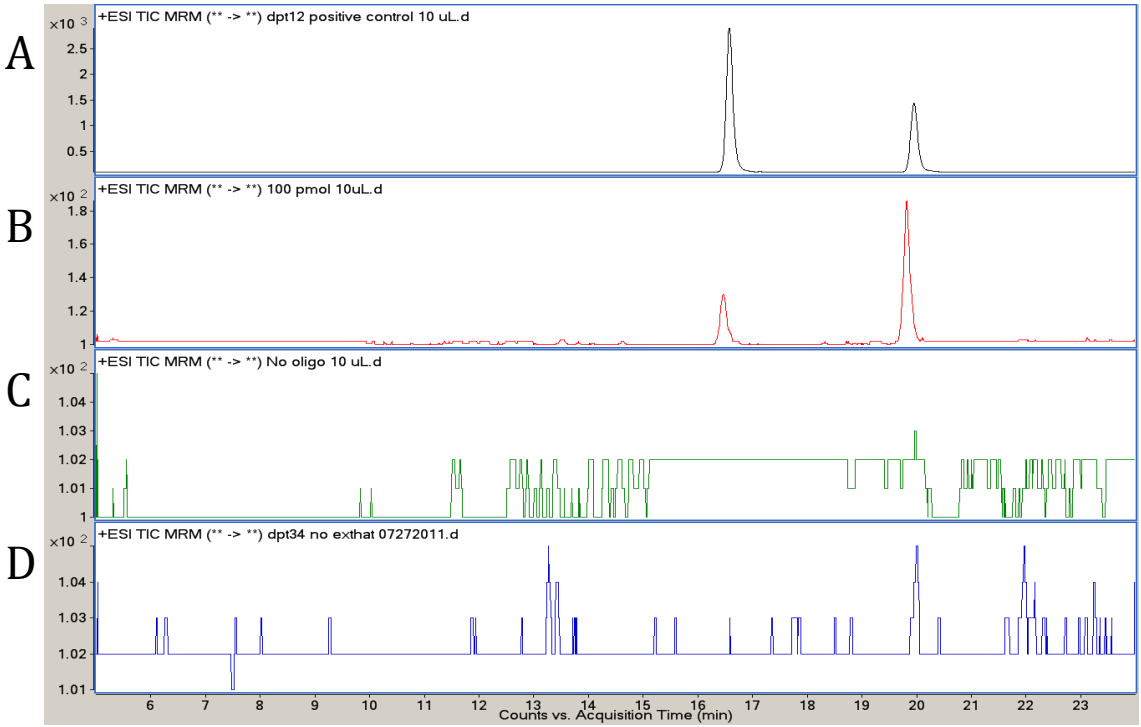
Supplementary Figure 4: MALDI-TOF analysis of iodine-ethanol cleavage of a PT-containing oligodeoxynucleotides and agarose gel analysis of iodine-ethanol cleavage of PT-containing pBluescript SK$^+$ and genomic DNA. (A) Iodine cleavage of complementary 48-mer oligonucleotides with site-specific PT. Products were interrogated by MALDI-TOF. Peaks were lettered based on composition of termini as follows A: hydroxyl, B: phosphate (+79.98), C: TRIS-phosphate (+183.1), and D: TRIS-diphosphate (+263.1). Contaminating TE buffer accounted for alternate termini involving TRIS. (B) Iodine cleavage of pBluescript SK+ isolated from *E. coli* B7A wild-type (PT$^+$) or mutant (PT$^-$) strains. After iodine treatment, plasmids were linearized with HindIII (lanes 1-4); or not linearized (lanes 5-8); M: 10k ladder (lane 9). (C) Iodine cleavage of genomic DNA from *Vibrio cyclitrophicus* FF75 wild type (PT$^+$) or mutant (PT$^-$) strains (lanes 1-3); and from *E. coli* B7A wild-type (PT$^+$) or mutant (PT$^-$) strains (lanes 4-6); M: 10k ladder (lane 7).



Supplementary Figure 5: Strategy for ICDS signal deconvolution for *E. coli* B7A. Individual read start positions were recorded for both strands, and positions shared by ≥ 50 reads were retained for analysis. Next, only positions where divergent reads mapped within 8 bp on opposite strands were retained, while any sites likewise enriched in the control were discarded. Finally, to rectify minor shifts in the read generation process, multiple call sites within close proximity (±2 bp of the dominant read peak) were collapsed into the centermost 8 bp region to define a unique ICDS call site. Overlaps with known locations of GAAC/GTTC and with SMRT-defined sites in the B7A genome were queried using bedtools to arrive at ICDS-only sites, unique SMRT-overlap sites, and rare instances (24 and 25, respectively, from each ICDS round) where multiple ICDS calls overlapped with a single SMRT site.



Supplementary Figure 6: Construction and PCR identification of *E. coli* B7A Δ*dndFGH* and *E. coli* B7A Δ*dndBCDEFGH*, and PT-dependent restriction phenotype in *E. coli* B7A. (A) The construction of recombinant plasmid pJTU6601 and mutant strain ZXQ-1. PCR confirmations of mutant strain ZXQ-1 with *ca.* 1.789-kb band using primers of B7A-FLL-L/ B7A-HRR-R, while the wild-type B7A with *ca.* 1.2-kb band using primers of B7A-T1/B7A-T2. (B) The construction of recombinant plasmid pJTU6602 and mutant strain ZXQ-2. PCR confirmations of mutant strain ZXQ-2 with *ca.* 2.346-kb band using primers of B7A-FLL-L/ B7A-BRR-R, while the wild-type B7A with *ca.* 1.2-kb band using primers of B7A-T1/B7A-T2. In lower panel, plasmid pBluescript SK^+ lacking PT was restricted by *E. coli* B7A wild-type cells (C), but not by Δ*dndF-H* mutant cells (E) during transformation. Plasmid pBluescript SK^+ with PT+ was transformed into *E. coli* B7A wild-type cells (D) and Δ*dndF-H* mutant cells (F) with a similar efficiency.



Supplementary Figure 7: LC-MS/MS analysis of PT-containing dinucleotides from *in vitro* reactions of a GAAC/GTTC-containing duplex oligodeoxynucleotide with cell-free extract from *S. enterica* serovar 87. *Salmonella enterica* serovar Cerro 87 has Dnd proteins highly homologous to those in *E. coli* B7A, with modification of a GAAC/GTTC consensus sequence. Cell-free extract prepared from *S. enterica* was reacted with a 31-mer duplex oligodeoxynucleotide containing a known GAAC/GTTC-containing PT modification site from *S. enterica*. Following annealing and immobilization of the biotinylated duplex oligo on streptavidin-agarose beads, cell-free extracts were added along with cofactors ATP, L-cysteine, and pyridoxal phosphate for a 1-hr reaction, followed by enzymatic release and LC-MS/MS analysis of PT-containing dinucleotides. (**A**) Reaction of cell-free extract with a PT-containing duplex oligodeoxynucleotide (SE1+PT/SE2+PT; see **Table S5**). (**B**) Reaction of cell-free extract with a duplex oligodeoxynucleotide (SE1/SE2; see **Table S5**). (**C**) Analysis of cell-free extract lacking oligodeoxynucleotides. (**D**) Analysis of a reaction lacking cell-free extract and containing the oligodeoxynucleotides SE1/SE2 (**Table S5**).

Supplementary Table 1: Kinetic signals from SMRT sequencing at 5'-GAAC-3'/3'-CTTG-5' sequence contexts for a pBluescript SK^+ plasmid grown in *Salmonella enterica* serovar Cerro 87, using a *Δdnd* knockout strain as the control.

Position	Strand	Motif	IPD ratio
141	+	GTTC	1.00
142	-	GAAC	0.97
152	+	GAAC	0.99
153	-	GTTC	1.00
173	+	GAAC	1.21
174	-	GTTC	1.14
234	+	GAAC	1.02
235	-	GTTC	0.98
290	+	GAAC	1.02
291	-	GTTC	1.09
338	+	GAAC	3.01
339	-	GTTC	2.17
730	+	GTTC	1.44
731	-	GAAC	1.77
770	+	GTTC	1.01
771	-	GAAC	1.03
1072	+	GTTC	1.63
1073	-	GAAC	3.41
1152	+	GAAC	1.04
1153	-	GTTC	1.04
1185	+	GAAC	1.21
1186	-	GTTC	1.18
1342	+	GTTC	1.19
1343	-	GAAC	1.37
1433	+	GTTC	1.27
1434	-	GAAC	1.34
1447	+	GTTC	1.06
1448	-	GAAC	1.02
1474	+	GAAC	1.33
1475	-	GTTC	1.27
1483	+	GTTC	1.33
1484	-	GAAC	1.75
1617	+	GTTC	1.16
1618	-	GAAC	1.13
1845	+	GAAC	1.08
1846	-	GTTC	1.03
2024	+	GTTC	1.04
2025	-	GAAC	1.04
2249	+	GTTC	1.44
2250	-	GAAC	1.71
2342	+	GTTC	1.75
2343	-	GAAC	2.66
2618	+	GAAC	2.24
2619	-	GTTC	2.17
2648	+	GTTC	1.10
2649	-	GAAC	1.17
2696	+	GTTC	1.34
2697	-	GAAC	1.45
2932	+	GTTC	1.08
2933	-	GAAC	1.05

Supplementary Table 2: Statistical analysis of the frequency of PT modification of GAAC/GTTC sites. Candidate PT-modified sites called from both methods were mapped with respect to genomic annotations reported in Supplementary Data 1 using bedtools. Pairwise comparisons between the number of sites mapping to genomic features were performed using a chi-square test for each method/run individually, and p-values were adjusted for multiple testing using the Benjamin-Hochberg procedure.

Gene class	# of genes in each class	Total DNA length (bp)	Average length per gene (bp)	# GAAC/GTTC sites	Average # GAAC/GTTC sites per gene	Average # GAAC/GTTC sites per kb of DNA
tRNA	86	6658	77	118	1.4	17.8
rRNA	22	32196	1463	274	12.5	11.0
ORF	5384	4558294	847	36637	6.8	7.8
Intergenic	1	672147	672147	3690	-	5.5

	SMRT-Seq			ICSD 1st run			ICSD 2nd run		
Gene class	# PT-modified GAAC/GTTC sites	Average # PT-modified sites per gene	Proportion modified	# PT-modified GAAC/GTTC sites	Average # PT-modified sites per gene	Proportion modified	# PT-modified GAAC/GTTC sites	Average # PT-modified sites per gene	Proportion modified
tRNA	3	0.03	0.03	6	0.07	0.05	3	0.03	0.03
rRNA	25	1.14	0.09	64	2.91	0.23	51	2.32	0.19
ORF	4499	0.84	0.12	6955	1.29	0.19	5933	1.10	0.16
Intergenic	329	-	0.09	524	-	0.14	461	-	0.12

Benjamin-Hochberg-corrected p-value for difference in PT modification between regions						
Region1	Region2	SMRT	ICDS 1st	ICDS 2nd	Pooled OR	Paired T-test p-value
tRNA	rRNA	0.062	9.61E-04	9.81E-04	0.18	0.094
tRNA	ORF	8.29E-03	2.27E-03	1.17E-03	0.20	0.035
tRNA	Intergenic	0.049	0.026	0.024	0.28	0.032
rRNA	ORF	0.194	0.18	0.19	0.99	0.64
rRNA	Intergenic	1	1.86E-03	1.10E-04	1.78	0.23
ORF	Intergenic	5.00E-07	1.96E-08	3.69E-15	1.41	0.069

Supplementary Table 3: Oligodeoxynucleotides used in gene inactivation, mutant confirmation, SMRT and ICDS sequencing and *in vitro* PT modification reactions

Name	Sequence
B7A-FLL	5′-CGGGATCCAATCCTCA AACAGGTTCACA-3′
B7A-FLR	5′-CTCCCAGAATGCTTTACACGATCGCGCTCATAACTTCAT-3′
B7A-HRL	5′-ATGAAGTTATGAGCGCGATCGTGATAAAGCATTCTGGGAG-3′
B7A-HRR	5′-CAGGTCGACCCTTCTCAAGCATGTAGCAA-3′
B7A-FLL-1	5′-CGGGATCCAATCCTCAAACAGGTTCACA-3′
B7A-FLR-1	5′-CTTAAAGCGTAGCCAGCAGATGTTCGATCGCGCTCATAACTTCAT-3′
B7A-BRL-1	5′-ATGAAGTTATGAGCGCGATCGAACATCTGCTGGCTACGCTTTAAG-3′
B7A-BRR-1	5′-CAGGTCGACGCATGTCTGCACGGTTGCTCT-3′
B7A-FLL-L	5′-TCCAGTAACCGTGTCCAT-3′
B7A-HRR-R	5′-ATCACTAAGGCTATCGTCTA-3′
B7A-BRR-R	5′-TAACCCTTACGGCAACGA-3′
B7A-T1	5′-CCCCGCTGTTCTTGCTGT-3′
B7A-T2	5′-CTCCGGTGCAATTTCTAG-3′
5′-FWD Tag	5′- /Phos/TTTAACCGCGAATTCCAG/dideoxyC/-3′
3′-REV Tag	5′- GCTGGAATTCGCGGTTAAAT-3′
15-TACCGC PCR Primer 2.0	5′-CAAGCAGAAGACGGCATACGAGATGCGGTACGGTCTCGGCATTCCTGCTGAACC GCTCTTCCGATCTGCTGGAATTCGCGGTTAAA-3′
16-ATGATA PCR Primer 2.0	5′-CAAGCAGAAGACGGCATACGAGATTATCATCGGTCTCGGCATTCCTGCTGAACC GCTCTTCCGATCTGCTGGAATTCGCGGTTAAA-3′
Paired-End PCR Primer 1.0	5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCT-3′
Custom Paired-End Read 2	5′-GCTGAACCGCTCTTCCGATCTGCTGGAATTCGCGGTTAAA-3′
PT-S4	5′-TCCAGTTTGG$_{PS}$AACAAGAGTC-3′
PT MALDI-Fwd	5′-GGCTCTTACGTCCAGTTTGG$_{PS}$AACAAGAGTCGCCGTTAAATTTCCGGCG-3′
PT MALDI-Rev	5′-CGCCGGAAATTTAACGGCGACTCTTG$_{PS}$TTCCAAACTGGACGTAAGAGCC-3′
SE1+PT	5′- (5-Biotin-G)TCGTGGTTGGCGACG$_{PS}$AACACCAGACCGTTA-3′
SE2+PT	5′- TAACGGTCTGGTG$_{PS}$TTCGTCGCCAACCACGA-3′
SE1	5′- (5-Biotin-G)TCGTGGTTGGCGACGAACACCAGACCGTTA-3′
SE2	5′- TAACGGTCTGGTGTTCGTCGCCAACCACGA-3′

Supplementary Table 4: ICDS mapping statistics for two replicate analyses of the *E. coli* B7A genome

	ICDS Replicate #1		ICDS Replicate #1		ICDS Replicate #2		ICDS Replicate #2	
	Iodine-treated		EtOH control		Iodine-treated		EtOH control	
Metrics	**Numbers**	**%**	**Numbers**	**%**	**Numbers**	**%**	**Numbers**	**%**
Number of Paired reads	15293948	100	12196661	100	8185279	100	7879595	100
Aligned concordantly 0 times	7076323	46.3	6081573	49.9	1794354	21.9	1604647	20.4
Aligned concordantly exactly 1 time	7654806	50.0	5682577	46.6	5958933	72.8	5790423	73.5
Aligned concordantly >1 times	562819	3.68	432511	3.55	431992	5.28	484525	6.15
Number of pairs aligned concordantly 0 times	7076323		6081573		1794354		1604647	
of which, aligned discordantly 1 time	55373	0.78	44906	0.74	12535	0.07	17626	1.1
----	----		----		----		----	
Number of pairs aligned 0 times concordantly or discordantly	7020950	46.3	6036667		1781819		1587021	
Number of mates making up the pairs	14041900		12073334		3563638		3174042	
Aligned 0 times	7271050	51.8	7054830	58.4	2402262	67.4	2137193	67.3
Aligned exactly 1 time	6298985	44.9	4700139	38.9	1077265	30.2	957836	30.2
Aligned >1 times	471865	3.36	318365	2.64	84111	2.36	79013	2.49
Overall alignment rate		76.2		71.1		85.3		86.4