OXFORD UNIVERSITY PRESS | Nucleic Acids Research

Some supplementary files may need to be viewed online via your Referee Centre at http://mc.manuscriptcentral.com/nar.

DNA target recognition domains in the Type I restriction and modification systems of Staphylococcus aureus

Journal:	*Nucleic Acids Research*
Manuscript ID	NAR-03656-H-2016.R1
Manuscript Type:	1 Standard Manuscript
Key Words:	DNA restriction, DNA modification, Staphylococcus aureus, epigenetics, endonuclease

DNA target recognition domains in the Type I restriction and modification systems of *Staphylococcus aureus*.

Laurie P. Cooper[1+], Gareth A. Roberts[1+], John H. White[1+], Yvette Luyten[2+], Edward K.M. Bower[1], Richard D. Morgan[2], Richard J. Roberts[2], Jodi A. Lindsay[3], David T.F. Dryden[4*]

1. EaStCHEM School of Chemistry, University of Edinburgh, The King's Buildings, Edinburgh EH9 3FJ, UK.
2. New England Biolabs, 240 County Road, Ipswich, MA 01938-2723, USA.
3. Institute of Infection and Immunity, St George's, University of London, Cranmer Terrace, London, SW17 0RE, UK.
4. Department of Biosciences, Durham University, Stockton Road, Durham, DH1 3LE, UK.

+ Joint first authors
* Author for correspondence:
david.t.dryden@durham.ac.uk, Tel. +44 (0)191 3341200

Keywords: DNA restriction, DNA modification, epigenetics, *Staphylococcus aureus*, Type I restriction system, endonuclease, methyltransferase.

Abbreviations: Restriction and Modification, RM; Horizontal Gene Transfer, HGT; hsd, host specificity for DNA; Methyltransferase, MTase; N6-methyl adenine, m6A; N4-methyl cytosine, m4C; C5-methyl cytosine, 5mC; Clonal Complex, CC; Sequence Type, ST; Target Recognition Domain, TRD; Single Molecule Real Time, SMRT; Enhanced Green Fluorescent Protein, EGFP; Polymerase Chain Reaction, PCR; ATP hydrolysis, ATPase; Lysogeny Broth, LB; S-adenosyl-L-methionine, SAM.

Abstract

Staphylococcus aureus displays a clonal population structure in which horizontal gene transfer between different lineages is extremely rare. This is due, in part, to the presence of a Type I DNA restriction and modification (RM) system given the generic name of *Sau1*, which maintains different patterns of methylation on specific target sequences on the genomes of different lineages. We have determined the target sequences recognised by the *Sau1* Type I RM systems present in a wide range of the most prevalent *S. aureus* lineages and assigned the sequences recognised to particular target recognition domains within the RM enzymes. We used a range of biochemical assays on purified enzymes and single molecule real-time sequencing on genomic DNA to determine these target sequences and their patterns of methylation. Knowledge of the main target sequences for *Sau1* will facilitate the synthesis of new vectors for transformation of the most prevalent lineages of this "untransformable" bacterium.

Introduction

Type I DNA restriction-modification (RM) systems are found in about half of the sequenced prokaryotic genomes (1-4). They present a formidable barrier to the invasion of the host cell by foreign DNA whether by transduction, transformation or conjugation and thus exercise control over horizontal gene transfer (HGT) (1,4-8). As an example of their effectiveness, less than 1 in 10^4 or 10^5 phage infections can successfully avoid the classical EcoKI Type I RM system of *Escherichia coli* K12. In some circumstances, such as when antirestriction systems are absent (9), when there are multiple target sites on the phage (10) or when RM expression is raised (11), the barrier due to this single RM system can be even greater. RM systems operate by methylating defined target sequences on the host genome and they maintain this methylation pattern through each round of DNA replication (modification). Foreign DNA entering the cell often contains the same target sequence but in an unmethylated state. These unmethylated target sequences are targeted for endonucleolytic cleavage by the RM system (restriction). The Type I RM system comprises three *hsd* (host specificity for DNA) genes, *hsdR*, *hsdM* and *hsdS* for restriction, modification and target sequence specificity respectively. The gene products form an $R_2M_2S_1$ complex in which HsdS (or S) recognises the target sequence, HsdM (or M) recognises the methylation status of the target and methylates hemimethylated targets while HsdR (or R) cleaves the DNA containing unmethylated targets after a complex reaction involving ATP hydrolysis and DNA translocation (12). An M_2S_1 complex can act solely as a methyltransferase (MTase) (13). Type I RM enzymes almost always recognise and methylate adenine nucleotides in their target sequences to form N6-methyl adenine (6mA) although a few forming N4-methyl cytosine (m4C) are now known (3,14). In addition to the protection offered by Type I, II and III RM systems, Type IV restriction systems can attack foreign DNA containing methylated sequences not found in the host (15).

The presence of multiple RM systems in a single host can increase the barrier to HGT still further. For instance, *Staphylococcus aureus* often contains two related Type I RM systems making its transformation extremely inefficient and hindering the genetic analysis of this organism (16-19). These genomes contain two *hsdM* and two *hsdS* and share a single *hsdR*, although some *S. aureus* strains have different numbers of *hsdM* and *hsdS* (Figure 1a). The presence of only a single *hsdR* is not a problem as it can interact with each *hsdMhsdS* pair. It has long been known that *S. aureus* displays a clonal population structure (20) in which HGT between different clonal complexes is exceedingly rare. Multi-locus sequence typing, microarray analysis and whole genome sequencing divides lineages of *S. aureus* and close relatives into the clonal complexes (CC) (20-23), each of which carries a different range of mobile genetic elements and antibiotic resistance genes on the genome (24-27). Each CC can be further subdivided into sequence types (ST) (22). Waldron and Lindsay (16) first realised that each CC of *S. aureus* contained a unique pair of Type I RM systems. A Type IV restriction system, SauUSI, was also identified later and recognised as a methyl-dependent restriction enzyme which would prevent the uptake of foreign DNA containing C5-methyl cytosine (5mC) (28,29). Thus most genetic manipulation of *S. aureus* is confined to strain RN4220, which has a defective Type I RM system due to a premature stop codon in *hsdR*. Furthermore, to avoid the Type IV system, DNA needs to be prepared from an *Escherichia coli* strain, such as *E. coli* ER2796, lacking the Dcm 5mC MTase (30).

The Type I RM systems in different strains of *S. aureus* were given the informal name of *Sau1* by Waldron and Lindsay (16) and it is clear from not only a comparison of the sequences of genes and proteins but also from the ability to use subunits from one strain to complement subunits from other strains (31) that the term *Sau1* describes a classical "family" of Type I RM systems. Type I RM families, Type IA to Type IE, were originally defined in *E. coli* and *Salmonella enterica* by DNA hybridisation, antibody cross reactivity and subunit complementation (32,33), although now it is more usual to use the high levels of sequence identity (over 90%) in HsdM and HsdR to define a family *in silico*. Although the name *Sau1* for this family of Type I RM systems in *S. aureus* is an

informal one not following the usual conventions (34), we retain it as it is established in the literature. However, it is important to note that some strains of *S. aureus* show additional Type I RM systems, which show limited amino acid sequence identity to the HsdR, HsdM and HsdS of *Sau1* (Figure 1a). For instance, Monk *et al.* (35) identified an active Type I RM system, SauJKDIII, in *S. aureus* JKD6159 which showed low sequence identity to members of the *Sau1* family. This is clearly a member of a new and different Type I RM family whose subunits will be unable to interact with the *Sau1* HsdM and HsdR (DTFD, JAL and MTG Holden, manuscript in preparation).

The *Sau1* Type I RM systems are so effective because they show great variability in the target sequences recognised thus preventing HGT between CC but allowing HGT between strains within a CC (31,35,36). This variability in target sequences is due to the modular construction of the Type I RM systems (Figure 1b). The S subunit contains two target recognition domains (TRD) each of which recognises one half of a bipartite target, for example the first Type I RM system in CC1, given the generic name CC1-1, recognises CCAYNNNNNTTAA (adenine methylation sites are underlined) (35,36). Swapping TRDs between S subunits generates new targets, for example the second Type I RM enzyme in CC1, termed CC1-2, couples the first TRD of CC1-1 with a different second TRD to recognise CCAYNNNNNNTGT. This swapping is easy because the DNA for S subunits contain conserved sequences bounding each TRD. Most *S. aureus* strains have two copies of *hsdS*, two of *hsdM* and one of *hsdR*. Thus there are often four TRDs in each CC, which define the restriction barrier against HGT. Some Type I RM enzymes have half-size HsdS incorporating only a single TRD. It has been shown that these products are often able to dimerise and recognise symmetric target sequences (37-39). We have been able to recapitulate these results on "half-HsdS" enzymes by manipulating the CC398-1 *S. aureus* system (EKM Bower and DTFD, unpublished results).

Previously we have identified the target sequences recognised by several common community-associated, hospital-associated and livestock-associated) MRSA clonal complexes (31,36) and recently several more have been identified (3,35,40). Monk *et al.* (35) and Jones *et al.* (40) have used this information to prepare DNA methylated by the MTase M_2S_1 component enzymes to aid the transformation of *S. aureus* strains that are usually resistant to transformation.

The identification of further targets recognised by the S subunits of *Sau1* Type I RM systems would in principle allow more CC to be transformed for genetic analysis. In addition, further understanding of the structural requirements for TRDs to recognise different specific DNA sequences is of intense interest as the Type I RM systems are very widespread in bacteria and archaea (1,4) and exert a considerable pressure on HGT and the evolution of prokaryotes. For instance, the use of multiple TRDs being exchanged between strains has been observed in *Helicobacter* (41), *Mycoplasma* (42,43), *Streptococci* (44,45) and *Bacteroides* (46).

Here we identify many further TRDs and their targets using both biochemical and PacBio single-molecule real-time (SMRT) sequencing methods to define the barriers to HGT in a wide range of *S. aureus* CC of global importance.

Materials and Methods

Nomenclature for expression plasmids encoding new MTases.

As each Type I S subunit contains two TRDs and we propose to determine the targets recognised by each TRD, we have given each TRD a single letter code, Table 1, and refer to the plasmids as pSauTRD1-TRD2, e.g. pSauBI expresses an S subunit containing TRD B and TRD I and the M subunit. If the TRD combination is the same as that found in a known clonal complex, then that CC is also given in brackets. The MTase would be called M.SauBI in this example and the S subunit S.SauBI and is from CC22. All sequences are given in the supplementary information.

Preparation of M.SauBI (CC22-1), M.SauCD (CC30-1), M.SauJK (CC30-2) and M.SauCL (CC45-1).
These four MTases were prepared as EGFP-His tag fusions as described in Roberts *et al.* (31). pSauBI-EGFP (CC22-1, genomic DNA from MRSA5906), pSauCD-EGFP (CC30-1, genomic DNA from MRSA252), pSauJK-EGFP (CC30-2, genomic DNA from MRSA252) and pSauCL-EGFP (CC45-1, genomic DNA from strain 70642) were all constructed by the polymerase chain reaction (PCR) with their *hsdS* fused to DNA encoding EGFP and a His-tag, with the following locus-specific oligonucleotides priming from the 3' end of the genes encoding the S subunits:
CC22-1 BI BS 5'GATCGAATTCCGGATCCAATAAACATCTTTTGTAAAAACAC3'
CC30-1 CD BS 5'GATCGAATTCCGGATCCTAAGAACATCTTTTGTAAAAAGG3'
CC30-2 JK BS 5'GATCGAATTCCGGATCCTATAAAAATTTTTTGAAGTAATCCTTG3'
CC45-1 CL R167K BS 5'GATCGAATTCCGGATCCAATAAACATCGATTTAAGTAAGGC3'
The sequence for CC45-1 introduced a single mutation R167K in the first TRD in the S subunit but since this change is found in other *S. aureus* isolates containing this TRD, the change is presumed to be completely neutral.

A new vector for MTase expression: pJF118his.
Although we had not experienced problems in examining the fusion proteins of S subunits and EGFP in biochemical work, we decided to construct a vector encoding *hsdS* with only a C-terminal His-tag. Vector pJF118his was made by PCR of the plasmid encoding the MTase CC5-1-EGFP constructed in Roberts *et al.* (31) with these two oligonucleotides:
pJFMShisTS 5'AGCTTCGAGAGGATCCCATCATCATCATCATCATTAAGAATTCAGCTTGGCTGTTTTGGCGG3'
and pJFMSEGFPhisBS 5'GAGTGAATCCCCGGGGATCCGTCGACC3'.
The resulting PCR product was cut with BamHI and unimolecular religation gave pJF118his into which the *hsdMS* operon could be ligated as BamHI fragments and from which all subsequent MTase clones were descended.

Construction of an MTase plasmid to allow TRD swaps: pSaudeltaXmaI.
A PCR-based strategy was devised to allow free pairwise assortment of desired TRDs in HsdS. Many, but not all of the HsdS subunits, including that encoded by the Type I system in CC398 (36), have a predicted proline-glycine sequence near the N-terminus. This dipeptide can be encoded by CCCGGG, which would be a target site for SmaI or XmaI. Oligonucleotides were designed which would introduce this motif in the N-terminus (a replacement with no amino acid changes) and at the C terminus (an insertion of two amino acids) of the S subunit of the CC398 system (36), by a two stage PCR fusion. Thus, primary PCR products were generated by reactions primed by: PromoterJF
5'GCTTCTGGCGTCAGGCAGCC3' with 398SmaIOligoBS
5'CCCATTCGCCTTCAAACCCGGGGAATCTCAACTCTGGCAC3' and 398SmaIOligoTS
5'GTGCCAGAGTTGAGATTCCCCGGGTTTGAAGGCGAATGGG3' with 398SmaIBamHI
5'GATCGATCGGATCCCCCGGGAATAAACATCTTTTGAAGTAATGAC3'.
The purified PCR products were then fused in a secondary PCR reaction primed by PromoterJF with 398SmaIBamHI. The product was then cut with BamHI, and ligated into the BamHI site of pJF118his as pSauNE-XmaI. This mutated form of the CC398-1 MTase, could assemble the complete restriction enzyme that proved to be active in endonucleolytic cleavage (36). This indicated that insertion of a proline and glycine towards the C-terminus did not affect the function of the enzyme. Subsequently, on reanalysing the DNA sequence, a single PCR mutation was discovered within the XmaI fragment. This caused a mutation A50S but this clearly did not affect the specificity or function of the S subunit in our assays. Digestion of pSauNE-XmaI with XmaI followed by intramolecular religation of the vector fragment generates pSaudeltaXmaI, into which any pairwise combination of TRDS with XmaI cohesive ends may be inserted.

Construction of MTases M.SauNI, M.SauND, M.SauNK, M.SauNL, M.SauBE, M.SauJE and M.SauCE (ST425-1) containing hybrid S subunits.

The DNA for each TRD of these S subunits was fused to the DNA for the reciprocal TRD of S.SauNE (CC398-1). This was achieved by creating primary PCRs with a short area of homology, which then allowed base pairing of single strands of each PCR, in a secondary PCR. For example, S.SauBE TRD B was generated from an appropriate plasmid template by PCR with oligonucleotides, TRD1FOR398SmaIOligoTS 5'GTGCCAGAGTTGAGATTCCCCGGGTTTGAAGGCGAATGGG3' paired with TRD1nearuniversal 5'GTTCTTCTAATTCAATTTGT3'. TRD E was similarly generated by PCR from plasmid template with oligonucleotides TRD2nearuniversal 5'ACAAATTGAATTAGAAGAAC3' and 398SmaIBamHI 5'GATCGATCGGATCCCCCGGGAATAAACATCTTTTGAAGTAATGAC3'. The final insert was then generated by PCR with the two gel-purified primary oligonucleotides and TRD1FOR398SmaIOligoTS 5'GTGCCAGAGTTGAGATTCCCCGGGTTTGAAGGCGAATGGG3' and 398SmaIBamHI 5'GATCGATCGGATCCCCCGGGAATAAACATCTTTTGAAGTAATGAC3'. S.SauCL was the only subunit for which we could not use the central universal oligonucleotides for PCR and required specific substitutes: TRDLFOR/CC45-1 5'ACAAATTGAATTAGAAGAACAAAAACTTGAATTACTTCAACAACAG3' and TRDC/CC45-1 5'GTTCTTCTAATTCAATTTGTCGATCGAGTTTGCTGAAGAAG3'. Each C-terminus is unique and where TRD2 was not TRD E, a specific oligonucleotide was employed: TRDIREV/CC22-1c-termsmaI 5'GATCGATCGGATCCCCCGGGAATAAACATCTTTTGTAAAAACAC3', TRDDREV/CC30-1c-termsmaI 5'GATCGATCGGATCCCCCGGGTAAGAACATCTTTTGTAAAAAGGATTG3', TRDKREV/CC30-2c-termsmaI 5'GATCGATCGGATCCCCCGGGTATAAAAATTTTTTGAAGTAATCCTTG3' and TRDLREV/CC45-1c-termsmaI 5'GATCGATCGGATCCCCCGGGAATAAACATCGATTTAAGTAAGGC3'. Each pure secondary PCR product was cut with XmaI and ligated into the XmaI site of pSaudeltaXmaI.

Construction of further MTases with further combinations of TRDs using synthetic genes.

Additional *hsdS* sequences were obtained as synthetic genes from GeneArt (ThermoFisher Scientific) with sequences optimised for expression in *E. coli* (Supplementary information). All the first TRDs begin with 5'CCCGGGTTTGAAGGCGAATGGGAG3', except that for CC80-2 which begins with 5'CCCGGGTTTGAAGGCGAATATTCT3'. All the first TRDs end with 5'CAAATTGAATTAGAAGAACAGAAG3'. All the second TRDs begin with 3'CAAATTGAATTAGAAGAACAGAAG5' and have a universal reverse oligonucleotide, Trd2unirev 5'GATCGATCGGATCCCCCGGG3'. These conserved sequences were used to create oligonucleotides to prime PCR reactions. Each pure secondary PCR product was cut with XmaI and ligated into the XmaI site of pSaudeltaXmaI. The orientation of the fragments was determined by PCR.

Expression and purification of MTases.

These new MTases and the R subunit of CC5 were expressed in *E. coli* BL21(DE3) and purified via HisTrap chromatography, size exclusion chromatography, diethylaminoethyl (DEAE) anion exchange chromatography and, if necessary, Heparin HiTrap chromatography (GE Healthcare, Uppsala, Sweden) as described previously (31).

Nuclease and ATPase assays.

Purified MTases were mixed with the CC5 R subunit and used in assays for ATP hydrolysis (ATPase) activity (coupled enzyme assay following a change in absorbance of NADH) and DNA cleavage activity (plasmid cutting assay with analysis via agarose gel electrophoresis) as previously described (31,36).

Preparation of genomic DNA for SMRT sequencing.

The expression plasmids harbouring the various MTases were used to transform a non-methylating (dam^- dcm^-) strain of *E. coli* ER2796 (30). Single colonies from the transformation plate of Lysogeny Broth (LB) agar medium supplemented with 10 µg/ml kanamycin, 10 µg/ml tetracycline as well as

100 µg/ml carbenicillin, which acted as a selection marker for the expression construct, were picked and used to inoculate 5 mL of LB containing the same cocktail of antibiotics. The cultures were incubated overnight with shaking at 37°C and 1 mL aliquots of the overnight culture were then pelleted by centrifugation (6000 g, 6 min, 4°C). The culture medium was carefully removed and the cell pellets stored at -20°C until required. Genomic DNA was prepared from each cell pellet using the Wizard Genomic DNA purification kit (Promega, Madison, WI) according to the manufacturer's instructions. The quality of the genomic DNA preparations was initially assessed by agarose gel electrophoresis and from the shape of the absorbance profile from 240 to 340 nm. Genomic DNA from *S. aureus* strains LGA251 (a kind gift from Mark Holmes) and NCTC13435 (a kind gift from Angela Kearns) was prepared by using the PurElut Bacterial Genomic Kit (EdgeBio, Gaithersburg, MD 20877, USA). The DNA library for SMRT sequencing was prepared and subsequently analysed as described in Anton *et al*. (30).

Methylation of plasmids using M.EcoGII.

M.EcoGII was kindly supplied by Dr. Iain Murray (New England Biolabs) and used to modify plasmids E2, E5, E10, E11 and E12 previously described (31) and plasmid pCN36 (47). 0.45 µg DNA was methylated using 2.0 U of M.EcoGII for 100 min at 37°C in a 50 µl volume. The reaction was in 1xNEB4 buffer (50 mM potassium acetate, 20 mM Tris acetate, 10 mM Mg acetate, 1 mM DTT (pH 7.9@25°C) supplemented with 320 µM S-adenosyl-L-methionine (SAM). As a negative control, DNA was incubated in the same buffer without M.EcoGII. The DNA samples were then supplemented with ATP (20 µM) and additional SAM (160 µM) and then digested with a Type I enzyme (CC5-1, CC5-2, CC30-1, CC45-1 or the NY TRD hybrid) for 14 min at 37°C. As a control, methylated and unmethylated DNA was digested with EcoRI.

Results and Discussion

Assigning TRDs to target sequences.

Each TRD was given a one letter code (A to Z and a* to f*), Table 1. There were 14 TRD1 examples and 18 TRD2 examples in our survey and these are found in 17 different CC or ST groups. Table 1 lists the target specificity and site of methylation for each TRD in our survey. These data were obtained by pairing TRDs and determining the complete target for each TRD pair as described in the next section and in full in the supplementary information. Of interest are the TRD pairs B and P and U and c*. These pairs recognise the same DNA sequence namely AGG and GAY respectively. Amino acid sequence comparisons of B with P and U with c* are shown in Figure 2.

TRD B and TRD P are virtually identical throughout the TRD region even though TRD B is the first TRD in the HsdS subunit and TRD P is the second TRD in the HsdS subunit, Figure 2a. While the high level of sequence identity is expected for Type I systems in the same family, the high level of identity between TRDs found in the first or second position in the HsdS subunit is more unusual. However, such a situation has previously been observed in comparisons of the Type I systems in *Salmonella blegdam* and *E. coli* R124 (48).

In contrast, TRDs U and c* are both examples of the second TRD in the HsdS subunit recognising 5'-GAY-3' but the level of identity between them is much lower (~36%) (Figure 2b). This level of identity between TRDs recognising the same target is expected if the TRDs are from different Type I RM families so the low level of identity observed here is unusual. Despite this low level of sequence identity, the predicted secondary structure elements are the same as expected from the early work of Sturrock and Dryden (49). In fact, all of the TRDs in the *Sau1* family of RM systems align well when secondary structure elements are taken into consideration (50) and they will have the same protein fold (Supplementary information: PROMALS alignments). Therefore, it should in future be possible

to predict the precise amino acid to nucleotide contacts involved in sequence recognition as was done for the Type IIG TRDs (51,52).

Determination of complete target sequences recognised by pairs of TRDs.

Tables 2, 3 and 4 show the TRD combinations investigated in this work and those investigated previously by ourselves and others along with their combined target sequences, methylation specificity and the methods used to determine these parameters. The full experimental data are given in the supplementary information. Many of the TRDs were investigated in more than one MTase and in more than one assay thus our set of data represents a self-consistent set. DNA cleavage and ATP hydrolysis assays were performed on purified MTases mixed with purified R subunit while SMRT data were collected from *E. coli* genomic DNA isolated after the hosts were transformed with a plasmid expressing the MTase or directly from *S. aureus* genomic DNA. The adenines targeted for methylation were determined easily by SMRT sequencing but for systems not examined in this manner, it was assumed if there was a single adenine in the site recognised by the TRD that this was the target for methylation.

Table 2 contains systems from a range of CC investigated previously as well as several examined in this study. It is important to note that in our work those systems containing M.SauMRSII plus S. SauMRSII, M.Sau133ORF1794P plus S.Sau133ORF1794P and M.SauMRSI plus S.SauMRSI are paired with the HsdR (SauN315ORF189P) from the N315 strain of CC5 in DNA cleavage and ATPase assays. Those shown in Tables 3 and 4 are studied as HsdS paired with the HsdM (M.SauSTORF499P) from strain S0385 of CC398 and the HsdR (SauN315ORF189P) from the N315 strain of CC5 (if used in DNA cleavage or ATPase assays). Therefore, these HsdS are not examined in the context of their natural genome, but since they are all from the *Sau1* family of Type I RM systems and the HsdM and HsdR of these RM systems are essentially identical in all of the strains, it is reasonable to assume that the target specificities identified are those that would be recognised in their natural host.

Identifying the complete target recognised by a member of the *Sau1* Type I RM family when both TRDs have unknown targets is difficult and ambiguous as either orientation may be correct. Hence, we combined TRDs with unknown targets with TRD E or TRD N to make a protein recognising a hybrid sequence in which one half of the target was already known(Table 3). A variety of methods were used to determine the target associated with each hybrid including DNA cleavage and ATP hydrolysis assays when the hybrid enzyme could be expressed and purified from *E. coli* and SMRT sequencing when the expression and purification levels were low, for example, the SauJK enzyme corresponding to the second Type I RM enzyme in CC30 did not express in *E. coli* despite its expression in *S. aureus* by Monk *et al.* (35). The ambiguity in assignment of targets in CC93 in Monk *et al.* (35) is resolved because the TRDs M and b* occur in more than one HsdS in our survey.

The DNA sequences for further pairs of TRDs found in a wide range of CC and ST groups were then inserted after the *hsdM* of CC398-1 in our expression vector and examined to ascertain the spacer sequence in the natural system (Table 4).

Genomic DNA from *S. aureus* strains NCTC13435 and LGA251 was prepared and examined using SMRT sequencing as these strains contain two TRD pairs, XY and e*f* respectively, which we could not express in *E. coli*. While SMRT signatures for the other Type I HsdS in these strains were very clear (Supplementary information) and in agreement with our results from *E. coli* (Table 4) and those of Monk *et al.* (35), these TRD pairs still showed no methylation activity even in their normal host. Thus these TRDs pairs are not active.

Analysis of spacer sequence length in *S. aureus* Type I RM systems.
It is apparent that the number of base pairs separating the adenines targeted for methylation and the number of base pairs in the non-specific spacer between the sequences recognised by the TRDs is not constant, with the former varying between 7 and 9 base pairs and the latter varying between 5 and 7 base pairs. This variation makes it very difficult to predict a Type I RM recognition sequence if one knows only the targets recognised by the two TRDs as the length of the spacer in the target is not recognised in any obvious manner by the TRDs. An example of this is the CC80-1 enzyme (Table 4) containing TRDs X and Y of known specificity. Since the enzyme did not methylate DNA *in vivo* for the SMRT analysis, the spacer and hence the complete target for CC80-1 remain unknown until the enzyme is purified and analysed biochemically. While it has been observed that insertions of multiples of four amino acids into the alpha helical spacers separating the TRDs can increase the length of the spacer in the target sequence in a predictable manner (65-67), it is clear from the structure of HsdS subunits (Figure 1b) that the junction between the TRDs and the alpha helical spacers in the conserved region is going to be of crucial importance for determining the fine details of the length of the spacer in the target sequence as was found for some Type IIB RM enzymes which contain a subunit equivalent to HsdS (68). Perhaps even single amino acid insertions or deletions will serve to rotate the TRD with respect to the rest of the subunit and thereby change the length of the spacer. Further progress in understanding the correlation between amino acid sequence and the length of the target spacer would be greatly aided by an accurate atomic structure of a Type I enzyme with DNA as the current models (12,13) lack sufficient resolution to be informative on this point.

Linking TRDs pairs to further clonal complexes and sequence types.
After determining the recognition sequences for all of the TRDs in Table 1 by creating artificial hybrids (Table 3) we also found that some of these TRD combinations do actually occur in natural systems as given in Table 5 (and supplementary information) (69). As sequence databases expand, more and more of the possible TRD combinations based on the TRDs in Table 1 will be found. As mentioned above, although the sequences recognised by the TRDs are known, the length of the non-specific spacer separating them is unknown so that the complete target cannot be specified accurately without experimentation.

Further TRDs in *S. aureus* Type I RM systems.
Searching the publicly available sequences in the NCBI database with individual TRD sequences revealed that some of those given in Table 1 can be found paired up with further novel TRDs. We have found four new TRDs shown in Table 6 in *S. aureus* strains 21343 and KPL1845. Strain 21343 contains “NOVEL 1” paired with TRD K and the TRD pair NQ described in Table 6. Strain KPL1845 also contains the TRD pair NQ and two further systems comprised of “NOVEL 2” paired with “NOVEL 3” and “NOVEL 4” paired with TRD f*. Undoubtedly further TRDs will be found as sequencing continues.

Improving transformation of *S. aureus* by avoiding targets recognised by the *Sau1* Type I RM family.
A general method of preparing DNA suitable for transformation of *S. aureus* which can overcome the RM barrier should be possible. Several DNA MTases belonging to Type II RM systems have been found which have extremely short target recognition sites, namely Hin1523, Nma1821 and Hia5 (70) and EcoGII recognising and methylating adenine in the targets 5'-A-3', 5'-AB-3' or 5'-BA-3'. The methylation performed by these enzymes should protect any DNA molecule from the RM enzymes described here (or indeed any RM barrier relying upon adenine methylation). Thus, DNA methylated *in vitro* with these unusual MTases could be used in subsequent transformation experiments even when major RM barriers are present.

We used the M.EcoGII adenine MTase (a kind gift from Iain Murray, New England Biolabs) to modify all adenines in several plasmids *in vitro*. The plasmids were from our collection of plasmids used to

determine the target sequences of the *S. aureus* Type I enzymes and have been previously described (31). These plasmids were then mixed with various purified *S. aureus* Type I restriction enzymes or, as a control, the EcoRI restriction enzyme. After one hour of methylation by M.EcoGII, the plasmids were completely resistant to digestion by EcoRI and by the *S. aureus* restriction enzymes (Figure 3). Furthermore, the shuttle vector pCN36 (47) was also protected from digestion by these same enzymes (data not shown). Subsequent experiments using the methylated pCN36 to transform *S. aureus* were unfortunately entirely unsuccessful (unpublished results by JAL using strains HO5096 (CC22), JE2 (CC8) and RN4220 (CC8, *hsdR*$^-$). The reason for the failure of transformation with the highly-methylated pCN36 when it should be resistant to all *Sau1* RM systems is not clear. This result may imply a further unrecognised barrier to transformation of *S. aureus* or some aspect of the physical properties of highly methylated DNA. Nevertheless, the method using MTases with very short target recognition sequences may be of use for transformation of other bacterial species.

CONCLUSIONS

In conclusion, we have determined the target recognition sequences of a considerable number of TRDs and HsdS specificity subunits of the Type I RM systems in *S. aureus*. This was achieved using a combination of gene synthesis, endonuclease activity, ATP hydrolysis activity and single molecule real-time genome sequencing. The systems analysed cover a large proportion of the known sequence types and clonal complexes of *S. aureus* and delineate more clearly the barrier to horizontal gene transfer within the *S. aureus* population.

The data obtained here will allow the construction of new *E. coli* strains for preparing methylated shuttle vectors (35) and MTase reagents for *in vitro* methylation of DNA (40) to assist transformation of further *S. aureus* strains. However, these approaches are time consuming and it is worth noting that the common shuttle vector used for transformation of *S. aureus*, pCN36 (47), contains a target site for almost every TRD pair investigated in this paper. This means that pCN36 is inevitably a poor vector for transformation of *S. aureus*. The construction of new shuttle vectors completely lacking *Sau1* targets via DNA synthesis, coupled with careful analysis of the fragments to be ligated into the vector so that they also lack targets, may be an effective way forward to improve transformation of *S. aureus* now that so many target specificities have been determined. Obviously, the avoidance of the sequence $AN_{6\text{-}9}T$, although difficult to achieve without altering protein coding sequences in a vector, would be a general method to negate the effect of the Type I RM systems in *S. aureus* and other prokaryotes.

Lastly, the determination of so many recognition sequences of Type I RM systems in different lineages of *S. aureus*, in effect a “Rosetta Stone”, means that now the population structure of *S. aureus* can be investigated from an epigenetic/evolutionary perspective (4) as performed previously with, for example, *H. pylori* (71) and *S. pneumoniae* (72).

Acknowledgements

DTFD thanks the Institute of Advanced Study, Durham University for providing a fellowship from January to April 2016 and an excellent environment for writing this paper. We thank Dr Iain Murray, New England Biolabs for supplying M.EcoGII, Mark Holmes for donating strain LGA251 and Angela Kearns for donating strain NCTC13435.

Source of funding

This work was supported by Biotechnology and Biological Sciences Research Council grant BB/K005804/1 to DTFD and the Wellcome Trust grants GR080463MA to D.T.F.D and 090288/Z/09/ZA to D.T.F.D. and J.A.L.

Figure legends

Figure 1. *S. aureus* genomes showing the genes and the typical organisation of TRDs in the HsdS DNA sequence specificity subunit. (a) Strain ED133 (CC133) has two *hsdS*; strain 11819-97 (CC80) has three *hsdS* (CC80) and strain JKD6159 (CC93) contains an extra Type I RM system from a different Type I RM family. From top to bottom: ED133, 11819-97, JKD6159. *hsdR* (red), *hsdM* (blue), *hsdS* (yellow). (b) The structural organisation of the HsdS specificity subunit. The conserved regions (cr) are common to all S subunits within a family. The two target recognition domains (TRD1 and TRD2) define the target sequences recognised by the RM enzyme and can be swapped between S subunits of the same family to generate new specificities.

Figure 2. Amino acid sequence and secondary structure alignment of two pairs of TRDs recognising the same DNA target. The TRD sequences are highlighted in yellow. Consensus secondary structure shows “h” for alpha helix and “e” for beta sheet. (a) TRDs B and P are examples of a first and a second TRD respectively recognising 5'-AGG-3'. (b) TRDs U and c* are both examples of second TRDs with the same specificity, 5'-GAY-3'. The long predicted alpha helices at the start and the end of the sequences are the conserved helical spacer regions in the HsdS subunits while the sequence between these helices makes up the TRD.

Figure 3. General protection from endonuclease activity using M.EcoGII MTase to methylate all adenines. Plasmid without M.EcoGII treatment is digested (- lanes) but plasmid with M.EcoGII treatment is protected from digestion (+ lanes). Panel (a) uses Sau347I (CC45-1, TRDs C and L) restriction enzyme against plasmids E2, E5 and E10 described in (31). Panel (b) uses SauNY (TRDs N and Y) against plasmids E10, E11 and E12 described in (31). Panel (c) uses three different enzymes, SauN315I (CC5-1, TRDs B and D), SauN315II (CC5-2, TRDs A and H) and SauMRSII (CC30-1, TRDs C and D), against plasmid E10. In each panel EcoRI restriction enzyme was used as a control and markers (M) are in kb.

REFERENCES

1) Oliveira, P.H., Touchon, M. and Rocha, E.P. (2014) The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res*. **42**, 10618-10631.
2) Roberts, R.J., Vincze, T., Posfai J. and Macelis, D. (2015) REBASE - a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res*. **43,** D298-D299.
3) Blow, M.J., Clark, T.A., Daum, C.G., Deutschbauer, A.M., Fomenkov, A., Fries, R., Froula, J., Kang, D.D., Malmstrom, R.R., Morgan, R.D., Posfai, J., Singh, K., Visel, A., Wetmore, K., Zhao, Z., Rubin, E.M., Korlach, J., Pennacchio, L.A. and Roberts, R.J. (2016) The Epigenomic Landscape of Prokaryotes. *PLoS Genet*.**12**, e1005854.
4) Oliveira, P.H., Touchon, M. and Rocha, E.P. (2016) Regulation of genetic flux between bacteria by restriction-modification systems. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5658-5663.
5) Loenen, W.A.M., Dryden, D.T.F, Raleigh, E.A., Wilson, G.G. and Murray, N.E. (2014) Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res*. **42,** 3-19.
6) Loenen, W.A.M., Dryden, D.T.F., Raleigh, E.A. and Wilson, G.G. (2014) Type I restriction enzymes and their relatives. *Nucleic Acids Res*. **42**, 20-44.
7) Pingoud, A., Wilson, G.G. and Wende, W. (2014) Type II restriction endonucleases--a historical perspective and more. *Nucleic Acids Res*. **42**, 7489-7527.
8) Rao, D.N., Dryden, D.T.F. and Bheemanaik, S. (2014) Type III restriction-modification enzymes: a historical perspective. *Nucleic Acids Res.* **42**, 45-55.
9) King, G. and Murray, N.E. (1995) Restriction alleviation and modification enhancement by the Rac prophage of Escherichia coli K-12. *Mol. Microbiol*. **16**, 769-777.
10) Murray, N.E., Batten, P.L. and Murray, K. (1973) Restriction of bacteriophage lambda by Escherichia coli K. *J. Mol. Biol*. **81**, 395-407.
11) Webb, J.L., King, G., Ternent, D., Titheradge, A.J.B. and Murray, N.E. (1996) Restriction by EcoKI is enhanced by co-operative interactions between target sequences and is dependent on DEAD box motifs. *EMBO J*. **15**, 2003-2009.
12) Kennaway, C.K., Taylor, J.E., Song, C.F., Potrzebowski, W., Nicholson, W., White, J.H., Swiderska, A., Obarska-Kosinska, A., Callow, P., Cooper, L.P., Roberts, G.A., Artero, J.B., Bujnicki, J.M., Trinick, J., Kneale, G.G. and Dryden, D.T.F. (2012) Structure and operation of the DNA-translocating Type I DNA restriction enzymes. *Genes Dev*. **26**, 92-104.
13) Kennaway, C.K., Obarska-Kosinska, A., White, J.H., Tuszynska, I., Cooper, L.P., Bujnicki, J.M., Trinick, J. and Dryden, D.T.F. (2009) The structure of M.EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein. *Nucleic Acids Res*. **37**, 762-770
14) Morgan, R.D., Luyten, Y.A., Johnson, S.A., Clough, E.M., Clark, T.A. and Roberts, R.J. (2016) Novel m4C modification in type I restriction-modification systems. *Nucleic Acids Res*. **44**, 9413-9425.
15) Loenen, W.A.M. and Raleigh, E.A. (2014) The other face of restriction: modification-dependent enzymes. *Nucleic Acids Res*. **42**, 56-69.
16) Waldron, D.E. and Lindsay, J.A. (2006) Sau1: a novel lineage-specific Type I Restriction-Modification system that blocks horizontal gene transfer into *Staphylococcus aureus*, and between *S. aureus* isolates of different lineages. *J. Bacteriol.,* **188**, 5578-5585.
17) Lindsay, J.A. (2010) Genomic variation and evolution of *Staphylococcus aureus*. *Intl J. Med. Microbiol.,* **300**, 98-103.
18) Monk, I.R., Shah, I.M., Xu, M., Tan, M.W. and Foster, T.J. (2012) Transforming the untransformable: application of direct transformation to manipulate genetically *Staphylococcus aureus* and *Staphylococcus epidermidis*. *Mbio.,* **3**, doi:pii: e00277-11.
19) Lindsay, J.A. (2014) *Staphylococcus aureus* genomics and the impact of horizontal gene transfer. *Intl. J. Med. Microbiol*. **304**, 103-109.

20) Feil, E.J., Cooper, J.E., Grundmann, H., Robinson, D.A., Enright, M.C., Berendt, T., Peacock, S.J., Smith, J.M., Murphy, M., Spratt, B.G., Moore, C.E. and Day, N.P. (2003) How clonal is *Staphylococcus aureus*? *J. Bacteriol.* **185**, 3307-3316.
21) Sung, J.M., Lloyd, D.H. and Lindsay, J.A. (2008) *Staphylococcus aureus* host specificity: comparative genomics of human versus animal isolates by multi-strain microarray. *Microbiol*. **154**, 1949-1959.
22) Monecke, S., Coombs, G., Shore, A.C., Coleman, D.C., Akpaka, P., Borg, M., Chow, H., Ip, M., Jatzwauk, L., Jonas, D., Kadlec, K., Kearns, A., Laurent, F., O'Brien, F.G., Pearson, J., Ruppelt, A., Schwarz, S., Scicluna, E., Slickers, P., Tan, H.L., Weber, S. and Ehricht, R.A. (2011) A field guide to pandemic, epidemic and sporadic clones of methicillin-resistant Staphylococcus aureus. *PLoS One*. **6**, e17936.
23) Méric, G., Miragaia, M., de Been, M., Yahara, K., Pascoe, B., Mageiros, L., Mikhail, J., Harris, L.G., Wilkinson, T.S., Rolo, J., Lamble, S., Bray, J.E., Jolley, K.A., Hanage, W.P., Bowden, R., Maiden, M.C., Mack, D., de Lencastre, H., Feil, E.J., Corander, J. and Sheppard, S.K. (2015) Ecological Overlap and Horizontal Gene Transfer in *Staphylococcus aureus* and Staphylococcus epidermidis. *Genome Biol. Evol*. **7**, 1313-1328.
24) McCarthy, A.J. and Lindsay, J.A. (2010) Genetic variation in *Staphylococcus aureus* surface and immune evasion genes is lineage associated: implications for vaccine design and host-pathogen interactions. *BMC Microbiol*. **10**, 173-187.
25) McCarthy, A.J., Lindsay, J.A. (2012) The distribution of plasmids that carry virulence and resistance genes in *Staphylococcus aureus* is lineage associated. *BMC Microbiol*. **12**, 104-111.
26) McCarthy, A.J., Witney, A.A. and Lindsay, J.A. (2012) *Staphylococcus aureus* temperate bacteriophage: carriage and horizontal gene transfer (HGT) is lineage associated. *Front. Cell. Infect. Microbiol*. **2**:6. doi: 10.3389/fcimb.2012.00006.
27) McCarthy, A.J., Loeffler, A., Witney, A.A., Gould, K.A., Lloyd, D.H. and Lindsay, J.A. (2014) Extensive horizontal gene transfer during *Staphylococcus aureus* co-colonization *in vivo*. *Genome Biol. Evol*. **6**, 2697-2708.
28) Corvaglia, A.R., François, P., Hernandez, D., Perron, K., Linder, P. and Schrenzel, J. (2010) A Type III-like restriction endonuclease functions as a major barrier to horizontal gene transfer in clinical *Staphylococcus aureus* strains. *Proc. Natl. Acad. Sci. U.S.A.,* **107**, 11954-11958.
29) Xu, S.Y., Corvaglia, A.R., Chan, S.H., Zheng, Y. and Linder, P. (2011) A type IV modification-dependent restriction enzyme SauUSI from *Staphylococcus aureus* subsp. aureus USA300. *Nucleic Acids Res*. **39**, 5597-5610.
30) Anton, B.P., Mongodin, E.F., Agrawal, S., Fomenkov, A., Byrd, D.R., Roberts, R.J. and Raleigh, E.A. (2015) Complete Genome Sequence of ER2796, a DNA Methyltransferase-Deficient Strain of Escherichia coli K-12. *PLoS One*. **10**, e0127446.
31) Roberts, G.A., Houston, P.J., White, J.H., Chen, K., Stephanou, A.S., Cooper, L.P., Dryden, D.T.F. and Lindsay, J.A. (2013) Impact of target site distribution for Type I restriction enzymes on the evolution of methicillin-resistant *Staphylococcus aureus* (MRSA) populations. *Nucleic Acids Res*. **41**, 7472-7484.
32) Barcus, V.A., Titheradge, A.J.B. and Murray, N.E. (1995) The diversity of alleles at the *hsd* locus in natural populations of *Escherichia coli*. *Genetics*. **140**, 1187-1197.
33) Titheradge, A.J.B., King, J., Ryu, J. and Murray, N.E. (2001) Families of restriction enzymes: an analysis prompted by molecular and genetic data for type ID restriction and modification systems. *Nucleic Acids Res.* **29**, 4195-4205.
34) Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., Blumenthal, R.M., Degtyarev, S.Kh., Dryden, D.T.F., Dybvig, K., Firman, K., Gromova, E.S., Gumport, R.I., Halford, S.E., Hattman, S., Heitman, J., Hornby, D.P., Janulaitis, A., Jeltsch, A., Josephsen, J., Kiss, A., Klaenhammer, T.R., Kobayashi, I., Kong, H., Krüger, D.H., Lacks, S., Marinus, M.G., Miyahara, M., Morgan, R.D., Murray, N.E., Nagaraja, V., Piekarowicz, A., Pingoud, A., Raleigh, E., Rao, D.N., Reich, N., Repin, V.E., Selker, E.U., Shaw, P.C., Stein, D.C., Stoddard, B.L., Szybalski, W.,

Trautner, T.A., Van Etten, J.L., Vitor, J.M., Wilson, G.G. and Xu, S.Y. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res*. **31**, 1805-1812.

35) Monk, I.R., Tree, J.J., Howden, B.P., Stinear, T.P. and Foster, T.J. (2015) Complete Bypass of Restriction Systems for Major *Staphylococcus aureus* lineages. *MBio*. **6**, e00308-15

36) Chen, K., Stephanou, A.S., Roberts, G.A., White, J.H., Cooper, L.P., Houston, P.J., Lindsay, J.A. and Dryden, D.T.F. (2016) The Type I Restriction Enzymes as Barriers to Horizontal Gene Transfer: Determination of the DNA Target Sequences Recognised by Livestock-Associated Methicillin-Resistant Staphylococcus aureus Clonal Complexes 133/ST771 and 398. *Adv. Exp. Med. Biol*. **915**, 81-97.

37) Abadjieva, A., Patel, J., Webb, M., Zinkevich, V. and Firman, K. (1993) A deletion mutant of the type IC restriction endonuclease EcoR1241 expressing a novel DNA specificity. *Nucleic Acids Res*. **21**, 4435-4443.

38) Meister, J., MacWilliams, M., Hübner, P., Jütte, H., Skrzypek, E., Piekarowicz, A. and Bickle, T.A. (1993) Macroevolution by transposition: drastic modification of DNA recognition by a type I restriction enzyme following Tn5 transposition. *EMBO J*. **12**, 4585-4591.

39) MacWilliams, M.P. and Bickle, T.A. (1996) Generation of new DNA binding specificity by truncation of the type IC EcoDXXI hsdS gene. *EMBO J*. **15**, 4775-4783.

40) Jones, M.J., Donegan, N.P., Mikheyeva, I.V. and Cheung, A.L. (2015) Improving transformation of *Staphylococcus aureus* belonging to the CC1, CC5 and CC8 clonal complexes. *PLoS One*.**10**, e0119487.

41) Furuta, Y., Kawai, M., Uchiyama, I. and Kobayashi, I. (2011) Domain movement within a gene: a novel evolutionary mechanism for protein diversification. *PLoS One*. **6**, e18819.

42) Dybvig, K., Sitaraman, R. and French, C.T. (1998) A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements. *Proc. Natl. Acad. Sci. U.S.A*. **95**, 13923-13928.

43) Xiao, L., Ptacek, T., Osborne, J.D., Crabb, D.M., Simmons, W.L., Lefkowitz, E.J., Waites, K.B., Atkinson, T.P. and Dybvig, K. (2015) Comparative genome analysis of *Mycoplasma pneumoniae*. *BMC Genomics*. **16**, 610. doi: 10.1186/s12864-015-1801-0.

44) Li, J., Li, J.W., Feng, Z., Wang, J., An, H., Liu, Y., Wang, Y., Wang, K., Zhang, X., Miao, Z., Liang, W., Sebra, R., Wang, G., Wang, W.C. and Zhang JR. (2016) Epigenetic Switch Driven by DNA Inversions Dictates Phase Variation in *Streptococcus pneumoniae*. *PLoS Pathog*. **12**, e1005762.

45) Willemse, N. and Schultsz, C. (2016) Distribution of Type I Restriction–Modification Systems in *Streptococcus suis*: An Outlook. *Pathogens*, **5**, 62; doi:10.3390/pathogens5040062.

46) Cerdeño-Tárraga, A.M., Patrick, S., Crossman, L.C., Blakely, G., Abratt, V., Lennard, N., Poxton, I., Duerden, B., Harris, B., Quail, M.A., Barron, A., Clark, L., Corton, C., Doggett, J., Holden, M.T.G., Larke, N., Line, A., Lord, A., Norbertczak, H., Ormond, D., Price, C., Rabbinowitsch, E., Woodward, J., Barrell, B. and Parkhill, J. (2005) Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. *Science* **307**, 1463-1465.

47) Charpentier E, Anton AI, Barry P, Alfonso B, Fang Y, Novick RP. Novel cassette-based shuttle vector system for gram-positive bacteria. Appl Environ Microbiol. 2004 Oct;70(10):6076-85.

48) Thorpe, P.H., Ternent, D. and Murray, N.E. (1997) The specificity of sty SKI, a type I restriction enzyme, implies a structure with rotational symmetry. *Nucleic Acids Res*. **25**, 1694-1700.

49) Sturrock, S.S. and Dryden, D.T.F. (1997) A prediction of the amino acids and structures involved in DNA recognition by type I DNA restriction and modification enzymes. *Nucleic Acids Res*. **25**, 3408-3414.

50) Pei, J., Kim, B.H., Tang, M. and Grishin, N.V. (2007) PROMALS web server for accurate multiple protein sequence alignments. *Nucleic Acids Res*. **35**, W649-652.

51) Morgan, R.D. and Luyten, Y.A. (2009) Rational engineering of type II restriction endonuclease

DNA binding and cleavage specificity. *Nucleic Acids Res*. **37**, 5222-5233.

52) Callahan, S.J., Luyten, Y.A., Gupta, Y.K., Wilson, G.G., Roberts, R.J., Morgan, R.D. and Aggarwal, A.K. (2016) Structure of Type IIL Restriction-Modification Enzyme MmeI in Complex with DNA Has Implications for Engineering New Specificities. *PLoS Biol*. **14**, e1002442.

53) Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K., Oguchi, A., Nagai, Y., Iwama, N., Asano, K., Naimi, T., Kuroda, H., Cui, L., Yamamoto, K. and Hiramatsu, K. (2002) Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet* **359,** 1819-1827.

54) Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Lian, J., Ito, T., Kanamori, M., Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A., Mizutani-Ui, Y., Takahashi, N.K., Sawano, T., Inoue, R., Kaito, C., Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J., Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C., Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H. and Hiramatsu, K. (2001) Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. *Lancet* **357**, 1225-1240.

55) Holden, M.T., Feil, E.J., Lindsay, J.A., Peacock, S.J., Day, N.P., Enright, M.C., Foster, T.J., Moore, C.E., Hurst, L., Atkin, R., Barron, A., Bason, N., Bentley, S.D., Chillingworth, C., Chillingworth, T., Churcher, C., Clark, L., Corton, C., Cronin, A., Doggett, J., Dowd, L., Feltwell, T., Hance, Z., Harris, B., Hauser, H., Holroyd, S., Jagels, K., James, K.D., Lennard, N., Line, A., Mayes, R., Moule, S., Mungall, K., Ormond, D., Quail, M.A., Rabbinowitsch, E., Rutherford, K., Sanders, M., Sharp, S., Simmonds, M., Stevens, K., Whitehead, S., Barrell, B.G., Spratt, B.G. and Parkhill, J. (2004) Complete genomes of two clinical Staphylococcus aureus strains: evidence for the rapid evolution of virulence and drug resistance. *Proc. Natl. Acad. Sci. USA* **101**, 9786-9791.

56) Chua, K., Seemann, T., Harrison, P.F., Davies, J.K., Coutts, S.J., Chen, H., Haring, V., Moore, R., Howden, B.P. and Stinear, T.P. (2010) Complete genome sequence of Staphylococcus aureus strain JKD6159, a unique Australian clone of ST93-IV community methicillin-resistant Staphylococcus aureus. *J. Bacteriol.* **192**, 5556-5557.

57) Guinane, C.M., Ben Zakour, N.L., Tormo-Mas, M.A., Weinert, L.A., Lowder, B.V., Cartwright, R.A., Smyth, D.S., Smyth, C.J., Lindsay, J.A., Gould, K.A., Witney, A., Hinds, J., Bollback, J.P., Rambaut, A., Penadés, J.R. and Fitzgerald, J.R. (2010) Evolutionary genomics of Staphylococcus aureus reveals insights into the origin and molecular basis of ruminant host adaptation. *Genome Biol Evol.* **2**, 454-466.

58) Sung, J.M., Lloyd, D.H. and Lindsay, J.A. (2008) Staphylococcus aureus host specificity: comparative genomics of human versus animal isolates by multi-strain microarray. *Microbiology* **154**, 1949-1959.

59) Schijffelen, M.J., Boel, C.H., van Strijp, J.A. and Fluit, A.C. (2010) Whole genome analysis of a livestock-associated methicillin-resistant *Staphylococcus aureus* ST398 isolate from a case of human endocarditis. *BMC Genomics* **11**, 376-386.

60) Holden, M.T., Hsu, L.Y., Kurt, K., Weinert, L.A., Mather, A.E., Harris, S.R., Strommenger, B., Layer, F., Witte, W., de Lencastre, H., Skov, R., Westh, H., Zemlickova, H., Coombs, G., Kearns, A.M., Hill, R.L., Edgeworth, J., Gould, I., Gant, V., Cooke, J., Edwards, G.F., McAdam, P.R., Templeton, K.E., McCann, A., Zhou, Z., Castillo-Ramierez, S., Feil, E.J., Hudson, L.O., Enright, M.C., Balloux, F., Aanensen, D.M., Spratt, B.G., Fitzgerald, J.R., Parkhill, J., Achtman, M., Bentley, S.D. and Nubel, U. (2013) A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Res.* **23**, 653-664.

61) Garcia-Alvarez, L., Holden, M.T., Lindsay, H., Webb, C.R., Brown, D.F., Curran, M.D., Walpole, E., Brooks, K., Pickard, D.J., Teale, C., Parkhill, J., Bentley, S.D., Edwards, G.F., Girvan, E.K., Kearns, A.M., Pichon, B., Hill, R.L., Larsen, A.R., Skov, R.L., Peacock, S.J., Maskell, D.J. and Holmes, M.A. (2011) Meticillin-resistant Staphylococcus aureus with a novel mecA

homologue in human and bovine populations in the UK and Denmark: a descriptive study. *Lancet Infect. Dis.* **11**, 595-603.

62) Chen, C.J., Unger, C., Hoffmann, W., Lindsay, J.A., Huang, Y.C. and Götz, F. (2013) Characterization and comparison of 2 distinct epidemic community-associated methicillin-resistant *Staphylococcus aureus* clones of ST59 lineage. *PLoS One* **8**, e63210.

63) Chen, Y., Chatterjee, S.S., Porcella, S.F., Yu, Y.S. and Otto, M. (2013) Complete genome sequence of a Pantón-Valentine leukocidin-negative community-associated methicillin-resistant *Staphylococcus aureus* strain of sequence type 72 from Korea. *PLoS One* **8**, e72803.

64) Holt, D.C., Holden, M.T., Tong, S.Y., Castillo-Ramirez, S., Clarke, L., Quail, M.A., Currie, B.J., Parkhill, J., Bentley, S.D., Feil, E.J. and Giffard, P.M. (2011) A very early-branching *Staphylococcus aureus* lineage lacking the carotenoid pigment staphyloxanthin. *Genome Biol. Evol*. **3**, 881-395.

65) Price, C., Lingner, J., Bickle, T.A., Firman, K. and Glover, S.W. (1989) Basis for changes in DNA recognition by the EcoR124 and EcoR124/3 type I DNA restriction and modification enzymes. *J. Mol. Biol*. **205**, 115-125.

66) Gubler, M., Braguglia, D., Meyer, J., Piekarowicz, A. and Bickle, T.A. (1992) Recombination of constant and variable modules alters DNA sequence recognition by type IC restriction-modification enzymes. *EMBO J*. **11**, 233-240.

67) Adamczyk-Popławska, M., Kondrzycka, A., Urbanek, K. and Piekarowicz, A. (2003) Tetra-amino-acid tandem repeats are involved in HsdS complementation in type IC restriction-modification systems. *Microbiology* **149**, 3311-3319.

68) Jurenaite-Urbanaviciene. S., Serksnaite, J., Kriukiene, E., Giedriene, J., Venclovas, C. and Lubys, A. (2007) Generation of DNA cleavage specificities of type II restriction endonucleases by reassortment of target recognition domains. *Proc. Natl. Acad. Sci. USA*. **104**, 10358-10363

69) Wattam, A.R., Abraham, D., Dalay, O., Disz, T.L., Driscoll, T., Gabbard, J.L., Gillespie, J.J., Gough, R., Hix, D., Kenyon, R., Machi, D., Mao, C., Nordberg, E.K., Olson, R., Overbeek, R., Pusch, G.D., Shukla, M., Schulman, J., Stevens, R.L., Sullivan, D.E., Vonstein, V., Warren, A., Will, R., Wilson, M.J., Yoo, H.S., Zhang, C., Zhang, Y. and Sobral, B.W. (2014) PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res*. **42**, D581-D591.

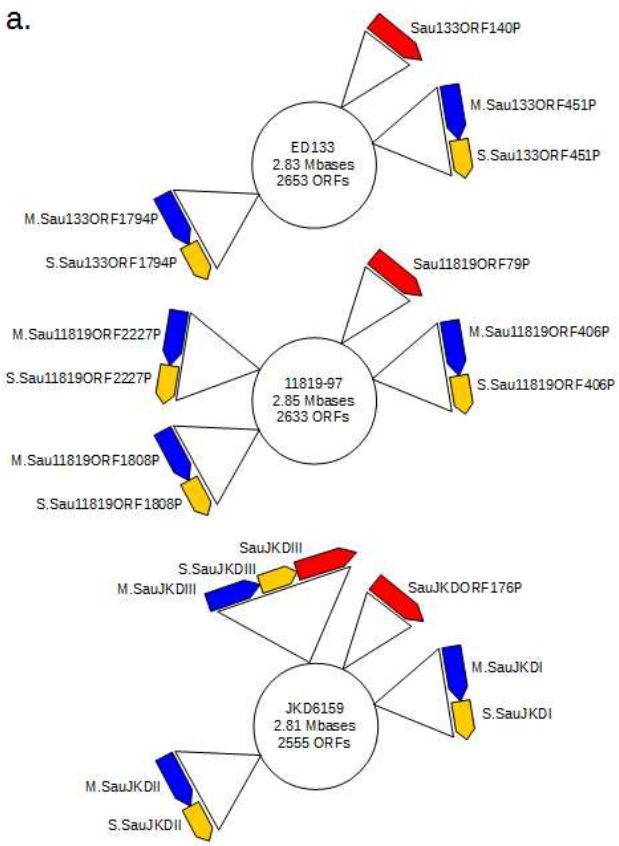
70) Drozdz, M., Piekarowicz, A., Bujnicki, J.M. and Radlinska, M. (2012) Novel non-specific DNA adenine methyltransferases. *Nucleic Acids Res*. **40**, 2119-2130.

71) Kojima, K.K., Furuta, Y., Yahara, K., Fukuyo, M., Shiwa, Y., Nishiumi, S., Yoshida, M., Azuma, T., Yoshikawa, H. and Kobayashi, I. (2016) Population evolution of *Helicobacter pylori* through diversification in DNA methylation and interstrain sequence homogenization. *Mol. Biol. Evol.* **33**, 2848-2859.

72) Croucher, N.J., Coupland, P.G., Stevenson, A.E., Callendrello, A., Bentley, S.D. and Hanage, W.P. (2014) Diversification of bacterial genome content through distinct mechanisms over different timescales. *Nature Commun.* **5**, 5471. doi: 10.1038/ncomms6471.

73) Larsen, M.V., Cosentino, S., Rasmusen, S., Hasman, H., Marvig, R.L., Jelsbak, L., Sicheritz-Ponten, T., Ussery, D.W., Aarestrup, F.M. and Lund, O. (2012) Multilocus sequence typing of total-genome-sequenced bacteria. *J. Clin. Microbiol*. **50**, 1355-1361.

Figure 1.


a.
Sau133ORF140P
M.Sau133ORF451P
S.Sau133ORF451P
ED133
2.83 Mbases
2653 ORFs
M.Sau133ORF1794P
S.Sau133ORF1794P
Sau11819ORF79P
M.Sau11819ORF2227P
M.Sau11819ORF406P
S.Sau11819ORF2227P
11819-97
2.85 Mbases
2633 ORFs
S.Sau11819ORF406P
M.Sau11819ORF1808P
S.Sau11819ORF1808P
SauJKDIII
S.SauJKDIII
M.SauJKDIII
SauJKDORF176P
M.SauJKDI
JKD6159
2.81 Mbases
2555 ORFs
S.SauJKDI
M.SauJKDII
S.SauJKDII

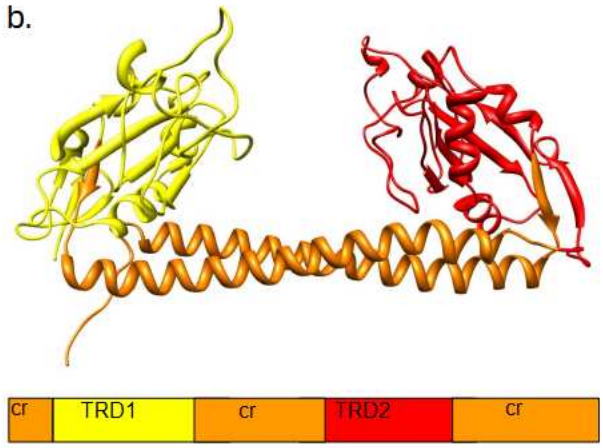

b.
cr
TRD1
cr
TRD2
cr

Figure 2.

a.

```
B (first TRD) and P (second TRD)
PROMALS alignment
Conservation:                      5555555555    999  555555 9999    95555 9999 99   999999999 9999999999
TRD P                          1   QIELEEQKLELFQQQKKGYMQKIFSQELRFKDESGNDYPDWEEKELGEVADRVIRKNKNFESKKPLTISG   70
TRD B                          1   ----------MSNTQKKNV------PELRFPGFEG----EWEEKKLGDLTDRVIRKNKNLESKKPLTISG   50
Consensus_ss:                          hhhhhhhhhhhhhhhhhhh                 eeeeeeheeeee           eeeee


Conservation:                      9999999999999999999999999999999999999999999999999999999999999999999999
TRD P                         71   QLGLIDQTEYFSKSVSSKNLENYTLIKNGEFAYNKSYSNGYPLGAIKRLTRYDSGVLSSLYICFSIKSEM  140
TRD B                         51   QLGLIDQTEYFSKSVSSKNLENYTLIKNGEFAYNKSYSNGYPLGAIKRLTRYDSGVLSSLYICFSIKSEM  120
Consensus_ss:                       eeee hhh              ee     eeeeee      eeeeeee     eeee  eeeee


Conservation:                      999999999999999999999999999999 999999999999999999 999 99 9999999999
TRD P                        141   SKDFMEAYFDSTHWYREVSGIAVEGARNHGLLNISVNDFFTILIKYPSLEEQRKIGDFFIKLDRQIELEE  210
TRD B                        121   SKDFMEAYFDSTHWYREVSGIAVEGARNHGLLNVSVNDFFTILIKYPSLEEQQKIGKFFSKLDRQIELEE  190
Consensus_ss:                        hhhhhhhh  hhhhhhhhhhhh       eeee hhhhh        hhhhhhhhhhhhhhhhhhhhh


Conservation:                      99999999 99
TRD P                        211   QKLELLQQRKKALLKSMLI  229
TRD B                        191   QKLELLQQQKKGYMQKIFS  210
Consensus_ss:                      hhhhhhhhhhhhhhhhh
```

b.

```
U and c* (both second TRDs)
PROMALS alignment
Conservation:                      9999999999999999999 999999999999999 999 9
TRD U                          1   QIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGEDYPDWEVTTIQNITKYTSSKKSSNQYADKDNSKG   70
TRD c*                         1   QIELEEQKLELLQQQKKGYLQKIFSQELRFKDENGNDYPEWRFARFKDFMYKPINIRPAINISKSELLTV   70
Consensus_ss:                          hhhhhhhhhhhhhhhhhhh                 eeeehhheeeee               ee


Conservation:                                    9   9    9   9     9     9     9                 9
TRD U                          71   YPVYDA--------VQEIGKDSNYDIEESYISILKDGAGVGRLNLRPG---KSSVIGTMGYIQSN--NVD  127
TRD c*                         71   KLHCKGIEKANINRVLKLGATNYYKRFEGQFIYGKQNFFNGAFDIVPKKFDGLYSSSDVPAFEINTEKIE  140
Consensus_ss:                       eeeee           ee    eeee    eeeeee   eeeeeee      eeee eeeeeee     h


Conservation:                                  9 9       9               9  9    99 99  9   9   99    9
TRD U                          128  IEFLYYRMKVVDFKKY----IIGSTIPHLYFKDYSKETLYIPSSIQEQAKIGMFISNLDKLIENKNLKLN  193
TRD c*                         141  PNYFISYISRPSFYKSKEKYSTGTGSKRIHENTVLNFSLHLPC-LNEQLKIASFVCFLNRKIELLERKIY  209
Consensus_ss:                       hhhhhhhh  hhhhhhhhhhh         hhhhh  ee     hhhhhhhhhhhhhhhhhhhhhhhhh


Conservation:                        9  99 999 999
TRD U                          194  CLKQLKQGLLQSMFI  208
TRD c*                         210  LIKKQKQALLQQMFI  224
Consensus_ss:                       hhhhhhhhhhhhh
```

Figure 3.

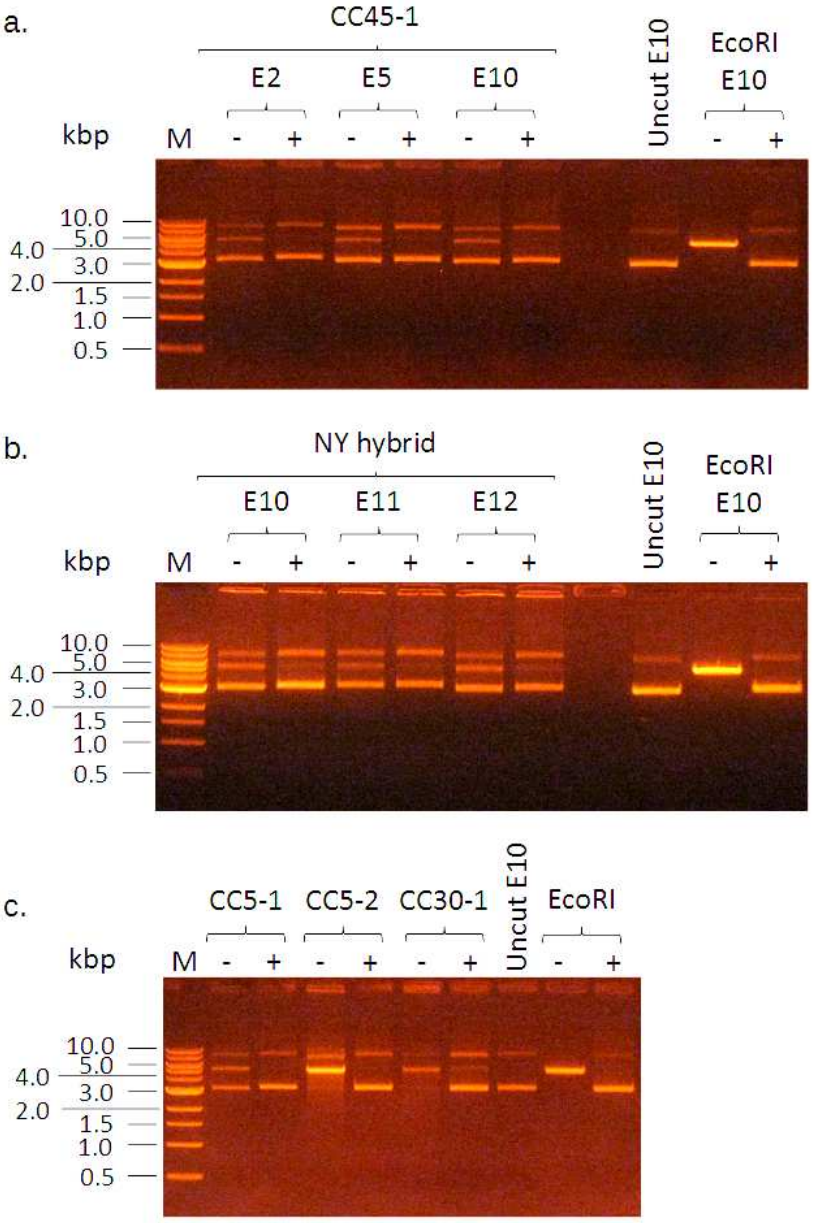

a.
CC45-1
E2
E5
E10
Uncut E10
EcoRI
E10
kbp
M
- + - + - + - +
10.0
5.0
4.0
3.0
2.0
1.5
1.0
0.5
b.
NY hybrid
E10
E11
E12
Uncut E10
EcoRI
E10
kbp
M
- + - + - + - +
10.0
5.0
4.0
3.0
2.0
1.5
1.0
0.5
c.
CC5-1
CC5-2
CC30-1
Uncut E10
EcoRI
kbp
M
- + - + - + - +
10.0
5.0
4.0
3.0
2.0
1.5
1.0
0.5

Table 1. TRD targets shown from 5' to 3'. The underlined A indicates the site of adenine methylation. TRD pair B and P, and TRD pair U and c recognise the same DNA sequence and are highlighted in red. In the targets, Y= C or T; R = A or G; W = A or T; H = A, T or C; N = A, G, T or C.*

TRD1 code letter and target		TRD2 code letter and target	
A	CCAY	D	ATC
B	AGG	E	TCAY
C	GWAG	F	TTAA
J	GGA	G	ACA
M	CAG	H	TAC
N	ACC	I	YTCA
O	CAAC	K	CGA
R	GARA	L	TTTA
T	CAAG	P	AGG
V	CNGA	Q	ACAY
X	TCTA	S	GCA
Z	GAC	U	GAY
b*	GGHA	W	CRAA
e*	GAG	Y	CTA
		a*	GAA
		c*	GAY
		d*	CYAA
		f*	GAAY

Table 2. The Sau1 RM systems with published recognition sequences. Target sites are shown from 5' to 3' with the length of the non-specific spacer shown as a number. Underlined A or T indicates the site of adenine methylation on the top or bottom strands respectively. The experimental methods used are indicated as g = target obtained by DNA cleavage with a purified enzyme, s = target obtained by SMRT sequencing of E. coli ER2796 genomic DNA, a = target obtained by ATPase assay with a purified enzyme. Full details are given in the supplementary information. S.Sau133ORF1794P is characterised in this work but is included here as it is part of the RM system found in strain ED133. SauMRSI and SauMRSII characterised by Monk et al. and S.SauSTORF499P characterised by Chen et al. are also further characterised in this work.

Strain name and genome reference	Clonal Complex or Sequence Type	S subunit name in REBASE	Recognition sequence	TRDs assigned	Suggested generic name	Experimental method	Reference for target specificity and method
MW2 (53)	CC1	S.SauMW2I	CCAY-5-TTAA	AF	CC1-1	g, s, a	g (31) a (36) s (CC8-1 and CC8-2 in strain NRS384 are from ref. 35)
		S.SauMW2II	CCAY-6-TGT	AG	CC1-2 (CC8-2)	g, s, a	
N315 (54)	CC5	S.SauN315II	CCAY-6-GTA	AH	CC5-2	g, s, a	
		S.SauN315I	AGG-5-GAT	BD	CC5-1 (CC8-1)	g, s, a	
MRSA252 (55)	CC30	S.SauMRSII	GWAG-5-GAT	CD	CC30-1	g, s	s (35) g, s (this work)
		S.SauMRSI	GGA-7-TCG	JK	CC30-2	s	s (35) s (this work)
JKD6159 (56)	CC93	S.SauJKDIII	GAAG-5-TAC or complement	Not a *Sau1* system	CC93-3	s	s (35) Note the ambiguity in assigning CC93-1 and CC93-3 is clarified with strains ED133 and 32320 and from Table 3.
		S.SauJKDII	GGHA-7-TCG	b*K	CC93-2	s	
		S.SauJKDI	CAG-6-TTC	Ma*	CC93-1	s	
ED133 (57)	CC133	S.Sau133ORF451P	CAG-5-RTGA	ME	CC133-1	g	g (36)
		S.Sau133ORF1794P	GGA-7-TTRG	Jd*	CC133-2	s	s (this work)
32320 (58)	CC133	S.Sau32320ORFAP	CAG-5-RTGA	ME	CC133-1	g	g (36)
S0385 (59)	CC398	S.SauSTORF499P	ACC-5-RTGA	NE	CC398-1	g, s	g (36) s (this work)

Table 3. The "artificial" Sau1 systems containing novel pairings of TRDs. Target sites are shown from 5' to 3' with the length of the non-specific spacer shown as a number. Underlined A or T indicates the site of adenine methylation on the top or bottom strands respectively. The experimental methods used are indicated as g = target obtained by DNA cleavage with a purified enzyme, s = target obtained by SMRT sequencing of E. coli ER2796 genomic DNA, a = target obtained by ATPase assay with a purified enzyme. Full details are given in the supplementary information.

"Artificial" Sau1 RM systems.					
Recognition sequence	TRDs assigned	Experimental method	Recognition sequence	TRDs assigned	Experimental method
AGG-5-RTGA	BE	a	ACC-6-TTC	Na*	s
GGA-6-RTGA	JE	g, s	ACC-6-RTC	Nc*	s
ACC-6-TGAR	NI	g	ACC-6-TTRG	Nd*	g, s
ACC-6-TCG	NK	g	GARA-6-RTGA	RE	s
ACC-6-TAAA	NL	g	CAAG-5-RTGA	TE	s
ACC-5-CCT	NP	s	CNGA-6-RTGA	VE	s
ACC-5-RTGT	NQ	g, s	TCTA-6-RTGA	XE	g, s
ACC-6-TGC	NS	s	GAC-5-RTGA	ZE	a
ACC-5-RTC	NU	g, s	GAC-6-TGC	ZS	a
ACC-6-TTYG	NW	g, s	GGHA-6-RTGA	b*E	s
ACC-6-TAG	NY	g, s	GAG-6-RTGA	e*E	g, s

*Table 4. The Sau1 RM systems investigated in this project. Target sites are shown from 5' to 3' with the length of the non-specific spacer shown as a number. Underlined A or T indicates the site of adenine methylation on the top or bottom strands respectively. TRD pair e*f* in strain LGA251 was not cloned in E. coli while TRD pair XY was cloned. However, no target modification was observed using SMRT on genomic DNA from either E. coli or S. aureus for these TRD pairs. If the genes are translated, their target is inferred from other TRDs in this table although the spacer length remains undefined. The experimental methods used are indicated as g = target obtained by DNA cleavage with a purified enzyme, s = target obtained by SMRT sequencing of E. coli ER2796 genomic DNA, s* = target obtained by SMRT sequencing of S. aureus genomic DNA, a = target obtained by ATPase assay with a purified enzyme. Full details are given in the supplementary information.*

Strain name and genome reference	Clonal Complex or Sequence Type	S subunit name in REBASE	Recognition sequence	TRDs assigned	Suggested generic name	Experimental method
CO1791 (58)	CC97	S.SauC01791ORFAP	CCAY-6-RTC	Ac*	CC97-1	s
HO5096 (60)	CC22	S.Sau5096I	AGG-6-TGAR	BI	CC22-1	g, s
LGA251 (61)	ST425	S.Sau251I	GWAG-5-RTGA	CE	ST425-1	g, s*
		S.Sau251ORF16900P	GAG-?-RTTC	e*f*	ST425-2	Not expressed, no signature with s*.
		S.Sau251II	GAAG-5-TAC or complement	Not a *Sau1* system	Same as CC93-3	s*
Isolate 3 (19)	CC51	S.SauL3ORFAP	GGA-6-CCT	JP	CC51-1	s
Isolate 3067 (19)	CC45	S.Sau347I	GWAG-6-TAAA	CL	CC45-1	g
Isolate 3150 (19)	CC15	S.SauL315ORFAP	CAAC-5-RTGA	OE	CC15-1	s
SA40 (62)	CC59	S.SauSA40ORF370P	GGA-6-RTGT	JQ	CC59-1	a
CN1 (63)	CC72	S.SauCN1ORF415P	GARA-6-RTGT	RQ	CC72-1	a
		S.SauCN1ORF1757P	GGA-7-TGC	JS	CC72-2	a
MSHR1132 (64)	CC75	S.Sau1132ORF3780P	CAAG-5-RTC	TU	CC75-1	g
		S.Sau1132ORF16570P	CNGA-7-TTYG	VW	CC75-2	s
NCTC13435 NCBI Biosample identifier: SAMEA2479566	ST80	S.Sau13435ORF394P	TCTA-?-TAG	XY	ST80-1	Not expressed, no signature with s or s*.
		S.Sau13435ORF1751P	GAC-6-TTYG	ZW	ST80-2	a, s*
		S.Sau13435ORF2165P	TCTA-6-RTTC	Xf*	ST80-3	s, s*
32326 (58)	CC873	S.Sau32326ORFAP	GAG-6-GAT	e*D	CC873-1	a

Table 5. Further TRD pairs found in sequenced strains of S. aureus. Every pair of TRD1 with TRD2 in table 1 was used in a BLASTP sequence search to identify HsdS subunit sequences in publicly accessible databases. Examples of strains containing these TRD pairs are shown. ST and CC are from the PATRIC database (69) or derived using www.cbs.dtu.dk/services/MLST (73). Some TRD pairs are present in many strains while others are rare.

TRD pair	Example Strain	Clonal Complex or Sequence Type of example strain	REBASE name
AD	FDAARGOS_159	ST5	S.Sau159ORF12345P
AL	K12S0375	ST692	S.Sau375ORFDP
AU	*S. schweitzeri* FSA084	-	S.SauFSA084ORF355P
AW	FDA209P	ST464	S.Sau209ORF1697P
BG	MRSN8611	ST8	S.Sau8611ORF11430P
BH	PLAC6019	ST5	S.Sau6019ORF851P
BU	SA-083	ST101	S.Sau083ORF9680P
BY	*S. argenteus* M260-MSHR	-	S.SarM260ORF2316P
Bf*	SA-083	ST101	S.Sau083ORF1720P
JE	Tager 104	ST49	S.Sau104ORF1102P
JL	W56227	ST45	S.Sau56227ORF970P
JW	CIG290	ST45	S.SauCIG290ORF2408P
JW	APS211	ST45	S.SauAPS211ORF9230P
MW	FSA037	ST1872	S.SauFSA037ORF2487P
NQ	KPL1845	ST96	S.Sau1845ORF2596P
Of*	USA300-TCH959	ST1159	S.SauTCH959ORF2844P
Rf*	Tager 104	ST49	S.Sau104ORF2433P
TY	M21126	ST2250	S.Sau21126ORF1065P
XF	21334	ST109, CC9	S.Sau21334ORF1353P
XF	RKI4	ST27	S.SauRKI4ORF1905P
XW	103564	ST80-PVL carrier	S.Sau103564ORF678P
ZY	D139	ST145	S.SauD139ORF2470P
b*W	ST20130941	CC15	S.Sau941ORF4310P
e*f*	SA-120	ST425	S.Sau120ORF4875P

Table 6. New TRD pairs associated with pairs shown in Tables 2, 3 and 4. The new TRDs of unknown specificity are termed NOVEL 1, NOVEL 2, NOVEL 3 and NOVEL 4. TRD NOVEL 3 is a second TRD while the others are first TRDs in the HsdS amino acid sequence. Subspecies 21343 and species KPL1845 also contain S.SauNQ (S.Sau21343ORF1169P and S.Sau1845ORF2596P respectively).

Subspecies 21343 Bioproject accession: PRJNA53699
> S.Sau21343ORF2597P TRD NOVEL 1 + TRD K MSNTQKKNVPELRFPGFEGEWEEKKLGEVATFAKGKLGAKKDVSQNGVPVILYGELYTKYGAIVSKIFSKTDIPENKLKMAKKNDVLIPSSGETAIDIATASCIYLNKGVAVGGDINILTPQKQDGRFISLSIN GINKNELSKYAQGKTVVHLYNNDIKNLKIAFPSEFEEQVRIGNFFSKLDRQIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPKWEEKKIEDIASQVYGGGTPNTKIKEFWNGDIPWIQSSDVKVNDL ILQQCNKFISKNSIELSSAKLIPANSIAIVTRVGVGKLCLVEFDYATSQDFLSLSSLKYDKLYSLYSLLYTMKKISANLQGTSIKGITKKELLDSIIKIPHNLEEQQKIGDLFYKIDKYISFNKCKIEILKSLK QGLLKKMFI
Species KPL1845. Bioproject accession: PRJNA169473
> S.Sau1845ORF1619P TRD NOVEL 2 + NOVEL 3 MTEQINTPELRFPEFKNEWSYDLVSDVVTNKSKKFDPKKEEAKKDIELDSIEQNTGRLLDTYISNDFTSQKNKFNKGNVLYSKLRPYLNKYYYATIDGVCSSEIWVLNTLNKDVLANKFLYYFIQTNRFSSVTN KSAGSKMPRADWELVKNIRLYKGSIEEQEKIGYFFSKLDRQIELEEKKLELLEQQKKGYMQKIFAQELRFKDENGNDYPDWVTKKLGDIGKVAMNKRIYKNETTENGEIPFYKIGNFGKNADTFITREKFDEYK EKYPYPNVGDILISASGSIGRTIEYTGEDAYYQDSNIVWLNHNDEVINKYLKYFYKIVKWSGIEGTTIKRLYNKNILNTKIELPTVEEQYKMANFLSKLDKIIDIQIEKIELLKQRKQGLLQKMFV
> S.Sau1845ORF2199P TRD NOVEL 4 + TRD f* MSNTQKKNVPELRFPEFEGEWKDVKFVSIFQEVSNKTSDLAKYPLFSLTVEKGITPKTERYKRDFLVKKSDNFKIVEPRDIVYNPMNVTLGAIDLSKYNYDIALSGYYHVMKIINSFNPDFISNFLKTEKMIIH YKKIATGSLMEKQRVHFSEFKNIIKKFPTNKEQQKIGDFFSKLDRQIELQVQKLELLQQQKKGYMQKIFSQELRFKDENGEDYPDWKEKKLGDITEQSMYGIGASATRFDSKNIYIRITDIDEKSRKLNYQNLT TPDELNNKYKLKRNDILFARTGASTGKSYIHKEEKDIYNYYFAGFLIKFEIDEQNNPLFIYQFTLTSKFNKWVKVMSVRSGQPGINSEEYAKLPLVLPNKLEQQKIAEFLDRFDQQIELEKQKIEILQQQKKGL LQSMFI

SUPPLEMENTARY INFORMATION
DNA target recognition domains in the Type I restriction/modification systems of Staphylococcus aureus.

Laurie P. Cooper, Gareth A. Roberts, John H. White, Yvette Luyten, Edward K.M. Bower, Richard D. Morgan, Richard J. Roberts, Jodi A. Lindsay, David T.F. Dryden.

SUPPLEMENTARY INFORMATION FOR TABLE 1.
The amino acid sequences in FASTA format of the first TRD with its letter code, DNA target (5′ to 3′) and methylation site underlined. The TRDs labelled as NOVEL 1, NOVEL 2 and NOVEL 4 were found once all of the other TRDs had been analysed but are included in this list for completeness.
The TRD sequences are flanked by the conserved regions so to obtain the amino acid sequence of any HsdS subunit simply paste the sequence for the second TRD directly on to the end of the sequence for the first TRD.

>A CCAY
MSNTQKKNVPELRFPGFEGEWEEKQLGDLTTKIGSGKTPKGGSENYTNKGIPFLRSQNIRNGKLNLNDLVYISKDIDDEMKNSRTYYGDVLLNITGASIG
RTAINSIVETHANLNQHVCIIRLKKEYYYIFFGQYLLSRKGKRKIFLAQSGGSREGLNFKEIANLKIFTPTIFEEQQKIGKFFSKLDRQIELEEQKLELL
QQQ
>B AGG
MSNTQKKNVPELRFPGFEGEWEEKQLGDLTDRVIRKNKNLESKKPLTISGQLGLIDQTEYFSKSVSSKNLENYTLIKNGEFAYNKSYSNGYPLGAIKRLT
RYDSGVLSSLYICFSIKSEMSKDFMEAYFDSTHWYREVSGIAVEGARNHGLLNVSVNDFFTILIKYPSLEEQQKIGKFFSKLDRQIELEEQKLELLQQQ
>C GWAG
MSNTQTKNVPELRFPGFEGEWEEKQVGELLEFKNGLNKGKEYFGSGSSIVNFKDVFNNRSINTNNLTGKVNVNSKELKNYSVEKGDVFFTRTSEVIGEIG
YPSVILNDPENTVFSGFVLRGRPKSGIDLINNNFKRYVFFTNSFRKEMITKSSMTTRALTSGTAINKMKVIYPVSAKEQKKIGDFFSKLDRQIELEEQKL
ELLQQQ
>J GGA
MSNTQKKNVPELRFPEFEGEWEERKLGDLIKVNSGKDYKHLDKGDIPVYGTGGYMTSVSEPLSEIDAVGIGRKGTINKPYLLEAPFWTVDTLFYCTPEKE
ADILFILSLFRKINWKLYDESTGVPSLSKQTINKINRLVPTNKEQQKIGEFFSKLDRQIELEEQKLELLQQQ
>M CAG
MSNTQTKNVPELRFPGFEGEWEEKKLEDLGLFQKSYSFSRAKEGNGKTKHIHYGDIHSKFKTVLDSDGNIPNIIEKAVFELIQKGDIVFADASEDYSDLG
KAVMIDFEPNSLISGLHTHLFRPLNNAISNFLIFYTKTLSYKKFIRQQGTGISVLGISKKSLLNLNVLIPRSELEQQKIGQFFSKLDRQIELEEQKLELL
QQQ
>N ACC
MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEISNIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRN
KLGKKGVMSPLYTVFKIQNIDLNFIEFYFKSSKWYRFMALNGDSGARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ

>O CAAC
MSNKQKKNVPELRFPGFEGEWEEKKLGEVGTFTSGGTPLKSKSEYWNGDIPWITTGDIHNIKRENITNFITEKGLNESSAKLITNEAILIAMYGQGKTRG
MSAILNFEATTNQACAIYQTNQNINFVFQYFQKLYEFLRSLSNEGSQKNLSLSLLKEITLNYPNEQEQKKIGDFFSKLDRQIELEEQKLELLQQQ
>R GARA
MSNTQKKNVPELRFPGFEGEWEEKKLGEVAKIYDGTHQTPKYTNEGIKFLSVENIKTLNSSKYISEEAFEKEFKIRPEFGDILMTRIGDIGTPNIVSSNE
KFAYYVSLALLKTKNLNSYFLKNLILSSSIQNELWRKTLHVAFPKKINKNEIGKIKINYPKKQEQQKIGQFFSKLDRQIELEEQKLELLQQQ
>T CAAG
MSNTQTKNVPELRFPGFEGEWEEKELGEIFQIISGSTPLKSNKEFYENGNINWVKTTDLNNSKVTHSKEKITEYAMKSLKLKLVPKNSVLIAMYGGFNQI
GRTGLLKIDATINQAISALLMNHETNPEFIQAFLNYQVKGWKRYAASSRKDPNITKKDIEQFKVPYVSINEQQKIGEFFSKIDHQIELEEQKLELLQQQ
>V CNGA
MSNTGKMNVPELRFPGFEGEWEEKELRELRNPKDKYSYTGGPFGSDLKKSDYTTDGIQIIQLQNIGDGYFYNSNKVFTSNEKAEVLKSCNVFPGDIVIAK
MADPIARAAIVPDNNIGKYLMASDGIRLSVDTVHFNTKFVLECINRKSFRKKVEDNSSGSTRMRIGLSTLGSLTLKTTTLKEQQKIGQFFSKLDRQIVLE
EQKLELLQQQ
>X TCTA
MSNTQKKNVPELRFPGFEGEWEEKQFADFTKINQGLQIAINERKTEYSPELYFYITNEFLRPNSQTKYFIENPPQSVIANKEDILMTRTGNTGKVVTNVF
GAFHNNFFKIKFDKNLYDRLFLVEVLNSSKIQNKILSLAGSSTIPDLNHSDFYSISSSYPLLREQQKIGKFFSKLDRQIELEEQKLELLQQQ
>Z GAC
MSNTQTKNVPELRFPGFEGEYSLDIFGNLATNKSEKFNPQNENASIDIELDCIEQNTGRLIKIYNSKEFSSQKNKFNPQNVLYGKLRPYLNKYYFTKKSG
VCSSEIWVLKSTKEDKLLNLFLYYFIQTKRYSDVASKSAGSKMPRADWGLIENIRVYFPELCEQQKIGQFFSKLDRQIELEEQKLELLQQQ
>b* GGHA
MSNTQKKNAPELRFPEFEGEWKEKKLEDTLEFIKDGTHGTHENVNNGPWLLSAKNIKNNKIIISSDDRKISESDYKKIYKNYKLEKGDLLLTIVGTIGRA
AIVKNPNNIAFQRSVAILKTKATYDVGFIFQLFQTKYFKNLLLRKQVVSAQPGLYLGDIRKIKISITNIIEEQRKIGIFFSKLDRQIELEEQKLELLQQQ
>e* GAG
MSNTQKKNVPELRFPGFEGEWEEKSISSFLKESKIKGSNGSHAKKLTVKLWGKGVVPKKETFKGSDNTQYYKRKAGQLMYGKLDFLNCAFGIVPDSLNNY
ESTIDSPSFDFINGDSKFLLERIKLKSFYKKFGDIANGSRKAKRINQDTFLSLPVFAPKYDEQLRIGEFFSKLDRQIELQKQKLELLQQQ
>NOVEL 1
MSNTQKKNVPELRFPGFEGEWEEKKLGEVATFAKGKLGAKKDVSQNGVPVILYGELYTKYGAIVSKIFSKTDIPENKLKMAKKNDVLIPSSGETAIDIAT
ASCIYLNKGVAVGGDINILTPQKQDGRFISLSINGINKNELSKYAQGKTVVHLYNNDIKNLKIAFPSEFEEQVRIGNFFSKLDRQIELEEQKLELLQQQ
>NOVEL 2
MSNTQKKNVPELRFPEFEGEWKDVKFVSIFQEVSNKTSDLAKYPLFSLTVEKGITPKTERYKRDFLVKKSDNFKIVEPRDIVYNPMNVTLGAIDLSKYNY
DIALSGYYHVMKIINSFNPDFISNFLKTEKMIIHYKKIATGSLMEKQRVHFSEFKNIIKKFPTNKEQQKIGDFFSKLDRQIELQVQKLELLQQQ
>NOVEL 4
MTEQINTPELRFPEFKNEWSYDLVSDVVTNKSKKFDPKKEEAKKDIELDSIEQNTGRLLDTYISNDFTSQKNKFNKGNVLYSKLRPYLNKYYYATIDGVC
SSEIWVLNTLNKDVLANKFLYYFIQTNRFSSVTNKSAGSKMPRADWELVKNIRLYKGSIEEQEKIGYFFSKLDRQIELEEKKLELLEQQQ

The amino acid sequences in FASTA format of the second TRD with its letter code, DNA target (5′ to 3′) and its methylation site underlined. The TRD labelled as NOVEL 3 was found once all of the other TRDs had been analysed but is included in this list for completeness.

>D ATC
KKGYMQKIFSQELRFKDENGEDYPHWENSKIEKYLKERNERSDKGQMLSVTINSGIIKFSELDRKDNSSKNKSNYKVVRKNDIAYNSMRMWQGASGKSNY
NGIVSPAYTVLYPTQNTSSLFIGYKFKTHRMIHKFKINSQGLTSDTWNLKYKQLKNINIDIPVLEEQEKIGDFFKKMDILISKQKIKIEILEKEKQSFLQ
KMFL
>E TCAY
KKGYMQKIFSQELRFKDENGNDYPEWEETTIKEIAQINXGKKDTKDAITNGSYDFYVRSPIVYKINTFSYEGEAILTVGDGVGVGKVFHYVNGKFDYHQR
VYKISDFKNYYGLLLFYYFSQNFLKETKKYSAKTSVDSVRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTKIQKQVIELLKQRKKALLQKMFI
>F TTAA
KKGYMQKIFSQELRFKDEEGKDYPDWKSKSIQEIFENKGGTALETEFNFDGNYKVISIGSYSINSTYNDQNIRVNKNKKTEKYILSKGDLAMVLNDKTKD
GKIIGRSIFIDKDNQYIYNQRTERLIPFAENDNKFLWFLMNTDLIRNKIKGMMQGATQVYINYSSIKLISIQLPLLEEQQKIRGFLEVLSGITTKQLHKI
DQLKERKKAFLQKMFI
>G ACA
KKGYMQKIFTQELRFKDENGEEYPEWENKFIKDIFIFENNRRKPITSSLREKGLYPYYGATGIIDYVKDYLFNNEERLLIGEDGAKWGQFETSSFIANGQ
YWVNNHAHVVKSNDHNLFFMNYYLNFKELRAFVTGNAPAKLTHANLCNINLKIPCLTEQDKVSALLKSIDNKMNNQMNRIELLKERKKELLQKMFI
>H TAC
KKCYIQKIFSQELRFKDEEGNYYKGWNKKQLKDVLEFSNKRTINENEYPVLTSSRQGLILQSDYYKDRKTFAESNIGYFILPKNHITYRSRSDDGIFKFN
LNLMIDVGIISKYYPVFKGIDANQYYLTLHLNYQLKKEYIKYATGTSQLVLSQKDLQNIKTKLPSYEEQQKIGDFFSEIDRLVEKQSSKVGRLKVRKKEL
LQKMFV
>I YTCA
KKGYMQKIFSQELRFKNENGNDYPDWERIKFFDVIDKVIDFRGRTPKKLNMEWSDEGYLALSAVNVKKGYIDFNVEAKYGNLDLYTRWMRGNELYKGQVL
FTTEAPMGNVAQVPDNKGYILSQRTIAFNSNEKITDNFLASLLSSENVYNDLLKLCSGATAKGVSQKNLNRLYVTIPHSISEQEEIAEFFRKINQLVELQ
KYKIEHTKSQKQVFLQKMFI
>K CGA
KKGYMQKIFSQELRFKDENGNDYPKWEEKKIEDIASQVYGGGTPNTKIKEFWNGDIPWIQSSDVKVNDLILQQCNKFISKNSIELSSAKLIPANSIAIVT
RVGVGKLCLVEFDYATSQDFLSLSSLKYDKLYSLYSLLYTMKKISANLQGTSIKGITKKELLDSIIKIPHNLEEQQKIGDLFYKIDKYISFNKCKIEMLK
SLKQGLLKKMFI
>L TTTA
KKGYMQKIFSQELRFKDENGNDYPNWRTIELKNILENIVDNRGKTPDNAPSEKYPLLEVNALGYYRPAYIKVSKFVSENTYNNWFREHLKENDILFSTVG
NTGIVSLMDNYKAVIAQNIVGLRVNNNNLPSFIYYMLSYKGNQKKIKRIQMGAVQPSVKVSQFKFIKYLVPIKDEQEKVAKLLIEIDKLVNKQLIKIELL
QQRKKALLKSMFI
>P AGG
KKGYMQKIFSQELRFKDESGNDYPDWEEKELGEVADRVIRKNKNFESKKPLTISGQLGLIDQTEYFSKSVSSKNLENYTLIKNGEFAYNKSYSNGYPLGA
IKRLTRYDSGVLSSLYICFSIKSEMSKDFMEAYFDSTHWYREVSGIAVEGARNHGLLNISVNDFFTILIKYPSLEEQRKIGDFFIKLDRQIELEEQKLEL
LQQRKKALLKSMLI
>Q ACAY
KKGYMQKIFSQELRFKDENGEDYSEWEERRFADIFKFHNKLRKPIKENLRVKGSYPYYGATGIIDYVDDFIFDGNYLLIGEDGANIITRSAPLVYLVNGK
FWVNNHAHILSPLNGNIQYLYQVAELVNYEKYNTGTAQPKLNIQNLKIINVVISTNLEEQQKIGSFLSKLDRQIDLEEQKLELLQQRKKALLKSMFV
>S GCA
KKGYMQKIFSQELRFKDENGNDYPDWTNERLGEVTTVTMGQSPKSVNYTDNSNDTVLIQGNADIENGLINPRIYTREVTKLIQKDEIILTVRAPVGKLAM
AQINACIGRGVCSIKGDKFLYYFLEWFATQNKWIRFSQGSTFESISGNDIRNIHIKIPVEDERTKIIKLLNSLDVLNSKTDLKIQNLKQRKQSLLQKIFV
>U GAY
KKGYMQKIFSQELRFKDENGEDYPDWEVTTIQNITKYTSSKKSSNQYADKDNSKGYPVYDAVQEIGKDSNYDIEESYISILKDGAGVGRLNLRPGKSSVI
GTMGYIQSNNVDIEFLYYRMKVVDFKKYIIGSTIPHLYFKDYSKETLYIPSSIQEQAKIGMFISNLDKLIENKNLKLNCLKQLKQGLLQSMFI
>W CRAA
KKGYMQKIFSQELRFKDENGNDYPDWEEKQLGELSQIVRGASPRPIKDPKWFNKESDIGWLRISDVTNQNGKIYHLEQKLSIEGQEKTRVLVTTHLLLSI
AASIGKPVMNFVKTGVHDGFLIFLKPKFNLFFMYYWLEYFKDKWSKYGQPGSQVNLNSEIVKSQTLNMPSNHEQEKVGQFFNRNEKLIELQQEKIMYIKR
CKQVLLQKMFI
>Y CTA
KKGYMQKIFSQELRFKDENGNDYPDWEKKKLKEIACVYTGNTPSKKENIYWNKGEYVWVTPTDINNSKNIYESENKLTQEGYKKARQLPENTLLVTCIAS
IGKNAILRKQGSCNQQINAVVPFENININDYLYYISDSLSTFMKSIAGKTATQIVNKNTFENLEIYLAPFEEQNKIADLISSLEELIEKQASKLIKMKSRK
QGMLQIMFI
>a* GAA
KKGYMQKIFSQELRFKDENGNDYPEWENKRIEDIANVNKGFTPSTNNNEYWDNNDKNWLSIAGMNQKYLYKGNKGISKDAAKNYMKVKNDTLIMSFKLTI
GKLAIVKAPLYTNEAICHFIWKVNKINTEFIYYYLNSLNISTFGVQAVKGVTLNNDSINSIIVKLPNEEEQNIIAKFLLEVDKTVNNQLVKTKLLKQRKK
GLLQRMFV
>c* GAY
KKGYLQKIFSQELRFKDENGNDYPEWRFARFKDFMYKPINIRPAINISKSELLTVKLHCKGIEKANINRVLKLGATNYYKRFEGQFIYGKQNFFNGAFDI
VPKKFDGLYSSSDVPAFEINTEKIEPNYFISYISRPSFYKSKEKYSTGTGSKRIHENTVLNFSLHLPCLNEQLKIASFVCFLNRKIELLERKIYLIKKQK
QALLQQMFI
>d* CYAA
KKGYMQKIFSQELRFKDENGNDYPEWENVMLQKVLKDKTEGIKRGPFGGALKKDIFVESGYAVYEQRNAIYDISNFRYYINENKYKEMQSFSVQPNDIIM
SCSGTIGRLALIPHNYTKGIINQALIRFRTNHKIRSEFFLIFMRSNQMQRKILEANPGSAITNLVPVKELKLIPFPLPVKFEQDKISQFIHIINRRIEQS
EKKIESLKNRKQGFLQKLFV
>f* GAAY
KKGYMQKIFSQELRFKDENGEDYPDWKEKKLGDITEQSMYGIGASATRFDSKNIYIRITDIDEKSRKLNYQNLTTPDELNNKYKLKRNDILFARTGASTG
KSYIHKEEKDIYNYYFAGFLIKFKINEQNSPLFIYQFTLTSKFNKWVKVMSVRSGQPGINSEEYAKLPLVLPNKLEQQKIAKFLDRFDRQIELEKQKIEI
LQQQKKGLLQSMFI
>NOVEL 3
KKGYMQKIFAQELRFKDENGNDYPDWVTKKLGDIGKVAMNKRIYKNETTENGEIPFYKIGNFGKNADTFITREKFDEYKEKYPYPNVGDILISASGSIGR
TIEYTGEDAYYQDSNIVWLNHNDEVINKYLKYFYKIVKWSGIEGTTIKRLYNKNILNTKIELPTVEEQYKMANFLSKLDKIIDIQIEKIELLKQRKQGLL
QKMFV

SUPPLEMENTARY INFORMATION FOR MATERIALS AND METHODS SECTION "CONSTRUCTION OF FURTHER MTASES WITH FURTHER COMBINATIONS OF TRDS USING SYNTHETIC GENES."

The DNA sequence is followed by the predicted amino acid sequence for each TRD. Some synthetic sequences encoded two TRDs.

CC15 TRD O

CCCGGGTTTGAAGGCGAATGGGAGGAAAAAAAACTGGGTGAAGTTGGCACCTTTACCAGCGGTGGC
ACTCCGCTGAAAAGCAAAAGCGAATATTGGAATGGTGATATTCCGTGGATTACCACAGGCGATATT
CATAACATTAAACGCGAAAACATCACCAACTTTATCACCGAAAAAGGCCTGAATGAAAGCAGCGCA
AAACTGATTACCAATGAAGCAATTCTGATTGCCATGTATGGTCAGGGTAAAACCCGTGGTATGAGC
GCCATTCTGAATTTTGAAGCAACCACCAATCAGGCCTGTGCAATTTATCAGACAAACCAGAACATC
AACTTCGTGTTCCAGTATTTCCAGAAACTGTATGAATTTCTGCGTAGCCTGAGCAATGAAGGTAGC
CAGAAAAATCTGAGCCTGAGCCTGCTGAAAGAAATTACCCTGAATTATCCGAACGAGCAAGAACAG
AAAAAAATCGGCGATTTCTTCAGCAAACTGGATCGTCAAATTGAATTAGAAGAACAGAAG

CC15 TRD O

PGFEGEWEEKKLGEVGTFTSGGTPLKSKSEYWNGDIPWITTGDIHNIKRENITNFITEKGLNESSA
KLITNEAILIAMYGQGKTRGMSAILNFEATTNQACAIYQTNQNINFVFQYFQKLYEFLRSLSNEGS
QKNLSLSLLKEITLNYPNEQEQKKIGDFFSKLDRQIELEEQK

CC51 TRD P

CAAATTGAATTAGAAGAACAGAAGCTGGAACTGTTTCAGCAGCAGAAAAAAGGCTATATGCAGAAA
ATCTTTAGCCAAGAGCTGCGCTTTAAAGATGAAAGCGGTAATGATTATCCGGATTGGGAAGAAAAA
GAACTGGGTGAAGTTGCAGATCGTGTGATTCGTAAAAACAAAAACTTTGAAAGCAAAAAACCGCTG
ACCATTAGCGGTCAGCTGGGTCTGATTGATCAGACAGAATATTTCAGCAAAAGCGTGAGCAGCAAA
AACCTGGAAAACTATACCCTGATTAAAAACGGCGAGTTCGCCTATAACAAAAGCTATAGCAATGGT
TATCCGCTGGGTGCAATTAAACGTCTGACCCGTTATGATAGCGGTGTTCTGAGCAGCCTGTATATT
TGCTTTAGCATCAAAAGCGAGATGAGCAAAGATTTCATGGAAGCCTATTTTGATAGCACCCATTGG
TATCGTGAAGTTAGCGGTATTGCAGTTGAAGGTGCACGTAATCATGGTCTGCTGAATATTAGCGTG
AACGATTTTTTCACCATCCTGATCAAATATCCGAGCCTGGAAGAACAGCGTAAAATCGGTGATTTC
TTCATTAAACTGGATCGCCAGATTGAGCTGGAAGAACAAAAACTGGAACTGCTGCAACAGCGCAAA
AAAGCACTGCTGAAAAGTATGCTGATCCCCGGGGGATCCGATCGATC

CC51 TRD P

QIELEEQKLELFQQQKKGYMQKIFSQELRFKDESGNDYPDWEEKELGEVADRVIRKNKNFESKKPL
TISGQLGLIDQTEYFSKSVSSKNLENYTLIKNGEFAYNKSYSNGYPLGAIKRLTRYDSGVLSSLYI
CFSIKSEMSKDFMEAYFDSTHWYREVSGIAVEGARNHGLLNISVNDFFTILIKYPSLEEQRKIGDF
FIKLDRQIELEEQKLELLQQRKKALLKSMLI

CC72-1 TRD R + CC59-1 TRD Q

CCCGGGTTTGAAGGCGAATGGGAGGAAAAAAAACTGGGTGAAGTGGCCAAAATCTATGATGGCACC
CATCAGACCCCGAAATATACCAATGAAGGTATCAAATTTCTGAGCGTGGAAAACATCAAAACCCTG
AATAGCAGCAAATACATTAGCGAAGAAGCCTTCGAGAAAGAATTCAAAATTCGTCCGGAATTTGGC
GATATTCTGATGACCCGTATTGGTGATATTGGCACCCCGAATATTGTTAGCAGCAATGAAAAATTC
GCCTACTATGTTAGCCTGGCACTGCTGAAAACCAAAAATCTGAACAGCTACTTCCTGAAAAACCTG
ATTCTGAGCAGCAGCATTCAGAATGAACTGTGGCGTAAAACCCTGCATGTTGCATTTCCGAAAAAA
ATCAACAAAAACGAGATCGGCAAAATCAAAATCAACTACCCGAAAAAACAAGAACAGCAGAAAATC
GGTCAGTTTTTCAGCAAACTGGATCGCCAAATTGAATTAGAAGAACAGAAGCTGGAACTGCTGCAA
CAGCAGAAAAAAGGTTATATGCAGAAAATCTTCAGCCAAGAGCTGCGCTTTAAAGATGAAAATGGT
GAAGATTATAGCGAGTGGGAAGAACGTCGTTTTGCCGATATTTTCAAATTTCACAACAAACTGCGC
AAACCGATCAAAGAAAATCTGCGTGTTAAAGGCAGCTATCCGTATTATGGTGCAACCGGCATTATT
GATTATGTGGATGATTTTATCTTCGATGGCAACTATCTGCTGATTGGCGAAGATGGTGCAAACATT
ATTACCCGTAGCGCACCGCTGGTTTATCTGGTTAATGGTAAATTTTGGGTGAACAACCATGCCCAT
ATTCTGAGTCCGCTGAATGGTAATATTCAGTATCTGTATCAGGTTGCCGAACTGGTGAACTATGAA

AAATACAATACCGGCACCGCACAGCCGAAACTGAACATTCAGAATCTGAAAATTATCAACGTGGTG
ATCAGCACCAATCTGGAAGAACAGCAAAAAATTGGTAGCTTCCTGAGCAAACTGGATCGTCAGATT
GACCTGGAAGAACAAAAACTGGAACTGCTGCAACAACGTAAAAAAGCACTGCTGAAAAGCATGTTC
GTGCCCGGGGGATCCGATCGATC

CC59-1 TRD Q

QIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGEDYSEWEERRFADIFKFHNKLRKPIKENLRV
KGSYPYYGATGIIDYVDDFIFDGNYLLIGEDGANIITRSAPLVYLVNGKFWVNNHAHILSPLNGNI
QYLYQVAELVNYEKYNTGTAQPKLNIQNLKIINVVISTNLEEQQKIGSFLSKLDRQIDLEEQKLEL
LQQRKKALLKSMFV

CC72-1 TRD R

PGFEGEWEEKKLGEVAKIYDGTHQTPKYTNEGIKFLSVENIKTLNSSKYISEEAFEKEFKIRPEFG
DILMTRIGDIGTPNIVSSNEKFAYYVSLALLKTKNLNSYFLKNLILSSSIQNELWRKTLHVAFPKK
INKNEIGKIKINYPKKQEQQKIGQFFSKLDRQIELEEQK

CC75-1 TRD T and TRD U

CCCGGGTTTGAAGGCGAATGGGAGGAAAAAGAACTGGGCGAAATCTTTCAGATTATTAGCGGTAGC
ACACCGCTGAAAAGCAACAAAGAATTTTATGAGAACGGCAACATCAACTGGGTTAAAACCACCGAT
CTGAATAATAGCAAAGTGACCCATAGCAAAGAAAAAATCACCGAGTATGCAATGAAAAGCCTGAAA
CTGAAACTGGTGCCGAAAAATAGCGTTCTGATTGCAATGTATGGTGGCTTTAATCAGATTGGTCGT
ACCGGTCTGCTGAAAATTGATGCAACCATTAATCAGGCAATTAGCGCACTGCTGATGAATCATGAA
ACCAACCCGGAATTTATTCAGGCCTTTCTGAATTATCAGGTGAAAGGTTGGAAACGTTATGCAGCA
AGCAGCCGTAAAGATCCGAATATCACCAAAAAAGATATCGAACAGTTCAAAGTGCCGTACGTGAGC
ATTAATGAACAGCAGAAAATTGGCGAGTTTTTTAGCAAAATCGATCATCAAATTGAATTAGAAGAA
CAGAAGCTGGAACTGCTGCAACAGCAGAAAAAAGGTTATATGCAGAAAATCTTCAGCCAAGAGCTG
CGCTTTAAAGATGAAAATGGTGAAGATTATCCGGATTGGGAAGTTACCACCATTCAGAACATTACC
AAATACACCAGCAGCAAAAAAAGCAGCAATCAGTATGCCGATAAAGACAACAGCAAAGGTTATCCG
GTTTATGATGCCGTTCAAGAAATTGGCAAAGATAGCAACTATGACATCGAAGAGAGCTATATCAGC
ATTCTGAAAGATGGTGCCGGTGTTGGTCGTCTGAATCTGCGTCCGGGTAAAAGCAGCGTTATTGGC
ACCATGGGTTATATTCAGAGCAACAACGTGGATATCGAGTTCCTGTATTATCGTATGAAAGTGGTG
GACTTCAAAAAATACATTATCGGTAGCACCATTCCGCACCTGTATTTCAAAGATTATAGCAAAGAA
ACCCTGTACATTCCGAGCAGCATTCAAGAACAGGCAAAAATTGGTATGTTCATCAGCAACCTGGAT
AAACTGATCGAGAACAAAAACCTGAAACTGAACTGTCTGAAACAACTGAAACAGGGATTGCTACAA
TCTATGTTTATTCCCGGGGGATCCGATCGATC

CC75-1 TRD T

PGFEGEWEEKELGEIFQIISGSTPLKSNKEFYENGNINWVKTTDLNNSKVTHSKEKITEYAMKSLK
LKLVPKNSVLIAMYGGFNQIGRTGLLKIDATINQAISALLMNHETNPEFIQAFLNYQVKGWKRYAA
SSRKDPNITKKDIEQFKVPYVSINEQQKIGEFFSKIDHQIELEEQK

CC75-1 TRD U

QIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGEDYPDWEVTTIQNITKYTSSKKSSNQYADKD
NSKGYPVYDAVQEIGKDSNYDIEESYISILKDGAGVGRLNLRPGKSSVIGTMGYIQSNNVDIEFLY
YRMKVVDFKKYIIGSTIPHLYFKDYSKETLYIPSSIQEQAKIGMFISNLDKLIENKNLKLNCLKQL
KQGLLQSMFI

CC75-2 TRD V

CCCGGGTTTGAAGGCGAATGGGAGGAAAAAGAACTGCGTGAACTGCGCAATCCGAAAGATAAATAC
AGCTATACCGGTGGTCCGTTTGGTAGCGATCTGAAAAAAAGCGATTATACCACCGATGGCATTCAG
ATTATTCAGCTGCAGAATATTGGTGACGGCTATTTCTATAACAGCAACAAAGTGTTTACCAGCAAC
GAAAAAGCCGAAGTTCTGAAAAGCTGTAATGTTTTTCCGGGTGATATTGTGATTGCCAAAATGGCA
GATCCGATTGCACGTGCCGCAATTGTTCCGGATAATAACATTGGTAAATACCTGATGGCCAGTGAT
GGTATTCGTCTGAGCGTTGATACCGTTCATTTTAACACCAAATTTGTGCTGGAATGCATCAACCGT
AAAAGCTTTCGTAAAAAAGTCGAGGATAATAGCAGCGGTAGCACCCGTATGCGTATTGGTCTGAGT
ACCCTGGGTAGCCTGACCCTGAAAACCACCACCCTGAAAGAACAGCAGAAAATTGGTCAGTTTTTC

AGCAAACTGGATCGTCAAATTGAATTAGAAGAACAGAAG

CC75-2 TRD V

PGFEGEWEEKELRELRNPKDKYSYTGGPFGSDLKKSDYTTDGIQIIQLQNIGDGYFYNSNKVFTSN
EKAEVLKSCNVFPGDIVIAKMADPIARAAIVPDNNIGKYLMASDGIRLSVDTVHFNTKFVLECINR
KSFRKKVEDNSSGSTRMRIGLSTLGSLTLKTTTLKEQQKIGQFFSKLDRQIELEEQK

CC75-2 TRD W

CAAATTGAATTAGAAGAACAGAAGCTGGAACTGCTGCAACAGCAGAAAAAAGGTTATATGCAGAAA
ATCTTCAGCCAAGAGCTGCGCTTTAAAGATGAAAATGGTAACGATTATCCGGATTGGGAAGAAAAA
CAGCTGGGTGAACTGAGCCAGATTGTTCGTGGTGCAAGTCCGCGTCCGATTAAAGATCCGAAATGG
TTTAACAAAGAAAGCGATATTGGTTGGCTGCGCATTAGTGATGTTACCAATCAGAATGGCAAAATC
TATCATCTGGAACAGAAACTGAGCATCGAAGGTCAAGAAAAAACCCGTGTTCTGGTTACCACCCAT
CTGCTGCTGAGCATTGCAGCAAGCATTGGTAAACCGGTTATGAACTTTGTGAAAACCGGTGTGCAT
GATGGCTTTCTGATTTTTCTGAAACCGAAATTCAACCTGTTCTTTATGTACTATTGGCTGGAATAT
TTCAAAGATAAATGGTCCAAATATGGTCAGCCTGGTAGCCAGGTTAATCTGAATAGCGAAATTGTT
AAAAGCCAGACCCTGAATATGCCGAGCAATCATGAACAAGAAAAAGTGGGCCAGTTTTTTAACCGC
AACGAAAAACTGATTGAACTGCAGCAAGAGAAAATCATGTATATCAAACGTTGCAAACAGGTGCTG
CTGCAAAAAATGTTTATTCCCGGGGGATCCGATCGATC

CC75-2 TRD W

QIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPDWEEKQLGELSQIVRGASPRPIKDPKW
FNKESDIGWLRISDVTNQNGKIYHLEQKLSIEGQEKTRVLVTTHLLLSIAASIGKPVMNFVKTGVH
DGFLIFLKPKFNLFFMYYWLEYFKDKWSKYGQPGSQVNLNSEIVKSQTLNMPSNHEQEKVGQFFNR
NEKLIELQQEKIMYIKRCKQVLLQKMFI

CC80-1 TRD X and TRD Y

CCCGGGTTTGAAGGCGAATGGGAGGAAAAACAGTTTGCCGACTTCACCAAAATTAACCAGGGTCTG
CAGATTGCCATTAATGAACGTAAAACCGAATATAGCCCTGAGCTGTATTTCTATATCACCAACGAA
TTTCTGCGTCCGAATAGCCAGACCAAATATTTCATTGAAAATCCGCCTCAGAGCGTGATTGCCAAC
AAAGAAGATATTCTGATGACCCGCACCGGTAATACCGGCAAAGTTGTTACCAATGTTTTTGGTGCC
TTCCACAACAACTTTTTCAAAATCAAATTCGATAAAAACCTGTATGATCGCCTGTTTCTGGTTGAA
GTTCTGAACAGCAGCAAAATCCAGAACAAAATTCTGAGCCTGGCAGGTAGCAGCACCATTCCGGAT
CTGAATCATAGCGATTTCTATAGCATTAGCAGCAGCTATCCGCTGCTGCGCGAACAGCAAAAAATT
GGCAAATTCTTTAGCAAACTGGATCGTCAAATTGAATTAGAAGAACAGAAGCTGGAACTGCTGCAA
CAGCAGAAAAAAGGTTATATGCAGAAAATCTTCAGCCAAGAGCTGCGCTTTAAAGATGAAAATGGT
AACGATTATCCGGATTGGGAGAAAAAAAAACTGAAAGAAATTGCCTGCGTGTATACCGGTAATACC
CCGAGCAAAAAAGAAAACATCTATTGGAACAAAGGCGAGTATGTTTGGGTTACCCCGACCGATATT
AACAACAGCAAAAACATTTATGAAAGCGAAAACAAACTGACCCAAGAAGGCTACAAAAAAGCACGT
CAGCTGCCGGAAAATACCCTGCTGGTTACCTGTATTGCAAGCATTGGTAAAAATGCCATTCTGCGT
AAACAGGGTAGCTGTAATCAGCAGATTAATGCAGTTGTGCCGTTTGAGAACATCAACATCGATTAT
CTGTATTATATCAGCGATAGCCTGAGCACCTTCATGAAAAGCATTGCAGGTAAAACCGCAACCCAG
ATTGTGAACAAAAACACCTTTGAAAACCTGGAAATTTACCTGGCACCTTTTGAGGAACAGAACAAA
ATTGCAGATCTGATTAGCAGCCTGGAAGAACTGATTGAAAAACAGGCAAGCAAACTGATCAAAATG
AAAAGCCGCAAACAGGGCATGCTGCAGATTATGTTTATTCCCGGGGGATCCGATCGATC

CC80-1 TRD X

PGFEGEWEEKQFADFTKINQGLQIAINERKTEYSPELYFYITNEFLRPNSQTKYFIENPPQSVIAN
KEDILMTRTGNTGKVVTNVFGAFHNNFFKIKFDKNLYDRLFLVEVLNSSKIQNKILSLAGSSTIPD
LNHSDFYSISSSYPLLREQQKIGKFFSKLDRQIELEEQK

CC80-1 TRD Y

QIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPDWEKKKLKEIACVYTGNTPSKKENIYW
NKGEYVWVTPTDINNSKNIYESENKLTQEGYKKARQLPENTLLVTCIASIGKNAILRKQGSCNQQI
NAVVPFENINIDYLYYISDSLSTFMKSIAGKTATQIVNKNTFENLEIYLAPFEEQNKIADLISSLE
ELIEKQASKLIKMKSRKQGMLQIMFI

CC80-2 TRD Z + CC72-2 TRD S

CCCGGGTTTGAAGGCGAATATTCTCTGGATATTTTTGGTAATCTGGCCACCAACAAAAGCGAAAAA
TTCAATCCGCAGAATGAAAACGCCAGCATTGATATTGAACTGGATTGCATTGAACAGAATACCGGT
CGTCTGATCAAAATCTATAACAGCAAAGAATTTAGCAGCCAGAAAAACAAATTTAACCCGCAGAAC
GTGCTGTATGGTAAACTGCGTCCGTATCTGAACAAATATTACTTCACCAAAAAAAGTGGTGTGTGC
AGCAGCGAAATTTGGGTTCTGAAAAGCACCAAAGAAGATAAACTGCTGAACCTGTTCCTGTACTAT
TTCATTCAGACCAAACGCTATAGTGATGTTGCAAGCAAAAGCGCAGGTAGCAAAATGCCTCGTGCA
GATTGGGGTCTGATTGAAAATATTCGTGTGTATTTTCCGGAACTGTGCGAACAGCAGAAAATTGGT
CAGTTTTTTAGCAAACTGGACCGTCAAATTGAATTAGAAGAACAGAAGCTGGAACTGCTGCAACAG
CAGAAAAAAGGTTATATGCAGAAAATCTTCAGCCAAGAGCTGCGCTTTAAAGATGAAAATGGTAAC
GATTATCCGGACTGGACCAATGAACGTCTGGGTGAAGTTACCACCGTTACCATGGGTCAGAGCCCG
AAAAGCGTGAATTATACCGATAATAGCAATGACACCGTTCTGATTCAGGGTAATGCCGATATTGAA
AACGGTCTGATTAATCCGCGTATCTATACCCGTGAAGTGACCAAACTGATTCAGAAAGATGAGATT
ATTCTGACCGTTCGTGCACCGGTTGGTAAACTGGCAATGGCACAGATTAATGCATGTATTGGTCGT
GGTGTTTGCAGCATTAAAGGCGATAAATTTCTGTATTATTTCCTGGAATGGTTCGCCACCCAGAAT
AAATGGATTCGTTTTAGCCAGGGTAGCACCTTTGAAAGCATTAGCGGTAATGATATTCGCAACATC
CATATCAAAATCCCGGTTGAAGATGAACGCACCAAAATTATCAAACTGCTGAATAGCCTGGATGTG
CTGAATTCAAAAACCGATCTGAAAATCCAGAATCTGAAACAGCGTAAACAGAGCCTGCTGCAAAAA
ATCTTTGTGCCCGGGGGATCCGATCGATC

CC80-2 TRD Z

PGFEGEYSLDIFGNLATNKSEKFNPQNENASIDIELDCIEQNTGRLIKIYNSKEFSSQKNKFNPQN
VLYGKLRPYLNKYYFTKKSGVCSSEIWVLKSTKEDKLLNLFLYYFIQTKRYSDVASKSAGSKMPRA
DWGLIENIRVYFPELCEQQKIGQFFSKLDRQIELEEQK

CC72-2 TRD S

QIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPDWTNERLGEVTTVTMGQSPKSVNYTDN
SNDTVLIQGNADIENGLINPRIYTREVTKLIQKDEIILTVRAPVGKLAMAQINACIGRGVCSIKGD
KFLYYFLEWFATQNKWIRFSQGSTFESISGNDIRNIHIKIPVEDERTKIIKLLNSLDVLNSKTDLK
IQNLKQRKQSLLQKIFV

CC93-2 TRD b*

CCCGGGTTTGAAGGCGAATGGGAGGAGAAAAAACTGGAAGATACCCTGGAATTCATTAAAGATGGC
ACCCATGGTACACATGAAAATGTTAATAATGGTCCGTGGCTGCTGAGCGCCAAAAACATTAAAAAC
AACAAAATCATCATCAGCAGCGACGATCGCAAAATTAGCGAAAGCGATTACAAAAAAATCTACAAA
AACTATAAACTGGAAAAAGGCGATCTGCTGCTGACCATTGTTGGCACCATTGGTCGTGCAGCAATT
GTTAAAAATCCGAACAATATTGCCTTTCAGCGTAGCGTTGCAATCCTGAAAACCAAAGCAACCTAT
GATGTGGGCTTTATCTTTCAGCTGTTCCAGACCAAATACTTTAAAAACCTGCTGCTGCGTAAACAG
GTTGTTAGCGCACAGCCTGGTCTGTATCTGGGTGATATTCGTAAAATCAAAATCAGCATTACCAAC
ATCATCGAAGAACAGCGCAAAATCGGTATCTTTTTCAGCAAACTGGATCGTCAAATTGAATTAGAA
GAACAGAAG

CC93-2 TRD b*

PGFEGEWEEKKLEDTLEFIKDGTHGTHENVNNGPWLLSAKNIKNNKIIISSDDRKISESDYKKIYK
NYKLEKGDLLLTIVGTIGRAAIVKNPNNIAFQRSVAILKTKATYDVGFIFQLFQTKYFKNLLLRKQ
VVSAQPGLYLGDIRKIKISITNIIEEQRKIGIFFSKLDRQIELEEQK

C93-3 TRD a*

CAAATTGAATTAGAAGAACAGAAGCTGGAACTGCTGCAACAGCAGAAAAAAGGTTATATGCAGAAA
ATCTTCAGCCAAGAGCTGCGCTTTAAAGATGAAAATGGTAACGATTATCCGGAATGGGAAAACAAA
CGCATTGAAGATATTGCCAATGTGAACAAAGGTTTTACCCCGAGCACCAACAATAACGAATATTGG
GATAACAACGATAAAAACTGGCTGAGCATTGCAGGCATGAATCAGAAATATCTGTATAAAGGCAAC
AAAGGCATCAGCAAAGATGCAGCCAAAAACTATATGAAAGTGAAAAACGACACCCTGATCATGTCC
TTTAAACTGACCATTGGTAAACTGGCGATTGTTAAAGCACCGCTGTATACCAATGAAGCCATTTGT
CATTTTATCTGGAAAGTGAACAAAATCAACACCGAGTTCATCTACTATTACCTGAACAGCCTGAAC
ATTAGCACCTTTGGTGTTCAGGCAGTTAAAGGTGTTACCCTGAATAACGATAGCATCAACAGCATT

ATTGTGAAACTGCCGAATGAAGAGGAACAGAACATTATCGCAAAATTTCTGCTGGAAGTGGACAAA
ACCGTTAATAATCAGCTGGTGAAAACCAAACTGCTGAAACAACGTAAAAAAGGCCTGCTGCAGCGT
ATGTTTGTTCCCGGGGGATCCGATCGATC

CC93-3 TRD a*

QIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPEWENKRIEDIANVNKGFTPSTNNNEYW
DNNDKNWLSIAGMNQKYLYKGNKGISKDAAKNYMKVKNDTLIMSFKLTIGKLAIVKAPLYTNEAIC
HFIWKVNKINTEFIYYYLNSLNISTFGVQAVKGVTLNNDSINSIIVKLPNEEEQNIIAKFLLEVDK
TVNNQLVKTKLLKQRKKGLLQRMFV

CC873 TRD e* + CC97 TRD c*

CCCGGGTTTGAAGGCGAATGGGAGGAAAAATCGATCAGCAGCTTTCTGAAAGAAAGCAAAATCAAA
GGTAGCAATGGTAGCCATGCAAAAAAACTGACCGTTAAACTGTGGGGTAAAGGTGTTGTTCCGAAA
AAAGAAACGTTTAAAGGCAGCGATAACACCCAGTATTACAAACGTAAAGCAGGTCAGCTGATGTAT
GGCAAACTGGATTTTCTGAATTGCGCCTTTGGTATTGTTCCGGATAGCCTGAATAACTATGAAAGC
ACCATTGATAGCCCGAGCTTTGATTTCATTAATGGCGATAGCAAATTTCTGCTGGAACGCATTAAA
CTGAAAAGCTTCTACAAAAAATTCGGCGATATTGCAAATGGCAGCCGTAAAGCAAAACGTATTAAT
CAGGATACCTTTCTGAGCCTGCCGGTTTTTGCACCGAAATATGATGAACAGCTGCGTATTGGTGAA
TTTTTCAGTAAACTGGATCGTCAAATTGAATTAGAAGAACAGAAGCTGGAACTGCTGCAACAGCAG
AAAAAAGGTTATCTGCAGAAAATCTTTAGCCAAGAGCTGCGCTTTAAAGATGAAAACGGTAATGAT
TATCCGGAATGGCGTTTTGCCCGTTTCAAAGATTTTATGTACAAACCGATTAATATCCGTCCGGCA
ATCAACATTAGCAAAAGCGAACTGCTGACCGTTAAACTGCATTGCAAAGGTATTGAAAAAGCCAAC
ATTAACCGTGTGCTGAAACTGGGTGCAACCAATTATTACAAACGTTTTGAAGGCCAGTTTATCTAT
GGCAAACAGAACTTTTTTAACGGTGCCTTTGATATCGTGCCGAAAAAATTCGATGGTCTGTATAGC
AGCAGTGATGTTCCGGCATTTGAAATCAATACCGAGAAAATTGAGCCGAACTACTTCATCAGCTAT
ATTAGCCGTCCGAGCTTCTATAAAAGCAAAGAGAAATATAGCACCGGCACCGGTAGCAAACGTATT
CATGAAAATACCGTGCTGAACTTTAGCCTGCATCTGCCGTGTCTGAATGAACAGCTGAAAATTGCA
AGCTTTGTGTGCTTTCTGAACCGTAAAATTGAACTGCTGGAACGCAAAATCTATCTGATCAAAAAA
CAGAAACAGGCCCTGCTGCAGCAAATGTTTATTCCCGGGGGATCCGATCGATC

CC873 TRD e*

PGFEGEWEEKSISSFLKESKIKGSNGSHAKKLTVKLWGKGVVPKKETFKGSDNTQYYKRKAGQLMY
GKLDFLNCAFGIVPDSLNNYESTIDSPSFDFINGDSKFLLERIKLKSFYKKFGDIANGSRKAKRIN
QDTFLSLPVFAPKYDEQLRIGEFFSKLDRQIELEEQK

CC97 TRD c*

QIELEEQKLELLQQQKKGYLQKIFSQELRFKDENGNDYPEWRFARFKDFMYKPINIRPAINISKSE
LLTVKLHCKGIEKANINRVLKLGATNYYKRFEGQFIYGKQNFFNGAFDIVPKKFDGLYSSSDVPAF
EINTEKIEPNYFISYISRPSFYKSKEKYSTGTGSKRIHENTVLNFSLHLPCLNEQLKIASFVCFLN
RKIELLERKIYLIKKQKQALLQQMFI

CC133-2 from ED133 TRD d*

CAAATTGAATTAGAAGAACAGAAGCTGGAACTGCTGCAACAGCAGAAAAAAGGTTATATGCAGAAA
ATCTTCAGCCAAGAGCTGCGCTTTAAAGATGAAAATGGTAACGATTATCCGGAATGGGAAAATGTG
ATGCTGCAGAAAGTTCTGAAAGATAAAACCGAAGGTATTAAACGTGGTCCGTTTGGTGGTGCACTG
AAAAAAGATATTTTTGTGGAAAGCGGCTATGCCGTTTATGAACAGCGTAATGCCATTTATGATATC
AGCAACTTCCGCTACTATATCAACGAGAACAAATACAAAGAGATGCAGAGCTTTAGCGTTCAGCCG
AATGATATTATCATGAGCTGTAGCGGCACCATTGGTCGTCTGGCACTGATTCCGCATAACTATACC
AAAGGTATTATCAACCAGGCCCTGATTCGTTTTCGTACCAATCATAAAATCCGCAGCGAATTCTTT
CTGATCTTTATGCGTAGCAATCAGATGCAGCGTAAAATTCTGGAAGCAAATCCGGGTAGCGCAATT
ACCAATCTGGTTCCGGTTAAAGAACTGAAACTGATCCCGTTTCCGCTGCCGGTTAAATTTGAACAG
GATAAAATCAGCCAGTTCATCCACATTATTAACCGTCGTATTGAACAGAGCGAGAAAAAAATCGAA
AGCCTGAAAAATCGCAAACAGGGTTTCCTGCAGAAACTGTTTGTTCCCGGGGGATCCGATCGATC

CC133-2 from ED133 TRD d*

QIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPEWENVMLQKVLKDKTEGIKRGPFGGAL
KKDIFVESGYAVYEQRNAIYDISNFRYYINENKYKEMQSFSVQPNDIIMSCSGTIGRLALIPHNYT

KGIINQALIRFRTNHKIRSEFFLIFMRSNQMQRKILEANPGSAITNLVPVKELKLIPFPLPVKFEQ
DKISQFIHIINRRIEQSEKKIESLKNRKQGFLQKLFV

ST80-3 TRD X + TRD f*

CCCGGGTTTGAAGGCGAATGGGAGGAAAAACAGTTTGCCGATTTTACCAAAATTAACCAGGGTCTG
CAGATTGCCATTAATGAACGTAAAACCGAATATAGCCCTGAGCTGTATTTCTATATCACCAACGAA
TTTCTGCGTCCGAATAGCCAGACCAAATATTTCATTGAAAATCCGCCTCAGAGCGTGATTGCCAAC
AAAGAAGATATTCTGATGACCCGCACCGGTAATACCGGCAAAGTTGTTACCAATGTTTTTGGTGCC
TTCCACAACAACTTTTTCAAAATCAAATTCGATAAAAACCTGTATGATCGCCTGTTTCTGGTTGAA
GTTCTGAACAGCAGCAAAATCCAGAACAAAATTCTGAGCCTGGCAGGTAGCAGCACCATTCCGGAT
CTGAATCATAGCGATTTCTATAGCATTAGCAGCAGCTATCCGCTGCTGCGCGAACAGCAAAAAATT
GGCAAATTCTTTAGCAAACTGGATCGCCAGATTGAACTGGAAGAACAGAAACTGGAACTGCTGCAA
CAGCAGAAAAAAGGCTATATGCAGAAAATCTTTAGCCAAGAGCTGCGCTTTAAAGATGAAAACGGT
GAAGATTATCCGGATTGGAAAGAAAAAAAACTGGGCGATATTACCGAGCAGAGCATGTATGGTATT
GGTGCAAGCGCAACCGTTTTGATAGCAAAAATATCTATATCCGCATCACCGACATCGATGAAAAAA
GCCGTAAACTGAATTATCAGAATCTGACCACACCGGATGAACTGAACAATAAATACAAACTGAAAC
GCAACGACATCCTGTTTGCACGTACCGGTGCAAGTACCGGTAAAAGCTATATTCATAAAGAAGAGA
AAGACATCTACAACTACTACTTTGCGGGTTTTCTGATCAAATTCAAAATTAACGAACAGAACAGTC
CGCTGTTCATCTATCAGTTTACCCTGACCAGCAAATTCAACAAATGGGTTAAAGTTATGAGCGTGC
GTAGCGGTCAGCCTGGTATTAATAGCGAAGAATATGCAAAACTGCCGCTGGTTCTGCCGAATAAAC
TGGAACAACAAAAAATCGCGAAATTCCTGGATCGTTTTGATCGTCAGATCGAGCTGGAAAAACAAA
AAATTGAAATTCTGCAGCAACAAAAAAAAGGCCTGCTGCAGAGTATGTTTATTCCCGGGGGATCCG
ATCGATC

ST80-3 TRD X

MSNTQKKNVPELRFPGFEGEWEEKQFADFTKINQGLQIAINERKTEYSPELYFYITNEFLRPNSQT
KYFIENPPQSVIANKEDILMTRTGNTGKVVTNVFGAFHNNFFKIKFDKNLYDRLFLVEVLNSSKIQ
NKILSLAGSSTIPDLNHSDFYSISSSYPLLREQQKIGKFFSKLDR

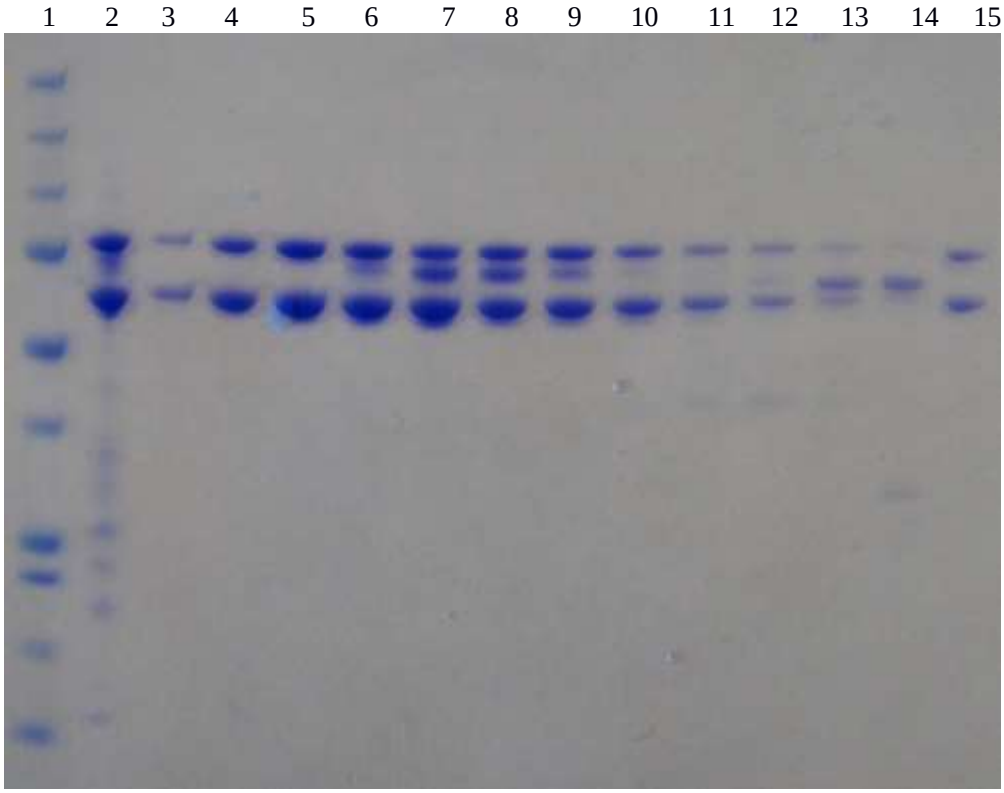
ST80-3 TRD f*

QIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGEDYPDWKEKKLGDITEQSMYGIGASATRFDS
KNIYIRITDIDEKSRKLNYQNLTTPDELNNKYKLKRNDILFARTGASTGKSYIHKEEKDIYNYYFA
GFLIKFKINEQNSPLFIYQFTLTSKFNKWVKVMSVRSGQPGINSEEYAKLPLVLPNKLEQQKIAKF
LDRFDRQIELEKQKIEILQQQKKGLLQSMFIPGGSHHHHHH

SUPPLEMENTARY INFORMATION FOR TABLE 2.

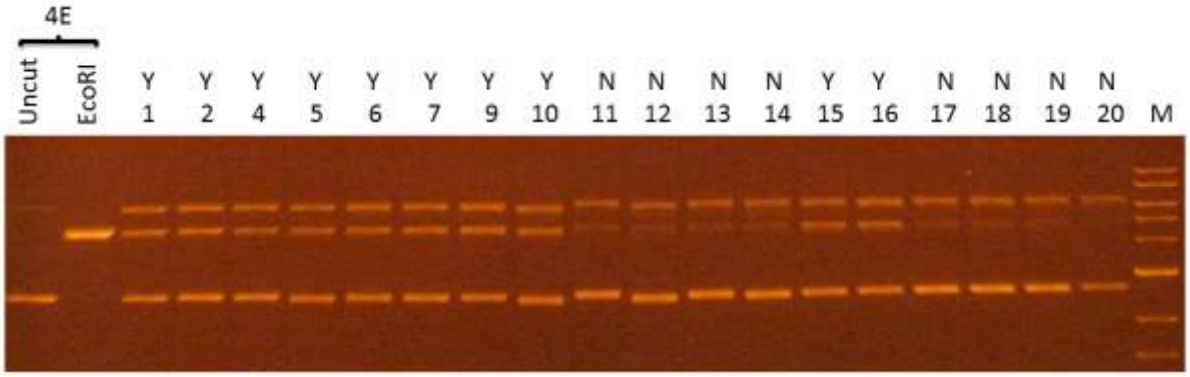
S.SauCD-EGFP
CC30-1 GWAG-5-GAT
This MTase was expressed and purified as a fusion with EGFP.

MSNTQTKNVPELRFPGFEGEWEEKKVGELLEFKNGLNKGKEYFGSGSSIVNFKDVFNNRSLNTNNL
TGKVNVNSKELKNYSVEKGDVFFTRTSEVIGEIGYPSVILNDPENTVFSGFVLRGRPKSGIDLINN
NFKRYVFFTNSFRKEMITKSSMTTRALTSGSAINKMKVIYPVSAKEQRKIGDFFSKLDRQIELEEQ
KLELLQQQKKGYMQKIFSQELRFKDENSEDYPHWENSKIEKYLKERNERSDKGQMLSVTINSGIIK
FSELDRKDNSSKDKSNYKVVRKNDIAYNSMRMWQGASGRSNYNGIVSPAYTVLYPTQNTSSLFIGY
KFKTHRMIHKFKINSQGLTSDTWNLKYKQLKNINIDIPVLEEQEKIGDFFKKMDILISKQKIKIEI
LEKEKQSFLQKMFLGSMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICT
TGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFE
GDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADH
YQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKHHHHHH



1- marker 2- Nickel column eluate 3-14 Fractions from gel filtration column
15- CC5-1 Purified protein marker

Nuclease assay on the plasmid library.

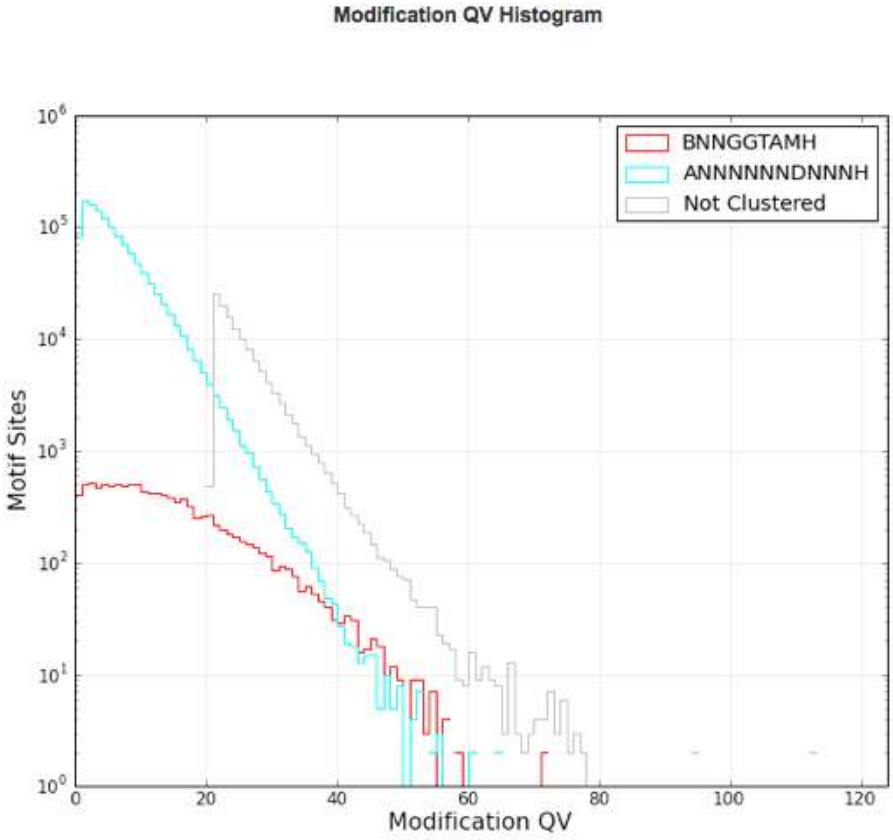


S.SauCD-EGFP
CC30-1 GWAG-5-GAT

SMRT did not work for the CC30-1 system when looking for methylation of genomic DNA from *E. coli*.

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
BNNGGTAMH	4	unknown	7.05	780	11058	37.7	88.6	
ANNNNNNDNNNH	1	unknown	0.11	1312	1235059	36.0	100.7	
Not Clustered	0		0.19	14583	7880091	36.1	107.4	

Modification QV Histogram By Motif



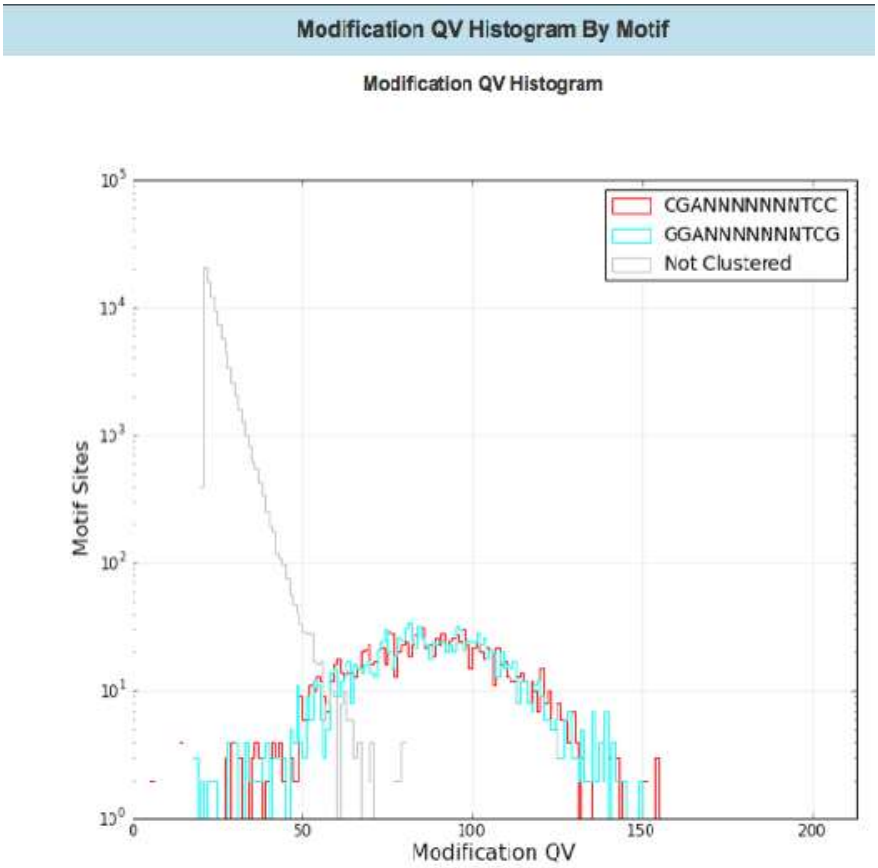
S.SauJK-EGFP
CC30-2 GGA-7-TCG
This MTase was a fusion with EGFP.

MSNTQKKNVPELRFPEFEGEWEERKLGDLIKVNSGKDYKHLDKGDIPVYGTGGYMTSVSEPLSEID
AVGIGRKGTINKPYLLEAPFWTVDTLFYCTPEKEADILFILSLFRKINWKLYDESTGVPSLSKQTI
NKINRLVPTNKEQQKIGEFFSKLDRQIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPKW
EEKKIEDIASQVYGGGTPNTKIKEFWNGDIPWIQSSDVKVNDLILQQCNKFISKNSIELSSAKLIP
ANSIAIVTRVGVGKLCLVEFDYATSQDFLSLSSLKYDKLYSLYSLLYTMKKISANLQGTSIKGITK
KELLDSIIKIPHNLEEQQKIGDLFYKIDKYISFNKCKIEMLKSLKQGLLKKMFIGSMVSKGEELFT
GVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYP
DHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKL
EYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSAL
SKDPNEKRDHMVLLEFVTAAGITLGMDELYKHHHHHH

This system could not be expressed so was used for SMRT sequencing only.

SMRT analysis of genomic DNA from *E. coli*.

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
CGANNNNNNNTCC	3	m6A	98.76	1439	1457	89.3	76.9	GGANNNNNNNTCG
GGANNNNNNNTCG	3	m6A	98.56	1436	1457	91.2	76.8	CGANNNNNNNTCC
Not Clustered	0		0.09	8260	9123294	35.7	87.7	



S.SauJd*
CC133-2 from ED133 GGA-7-TTRG
This enzyme was studied using the SMRT assay. There are minor variations in S subunit sequence in CC133-2.
Recombinant S.SauJd* CC133-2

MSNTQKKNVPELRFPGFEGEWEEKKLGDLIKVNSGKDYKHLEKGDIPVYGTGGYMTSVSEPLSEID
AVGIGRKGTINKPYLLEAPFWTVDTLFYCTPKKETDILFILSLFRKINWKVYDESTGVPSLSKQTI
NKINRFVPSNKEQQKIGEFFIKLDRQIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPEW
ENVMLQKVLKDKTEGIKRGPFGGALKKDIFVESGYAVYEQRNAIYDISNFRYYINENKYKEMQSFS
VQPNDIIMSCSGTIGRLALIPHNYTKGIINQALIRFRTNHKIRSEFFLIFMRSNQMQRKILEANPG
SAITNLVPVKELKLIPFPLPVKFEQDKISQFIHIINRRIEQSEKKIESLKNRKQGFLQKLFVPGGS
HHHHHH

Wild type S.SauJd*

MSNTQKKNVPELRFPGFEGEWEEKKLESIIKVNSGKDYKHLDKGDIPVYGTGGYMTSVSEPLSEID
AVGIGRKGTINKPYLLEAPFWTVDTLFYCTPKKETDILFILSLFRKINWKVYDESTGVPSLSKQTI
NKINRFVPTNKEQQKIGKFFSKLDRQIELQEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPEW
ENVMLQKVLKDKTEGIKRGPFGGALKKDIFVESGYAVYEQRNAIYDISNFRYYINENKYKEMQSFS
VQPNDIIMSCSGTIGRLALIPHNYTKGIINQALIRFRTNHKIRSEFFLIFMRSNQMQRKILEANPG
SAITNLVPVKELKLIPFPLPVKFEQDKISQFIHIINRRIEQSEKKIESLKNRKQGFLQKLFV*

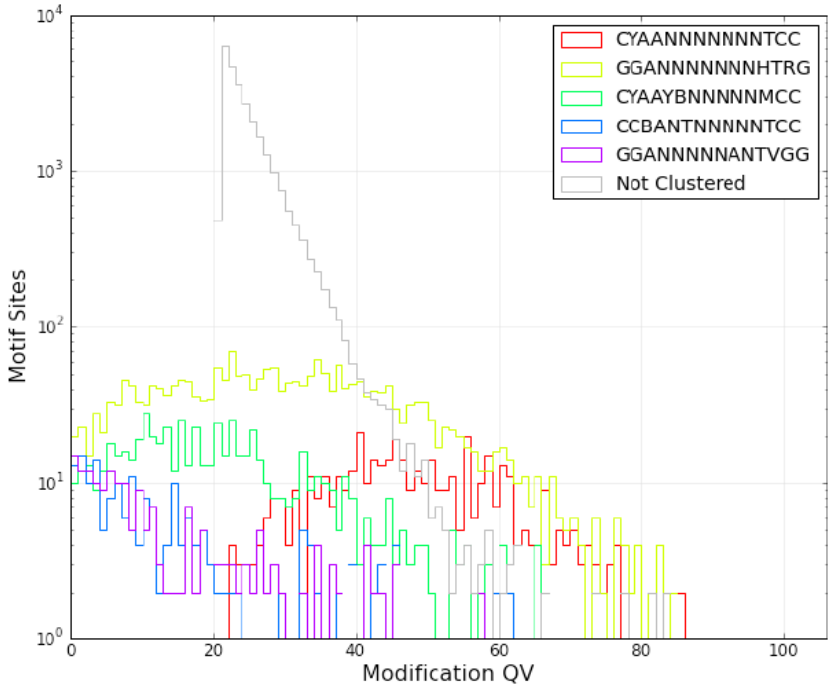
Reports for Job Dryden_J_delta_MODs

SMRT Cells: 1 Movies: 1

Motif Summary

Motifs	Modified Position	Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
CYAANNNNNNNTCC	4	m6A	90.36%	422	467	51.19	30.33	
GGANNNNNNNHTRG	3	m6A	47.67%	1114	2337	45.72	32.24	
CYAAYBNNNNNMCC	4	m6A	25.68%	169	658	42.89	34.51	
CCBANTNNNNNTCC	4	m6A	20.39%	42	206	44.40	32.14	GGANNNNNANTVGG
GGANNNNNANTVGG	3	m6A	18.45%	38	206	44.76	31.37	CCBANTNNNNNTCC

Modification QVs



S.SauNE
CC398-1 ACC-5-RTGA

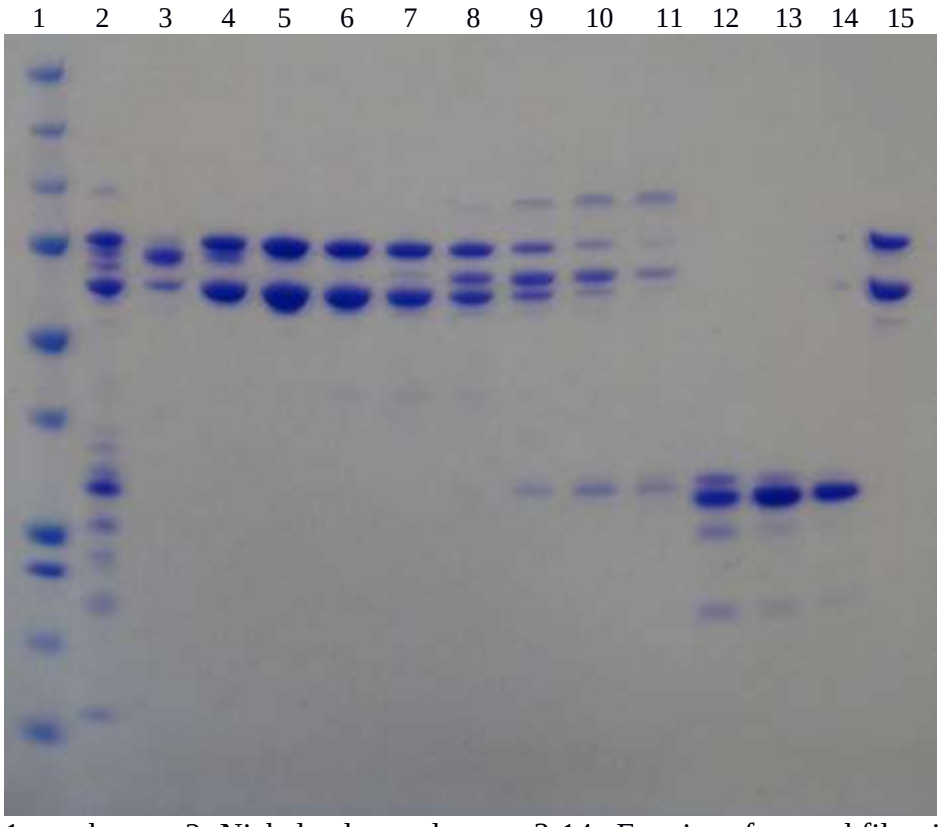
The clone obtained contained a single amino acid substitution A50S which did not affect activity. The enzyme was expressed using plasmid pSauNE-XmaI.

S.SauNEXmaI "Expected" sequence

MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNS**A**ELGIISQKDYFDKEIS
NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGKKGVMSPLYTVFKIQNIDLNFIEFYFKSS
KWYRFMALNGDSGARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
KKGYMQKIFSQELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKDAITNGSYDFYVRSPIVYKIN
TFSYEGEAILTVGDGVGVGKVFHYVNGKFDYHQRVYKISDFKNYYGLLLFYYFSQNFLKETKKYSA
KTSVDSVRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTKIQKQVIELLKQRKKSLLQKMFIPGGS
HHHHHH*

S.SauNEXmaI "Actual" sequence

MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNS**S**ELGIISQKDYFDKEIS
NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGKKGVMSPLYTVFKIQNIDLNFIEFYFKSS
KWYRFMALNGDSGARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
KKGYMQKIFSQELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKDAITNGSYDFYVRSPIVYKIN
TFSYEGEAILTVGDGVGVGKVFHYVNGKFDYHQRVYKISDFKNYYGLLLFYYFSQNFLKETKKYSA
KTSVDSVRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTKIQKQVIELLKQRKKSLLQKMFIPGGS
HHHHHH*



1- marker 2- Nickel column eluate 3-14 Fractions from gel filtration column
15- CC5-1 purified protein marker

S.SauNE
CC398-1 ACC-5-RTGA

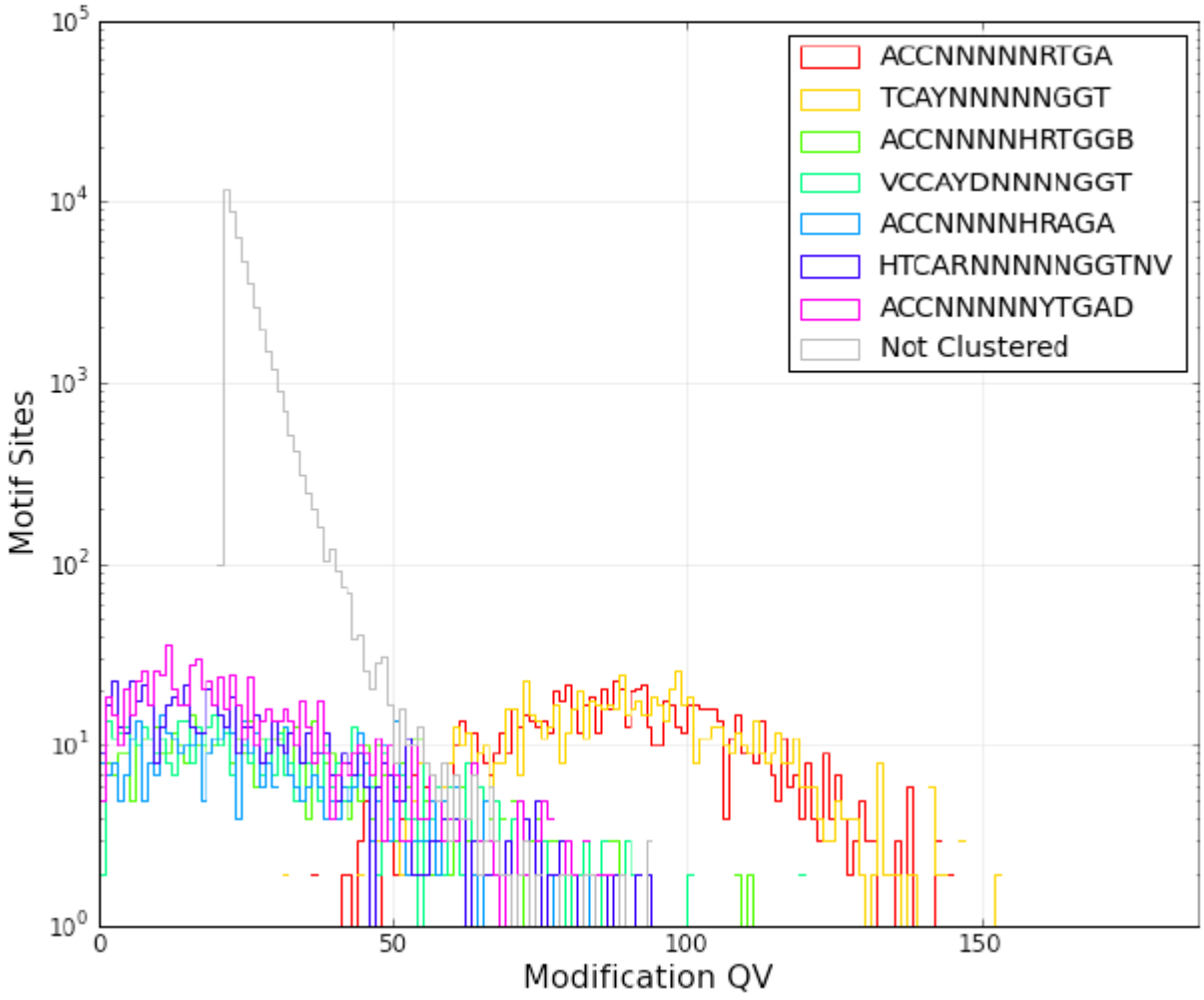
Reports for Job Ed_1_Dryden_MODs

SMRT Cells: 1 Movies: 1

Motif Summary

Motifs	Modified Position	Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACCNNNNNRTGA	1	m6A	99.69%	971	974	89.04	57.17	TCAYNNNNNGGT
TCAYNNNNNGGT	3	m6A	99.69%	971	974	90.00	57.86	ACCNNNNNRTGA
ACCNNNNHRTGGB	1	m6A	49.07%	291	593	54.17	60.71	VCCAYDNNNNGGT
VCCAYDNNNNGGT	4	m6A	45.36%	269	593	54.62	61.85	ACCNNNNHRTGGB
ACCNNNNHRAGA	1	m6A	41.75%	200	479	48.38	61.76	
HTCARNNNNNGGTNV	4	m6A	36.31%	264	727	51.22	62.33	
ACCNNNNNYTGAD	1	m6A	34.9%	320	917	49.93	60.88	

Modification QVs

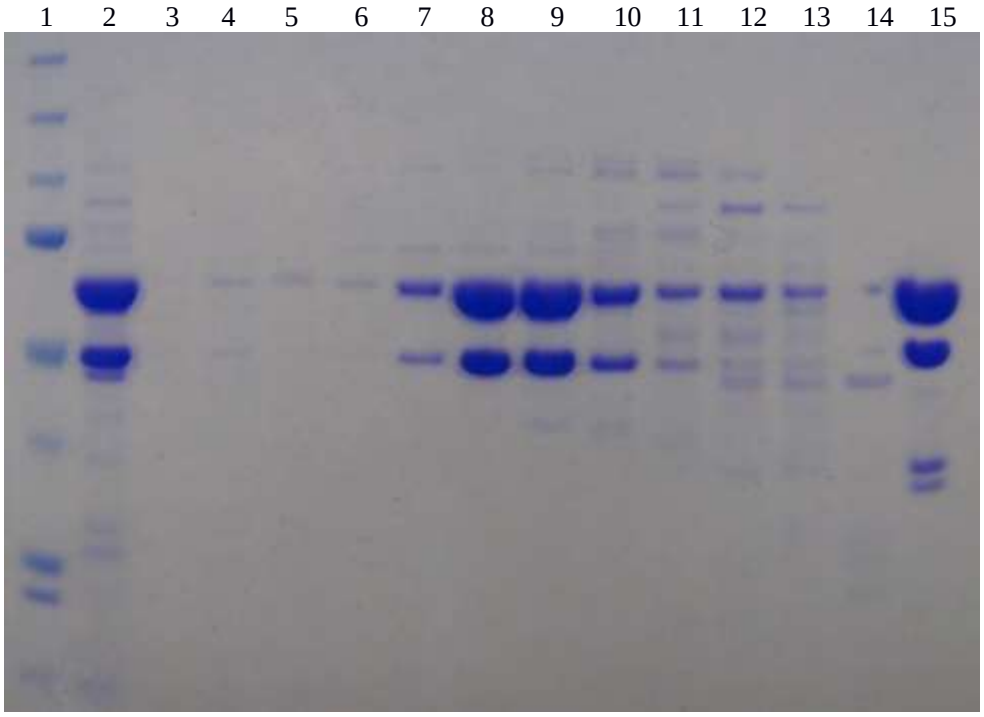


SUPPLEMENTARY INFORMATION FOR TABLE 3.

S.SauBE AGG-5-RTGA

This MTase was purified but cut all the plasmids in the nuclease assay. Therefore once the targets for each TRD had been determined from other MTases, we used the ATPase assay to verify the length of the non-specific spacer.

MSNTQKKNVPELRFPGFEGEWEEKKLGDLTDRVIRKNKNLESKKPLTISGQLGLIDQTEYFSKSVS
SKNLENYTLIKNGEFAYNKSYSNGYPLGAIKRLTRYDSGVLSSLYICFSIKSEMSKDFMEAYFDST
HWYREVSGIAVEGARNHGLLNVSVNDFFTILIKYPSLEEQQKIGKFFSKLDRQIELEEQKLELLQQ
QKKGYMQKIFSQELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKDAITNGSYDFYVRSPIVYKI
NTFSYEGEAILTVGDGVGVGKVFHYVNGKFDYHQRVYKISDFKNYYGLLLFYYFSQNFLKETKKYS
AKTSVDSVRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTKIQKQVIELLKQRKKSLLQKMFIPGG
SHHHHHH



1- marker 2- Nickel column eluate 3-14 Fractions from gel filtration column
15- CC398-1 purified protein marker

Oligonucleotides for checking BE target site using ATPase assay.

Underlined refers to methylated bases.

5′-AGG-N-RTGA-3′

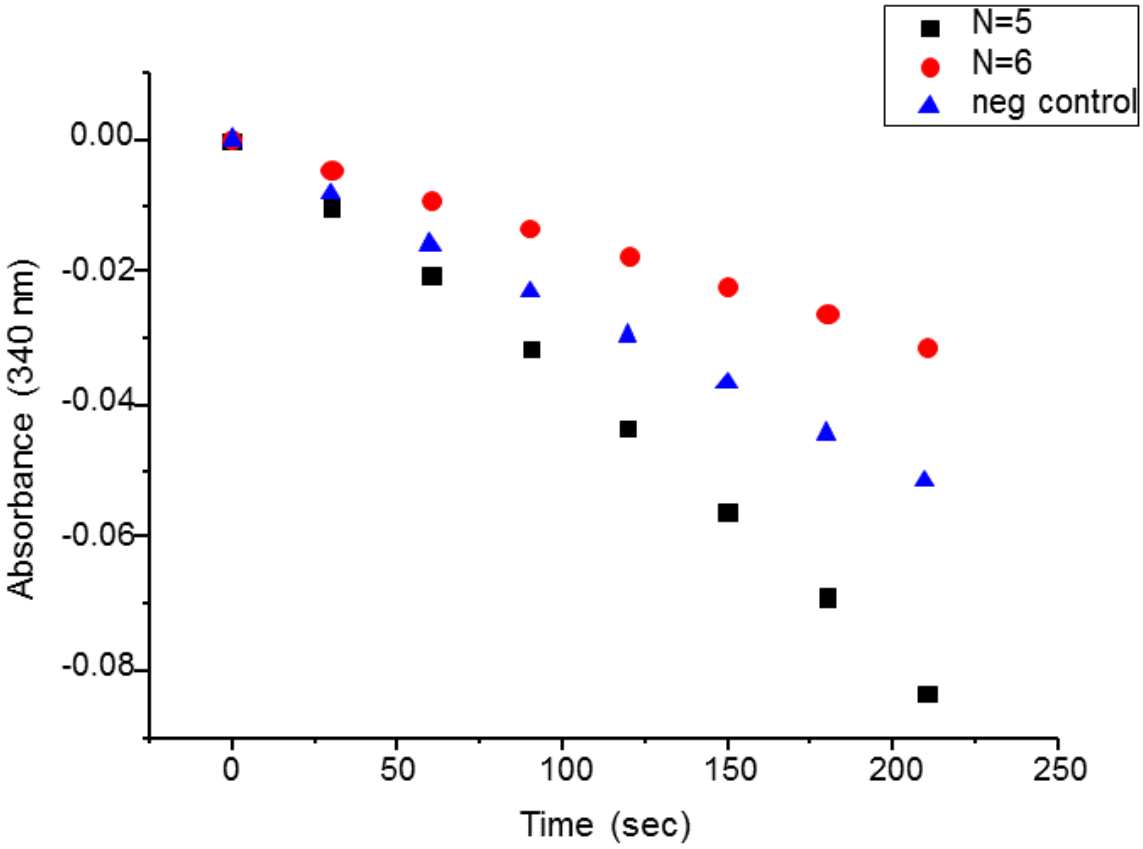
3′-TCC-N-YACT -5′

N values may be 4-6 i.e., number of base pairs between methylated adenines of 7-9. However, DNA digests show that pUC19 contains the site. This rules out the possibility of N=4 (i.e., no site in pUC19 for N value of 4). Therefore we checked for N5 and N6 only.

Oligonucleotide name	DNA sequence (5' to 3')
BE5for	AGATGATGGAATCAATGCAGGTTCCAGTGAGCCCTATACGATATAA
BE5rev	TTATATCGTATAGGGCTCACTGGAACCTGCATTGATTCCATCATCT
BE6for	AGATGATGGAATCAATGCAGGTTCACAGTGAGCCCTATACGATATAA
BE6rev	TTATATCGTATAGGGCTCACTGTGAACCTGCATTGATTCCATCATCT

S.SauBE AGG-5-RTGA

N=5 gives the most activity therefore we conclude from the ATPase assay that the site for the BE TRD combination is AGG-5-RTGA.

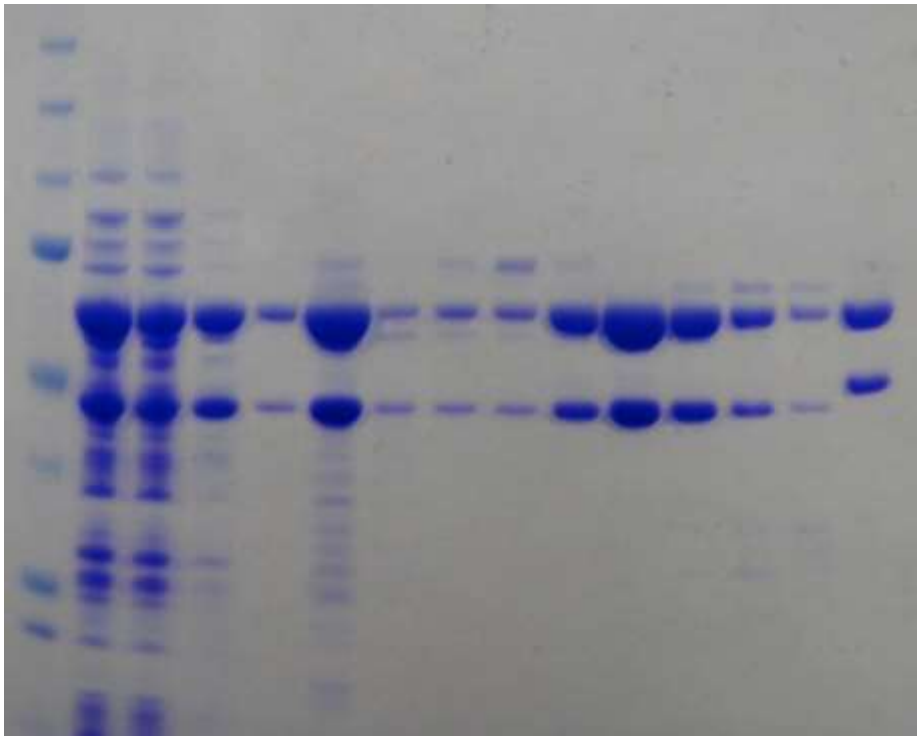


S.SauJE GGA-6-RTGA

This MTase was used in both nuclease and SMRT assays. The TRD pair JE occurs in other ST groups namely ST49 and ST50.

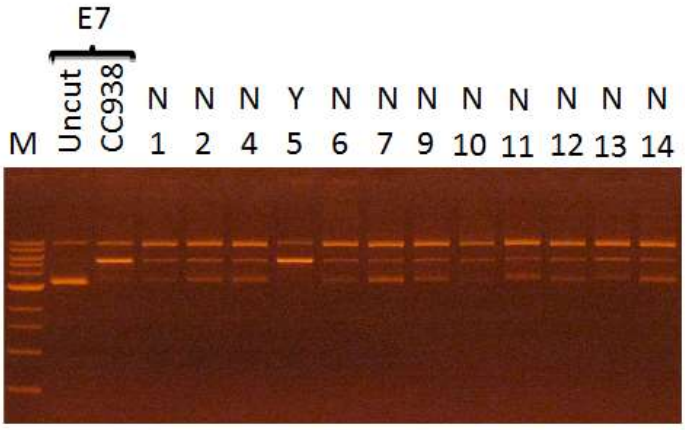
MSNTQKKNVPELRFPGFEGEWEEKKLGDLIKVNSGKDYKHLEKGDIPVYGTGGYMTSVSEPLSEID
AVGIGRKGTINKPYLLEAPFWTVDTLFYCTPKKETDILFILSLFRKINWKVYDESTGVPSLSKQTI
NKINRFVPSNKEQQKIGEFFIKLDRQIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGKDYPEW
EETTIKEIAQINTGKKDTKDAITNGSYDFYVRSPIVYKINTFSYEGEAILTVGDGVGVGKVFHYVN
GKFDYHQRVYKISDFKNYYGLLLFYYFSQNFLKETKKYSAKTSVDSVRKDMIANMKVPRPIYIEQK
KIGQFIKRVDNKTKIQKQVIELLKQRKKSLLQKMFIPGGSHHHHHH

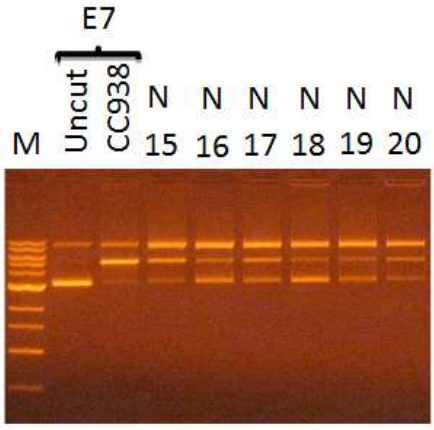
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15



1- marker 2- soluble cell extract 3- Nickel column flow through 4- Nickel column wash 1 5- Nickel column wash 2 6- Nickel column eluate 7-14 Fractions from gel filtration column 15- CC398-1 purified protein marker

Possible site: GGANNNNNNRTGA Note that the background linearisation may be due to the enzyme displaying star activity against a similar site (i.e., a single GGAN7RTGA site is found in pUC19) to the real site (GGAN6RTGA). Repeated digests generate an identical pattern of digestion.





S.SauJE GGA-6-RTGA

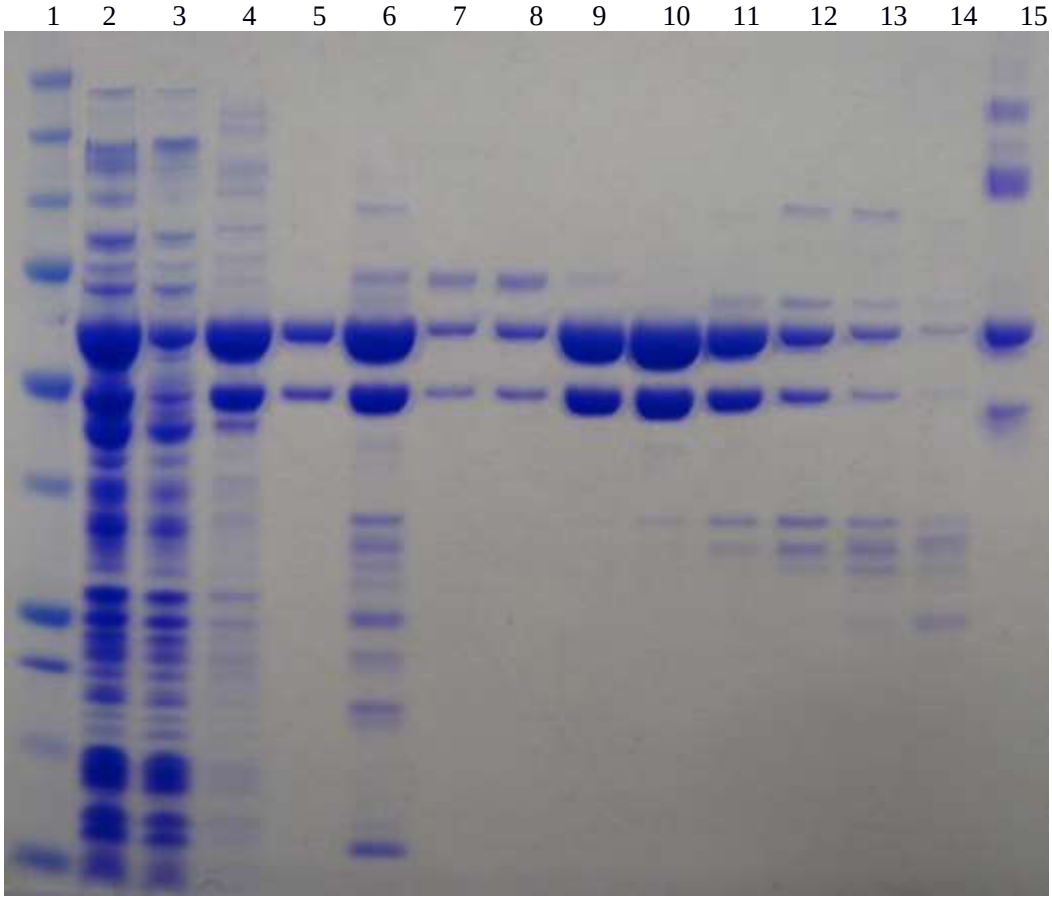
SMRT data showed only the N=6 spacer giving modification.

SMRT Cells: 2 Movies: 2 Print Download Email

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
TCAYNNNNNNTCC	3	m6A	31.44	305	970	38.3	17.1	GGANNNNNNRTGA
GGANNNNNNRTGA	3	m6A	24.43	237	970	38.3	17.5	TCAYNNNNNNTCC
Not Clustered	0		0.00	324	9,124,268	34.2	15.8	

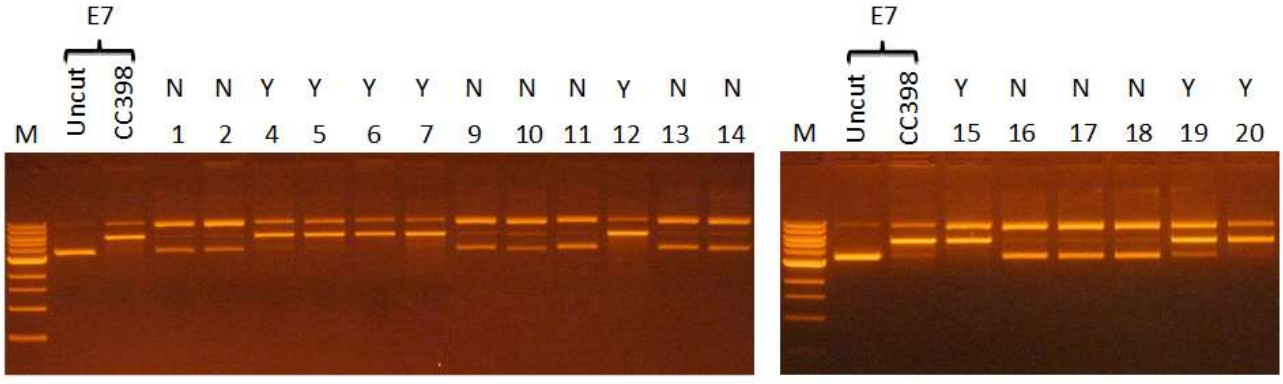
S.SauNI ACC-6-TGAR

MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGKKGVMSPLYTVFKIQNIDLNFIEFYFKSS
KWYRFMALNGDSGARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
KKGYMQKIFSQELRFKNENGNDYPDWERIKFFDVIDKVIDFRGRTPKKLNMEWSDEGYLALSAVNV
KKGYIDFNVEAKYGNLDLYTRWMRGNELYKGQVLFTTEAPMGNVAQVPDNKGYILSQRTIAFNSNE
KITDNFLASLLSSENVYNDLLKLCSGATAKGVSQKNLNRLYVTIPHSISEQEEIAEFFRKINQLVE
LQKYKIEHTKSQKQVFLQKMFIPGGSHHHHHH



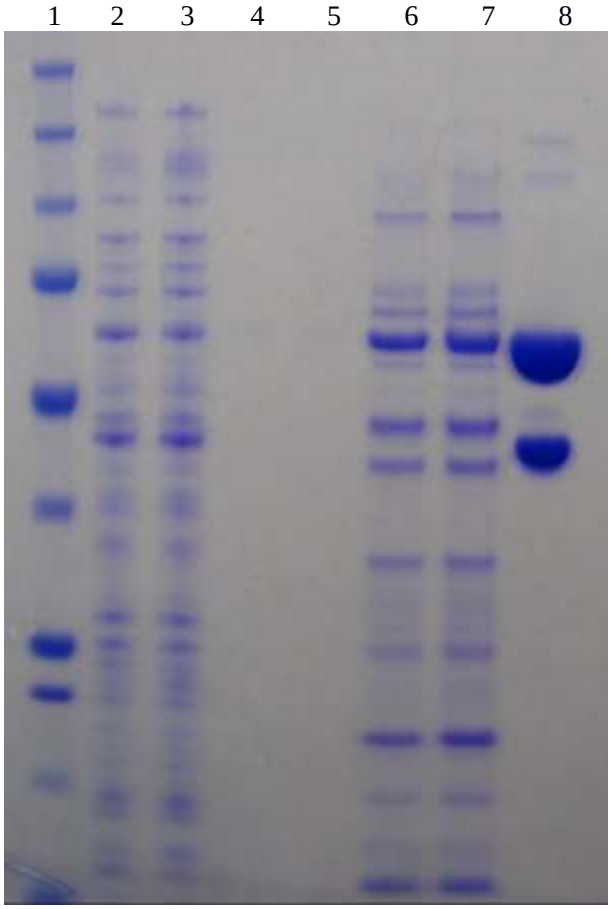
1- marker 2- soluble cell extract 3- Nickel column flow through 4- Nickel column wash 1
5- Nickel column wash 2 6- Nickel column eluate 7-14 Fractions from gel filtration column
15- CC398-1 purified protein marker

Nuclease assay on the plasmid library gave a clear result.



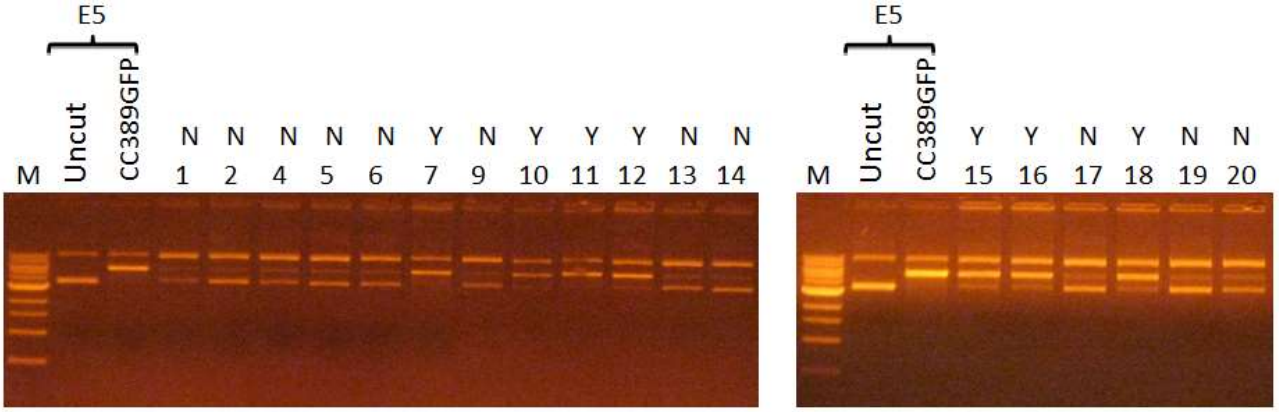
S.SauNK ACC-6-TCG

MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGKKGVMSPLYTVFKIQNIDLNFIEFYFKSS
KWYRFMALNGDSGARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
KKGYMQKIFSQELRFKDENGNDYPNWEEKKIEDIASQVYGGGTPNTKIKEFWNGDIPWIQSSDVKV
NDLILRQCNKFISKNSIELSSAKLIPANSIAIVTRVGVGKLCLVEFDYATSQDFLSLSSLKYDKLY
SLYSLLYTMKKISANLQGTSIKGITKKELLDSIIKIPHNLEEQQKIGDLFYKIDKYISFNKCKIEI
LKSLKQGLLQKIFIPGGSHHHHHH



1- marker 2- soluble cell extract 3- Nickel column flow through
4- Nickel column wash 1 5- Nickel column wash 2 6- Nickel column eluate
7- Eluate after PD10 desalting 8- CC398-1 purified protein marker

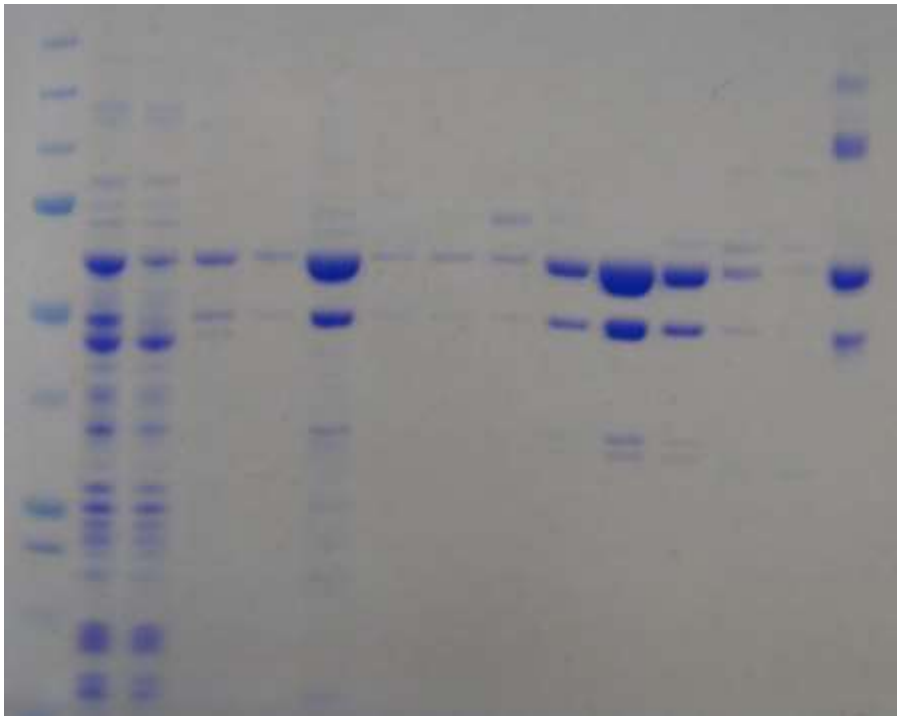
Nuclease assay gave a clear result.



S.SauNL ACC-6-TAAA

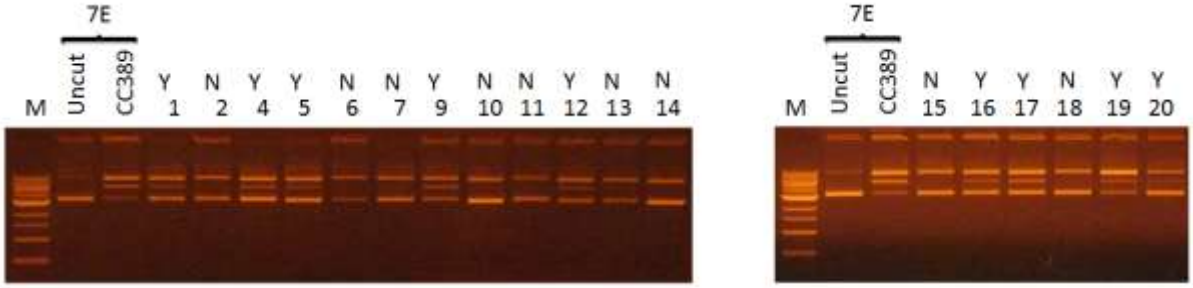
MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGKKGVMSPLYTVFKIQNIDLNFIEFYFKSS
KWYRFMALNGDSGARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
KKGYMQKIFSQELRFKDENGNDYPNWRTIELKNILENIVDNRGKTPDNAPSEKYPLLEVNALGYYR
PAYIKVSKFVSENTYNNWFREHLKENDILFSTVGNTGIVSLMDNYKAVIAQNIVGLRVNNNNLPSF
IYYMLSYKGNQKKIKRIQMGAVQPSVKVSQFKFIKYLVPIKDEQEKVAKLLIEIDKLVNKQLIKIE
LLQQRKKALLKSMFIPGGSHHHHHH

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15



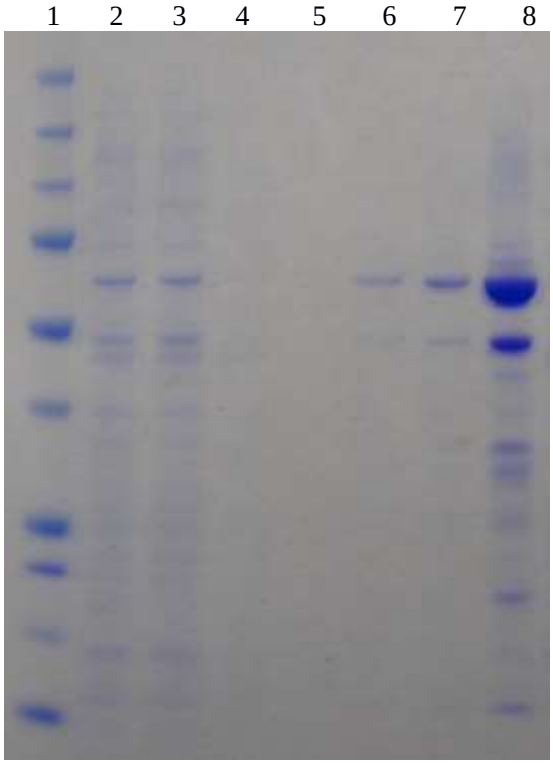
1- marker 2- soluble cell extract 3- Nickel column flow through 4- Nickel column wash 1
5- Nickel column wash 2 6- Nickel column eluate 7-14 Fractions from gel filtration column
15- CC398-1 purified protein marker

Nuclease assay gave a clear result.



S.SauNP ACC-5-CCT

MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGKKGVMSPLYTVFKIQNIDLNFIEFYFKSS
KWYRFMALNGDSGARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELFQQQ
KKGYMQKIFSQELRFKDESGNDYPDWEEKELGEVADRVIRKNKNFESKKPLTISGQLGLIDQTEYF
SKSVSSKNLENYTLIKNGEFAYNKSYSNGYPLGAIKRLTRYDSGVLSSLYICFSIKSEMSKDFMEA
YFDSTHWYREVSGIAVEGARNHGLLNISVNDFFTILIKYPSLEEQRKIGDFFIKLDRQIELEEQKL
ELLQQRKKALLKSMLIPGGSHHHHHH



1- marker, 2- soluble cell extract, 3- Nickel column flow through, 4- Nickel column wash 1, 5- Nickel column wash 2, 6- Nickel column eluate, 7- eluate after conc. and PD10 desalting, 8- Final concentrated protein

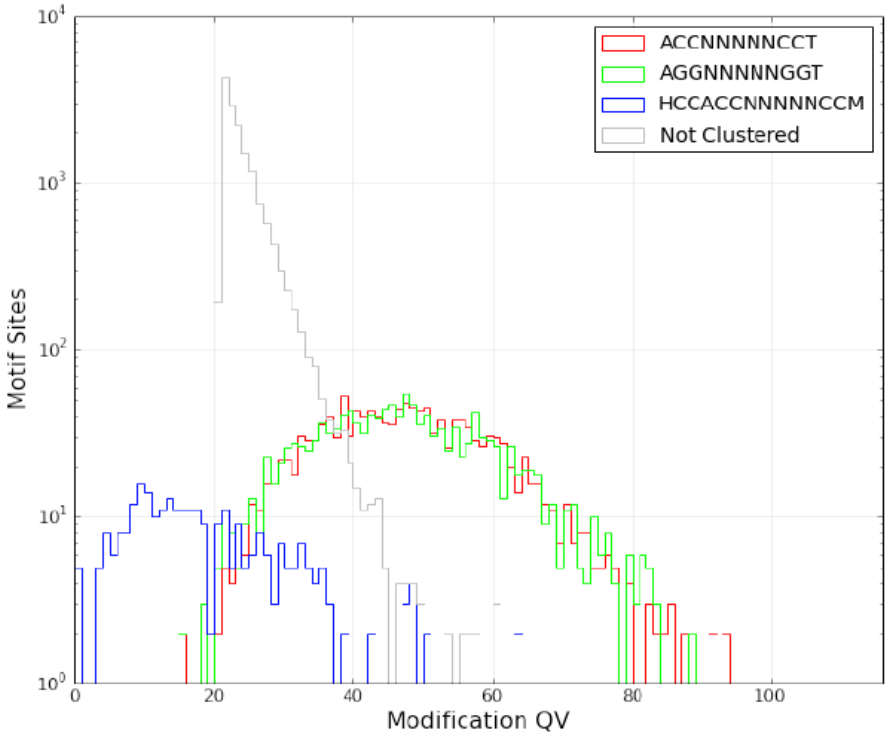
Although purified this MTase was only assayed via SMRT.

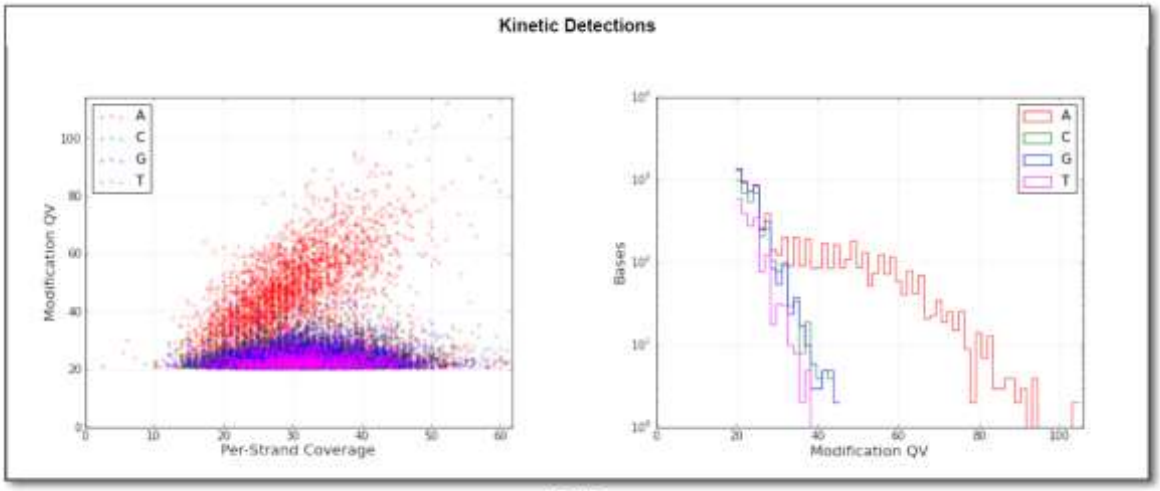
S.SauNP ACC-5-CCT

Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACCNNNNNCCT	1	m6A	91.03	1320	1450	49.8	29.6	AGGNNNNNGGT
AGGNNNNNGGT	1	m6A	89.79	1302	1450	50.0	29.7	ACCNNNNNCCT
HCCACCNNNNNCCM	4	m6A	17.39	52	299	40.0	34.2	
Not Clustered	0		0.01	737	9114127	34.7	34.2	

Modification QV Histogram By Motif



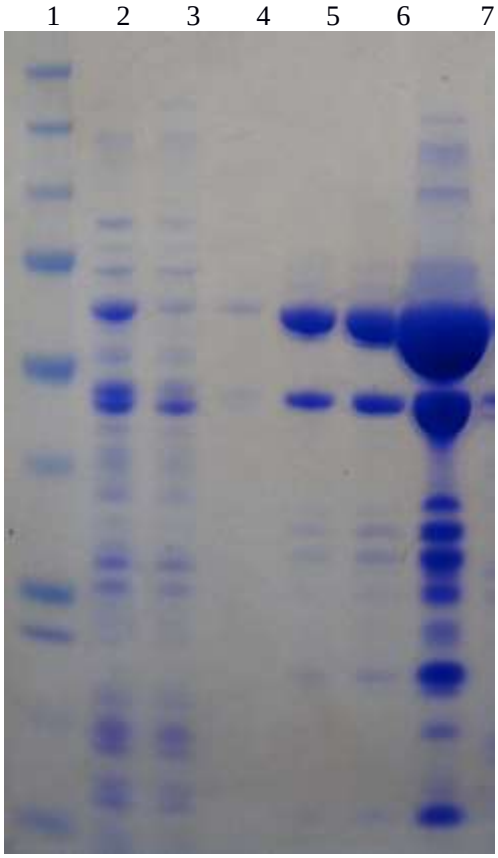


Motifs

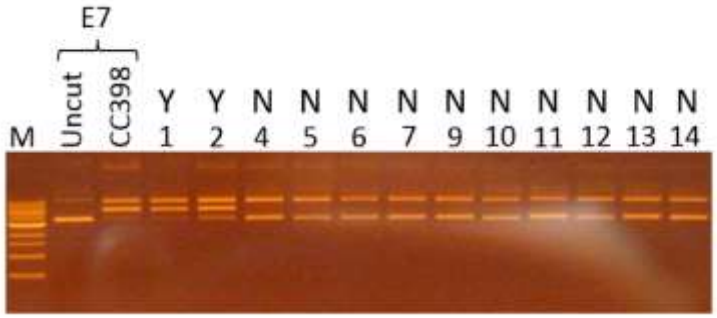
Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACCNNNNNCCT	1	m6A	91.03	1320	1450	49.8	29.6	AGGNNNNNGGT
AGGNNNNNGGT	1	m6A	89.79	1302	1450	50.0	29.7	ACCNNNNNCCT
HCCACCNNNNNCCM	4	m6A	17.39	52	299	40.0	34.2	
Not Clustered	0		0.01	737	9114127	34.7	34.2	

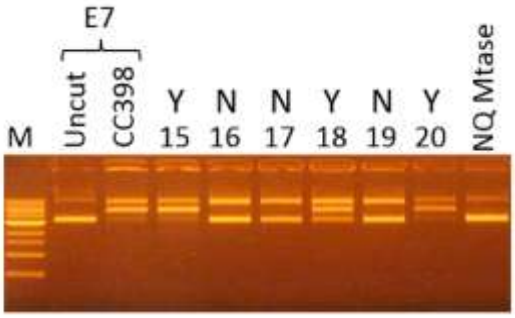
S.SauNQ ACC-5-RTGT

MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGKKGVMSPLYTVFKIQNIDLNFIEFYFKSS
KWYRFMALNGDSGARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
KKGYMQKIFSQELRFKDENGEDYSEWEERRFADIFKFHNKLRKPIKENLRVKGSYPYYGATGIIDY
VDDFIFDGNYLLIGEDGANIITRSAPLVYLVNGKFWVNNHAHILSPLNGNIQYLYQVAELVNYEKY
NTGTAQPKLNIQNLKIINVVISTNLEEQQKIGSFLSKLDRQIDLEEQKLELLQQRKKALLKSMFVP
GGSHHHHHH



1- marker 2- soluble cell extract 3- Nickel column flow through
4- Nickel column wash 5- Nickel column eluate 6- eluate after conc. and PD10 desalting
7- Final concentrated protein
DNA cleavage assay and SMRT assay agreed.



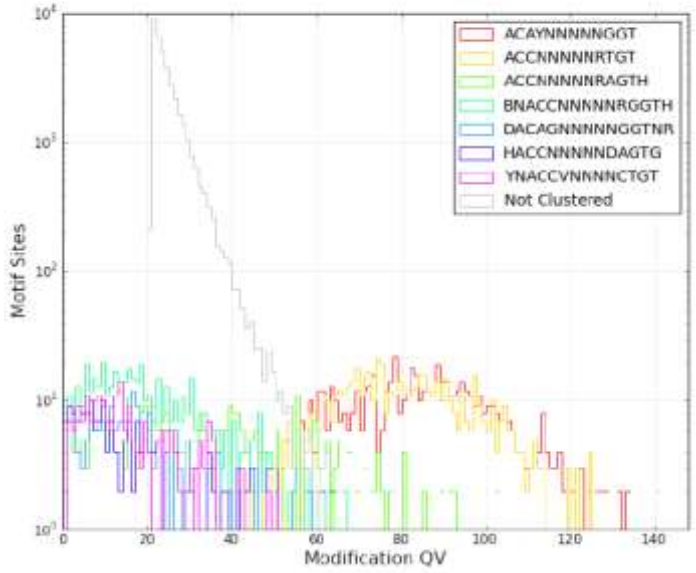


S.SauNQ ACC-5-RTGT

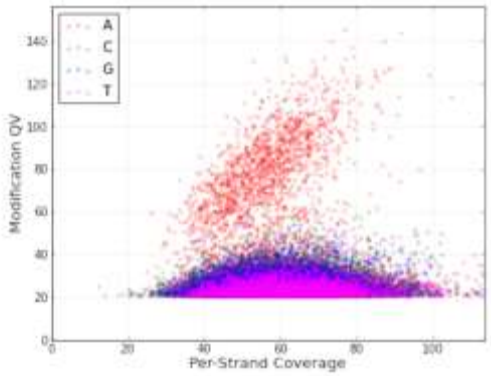
Motifs

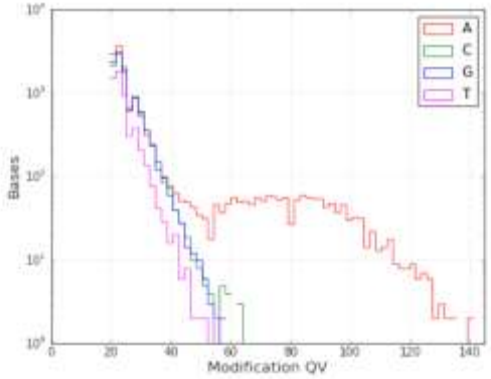
Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACAYNNNNNGGT	3	m6A	99.85	654	655	83.5	56.3	ACCNNNNNRTGT
ACCNNNNNRTGT	1	m6A	99.85	654	655	80.7	55.5	ACAYNNNNNGGT
ACCNNNNNRAGTH	1	m6A	55.56	215	387	54.3	56.5	
BNACCNNNNNRGGTH	3	m6A	23.74	118	497	45.8	57.6	
DACAGNNNNNGGTNR	4	m6A	21.65	50	231	42.8	55.5	
HACCNNNNNDAGTG	2	m6A	21.03	41	195	44.8	57.9	
YNACCVNNNNCTGT	3	m6A	20.52	47	229	42.9	57.7	
Not Clustered	0		0.03	3095	9114477	35.3	62.5	

Modification QV Histogram By Motif



Kinetic Detections





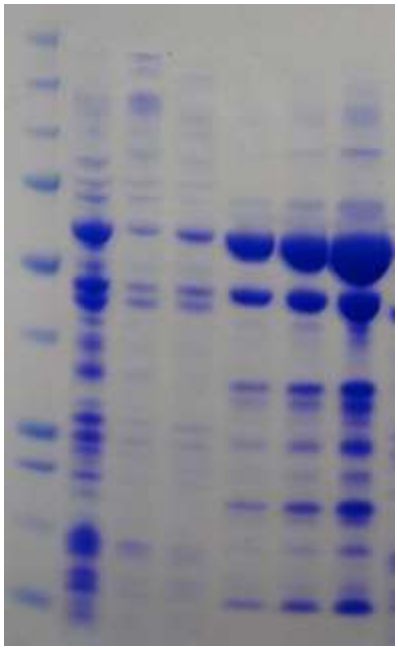
Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACAYNNNNNGGT	3	m6A	99.85	654	655	83.5	56.3	ACCNNNNNRTGT
ACCNNNNNRTGT	1	m6A	99.85	654	655	80.7	55.5	ACAYNNNNNGGT
ACCNNNNNRAGTH	1	m6A	55.56	215	387	54.3	56.5	
BNACCNNNNNRGGTH	3	m6A	23.74	118	497	45.8	57.6	
DACAGNNNNNGGTNR	4	m6A	21.65	50	231	42.8	55.5	
HACCNNNNNDAGTG	2	m6A	21.03	41	195	44.8	57.9	
YNACCVNNNNCTGT	3	m6A	20.52	47	229	42.9	57.7	
Not Clustered	0		0.03	3095	9114477	35.3	62.5	

S.SauNS ACC-6-TGC

MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGKKGVMSPLYTVFKIQNIDLNFIEFYFKSS
KWYRFMALNGDSGARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
KKGYMQKIFSQELRFKDENGNDYPDWTNERLGEVTTVTMGQSPKSVNYTDNSNDTVLIQGNADIEN
GLINPRIYTREVTKLIQKDEIILTVRAPVGKLAMAQINACIGRGVCSIKGDKFLYYFLEWFATQNK
WIRFSQGSTFESISGNDIRNIHIKIPVEDERTKIIKLLNSLDVLNSKTDLKIQNLKQRKQSLLQKI
FVPGGSHHHHHH

1 2 3 4 5 6 7



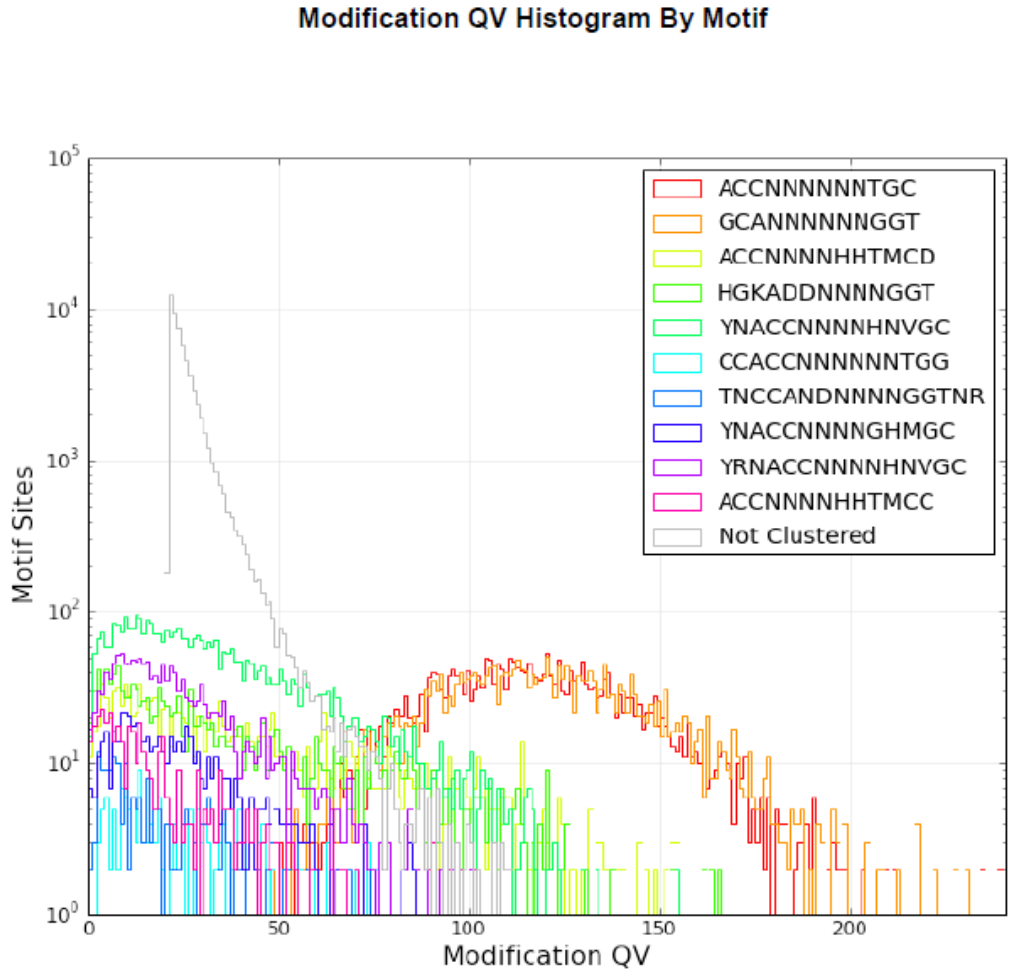
1- marker 2- soluble cell extract
3- Nickel column flow through 4- Nickel column wash
5- Nickel column eluate 6- eluate after conc. and PD10 desalting
7- final protein after concentration

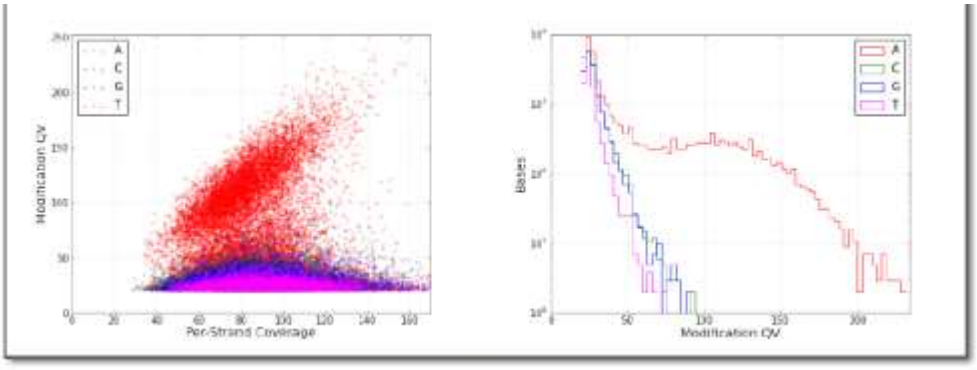
Although purified this MTase was only assayed via SMRT.

Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACCNNNNNNTGC	1	m6A	100.00	2938	2938	118.8	81.7	GCANNNNNNGGT
GCANNNNNNGGT	3	m6A	99.90	2935	2938	120.7	83.9	ACCNNNNNNTGC
ACCNNNNHHTMCD	1	m6A	57.03	925	1622	71.1	83.7	HGKADDNNNNGGT
HGKADDNNNNGGT	4	m6A	48.83	792	1622	68.8	86.4	ACCNNNNHHTMCD
YNACCNNNNHNVGC	3	m6A	46.49	1925	4141	57.7	84.6	
CCACCNNNNNNTGG	3	m6A	39.15	74	189	55.7	85.9	
TNCCANDNNNNGGTNR	5	m6A	31.60	73	231	53.5	82.5	
YNACCNNNNGHMGC	3	m6A	31.35	195	622	53.2	87.4	
YRNACCNNNNHNVGC	4	m6A	28.65	465	1623	48.9	86.3	
ACCNNNNHHTMCC	1	m6A	27.58	131	475	58.1	84.3	
Not Clustered	0		0.09	8284	9100925	38.6	92.5	

S.SauNS ACC-6-TGC



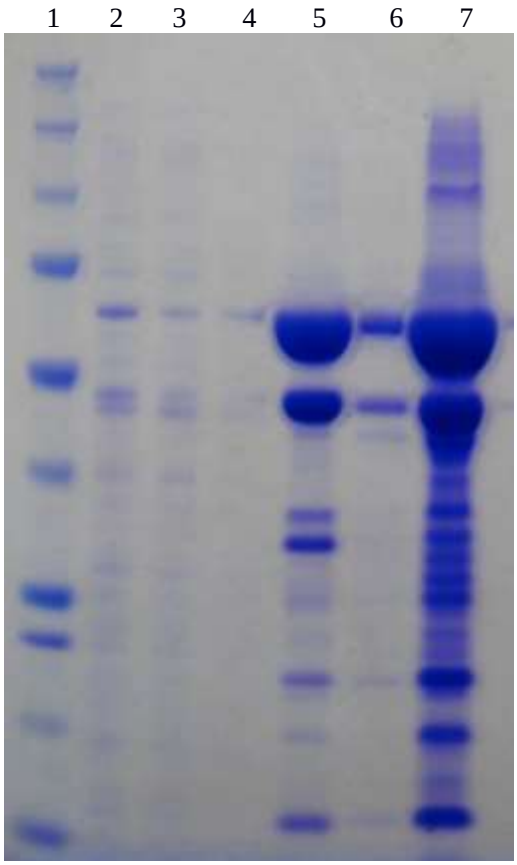


Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACCNNNNNNTGC	1	m6A	100.00	2938	2938	118.8	81.7	GCANNNNNNGGT
GCANNNNNNGGT	3	m6A	99.90	2935	2938	120.7	83.9	ACCNNNNNNTGC
ACCNNNNHHTMCD	1	m6A	57.03	925	1622	71.1	83.7	HGKADDNNNNGGT
HGKADDNNNNGGT	4	m6A	48.83	792	1622	68.8	86.4	ACCNNNNHHTMCD
YNACCNNNNHNVGC	3	m6A	46.49	1925	4141	57.7	84.6	
CCACCNNNNNNTGG	3	m6A	39.15	74	189	55.7	85.9	
TNCCANDNNNNGGTNR	5	m6A	31.60	73	231	53.5	82.5	
YNACCNNNNGHMGC	3	m6A	31.35	195	622	53.2	87.4	
YRNACCNNNNHNVGC	4	m6A	28.65	465	1623	48.9	86.3	
ACCNNNNHHTMCC	1	m6A	27.58	131	475	58.1	84.3	
Not Clustered	0		0.09	8204	9100925	38.6	92.5	

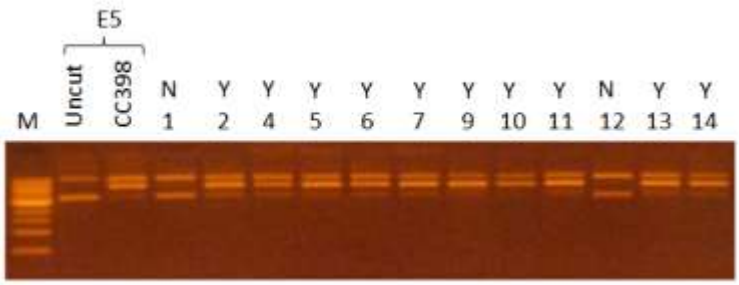
S.SauNU ACC-5-RTC

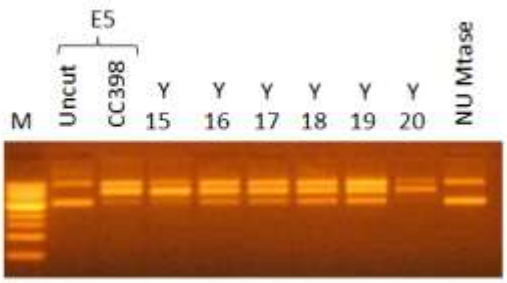
MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGKKGVMSPLYTVFKIQNIDLNFIEFYFKSS
KWYRFMALNGDSGARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
KKGYMQKIFSQELRFKDENGEDYPDWEVTTIQNITKYTSSKKSSNQYADKDNSKGYPVYDAVQEIG
KDSNYDIEESYISILKDGAGVGRLNLRPGKSSVIGTMGYIQSNNVDIEFLYYRMKVVDFKKYIIGS
TIPHLYFKDYSKETLYIPSSIQEQAKIGMFISNLDKLIENKNLKLNCLKQLKQGLLQSMFIPGGSH
HHHHH



1- marker 2- soluble cell extract
3- Nickel column flow through 4- Nickel column wash
5- Nickel column eluate 6- eluate after PD10 desalting
7- final protein after concentration

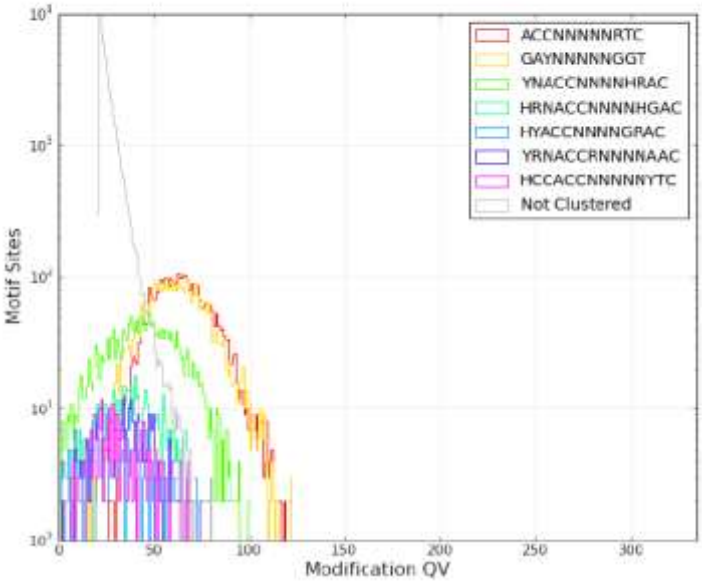
DNA cleavage assay worked despite there being one site in pUC19 but this site was subject to dam methylation and therefore not cut.





S.SauNU ACC-5-RTC

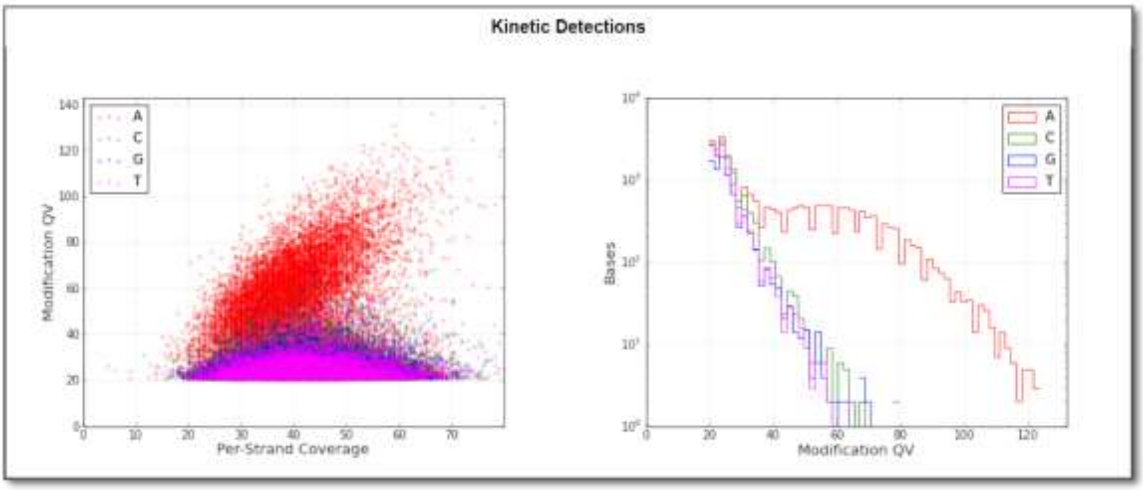
Modification QV Histogram By Motif



Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACCNNNNNRTC	1	m6A	81.87	3162	3862	70.1	41.5	GAYNNNNNGGT
GAYNNNNNGGT	2	m6A	75.14	2902	3862	70.8	41.9	ACCNNNNNRTC
YNACCNNNNHRAC	3	m6A	37.48	820	2188	64.9	43.3	
HRNACCNNNNHGAC	4	m6A	26.22	140	534	63.7	43.0	
HYACCNNNNGRAC	3	m6A	19.23	50	260	63.3	44.4	
YRNACCRNNNNAAC	4	m6A	17.15	59	344	63.7	45.8	
HCCACCNNNNNYTC	4	m6A	16.81	39	232	61.9	45.9	
Not Clustered	0		0.00	229	9106044	58.3	48.5	

Kinetic Detections

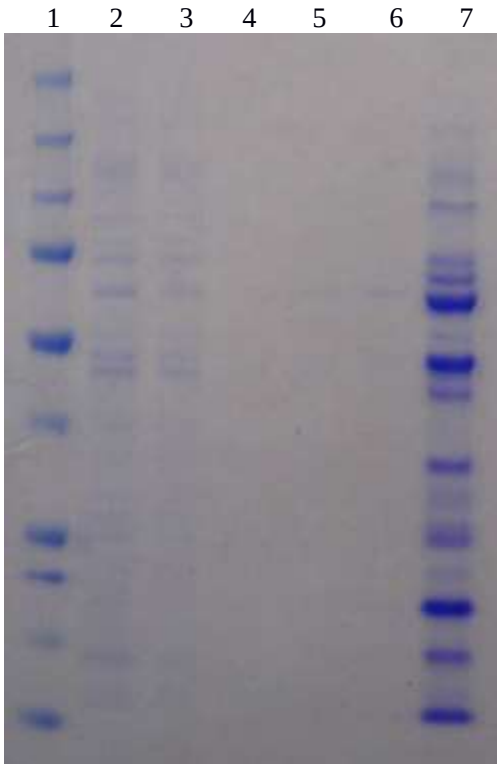


Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACCNNNNNRTC	1	m6A	81.87	3162	3862	70.1	41.5	GAYNNNNNGGT
GAYNNNNNGGT	2	m6A	75.14	2902	3862	70.8	41.9	ACCNNNNNRTC
YNACCNNNNHRAC	3	m6A	37.48	820	2188	64.9	43.3	
HRNACCNNNNHGAC	4	m6A	26.22	140	534	63.7	43.0	
HYACCNNNNGRAC	3	m6A	19.23	50	260	63.3	44.4	
YRNACCRNNNNAAC	4	m6A	17.15	59	344	63.7	45.8	
HCCACCNNNNNYTC	4	m6A	16.81	39	232	61.9	45.9	
Not Clustered	0		0.00	229	9106044	58.3	48.5	

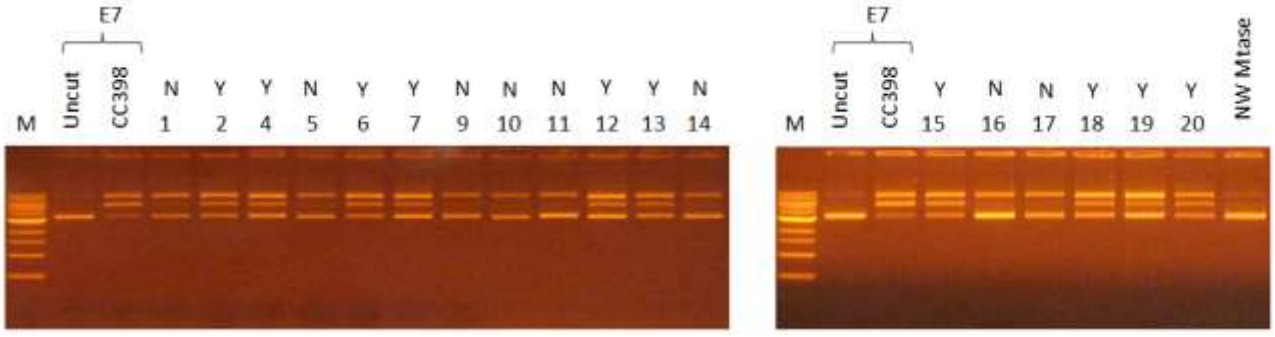
S.SauNW ACC-6-TTYG

MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGKKGVMSPLYTVFKIQNIDLNFIEFYFKSS
KWYRFMALNGDSGARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
KKGYMQKIFSQELRFKDENGNDYPDWEEKQLGELSQIVRGASPRPIKDPKWFNKESDIGWLRISDV
TNQNGKIYHLEQKLSIEGQEKTRVLVTTHLLLSIAASIGKPVMNFVKTGVHDGFLIFLKPKFNLFF
MYYWLEYFKDKWSKYGQPGSQVNLNSEIVKSQTLNMPSNHEQEKVGQFFNRNEKLIELQQEKIMYI
KRCKQVLLQKMFIPGGSHHHHHH



1- marker 2- soluble cell extract
3- Nickel column flow through 4- Nickel column wash
5- Nickel column eluate 6- eluate after PD10 desalting
7- final protein after concentration

DNA cleavage assay.

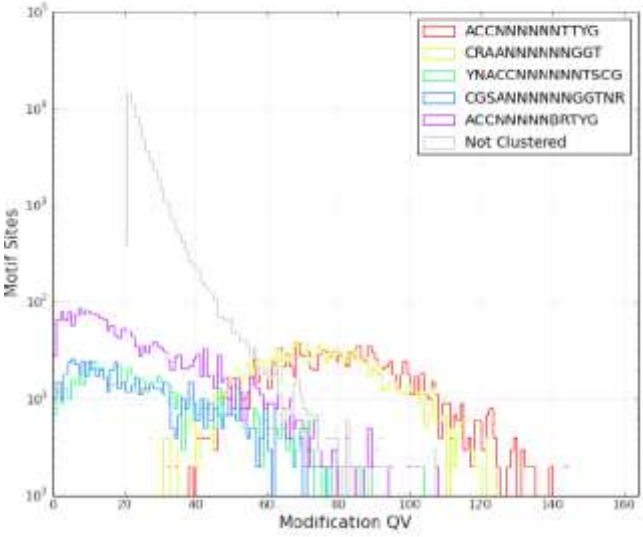


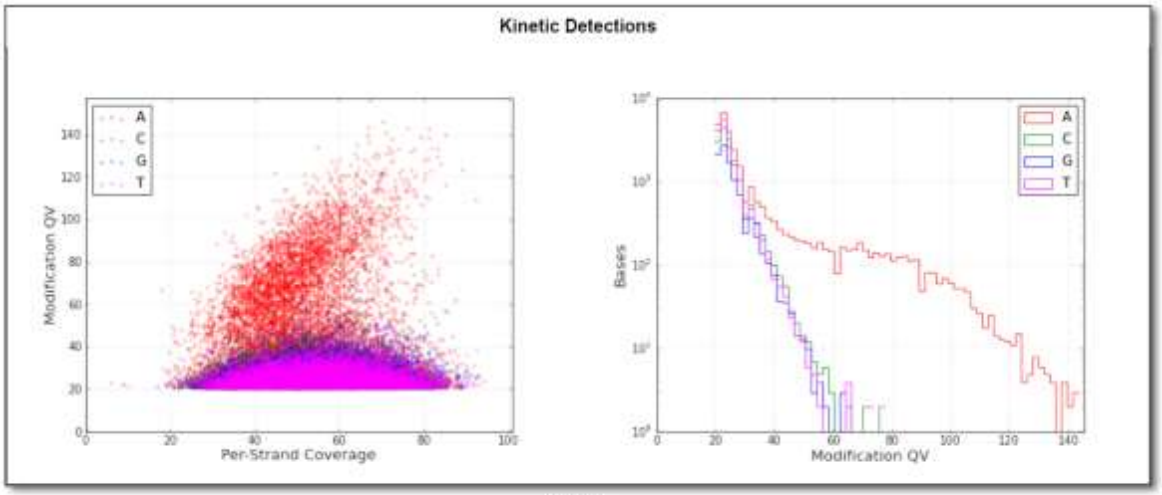
S.SauNW ACC-6-TTYG

Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACCNNNNNNTTYG	1	m6A	99.86	1461	1463	80.7	49.6	CRAANNNNNNGGT
CRAANNNNNNGGT	4	m6A	99.59	1457	1463	74.7	48.6	ACCNNNNNNTTYG
YNACCNNNNNNTSCG	3	m6A	39.52	313	792	52.0	52.7	CGSANNNNNNGGTNR
CGSANNNNNNGGTNR	4	m6A	35.23	279	792	50.1	52.3	YNACCNNNNNNTSCG
ACCNNNNNBRTYG	1	m6A	28.16	680	2415	49.3	51.3	
Not Clustered	0		0.08	6917	9110401	37.6	55.8	

Modification QV Histogram By Motif



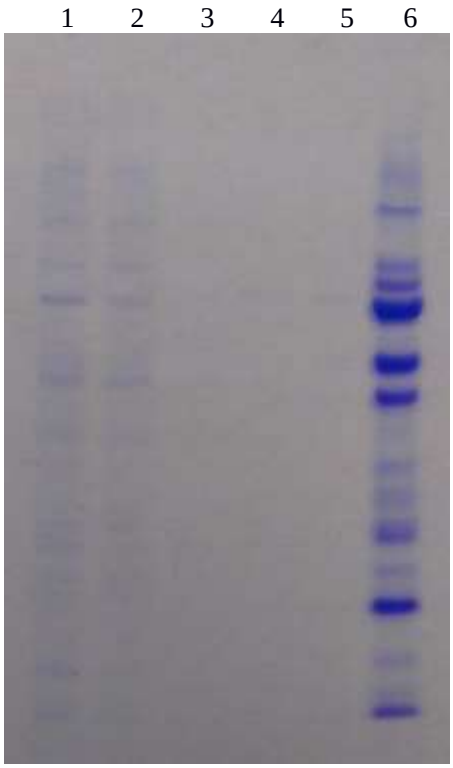


Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACCNNNNNNTTYG	1	m6A	99.86	1461	1463	80.7	49.6	CRAANNNNNNGGT
CRAANNNNNNGGT	4	m6A	99.59	1457	1463	74.7	48.6	ACCNNNNNNTTYG
YNACCNNNNNNTSCG	3	m6A	39.52	313	792	52.0	52.7	CGSANNNNNNGGTNR
CGSANNNNNNGGTNR	4	m6A	35.23	279	792	50.1	52.3	YNACCNNNNNNTSCG
ACCNNNNNBRTYG	1	m6A	28.16	680	2415	49.3	51.3	
Not Clustered	0		0.08	6917	9110401	37.6	55.8	

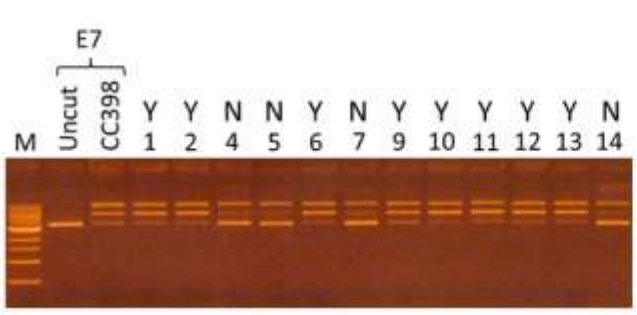
S.SauNY ACC-6-TAG

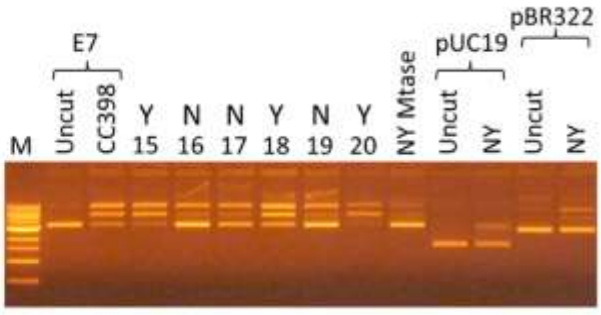
MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGKKGVMSPLYTVFKIQNIDLNFIEFYFKSS
KWYRFMALNGDSGARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
KKGYMQKIFSQELRFKDENGNDYPDWEKKKLKEIACVYTGNTPSKKENIYWNKGEYVWVTPTDINN
SKNIYESENKLTQEGYKKARQLPENTLLVTCIASIGKNAILRKQGSCNQQINAVVPFENINIDYLY
YISDSLSTFMKSIAGKTATQIVNKNTFENLEIYLAPFEEQNKIADLISSLEELIEKQASKLIKMKS
RKQGMLQIMFIPGGSHHHHHH



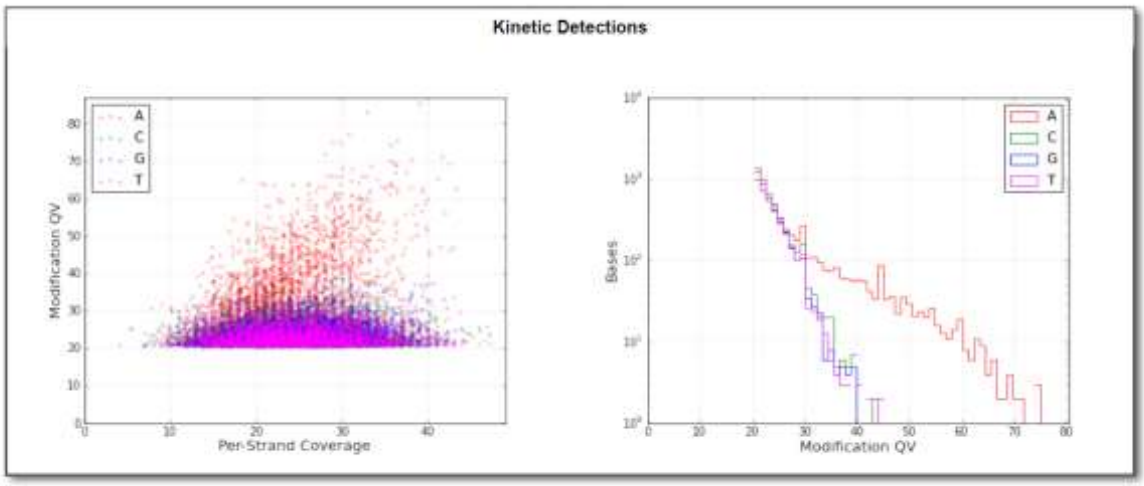
1- soluble cell extract 2- Nickel column flow through
3- Nickel column wash 1 4- Nickel column wash 2 5- Nickel column eluate
6- final protein after PD10 desalting and concentration of eluate

DNA cleavage assay.





S.SauNY A<u>A</u>CC-6-T<u>T</u>AG



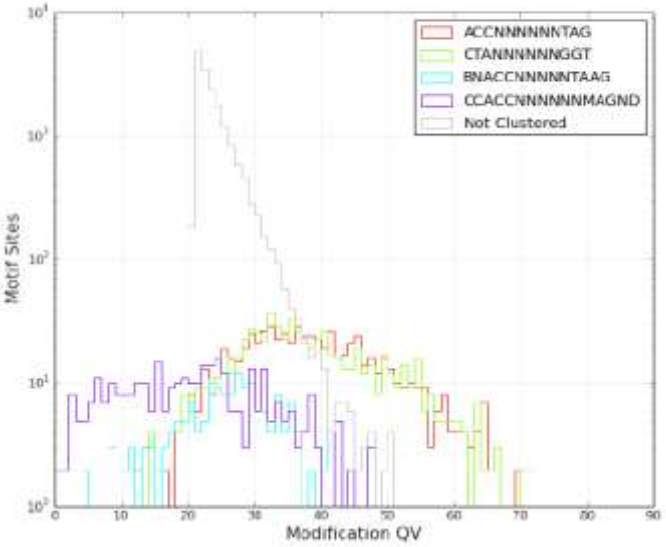
Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACCNNNNNNTAG	1	m6A	75.92	539	710	43.3	24.7	CTANNNNNNGGT
CTANNNNNNGGT	3	m6A	72.39	514	710	42.9	24.7	ACCNNNNNNTAG
BNACCNNNNNTAAG	3	m6A	34.00	68	200	39.3	25.0	
CCACCNNNNNNMAGND	3	m6A	23.68	85	359	38.3	26.1	
Not Clustered	0		0.01	622	9115347	34.4	27.6	

Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACCNNNNNNTAG	1	m6A	75.92	539	710	43.3	24.7	CTANNNNNNGGT
CTANNNNNNGGT	3	m6A	72.39	514	710	42.9	24.7	ACCNNNNNNTAG
BNACCNNNNNTAAG	3	m6A	34.00	68	200	39.3	25.0	
CCACCNNNNNNMAGND	3	m6A	23.68	85	359	38.3	26.1	
Not Clustered	0		0.01	622	9115347	34.4	27.6	

Modification QV Histogram By Motif



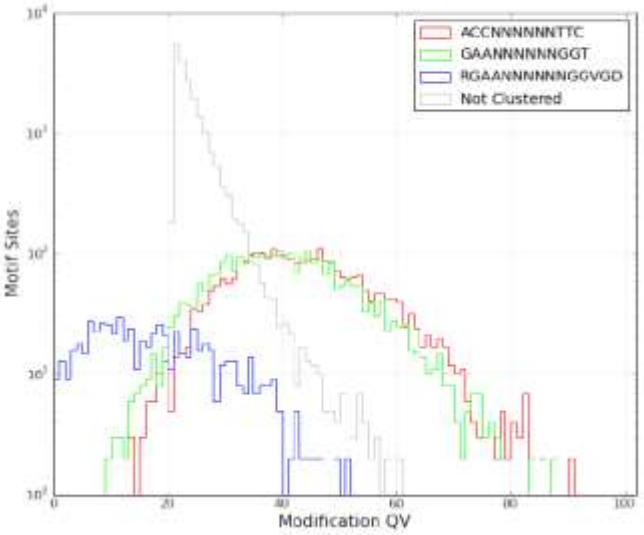
S.SauNa* ACC-6-TTC

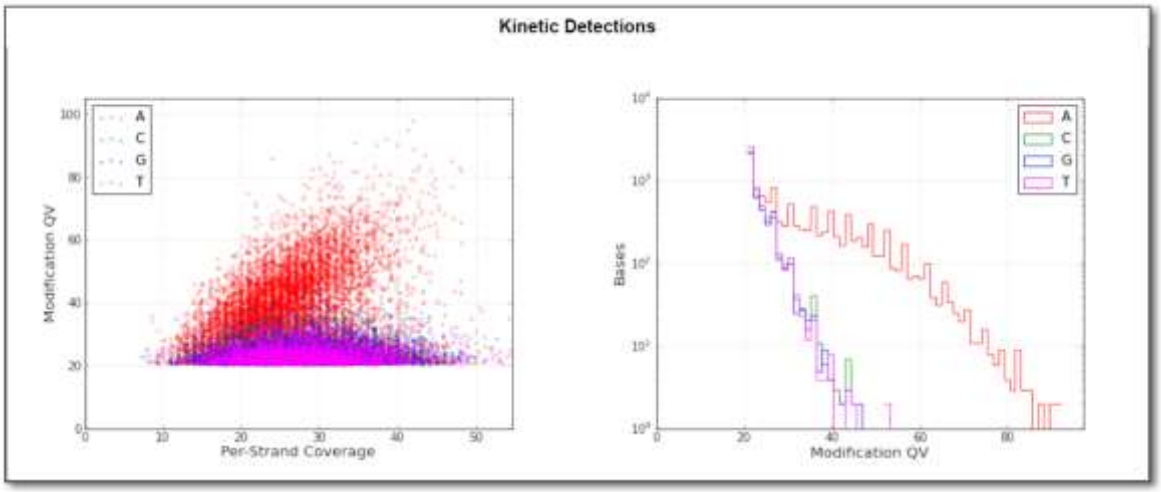
Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACCNNNNNNTTC	1	m6A	85.91	2567	2988	46.4	26.2	GAANNNNNNGGT
GAANNNNNNGGT	3	m6A	78.11	2334	2988	44.7	26.2	ACCNNNNNNTTC
RGAANNNNNNGGVGD	4	m6A	16.09	107	665	37.1	27.8	
Not Clustered	0		0.01	1034	9110685	35.6	31.8	

MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGKKGVMSPLYTVFKIQNIDLNFIEFYFKSS
KWYRFMALNGDSGARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
KKGYMQKIFSQELRFKDENGNDYPEWENKRIEDIANVNKGFTPSTNNNEYWDNNDKNWLSIAGMNQ
KYLYKGNKGISKDAAKNYMKVKNDTLIMSFKLTIGKLAIVKAPLYTNEAICHFIWKVNKINTEFIY
YYLNSLNISTFGVQAVKGVTLNNDSINSIIVKLPNEEEQNIIAKFLLEVDKTVNNQLVKTKLLKQR
KKGLLQRMFVPGGSHHHHHH

Modification QV Histogram By Motif



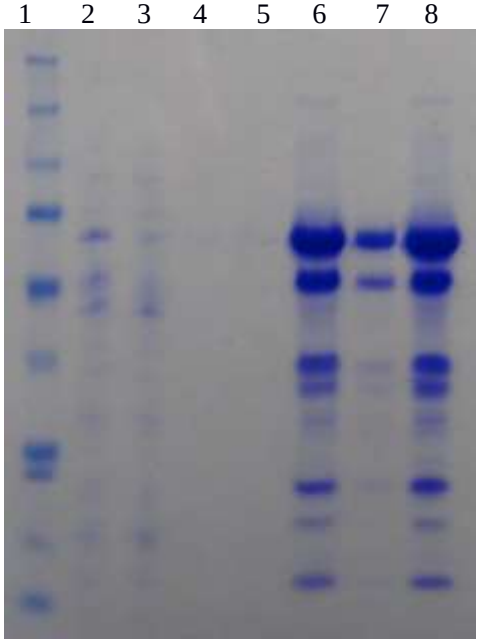


Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACCNNNNNNTTC	1	m6A	85.91	2567	2988	46.4	26.2	GAANNNNNNGGT
GAANNNNNNGGT	3	m6A	78.11	2334	2988	44.7	26.2	ACCNNNNNNTTC
RGAANNNNNNGGVGD	4	m6A	16.09	107	665	37.1	27.8	
Not Clustered	0		0.01	1034	9110685	35.6	31.8	

S.SauNc* ACC-6-RTC

MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGKKGVMSPLYTVFKIQNIDLNFIEFYFKSS
KWYRFMALNGDSGARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
KKGYLQKIFSQELRFKDENGNDYPEWRFARFKDFMYKPINIRPAINISKSELLTVKLHCKGIEKAN
INRVLKLGATNYYKRFEGQFIYGKQNFFNGAFDIVPKKFDGLYSSSDVPAFEINTEKIEPNYFISY
ISRPSFYKSKEKYSTGTGSKRIHENTVLNFSLHLPCLNEQLKIASFVCFLNRKIELLERKIYLIKK
QKQALLQQMFIPGGSHHHHHH



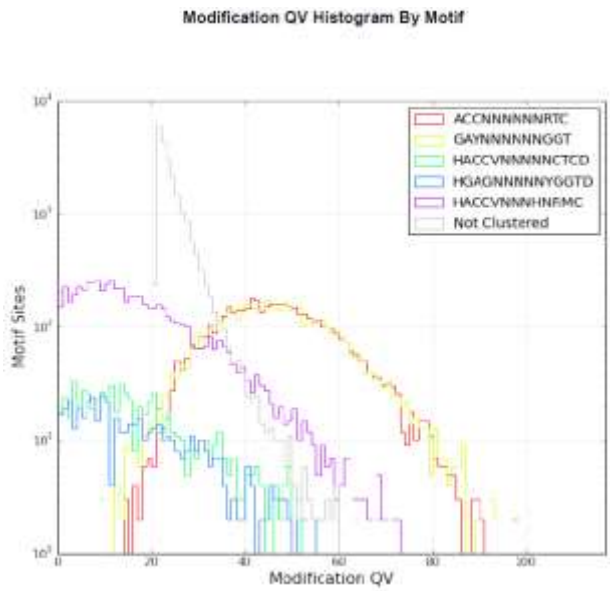
1- marker 2- soluble cell extract
3- Nickel column flow through 4- Nickel column wash 1
5- Nickel column wash 2 6- Nickel column eluate
7- eluate after PD10 desalting
8- final protein after concentration

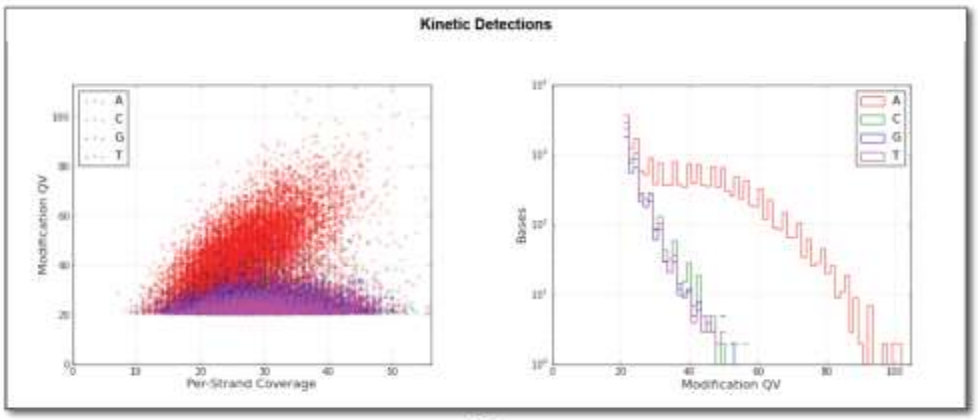
Although purified, this MTase was only assayed by SMRT.

Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACCNNNNNNRTC	1	m6A	90.98	4680	5144	48.9	27.4	GAYNNNNNNGGT
GAYNNNNNNGGT	2	m6A	90.18	4639	5144	50.0	27.8	ACCNNNNNNRTC
HACCVNNNNNCTCD	2	m6A	16.64	117	703	40.7	30.2	
HGAGNNNNNYGGTD	3	m6A	16.60	86	518	41.5	29.5	
HACCVNNNHNRMC	2	m6A	14.94	936	6265	40.2	29.1	
Not Clustered	0		0.01	1163	9099552	35.6	31.2	

S.SauNc* ACC-6-RTC



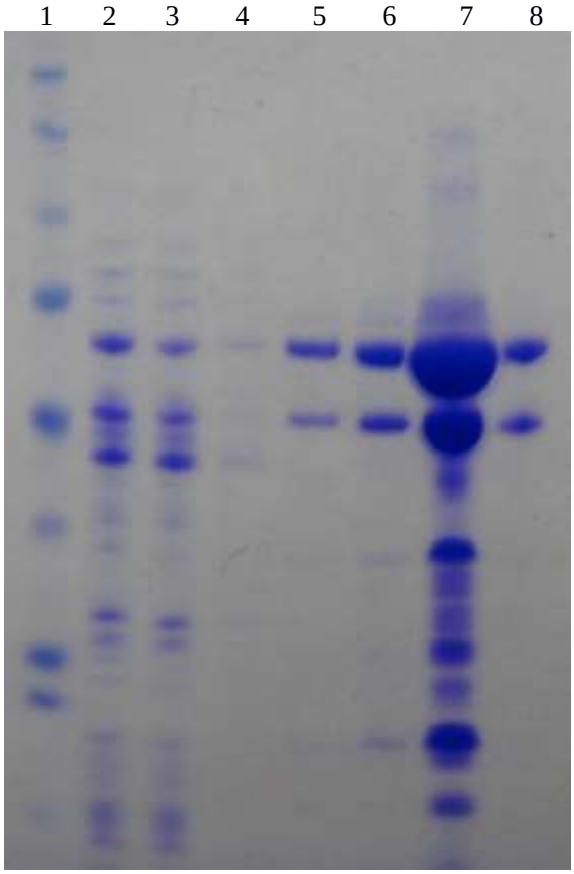


Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACCNNNNNNRTC	1	m6A	90.98	4680	5144	48.9	27.4	GAYNNNNNNGGT
GAYNNNNNNGGT	2	m6A	90.18	4639	5144	50.0	27.8	ACCNNNNNNRTC
HACCVNNNNNCTCD	2	m6A	16.64	117	703	40.7	30.2	
HGAGNNNNNYGGTD	3	m6A	16.60	86	518	41.5	29.5	
HACCVNNNHNRMC	2	m6A	14.94	936	6265	40.2	29.1	
Not Clustered	0		0.01	1163	9099552	35.6	31.2	

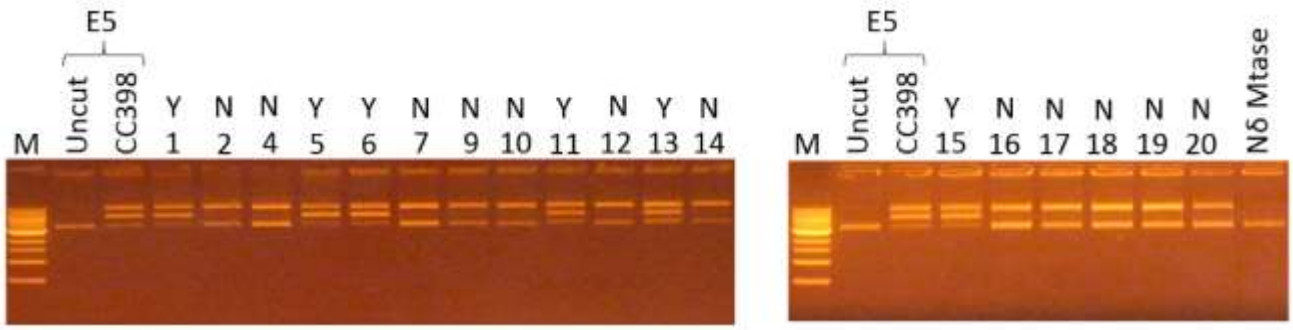
S.SauNd* ACC-6-TTRG

MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGKKGVMSPLYTVFKIQNIDLNFIEFYFKSS
KWYRFMALNGDSGARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
KKGYMQKIFSQELRFKDENGNDYPEWENVMLQKVLKDKTEGIKRGPFGGALKKDIFVESGYAVYEQ
RNAIYDISNFRYYINENKYKEMQSFSVQPNDIIMSCSGTIGRLALIPHNYTKGIINQALIRFRTNH
KIRSEFFLIFMRSNQMQRKILEANPGSAITNLVPVKELKLIPFPLPVKFEQDKISQFIHIINRRIE
QSEKKIESLKNRKQGFLQKLFVPGGSHHHHHH



1- marker 2- soluble cell extract 3- Nickel column flow through
4- Nickel column wash 5- Nickel column eluate
6- eluate after conc. and PD10 desalting
7- final protein after concentration 8- NP purified protein marker

DNA cleavage assay.



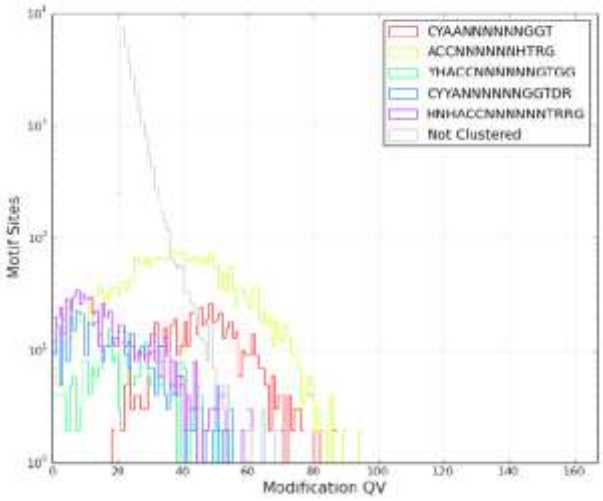
Site determined to be ACC-6-TTRG or ACC-6-YTRG. Note that the underlined site was determined by SMRT and is accepted since if Y is a cytosine, then it can't be methylated.

S.SauNd* ACC-6-TTRG

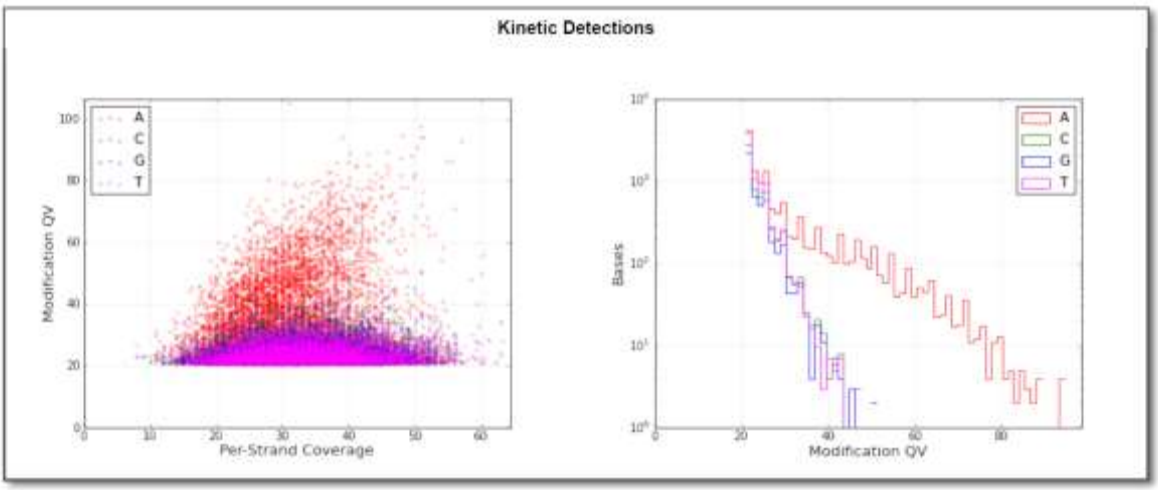
Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
CYAANNNNNNGGT	4	m6A	89.98	557	619	50.0	29.9	
ACCNNNNNNHTRG	1	m6A	67.06	2013	3002	47.5	31.0	
YHACCNNNNNNGTGG	3	m6A	30.03	88	293	40.5	34.0	
CYYANNNNNNGGTDR	4	m6A	20.24	102	504	42.0	32.9	
HNHACCNNNNNNTRRG	4	m6A	17.94	127	708	41.5	32.6	
Not Clustered	0		0.02	1435	9112200	35.8	37.4	

Modification QV Histogram By Motif



Kinetic Detections



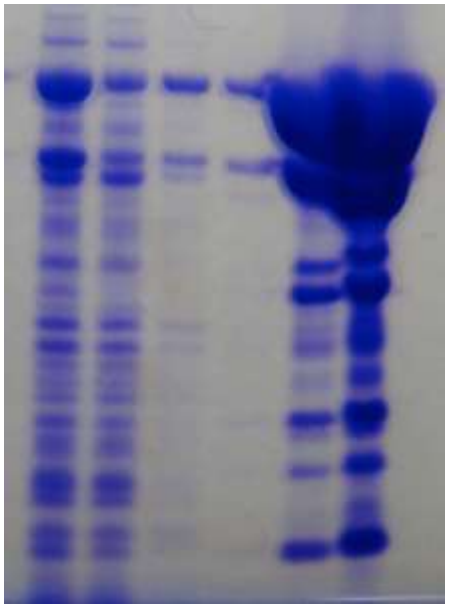
Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
CYAANNNNNNGGT	4	m6A	89.98	557	619	50.0	29.9	
ACCNNNNNNHTRG	1	m6A	67.06	2013	3002	47.5	31.0	
YHACCNNNNNNGTGG	3	m6A	30.03	88	293	40.5	34.0	
CYYANNNNNNGGTDR	4	m6A	20.24	102	504	42.0	32.9	
HNHACCNNNNNNTRRG	4	m6A	17.94	127	708	41.5	32.6	
Not Clustered	0		0.02	1435	9112200	35.8	37.4	

S.SauRE GARA-6-RTGA

MSNTQKKNVPELRFPGFEGEWEEKKLGEVAKIYDGTHQTPKYTNEGIKFLSVENIKTLNSSKYISE
EAFEKEFKIRPEFGDILMTRIGDIGTPNIVSSNEKFAYYVSLALLKTKNLNSYFLKNLILSSSIQN
ELWRKTLHVAFPKKINKNEIGKIKINYPKKQEQQKIGQFFSKLDRQIELEEQKLELLQQQKKGYMQ
KIFSQELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKDAITNGSYDFYVRSPIVYKINTFSYEG
EAILTVGDGVGVGKVFHYVNGKFDYHQRVYKISDFKNYYGLLLFYYFSQNFLKETKKYSAKTSVDS
VRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTKIQKQVIELLKQRKKSLLQKMFIPGGSHHHHHH

1 2 3 4 5 6



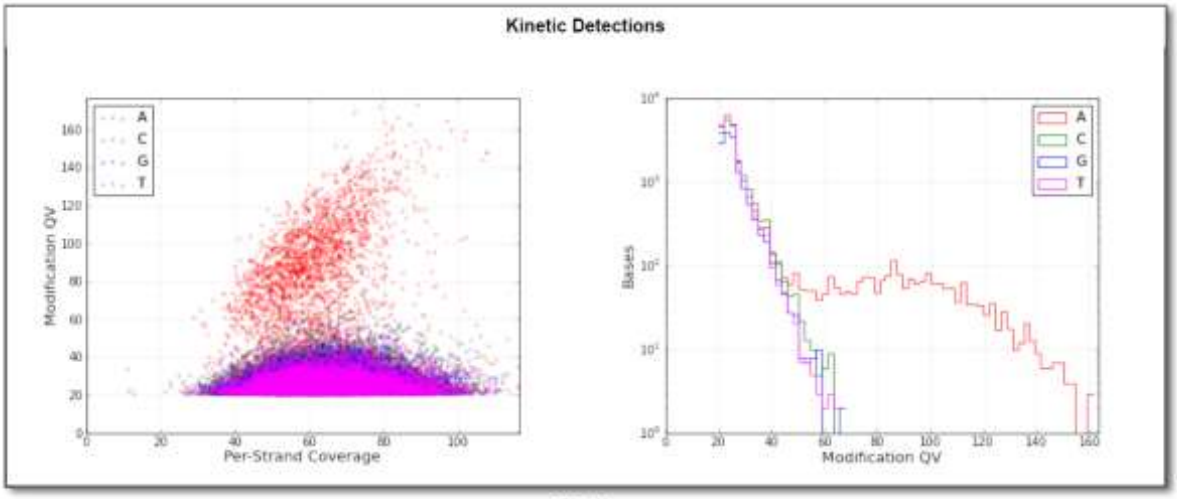
1- soluble cell extract, 2- Nickel column flow through, 3-Nickel column wash 1, 4- Nickel column wash 2, 5- Nickel column eluate, 6- Final protein after PD10 desalting and concentration

Although purified, this MTase was only used in SMRT.

Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
TCAYNNNNNNTYTC	3	m6A	91.88	792	862	104.6	62.5	GARANNNNNNRTGA
GARANNNNNNRTGA	4	m6A	79.58	686	862	96.7	61.7	TCAYNNNNNNTYTC
CCACDNNNNNTYTC	3	m6A	15.87	30	189	90.4	65.5	
TCAGNNTNNNTNTCNB	3	m6A	15.03	29	193	91.0	67.7	
Not Clustered	0		0.00	129	9115220	85.4	67.9	

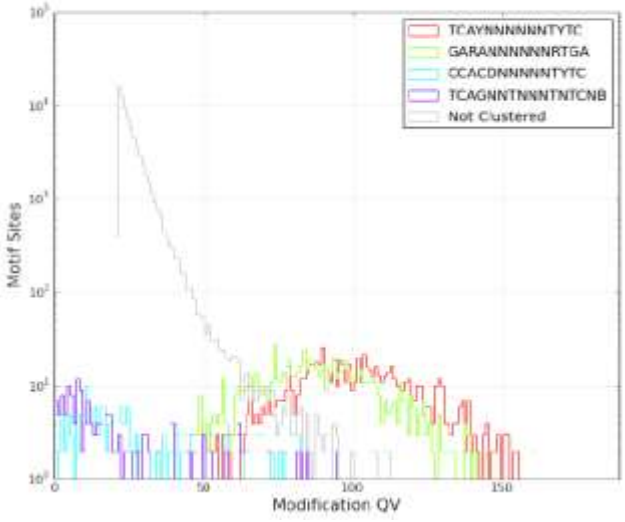
S.SauRE GARA-6-RTGA



Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
TCAYNNNNNNTYTC	3	m6A	91.88	792	862	104.6	62.5	GARANNNNNNRTGA
GARANNNNNNRTGA	4	m6A	79.58	686	862	96.7	61.7	TCAYNNNNNNTYTC
CCACDNNNNNTYTC	3	m6A	15.87	30	189	90.4	65.5	
TCAGNNTNNNTNTCNB	3	m6A	15.03	29	193	91.0	67.7	
Not Clustered	0		0.00	129	9115220	85.4	67.9	

Modification QV Histogram By Motif



S.SauTE CAAG-5-RTGA

MSNTQKKNVPELRFPGFEGEWEEKELGEIFQIISGSTPLKSNKEFYENGNINWVKTTDLNNSKVTH
SKEKITEYAMKSLKLKLVPKNSVLIAMYGGFNQIGRTGLLKIDATINQAISALLMNHETNPEFIQA
FLNYQVKGWKRYAASSRKDPNITKKDIEQFKVPYVSINEQQKIGEFFSKIDHQIELEEQKLELLQQ
QKKGYMQKIFSQELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKDAITNGSYDFYVRSPIVYKI
NTFSYEGEAILTVGDGVGVGKVFHYVNGKFDYHQRVYKISDFKNYYGLLLFYYFSQNFLKETKKYS
AKTSVDSVRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTKIQKQVIELLKQRKKSLLQKMFIPGG
SHHHHHH

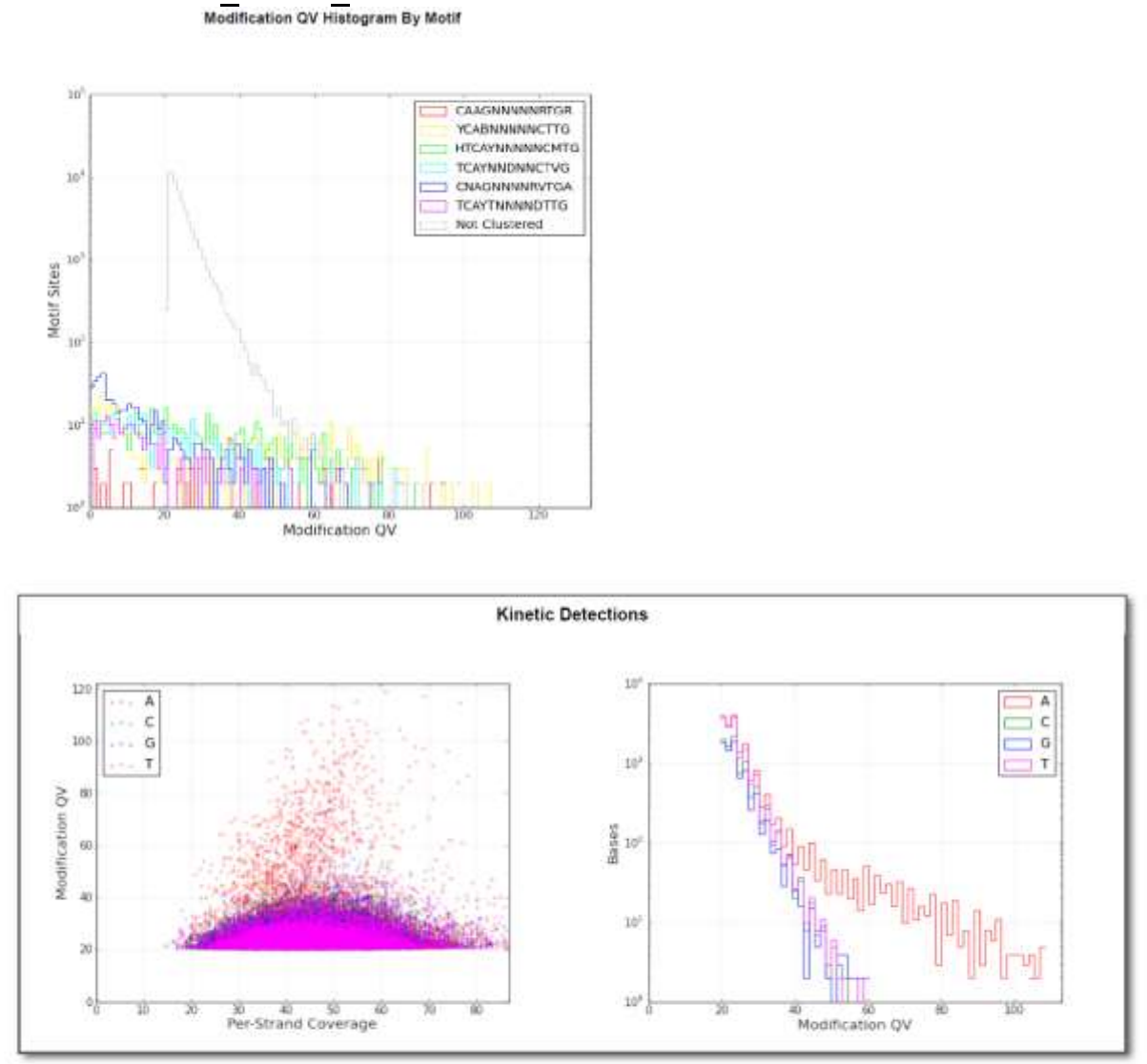
1- soluble cell extract 2- Nickel column flow through
3-Nickel column wash 1 4- Nickel column wash 2
5- Nickel column eluate 6- eluate after concentrating and PD10 step
7- Final concentrated protein

This MTase did not purify well and was only analysed by SMRT sequencing. The degeneracy in the target determined by SMRT sequencing can be removed using results from other systems.

Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
CAAGNNNNNRTGR	3	m6A	80.15	214	267	63.8	42.7	
YCABNNNNNCTTG	3	m6A	55.08	260	472	65.1	43.1	
HTCAYNNNNNCMTG	4	m6A	47.41	238	502	51.7	44.4	
TCAYNNDNNCTVG	3	m6A	31.07	119	383	51.0	44.3	
CNAGNNNNRVTGA	3	m6A	28.62	170	594	59.5	43.6	
TCAYTNNNNDTTG	3	m6A	16.89	38	225	45.4	44.3	
Not Clustered	0		0.04	3962	9114883	35.8	50.6	

S.SauTE CAAG-5-RTGA



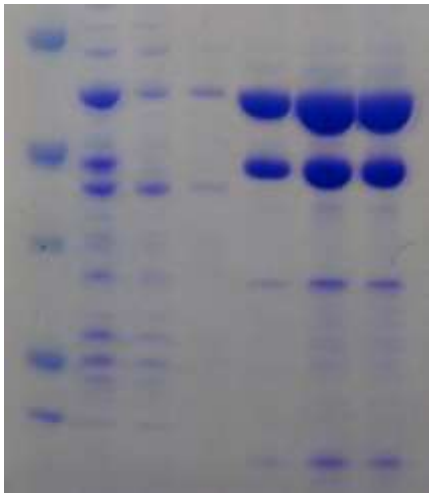
Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
CAAGNNNNNRTGR	3	m6A	80.15	214	267	63.8	42.7	
YCABNNNNNCTTG	3	m6A	55.08	260	472	65.1	43.1	
HTCAYNNNNNCMTG	4	m6A	47.41	238	502	51.7	44.4	
TCAYNNDNNCTVG	3	m6A	31.07	119	383	51.0	44.3	
CNAGNNNNRVTGA	3	m6A	28.62	170	594	59.5	43.6	
TCAYTNNNNDTTG	3	m6A	16.89	38	225	45.4	44.3	
Not Clustered	0		0.04	3962	9114883	35.8	50.6	

S.SauVE CNGA-6-RTGA

MSNTQKKNVPELRFPGFEGEWEEKELRELRNPKDKYSYTGGPFGSDLKKSDYTTDGIQIIQLQNIG
DGYFYNSNKVFTSNEKAEVLKSCNVFPGDIVIAKMADPIARAAIVPDNNIGKYLMASDGIRLSVDT
VHFNTKFVLECINRKSFRKKVEDNSSGSTRMRIGLSTLGSLTLKTTTLKEQQKIGQFFSKLDRQIE
LEEQKLELLQQQKKGYMQKIFSQELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKDAITNGSYD
FYVRSPIVYKINTFSYEGEAILTVGDGVGVGKVFHYVNGKFDYHQRVYKISDFKNYYGLLLFYYFS
QNFLKETKKYSAKTSVDSVRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTKIQKQVIELLKQRKK
SLLQKMFIPGGSHHHHHH

1 2 3 4 5 6 7



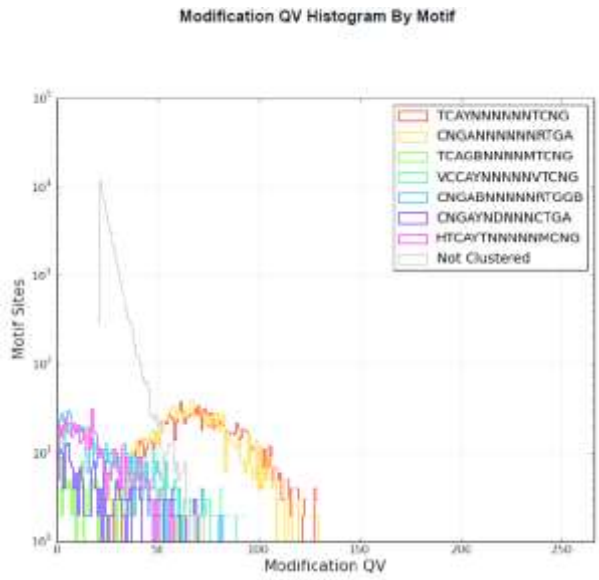
1- marker, 2- soluble cell extract, 3- Nickel column flow through, 4- Nickel column wash, 5- Nickel column eluate, 6- eluate after conc. and PD10 desalting, 7- Final protein after concentration

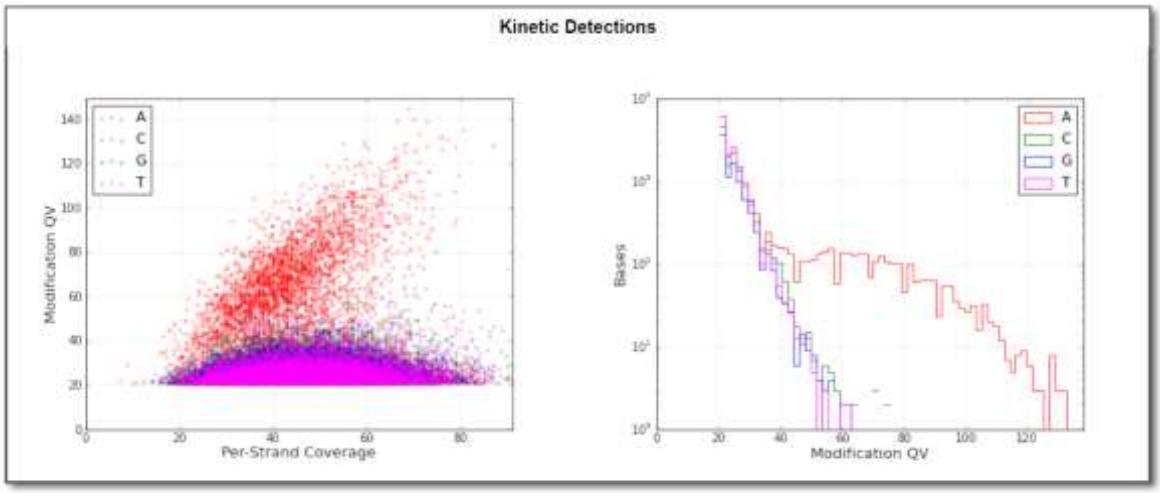
Although the MTase was purified, it was only analysed via SMRT sequencing.

Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
TCAYNNNNNNTCNG	3	m6A	99.34	1354	1363	73.0	43.2	CNGANNNNNNRTGA
CNGANNNNNNRTGA	4	m6A	98.09	1337	1363	71.4	43.4	TCAYNNNNNNTCNG
TCAGBNNNNNMTCNG	3	m6A	41.58	79	190	53.7	43.2	
VCCAYNNNNNVTCNG	4	m6A	33.12	211	637	50.8	44.7	CNGABNNNNNRTGGB
CNGABNNNNNRTGGB	4	m6A	29.98	191	637	50.2	45.8	VCCAYNNNNNVTCNG
CNGAYNDNNNCTGA	4	m6A	31.90	74	232	53.0	44.9	
HTCAYTNNNNNMCNG	4	m6A	23.31	131	562	44.6	45.5	
Not Clustered	0		0.05	4345	9112342	35.9	49.9	

S.SauVE CNGA-6-RTGA



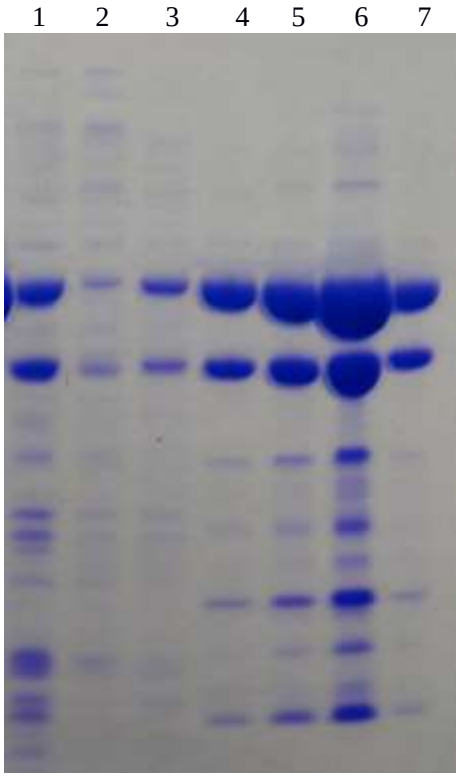


Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
TCAYNNNNNNTCNG	3	m6A	99.34	1354	1363	73.0	43.2	CNGANNNNNNRTGA
CNGANNNNNNRTGA	4	m6A	98.09	1337	1363	71.4	43.4	TCAYNNNNNNTCNG
TCAGBNNNNNMTCNG	3	m6A	41.58	79	190	53.7	43.2	
VCCAYNNNNNVTCNG	4	m6A	33.12	211	637	50.8	44.7	CNGABNNNNNRTGGB
CNGABNNNNNRTGGB	4	m6A	29.98	191	637	50.2	45.8	VCCAYNNNNNVTCNG
CNGAYNDNNNCTGA	4	m6A	31.90	74	232	53.0	44.9	
HTCAYTNNNNNMCNG	4	m6A	23.31	131	562	44.6	45.5	
Not Clustered	0		0.05	4345	9112342	35.9	49.9	

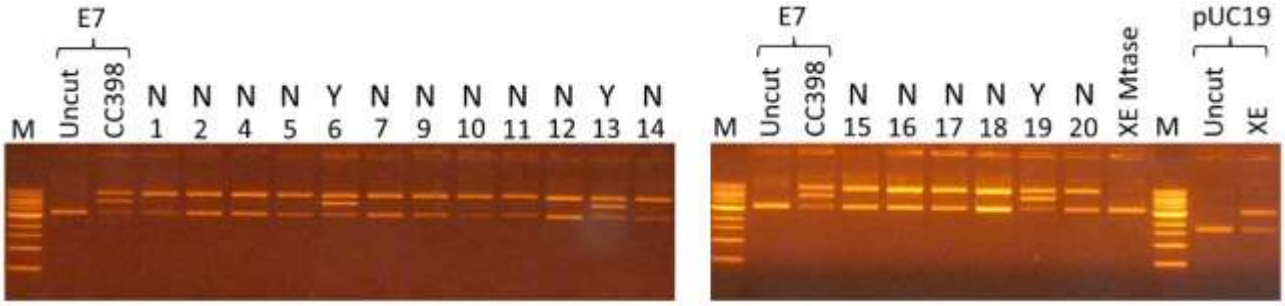
S.SauXE TCTA-6-RTGA

MSNTQKKNVPELRFPGFEGEWEEKQFADFTKINQGLQIAINERKTEYSPELYFYITNEFLRPNSQT
KYFIENPPQSVIANKEDILMTRTGNTGKVVTNVFGAFHNNFFKIKFDKNLYDRLFLVEVLNSSKIQ
NKILSLAGSSTIPDLNHSDFYSISSSYPLLREQQKIGKFFSKLDRQIELEEQKLELLQQQKKGYMQ
KIFSQELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKDAITNGSYDFYVRSPIVYKINTFSYEG
EAILTVGDGVGVGKVFHYVNGKFDYHQRVYKISDFKNYYGLLLFYYFSQNFLKETKKYSAKTSVDS
VRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTKIQKQVIELLKQRKKSLLQKMFIPGGSHHHHHH



1- soluble cell extract 2- Nickel column flow through
3- Nickel column wash 4- Nickel column eluate
5- eluate after conc. and PD10 desalting
6- final protein after concentration
7- CC398-1 purified protein marker

DNA cleavage assay.



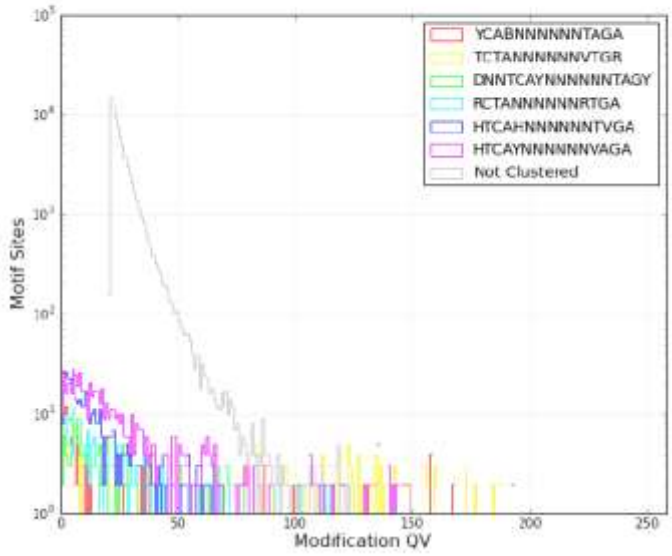
S.SauXE TCTA-6-RTGA

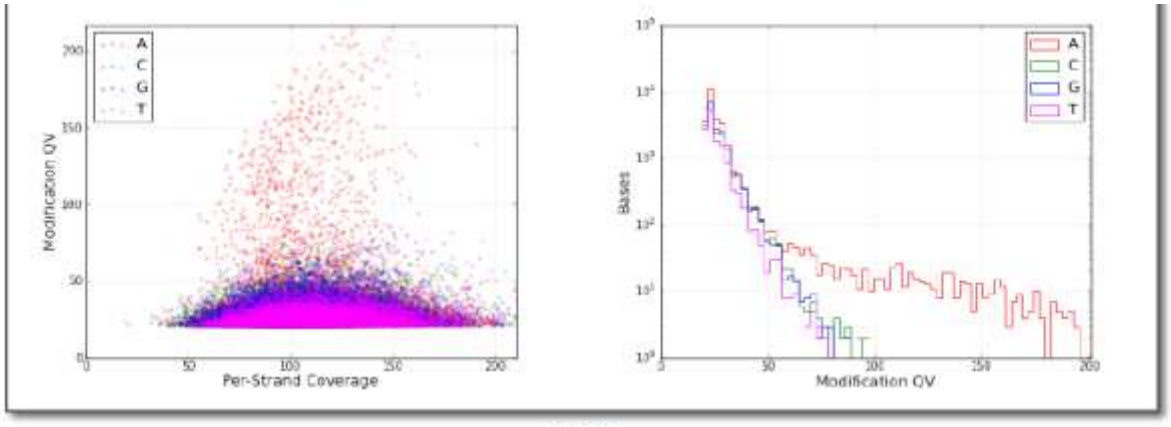
The degeneracy in the target determined by SMRT sequencing can be resolved by reference to targets from other systems.

Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
YCABNNNNNNNTAGA	3	m6A	68.77	196	285	121.6	104.1	TCTANNNNNNVTGR
TCTANNNNNNVTGR	4	m6A	68.07	194	285	120.0	106.6	YCABNNNNNNNTAGA
DNNTCAYNNNNNNNTAGY	6	m6A	44.57	82	184	79.3	103.0	
RCTANNNNNNNRTGA	4	m6A	40.41	99	245	71.8	104.1	
HTCAHNNNNNNNTVGA	4	m6A	32.43	156	481	103.9	104.2	
HTCAYNNNNNNNVAGA	4	m6A	32.07	211	658	67.2	107.8	
Not Clustered	0		0.12	10813	9115188	38.1	115.4	

Modification QV Histogram By Motif



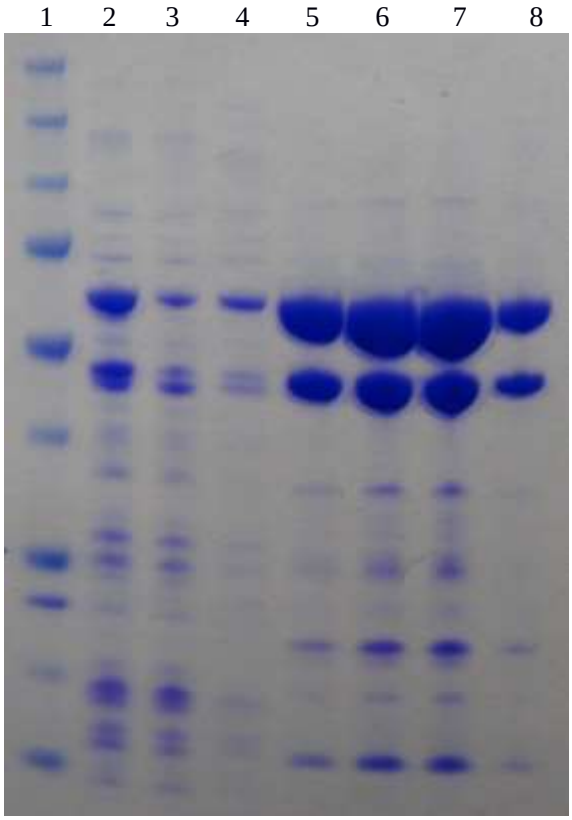


Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
YCABNNNNNNNTAGA	3	m6A	68.77	196	285	121.6	104.1	TCTANNNNNNVTGR
TCTANNNNNNVTGR	4	m6A	68.07	194	285	120.0	106.6	YCABNNNNNNNTAGA
DNNTCAYNNNNNNNTAGY	6	m6A	44.57	82	184	79.3	103.0	
RCTANNNNNNNRTGA	4	m6A	40.41	99	245	71.8	104.1	
HTCAHNNNNNNNTVGA	4	m6A	32.43	156	481	103.9	104.2	
HTCAYNNNNNNNVAGA	4	m6A	32.07	211	658	67.2	107.8	
Not Clustered	0		0.12	10813	9115188	38.1	115.4	

S.SauZE GAC-5-RTGA

MSNTQKKNVPELRFPGFEGEYSLDIFGNLATNKSEKFNPQNENASIDIELDCIEQNTGRLIKIYNS
KEFSSQKNKFNPQNVLYGKLRPYLNKYYFTKKSGVCSSEIWVLKSTKEDKLLNLFLYYFIQTKRYS
DVASKSAGSKMPRADWGLIENIRVYFPELCEQQKIGQFFSKLDRQIELEEQKLELLQQQKKGYMQK
IFSQELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKDAITNGSYDFYVRSPIVYKINTFSYEGE
AILTVGDGVGVGKVFHYVNGKFDYHQRVYKISDFKNYYGLLLFYYFSQNFLKETKKYSAKTSVDSV
RKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTKIQKQVIELLKQRKKSLLQKMFIPGGSHHHHHH*



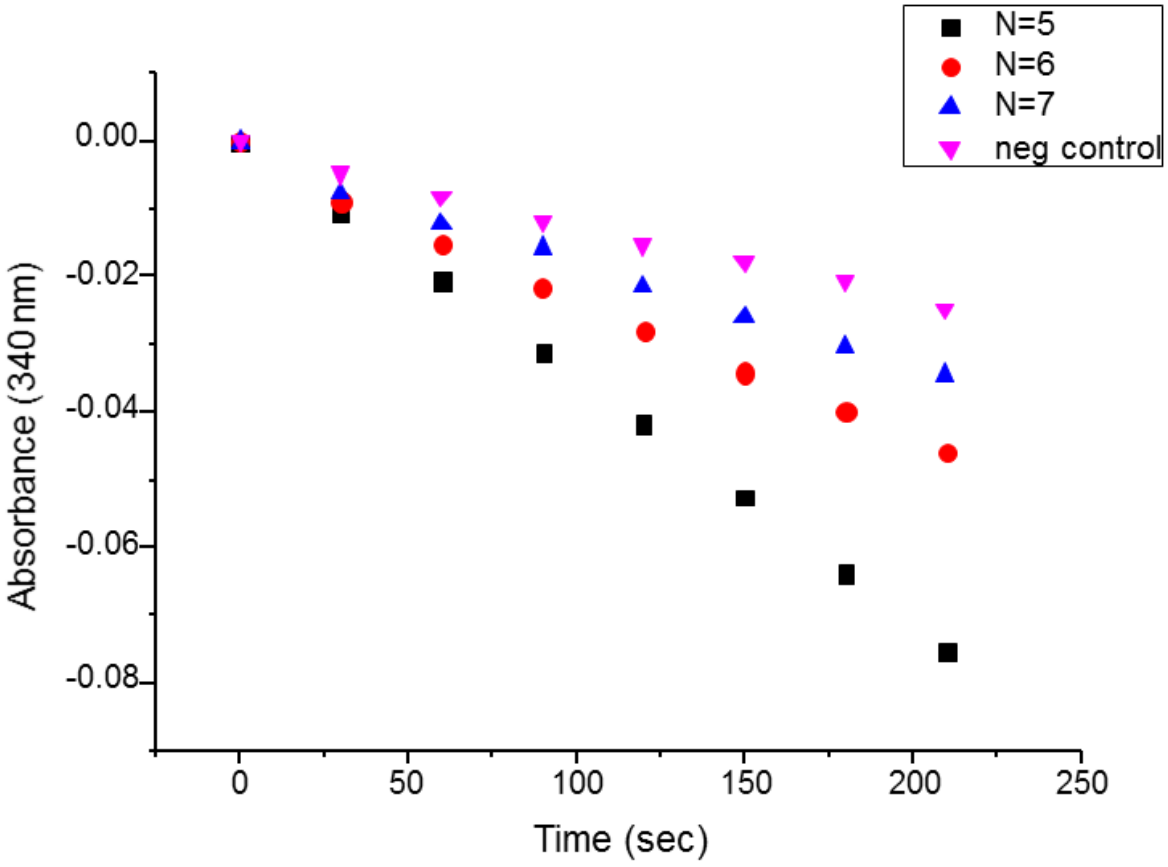
1- marker 2- soluble cell extract 3- Nickel column flow through
4- Nickel column wash 5- Nickel column eluate 6- eluate after conc. step and PD10 desalting
7- final concentrated protein 8- CC398-1 purified protein marker

DNA cleavage assay showed cutting of all plasmids so the ATPase assay was used given that we knew the individual TRD specificities.

Oligonucleotide name	DNA sequence (5' to 3')
ZE5for	AGATGATGGAATCAATGCGACTTCCAGTGAGCCCTATACGATATAA
ZE5rev	TTATATCGTATAGGGCTCACTGGAAGTCGCATTGATTCCATCATCT
ZE6for	AGATGATGGAATCAATGCGACTTCCATGTGAGCCCTATACGATATAA
ZE6rev	TTATATCGTATAGGGCTCACATGGAAGTCGCATTGATTCCATCATCT
ZE7for	AGATGATGGAATCAATGCGACTTCACATGTGAGCCCTATACGATATAA
ZE7rev	TTATATCGTATAGGGCTCACATGTGAAGTCGCATTGATTCCATCATCT

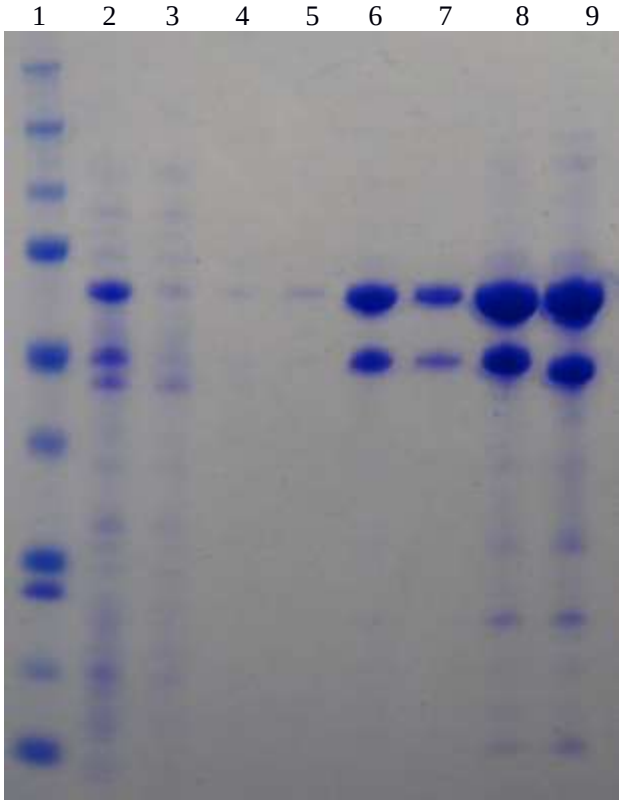
S.SauZE GAC-5-RTGA

N=5 gives the clearest signal.



S.SauZS GAC-6-TGC

MSNTQKKNVPELRFPGFEGEYSLDIFGNLATNKSEKFNPQNENASIDIELDCIEQNTGRLIKIYNS
KEFSSQKNKFNPQNVLYGKLRPYLNKYYFTKKSGVCSSEIWVLKSTKEDKLLNLFLYYFIQTKRYS
DVASKSAGSKMPRADWGLIENIRVYFPELCEQQKIGQFFSKLDRQIELEEQKLELLQQQKKGYMQK
IFSQELRFKDENGNDYPDWTNERLGEVTTVTMGQSPKSVNYTDNSNDTVLIQGNADIENGLINPRI
YTREVTKLIQKDEIILTVRAPVGKLAMAQINACIGRGVCSIKGDKFLYYFLEWFATQNKWIRFSQG
STFESISGNDIRNIHIKIPVEDERTKIIKLLNSLDVLNSKTDLKIQNLKQRKQSLLQKIFVPGGSH
HHHHH*



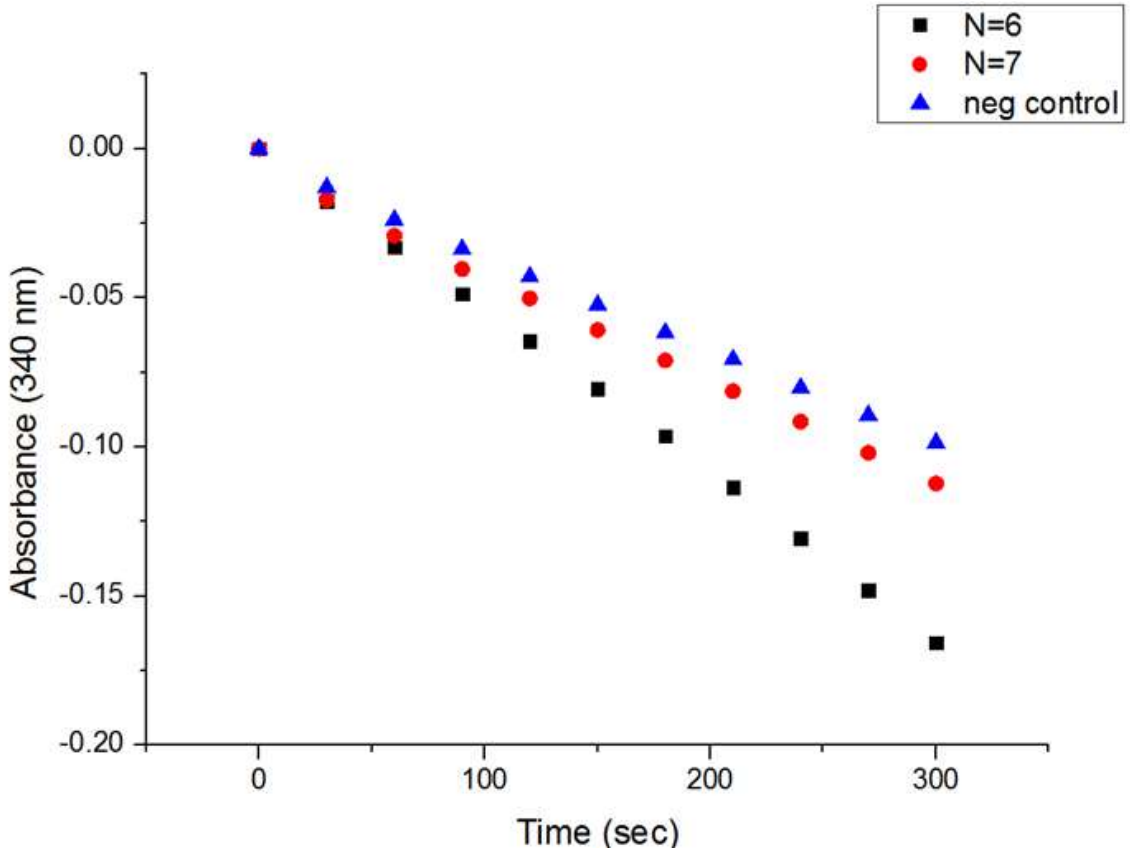
1- marker 2- soluble cell extract 3- Nickel column flow through
4- Nickel column wash 1 5- Nickel column wash 2 6- Nickel column eluate
7- eluate after PD10 desalting 8- final protein after concentration
9- NQ purified protein marker

The DNA cleavage assay showed cutting of all plasmids so the ATPase assay was used since we knew the TRD specificities.

Oligonucleotide name	DNA sequence (5' to 3')
ZS6for	AGATGATGGAATCAATGCGACTTCCATTGCGCCCTATACGATATAA
ZS6rev	TTATATCGTATAGGGCGCAATGGAAGTCGCATTGATTCCATCATCT
ZS7for	AGATGATGGAATCAATGCGACTTCACATTGCGCCCTATACGATATAA
ZS7rev	TTATATCGTATAGGGCGCAATGTGAAGTCGCATTGATTCCATCATCT

S.SauZS GAC-6-TGC

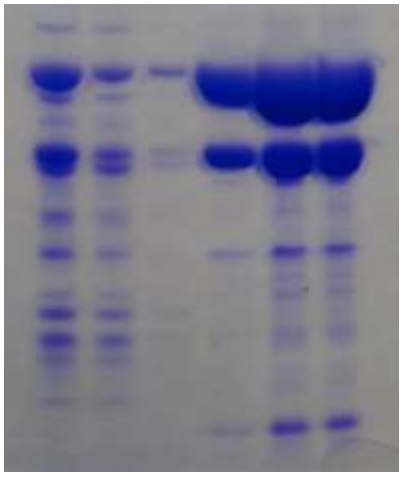
N=6 gives the greatest activity.



S.Saub*E GGHA-6-RTGA

MSNTQKKNVPELRFPGFEGEWEEKKLEDTLEFIKDGTHGTHENVNNGPWLLSAKNIKNNKIIISSD
DRKISESDYKKIYKNYKLEKGDLLLTIVGTIGRAAIVKNPNNIAFQRSVAILKTKATYDVGFIFQL
FQTKYFKNLLLRKQVVSAQPGLYLGDIRKIKISITNIIEEQRKIGIFFSKLDRQIELEEQKLELLQ
QQKKGYMQKIFSQELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKDAITNGSYDFYVRSPIVYK
INTFSYEGEAILTVGDGVGVGKVFHYVNGKFDYHQRVYKISDFKNYYGLLLFYYFSQNFLKETKKY
SAKTSVDSVRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTKIQKQVIELLKQRKKSLLQKMFIPG
GSHHHHHH

1 2 3 4 5 6



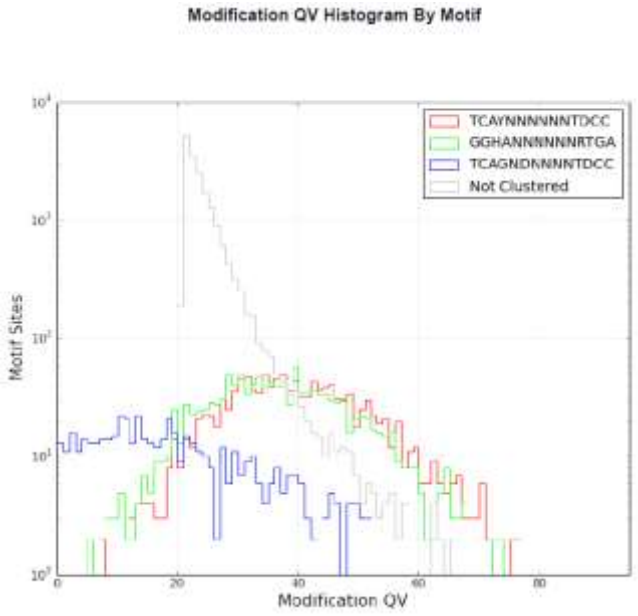
1- soluble cell extract, 2- Nickel column flow through, 3- Nickel column wash, 4- Nickel column eluate, 5- eluate after conc. step and PD10 desalting, 6- Final concentrated protein

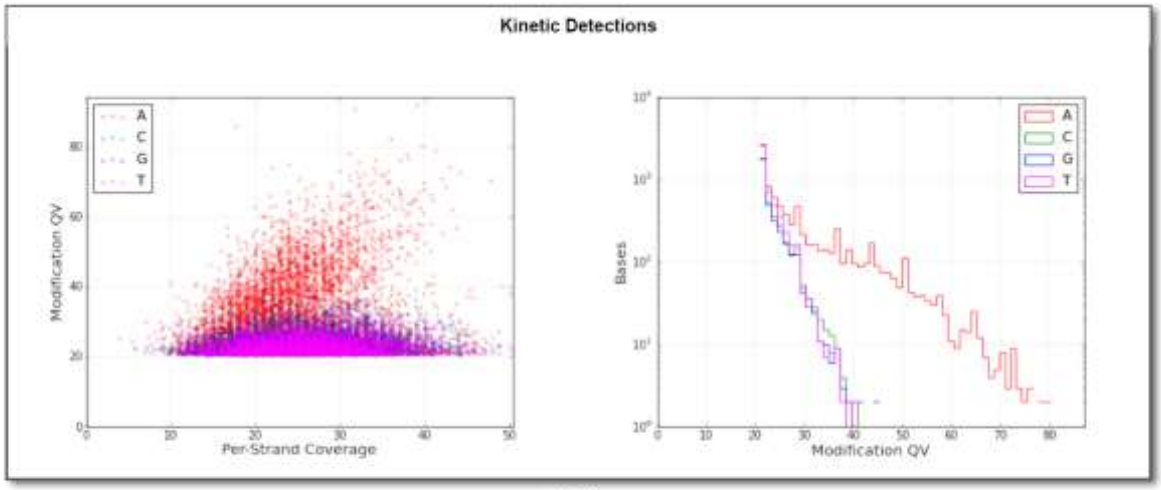
Although this MTase was purified, it was only used in the SMRT sequencing assay.

Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
TCAYNNNNNNTDCC	3	m6A	77.39	989	1278	44.6	24.6	GGHANNNNNNRTGA
GGHANNNNNNRTGA	4	m6A	67.68	865	1278	43.2	24.7	TCAYNNNNNNTDCC
TCAGNDNNNNTDCC	3	m6A	21.57	110	510	41.0	25.2	
Not Clustered	0		0.01	928	9114260	36.7	27.3	

S.Saub*E GGHA-6-RTGA





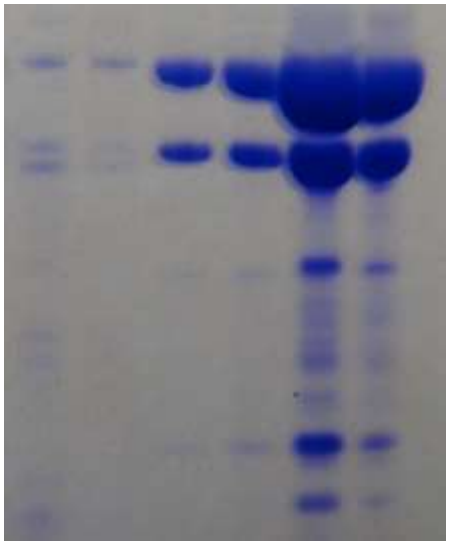
Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
TCAYNNNNNNTDCC	3	m6A	77.39	989	1278	44.6	24.6	GGHANNNNNNRTGA
GGHANNNNNNRTGA	4	m6A	67.68	865	1278	43.2	24.7	TCAYNNNNNNTDCC
TCAGNDNNNNTDCC	3	m6A	21.57	110	510	41.0	25.2	
Not Clustered	0		0.01	928	9114260	36.7	27.3	

S.Saue*E GAG-6-RTGA

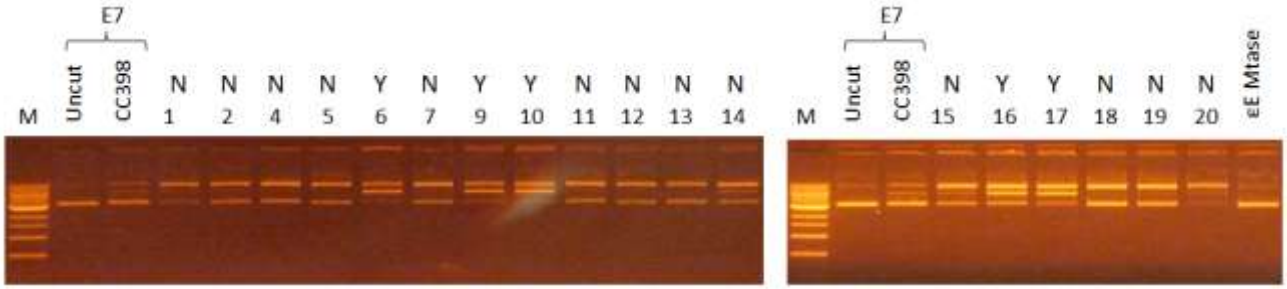
MSNTQKKNVPELRFPGFEGEWEEKSISSFLKESKIKGSNGSHAKKLTVKLWGKGVVPKKETFKGSD
NTQYYKRKAGQLMYGKLDFLNCAFGIVPDSLNNYESTIDSPSFDFINGDSKFLLERIKLKSFYKKF
GDIANGSRKAKRINQDTFLSLPVFAPKYDEQLRIGEFFSKLDRQIELEEQKLELLQQQKKGYMQKI
FSQELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKDAITNGSYDFYVRSPIVYKINTFSYEGEA
ILTVGDGVGVGKVFHYVNGKFDYHQRVYKISDFKNYYGLLLFYYFSQNFLKETKKYSAKTSVDSVR
KDMIANMKVPRPIYIEQKKIGQFIKRVDNKTKIQKQVIELLKQRKKSLLQKMFIPGGSHHHHHH

1 2 3 4 5 6



1- soluble cell extract, 2- Nickel column flow through, 3- Nickel column eluate, 4- eluate after PD10 desalting, 5- Final concentrated protein, 6- RE purified protein as marker

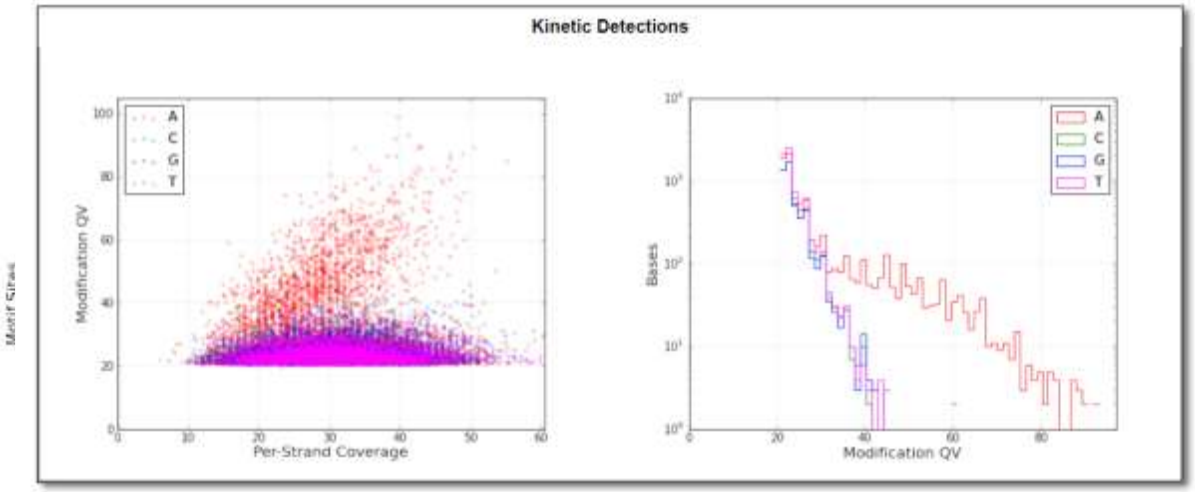
DNA cleavage assay.



S.Saue*E GAG-6-RTGA

Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
TCAYNNNNNNCTC	3	m6A	88.60	762	860	49.4	28.5	GAGNNNNNNRTGA
GAGNNNNNNRTGA	2	m6A	87.33	751	860	50.0	28.2	TCAYNNNNNNCTC
GAGNDNNNNGTGGB	2	m6A	20.22	37	183	40.9	29.5	
DNNGAGNDNNNNGAGA	5	m6A	18.56	36	194	39.8	32.7	
Not Clustered	0		0.01	914	9115229	34.6	36.3	



Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
TCAYNNNNNNCTC	3	m6A	88.60	762	860	49.4	28.5	GAGNNNNNNRTGA
GAGNNNNNNRTGA	2	m6A	87.33	751	860	50.0	28.2	TCAYNNNNNNCTC
GAGNDNNNNGTGGB	2	m6A	20.22	37	183	40.9	29.5	
DNNGAGNDNNNNGAGA	5	m6A	18.56	36	194	39.8	32.7	
Not Clustered	0		0.01	914	9115229	34.6	36.3	

SUPPLEMENTARY INFORMATION FOR TABLE 4.

S.SauAc* CCAY-6-RTC

The Ac* TRD combination is found in CC97-1. The MTase was not purified but instead used to methylate the genome of E. coli ER2796 for SMRT analysis. The target is CCAYNNNNNNRTC. There are a few minor amino acid differences in the S.SauAc* between members of CC97.

CC97

Recombinant S.SauAc* CC97-1

MSNTQKKNVPELRFPGFEGEWEEKKLGDLTTKIGSGKTPKGGSENYTNKGIPFLRSQNIRNGKLNLNDLVYISKDIDDEM
KNSRTYYGDVLLNITGASIGRTAINSIVEIHANLNQHVCIIRLKKEYYYNFFGQYLLSRKGKRKIFLAQSGGSREGLNFK
EIANLKIFTPTIFEEQQKIGEFISKLDRQIELEEQKLELLQQQKKGYLQKIFSQELRFKDENGNDYPEWRFARFKDFMYK
PINIRPAINISKSELLTVKLHCKGIEKANINRVLKLGATNYYKRFEGQFIYGKQNFFNGAFDIVPKKFDGLYSSSDVPAF
EINTEKIEPNYFISYISRPSFYKSKEKYSTGTGSKRIHENTVLNFSLHLPCLNEQLKIASFVCFLNRKIELLERKIYLIK
KQKQALLQQMFIPGGSHHHHHH

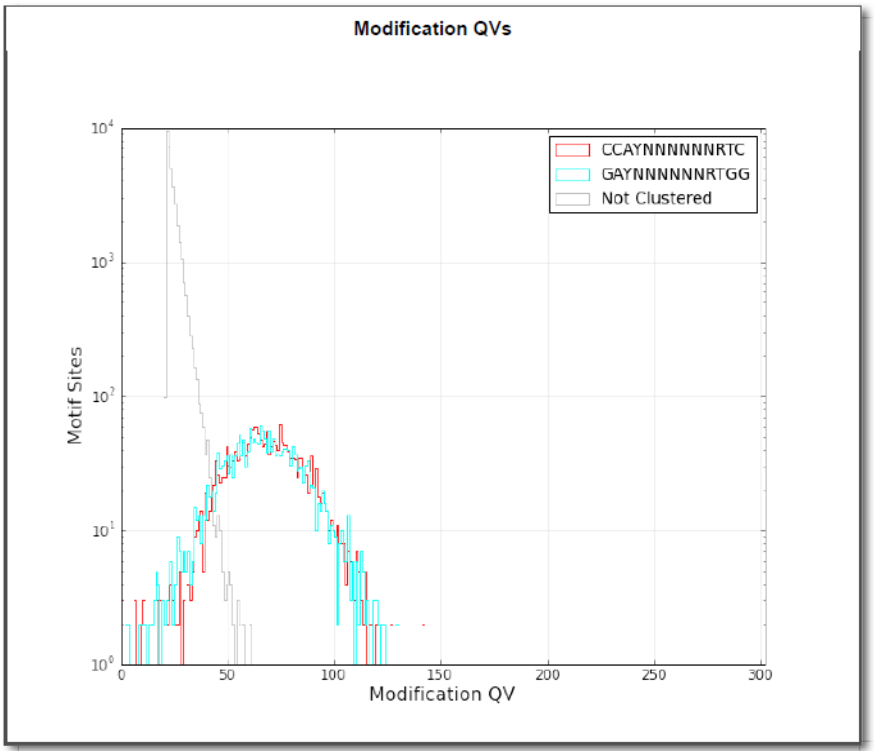
Wild Type S.SauAc*

MSNTQKKNVPELRFPGFEGEWEEKQLGDLTTKIGSGKTPKGGSENYTNKGIPFLRSQNIRNGKLNLNDLVYISKDIDDEM
KNSRTYYGDVLLNITGASIGRTAINSIVETHANLNQHVCIIRLKKEYYYIFFGQYLLSRKGKRKIFLAQSGGSREGLNFK
EIANLKIFTPTIFEEQQKIGKFFSKLDRQIELEEQKLELLQQQKKGYLQKIFSQELRFKDENGNDYPEWRFARFKDFMYK
PINIRPAINISKSELLTVKLHCKGIEKANINRVLKLGATNYYKRFEGQFIYGKQNFFNGAFDIVPKKFDGLYSSSDVPAF
EINTEKIEPNYFISYISRPSFYKSKEKYSTGTGSKRIHENTVLNFSLHLPCLNEQLKIASFVCFLNRKIELLERKIYLIK
KQKQALLQQMFI*

SMRT Cells: 1 Movies: 1

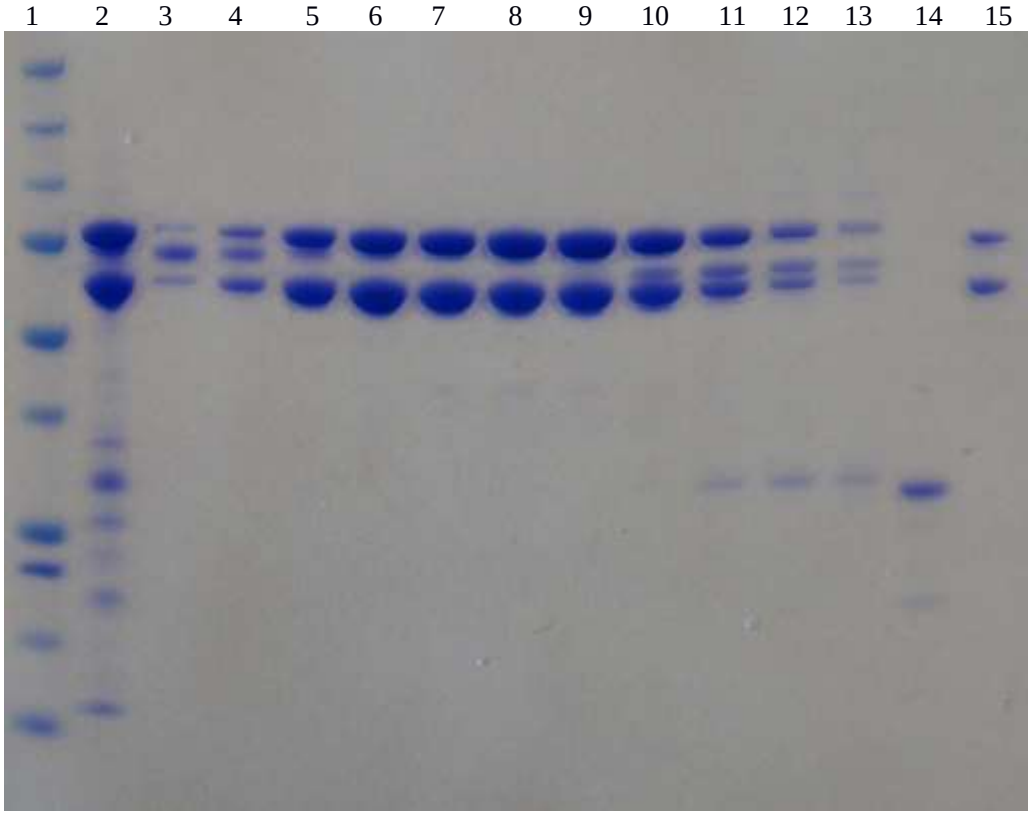
Motif Summary

Motifs	Modified Position	Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
CCAYNNNNNNRTC	3	m6A	97.52%	2199	2255	68.95	50.78	GAYNNNNNNRTGG
GAYNNNNNNRTGG	2	m6A	96.01%	2165	2255	68.50	51.01	CCAYNNNNNNRTC



S.SauBI-EGFP
CC22-1 AGG-6-TGAR
This MTase was expressed and purified as a fusion with EGFP.
Nuclease assays and SMRT analysis gave the same target site.

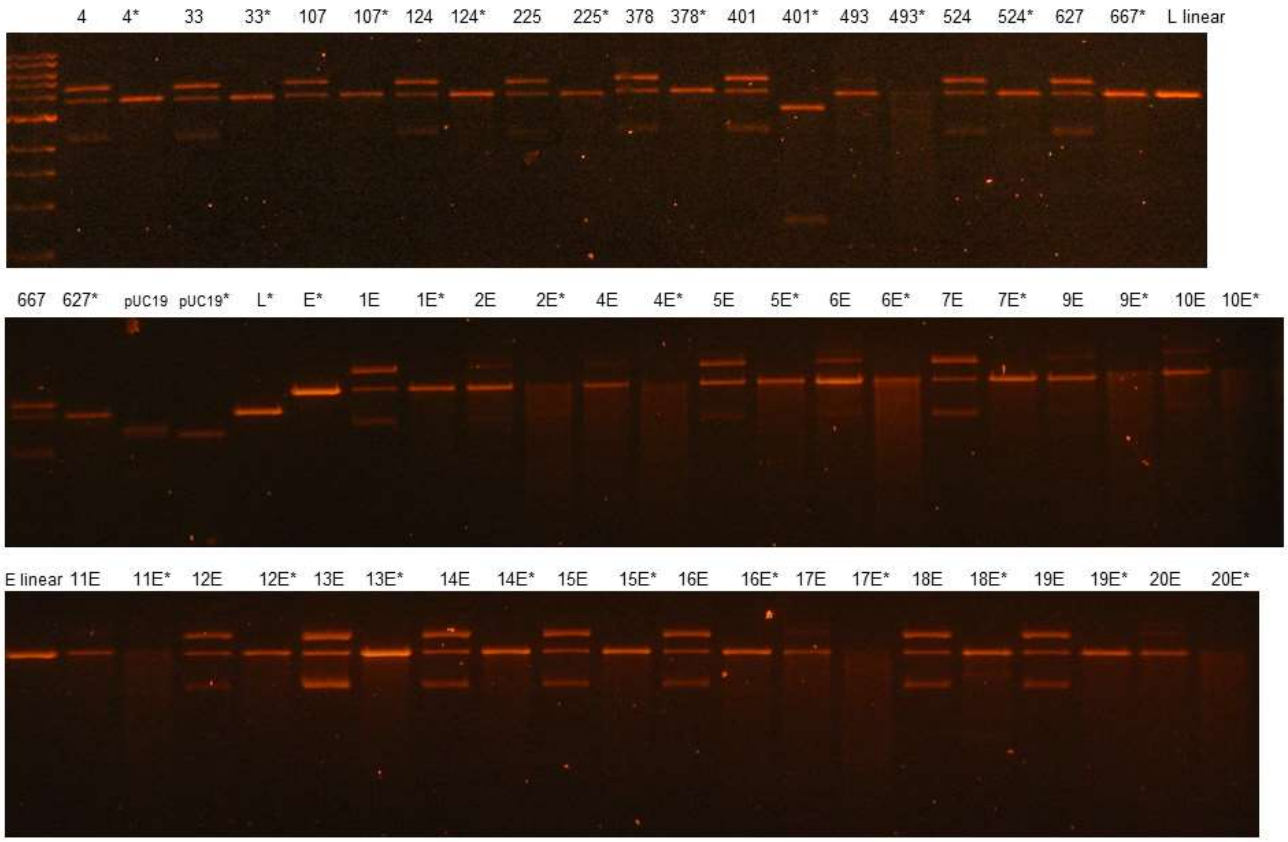
MSNTQKKNVPELRFPGFEGEWEEKKLGDLTDRVIRKNKNLESKKPLTISGQLGLIDQTEYFSKSVS
SKNLENYTLIKNGEFAYNKSYSNGYPLGAIKRLTRYDSGVLSSLYICFSIKSEMSKDFMEAYFDST
HWYREVSGIAVEGARNHGLLNVSVNDFFTILIKYPSLEEQQKIGKFFSKLDRQIELEEQKLELLQQ
QKKGYMQKIFSQELRFKNENGNDYPDWERIKFFDVIDKVIDFRGRTPKKLNMEWSDEGYLALSAVN
VKKGYIDFNVEAKYGNLDLYTRWMRGNELYKGQVLFTTEAPMGNVAQVPDNKGYILSQRTIAFNSN
EKITDNFLASLLSSENVYNDLLKLCSGATAKGVSQKNLNRLYVTIPHSISEQEEIAEFFRKINQLV
ELQKYKIEHTKSQKQVFLQKMFIGSMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYG
KLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNY
KTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIE
DGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYK
HHHHHH



1- marker 2- Nickel column eluate 3-14 Fractions from gel filtration column
15- CC5-1 Purified protein marker

S.SauBI-EGFP
CC22-1 AGG-6-TGAR

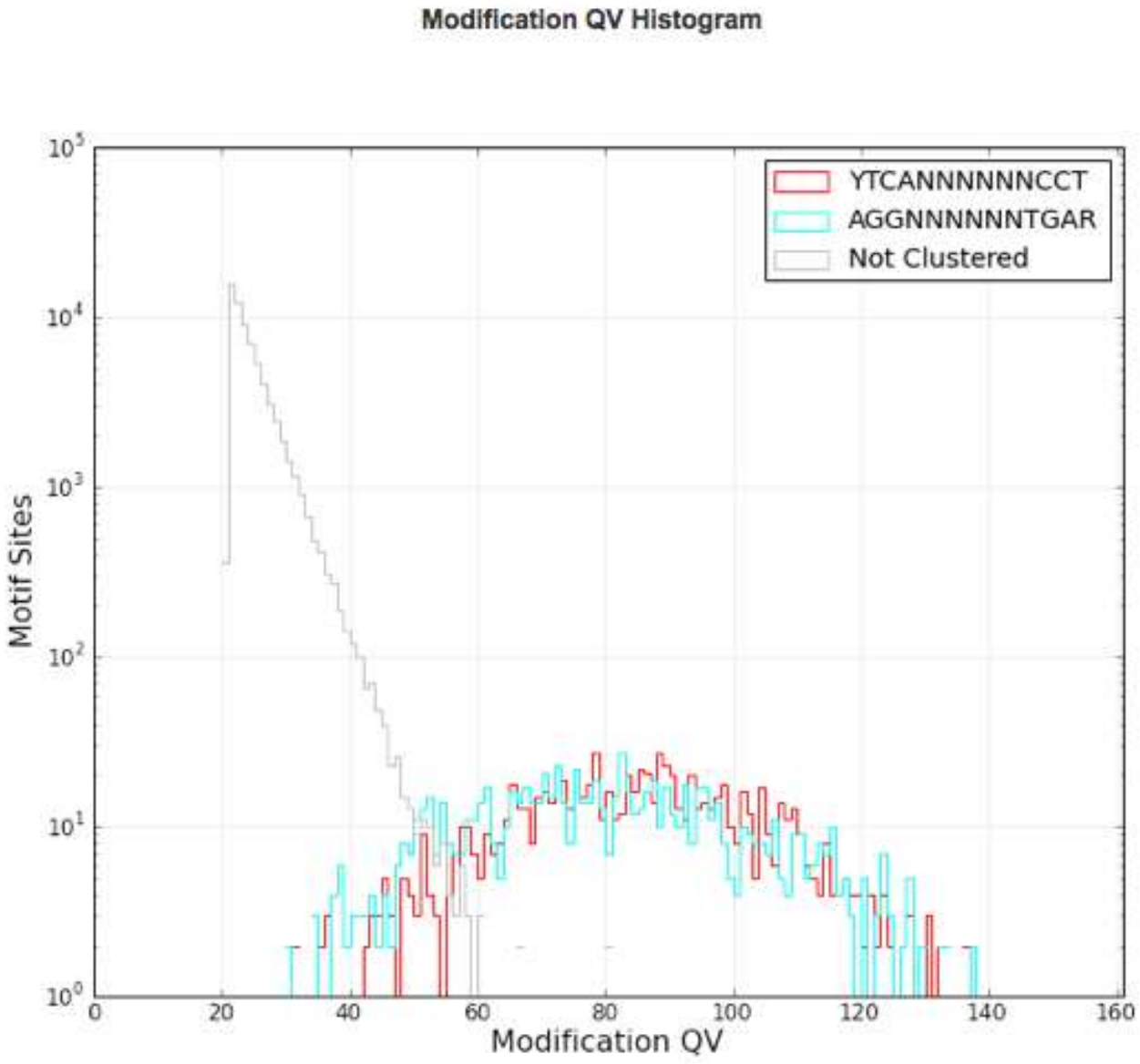
DNA cleavage assay



S.SauBI-EGFP
CC22-1 AGG-6-TGAR

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
YTCANNNNNCCT	4	m6A	99.24	919	926	86.2	56.3	AGGNNNNNTGAR
AGGNNNNNTGAR	1	m6A	99.24	919	926	83.9	55.7	YTCANNNNNCCT
Not Clustered	0		0.06	5230	9124356	34.8	61.5	

Modification QV Histogram By Motif



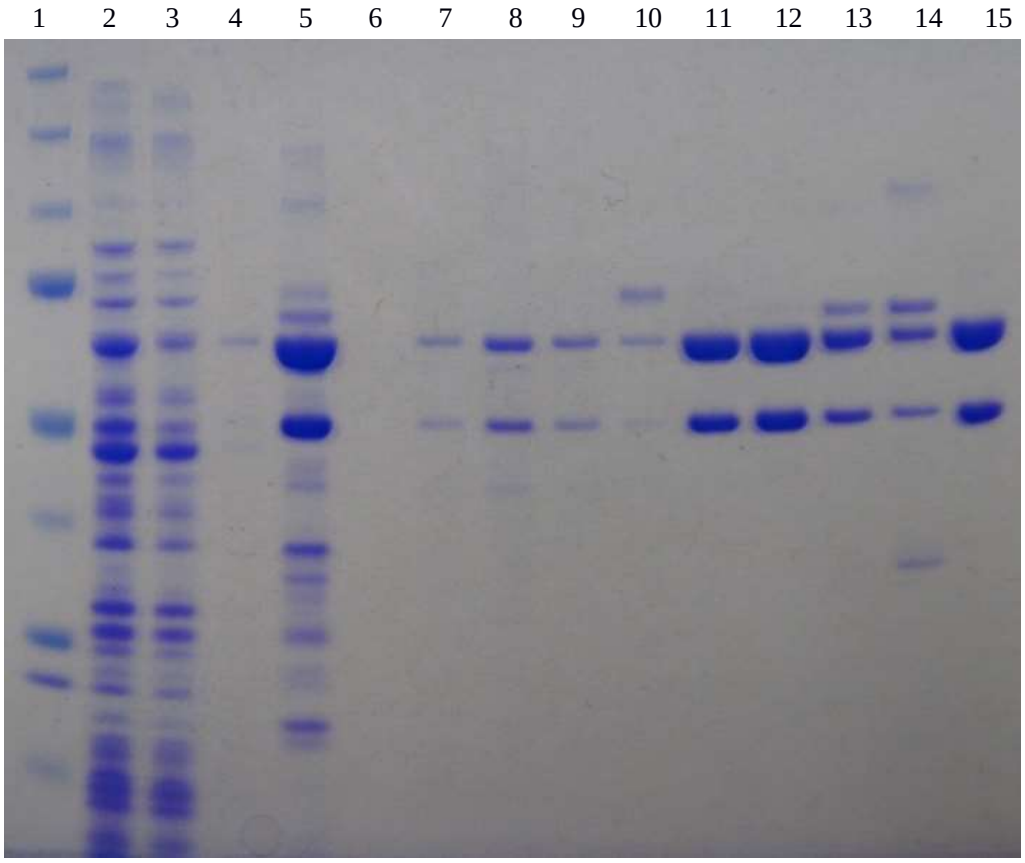
S.SauCE
ST425-1 GWAG-5-RTGA
The recombinant enzyme with TRDs C and E was purified and used in the nuclease assay. There are minor differences in amino acid sequence between members of ST425-1.

Recombinant S.SauCE CC425-1 GWAG-5-RTGA

MSNTQKKNVPELRFPGFEGEWEEKKVGELLEFKNGLNKGKEYFGSGSSIVNFKDVFNNRSINTNNL
TGKVNVNSKELKNYSVEKGDVFFTRTSEVIGEIGYPSVILNDPENTVFSGFVLRGRPKSGIDLINN
NFKRYVFFTNSFRKEMITKSSMTTRALTSGTAINRMKVIYPVSAKEQKKIGDFFSKLDRQIELEEQ
KLELLQQQKKGYMQKIFSQELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKDAITNGSYDFYVR
SPIVYKINTFSYEGEAILTVGDGVGVGKVFHYVNGKFDYHQRVYKISDFKNYYGLLLFYYFSQNFL
KETKKYSAKTSVDSVRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTKIQKQVIELLKQRKKSLLQ
KMFIPGGSHHHHHH

Wild type S.SauCE

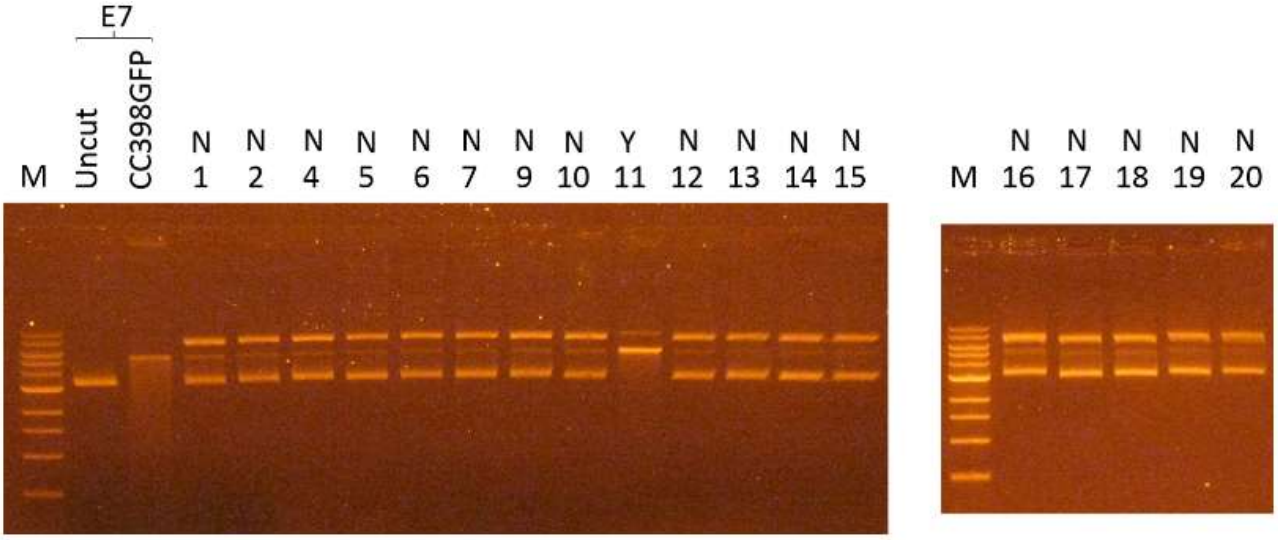
MSNTQTKNVPELRFPGFEGEWEEKQVGELLEFKNGLNKGKEYFGSGSSIVNFKDVFNNRSINTNNL
TGKVNVNSKELKNYSVEKGDVFFTRTSEVIGEIGYPSVILNDPENTVFSGFVLRGRPKSGIDLINN
NFKRYVFFTNSFRKEMITKSSMTTRALTSGTAINKMKVIYPVSAKEQKKIGDFFSKLDRQIELEEQ
KLELLQQQKKGYMQKIFTQELRFKDENGNDYPEWEETTIKEIAQINTGKKDTKDAITNGSYDFYVR
SPIVYKINTFSYEGEAILTVGDGVGVGKVFHYVNGKFDYHQRVYKISDFKNYYGLLLFYYFSQNFL
KETKKYSAKTSVDSVRKDMVANMKVPRPIYIEQEKIGQFIKKVDNKIKIQKQVIELLKQRKKALLQ
KMFI*



1- marker 2- soluble cell extract 3- flow through from Nickel column 4- wash from Nickel column
5- eluate from Nickel column 6-14 Fractions from gel filtration column 15- CC398-1 purified protein marker

S.SauCE
ST425-1 GWAG-5-RTGA

DNA cleavage assay.



S.SauJP
CC51 GGA-6-CCT
This MTase was used in the SMRT analysis of *E. coli* ER2796.
There are minor variations in the sequences of the S subunits in CC51.
Recombinant S.SauJP CC51-1

MSNTQKKNVPELRFPGFEGEWEEKKLGDLIKVNSGKDYKHLEKGDIPVYGTGGYMTSVSEPLSEID
AVGIGRKGTINKPYLLEAPFWTVDTLFYCTPKKETDILFILSLFRKINWKVYDESTGVPSLSKQTI
NKINRFVPSNKEQQKIGEFFIKLDRQIELEEQKLELFQQQKKGYMQKIFSQELRFKDESGNDYPDW
EEKELGEVADRVIRKNKNFESKKPLTISGQLGLIDQTEYFSKSVSSKNLENYTLIKNGEFAYNKSY
SNGYPLGAIKRLTRYDSGVLSSLYICFSIKSEMSKDFMEAYFDSTHWYREVSGIAVEGARNHGLLN
ISVNDFFTILIKYPSLEEQRKIGDFFIKLDRQIELEEQKLELLQQRKKALLKSMLIPGGSHHHHHH

Wild Type S.SauJP

MSNTQTKNVPELRFPGFEGEWEEKKLEDIIKVNSGKDYKHLDKGDIPVYGTGGYMTSVSEPLSEID
AVGIGRKGTINKPYLLEAPFWTVDTLFYCTPKKETDILFILSLFRKINWKVYDESTGVPSLSKQTI
NKINRFVPTNKEQQKIGKFFSKLDRQIELEEQKLELFQQQKKGYMQKIFSQELRFKDESGNDYPDW
EEKELGEVADRVIRKNKNFESKKPLTISGQLGLIDQTEYFSKSVSSKNLENYTLIKNGEFAYNKSY
SNGYPLGAIKRLTRYDSGVLSSLYICFSIKSEMSKDFMEAYFDSTHWYREVSGIAVEGARNHGLLN
ISVNDFFTILIKYPSLEEQRKIGDFFIKLDRQIELEEQKLELLQQRKKALLKSMLI

Reports for Job Dryden_J_P_MODs

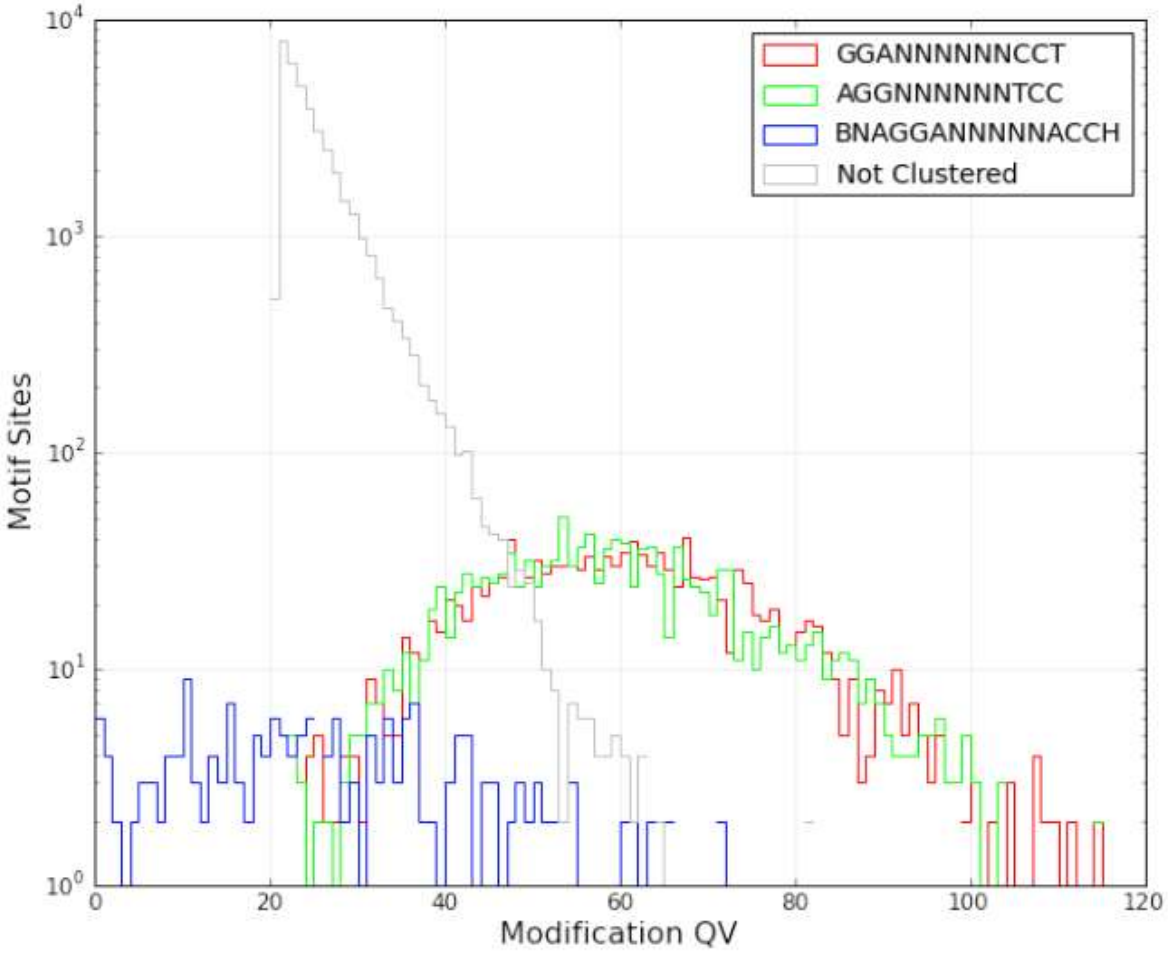
PACIFIC BIOSCIENCES

SMRT Cells: 1 Movies: 1

Motif Summary

Motifs	Modified Position	Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
GGANNNNNNCCT	3	m6A	98.1%	1340	1366	62.31	39.31	AGGNNNNNNTCC
AGGNNNNNNTCC	1	m6A	97.58%	1333	1366	61.04	39.18	GGANNNNNNCCT
BNAGGANNNNNACCH	3	m6A	46.26%	99	214	47.92	39.46	

S.SauJP
CC51 GGA-6-CCT



S.SauCL-EGFP
CC45-1 GWAG-6-TAAA

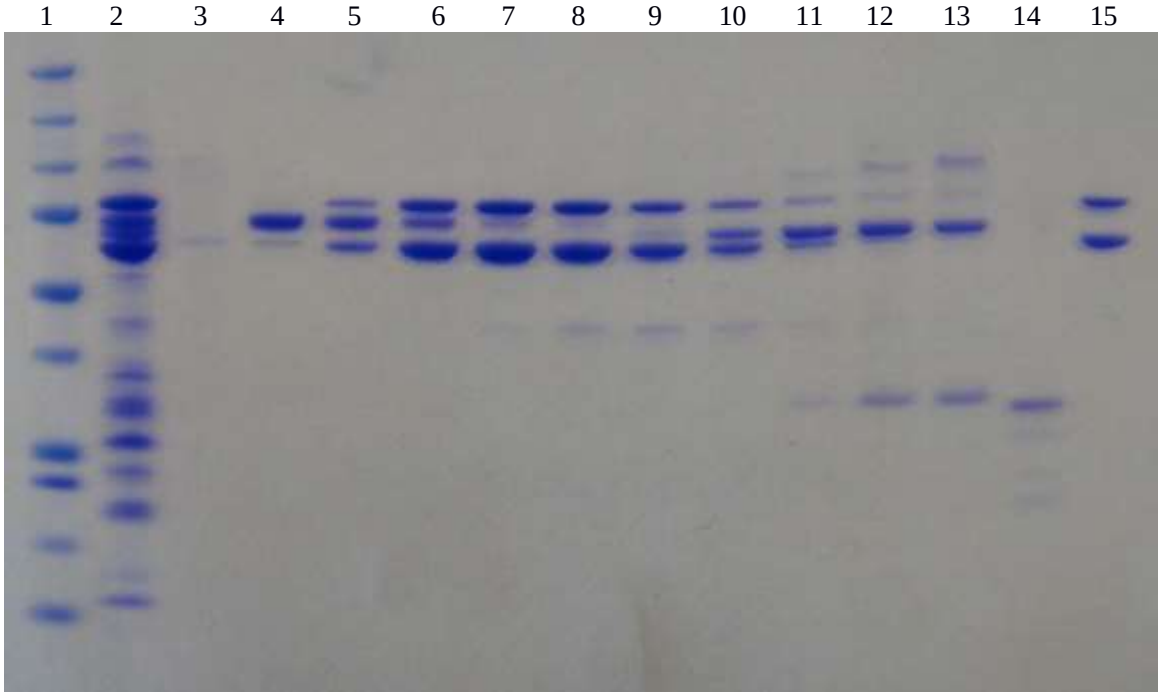
Two separate clones of pSauCL-EGFP encode residue 167 as Lysine (K) instead of arginine (R), but this does not affect the specificity as identical sequences are recognised in Trd C from CC30-1.

S.SauCL-EGFP "Expected" sequence

MSNTQKKNVPELRFPGFEGEWEEKKVGELLEFKNGLNKGKEYFGSGSSIVNFKDVFNNRSINTNNL
TGKVNVNSKELKNYSVEKGDVFFTRTSEVIGEIGYPSVILNDPENTVFSGFVLRGRPKSGIDLINN
NFKRYVFFTNSFRKEMITKSSMTTRALTSGTAIN**R**MKVIYPVSAKEQKKIGDFFSKLDRQIELEEQ
KLELLQQQKKGYMQKIFSQELRFKDENGNDYPNWRTIELKNILENIVDNRGKTPDNAPSEKYPLLE
VNALGYYRPAYIKVSKFVSENTYNNWFREHLKENDILFSTVGNTGIVSLMDNYKAVIAQNIVGLRV
NNNNLPSFIYYMLSYKGNQKKIKRIQMGAVQPSVKVSQFKFIKYLVPIKDEQEKVAKLLIEIDKLV
NKQLIKIELLQQRKKALLKSMFIGSMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYG
KLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNY
KTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIE
DGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYK
HHHHHH

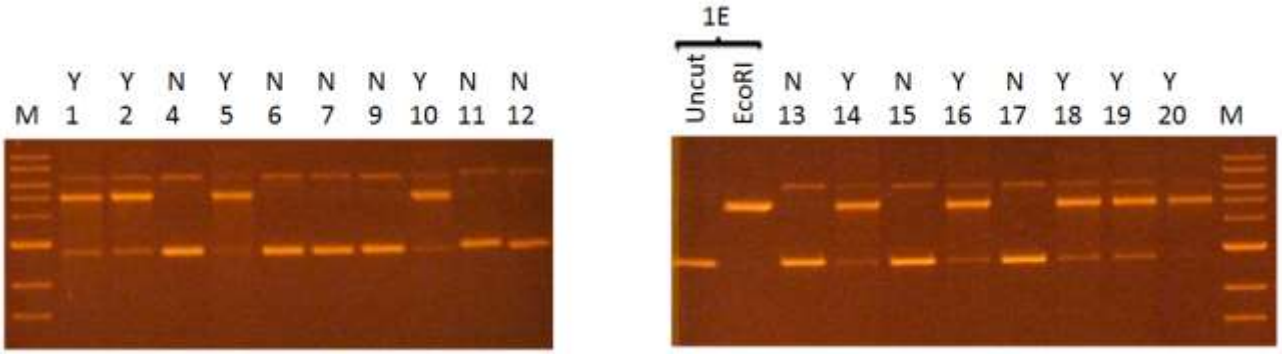
S.SauCL-EGFP "Actual" sequence

MSNTQKKNVPELRFPGFEGEWEEKKVGELLEFKNGLNKGKEYFGSGSSIVNFKDVFNNRSINTNNL
TGKVNVNSKELKNYSVEKGDVFFTRTSEVIGEIGYPSVILNDPENTVFSGFVLRGRPKSGIDLINN
NFKRYVFFTNSFRKEMITKSSMTTRALTSGTAIN**K**MKVIYPVSAKEQKKIGDFFSKLDRQIELEEQ
KLELLQQQKKGYMQKIFSQELRFKDENGNDYPNWRTIELKNILENIVDNRGKTPDNAPSEKYPLLE
VNALGYYRPAYIKVSKFVSENTYNNWFREHLKENDILFSTVGNTGIVSLMDNYKAVIAQNIVGLRV
NNNNLPSFIYYMLSYKGNQKKIKRIQMGAVQPSVKVSQFKFIKYLVPIKDEQEKVAKLLIEIDKLV
NKQLIKIELLQQRKKALLKSMFIGSMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYG
KLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNY
KTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIE
DGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYK
HHHHHH



1- marker 2- Nickel column eluate 3-14 Fractions from gel filtration column
15- CC5-1 purified protein marker

S.SauCL-EGFP
CC45-1 GWAG-6-TAAA
DNA cleavage assay.



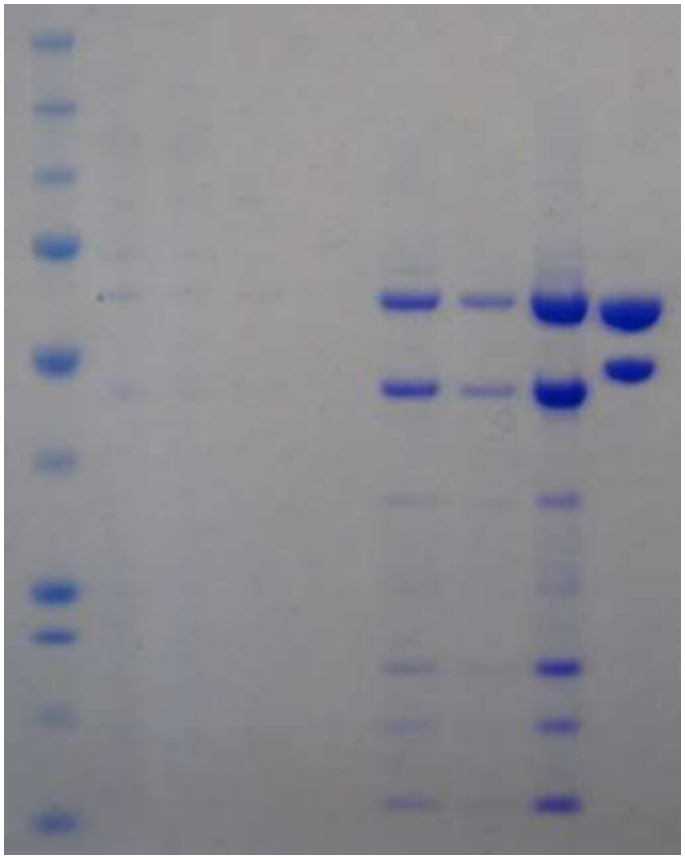
S.SauOE
CC15
Recombinant S.SauOE CC15-1 CAAC-5-RTGA

MSNTQKKNVPELRFPGFEGEWEEKKLGEVGTFTSGGTPLKSKSEYWNGDIPWITTGDIHNIKRENI
TNFITEKGLNESSAKLITNEAILIAMYGQGKTRGMSAILNFEATTNQACAIYQTNQNINFVFQYFQ
KLYEFLRSLSNEGSQKNLSLSLLKEITLNYPNEQEQKKIGDFFSKLDRQIELEEQKLELLQQQKKG
YMQKIFSQELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKDAITNGSYDFYVRSPIVYKINTFS
YEGEAILTVGDGVGVGKVFHYVNGKFDYHQRVYKISDFKNYYGLLLFYYFSQNFLKETKKYSAKTS
VDSVRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTKIQKQVIELLKQRKKSLLQKMFIPGGSHHH
HHH

Wild Type S.SauOE

MSNKQKKNVPELRFPGFEGEWEEKKLGEVGTFTSGGTPLKSKSEYWNGDIPWITTGDIHNIKRENI
TNFITEKGLNESSAKLITNEAILIAMYGQGKTRGMSAILNFEATTNQACAIYQTNQNINFVFQYFQ
KLYEFLRSLSNEGSQKNLSLSLLKEITLNYPNEQEQKKIGDFFSKLDRQIELEEQKLELLQQQKKG
YMQKIFSQELRFKDENGNDYPEWEETTIKEIAQINXGKKDTKDAITNGSYDFYVRSPIVYKINTFS
YEGEAILTVGDGVGVGKVFHYVNGKFDYHQRVYKISDFKNYYGLLLFYYFSQNFLKETKKYSAKTS
VDSVRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTKIQKQVIELLKQRKKALLQKMFI

1 2 3 4 5 6 7 8 9



1- marker 2- soluble cell extract 3- Nickel column flow through
4- Nickel column wash 1 5- Nickel column wash 2 6- Nickel column eluate
7- eluate after PD10 desalting 8- Final concentrated protein
9- CC398-1 purified protein marker

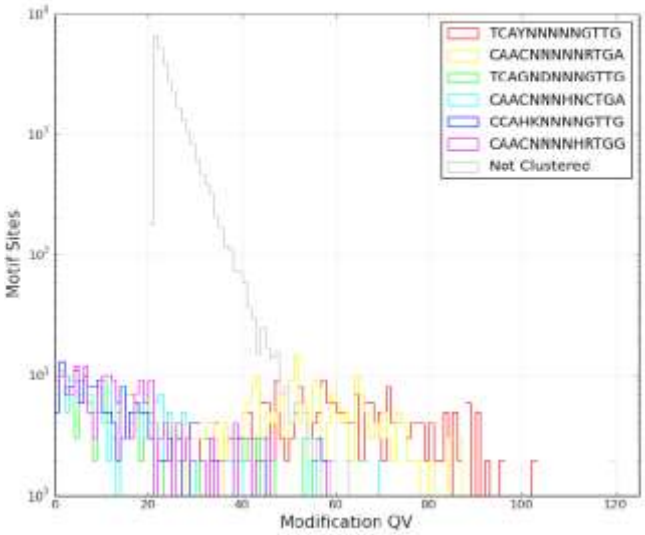
Although purified, this MTase was only used in SMRT sequencing.

S.SauOE
CC15
Recombinant S.SauOE CC15-1 CAAC-5-RTGA

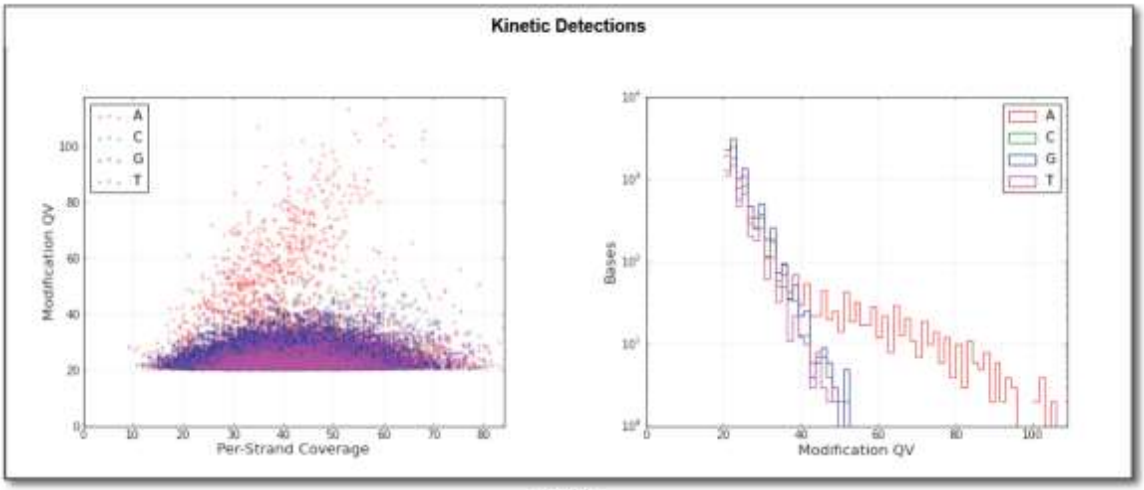
Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
TCAYNNNNNGTTG	3	m6A	96.67	261	270	63.4	38.3	CAACNNNNNRTGA
CAACNNNNNRTGA	3	m6A	92.22	249	270	57.7	36.1	TCAYNNNNNGTTG
TCAGNDNNNGTTG	3	m6A	23.86	47	197	48.5	38.3	CAACNNNHNCTGA
CAACNNNHNCTGA	3	m6A	17.77	35	197	49.9	38.6	TCAGNDNNNGTTG
CCAHKNNNNGTTG	3	m6A	16.93	32	189	50.6	40.4	
CAACNNNNHRTGG	3	m6A	16.28	35	215	45.3	38.2	
Not Clustered	0		0.02	2201	9115988	34.9	47.5	

Modification QV Histogram By Motif



Kinetic Detections



Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
TCAYNNNNNGTTG	3	m6A	96.67	261	270	63.4	38.3	CAACNNNNNRTGA
CAACNNNNNRTGA	3	m6A	92.22	249	270	57.7	36.1	TCAYNNNNNGTTG
TCAGNDNNNGTTG	3	m6A	23.86	47	197	48.5	38.3	CAACNNNHNCTGA
CAACNNNHNCTGA	3	m6A	17.77	35	197	49.9	38.6	TCAGNDNNNGTTG
CCAHKNNNNGTTG	3	m6A	16.93	32	189	50.6	40.4	
CAACNNNNHRTGG	3	m6A	16.28	35	215	45.3	38.2	
Not Clustered	0		0.02	2201	9115988	34.9	47.5	

S.SauJQ
CC59

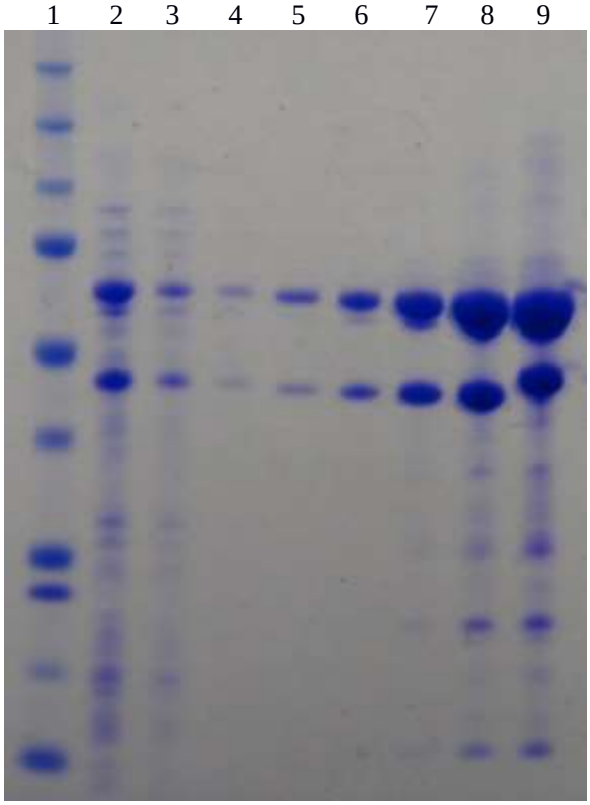
This enzyme was purified and analysed using the ATPase assay as both TRD specificities were known and the DNA cleavage assay showed cutting of all plasmids.

Recombinant S.SauJQ CC59-1 GGA-6-RTGT

MSNTQKKNVPELRFPGFEGEWEEKKLGDLIKVNSGKDYKHLEKGDIPVYGTGGYMTSVSEPLSEID
AVGIGRKGTINKPYLLEAPFWTVDTLFYCTPKKETDILFILSLFRKINWKVYDESTGVPSLSKQTI
NKINRFVPSNKEQQKIGEFFIKLDRQIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGEDYSEW
EERRFADIFKFHNKLRKPIKENLRVKGSYPYYGATGIIDYVDDFIFDGNYLLIGEDGANIITRSAP
LVYLVNGKFWVNNHAHILSPLNGNIQYLYQVAELVNYEKYNTGTAQPKLNIQNLKIINVVISTNLE
EQQKIGSFLSKLDRQIDLEEQKLELLQQRKKALLKSMFVPGGSHHHHHH

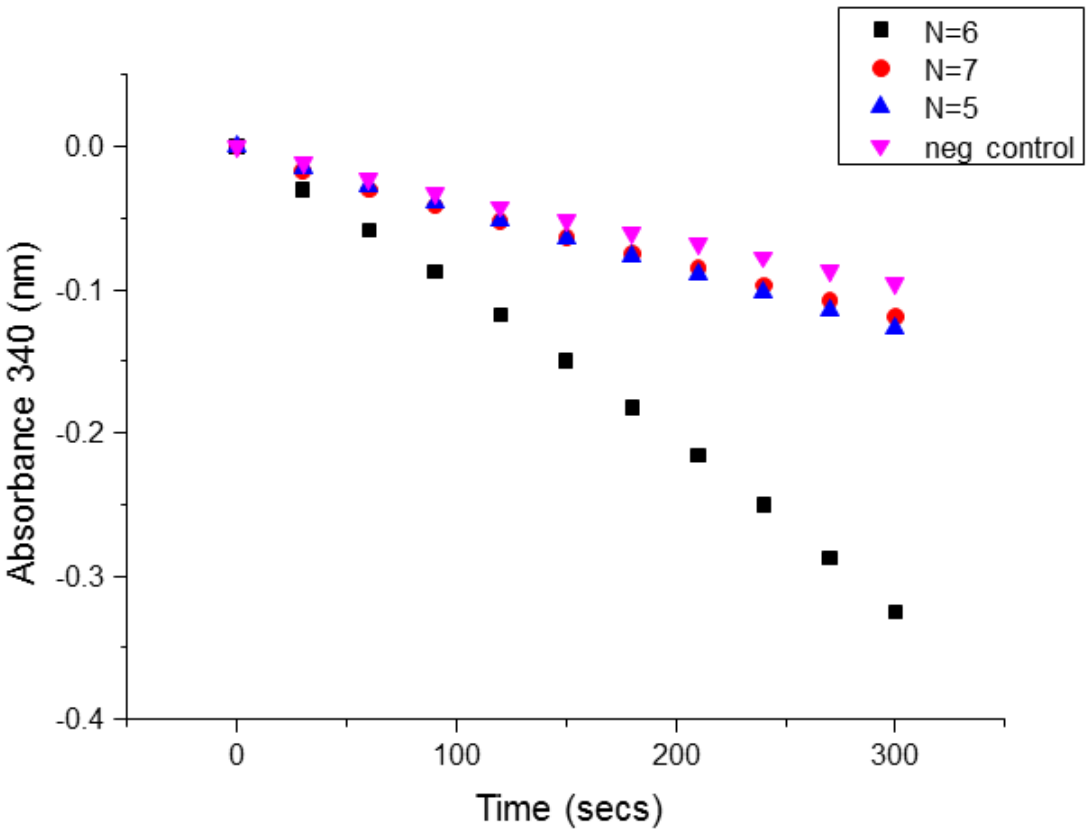
Wild type S.SauJQ

MSNTQKKNVPELRFPEFEGEWEERKLGDLIKVNSGKDYKHLDKGDIPVYGTGGYMTSVSEPLSEID
AVGIGRKGTINKPYLLEAPFWTVDTLFYCTPEKEADILFILSLFRKINWKLYDESTGVPSLSKQTI
NKINRLVPTNKEQQKIGEFFSKLDRQIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGEDYSEW
EERRFADIFKFHNKLRKPIKENLRVKGSYPYYGATGIIDYVDDFIFDGNYLLIGEDGANIITRSAP
LVYLVNGKFWVNNHAHILSPLNGNIQYLYQVAELVNYEKYNTGTAQPKLNIQNLKIINVVISTNLE
EQQKIGSFLSKLDRQIDLEEQKLELLQQRKKALLKSMFV*



1- marker 2- soluble cell extract 3- Nickel column flow through
4- Nickel column wash 1 5- Nickel column wash 2 6- Nickel column eluate
7- eluate after conc. and PD10 desalting
8- final concentrated protein 9- XE purified protein marker

S.SauJQ
CC59
Recombinant S.SauJQ CC59-1 GGA-6-RTGT
ATPase assay shows that N=6.



Oligonucleotide name	DNA sequence (5' to 3')
JQ5for	AGATGATGTCATCAATGCGGATTACAGTGTGCCCTATACGATATAA
JQ5rev	TTATATCGTATAGGGCACACTGTAATCCGCATTGATGACATCATCT
JQ6for	AGATGATGTCATCAATGCGGATTGACAGTGTGCCCTATACGATATAA
JQ6rev	TTATATCGTATAGGGCACACTGTCAATCCGCATTGATGACATCATCT
JQ7for	AGATGATGTCATCAATGCGGATTAGACAGTGTGCCCTATACGATATAA
JQ7rev	TTATATCGTATAGGGCACACTGTCTAATCCGCATTGATGACATCATCT

S.SauRQ
CC72

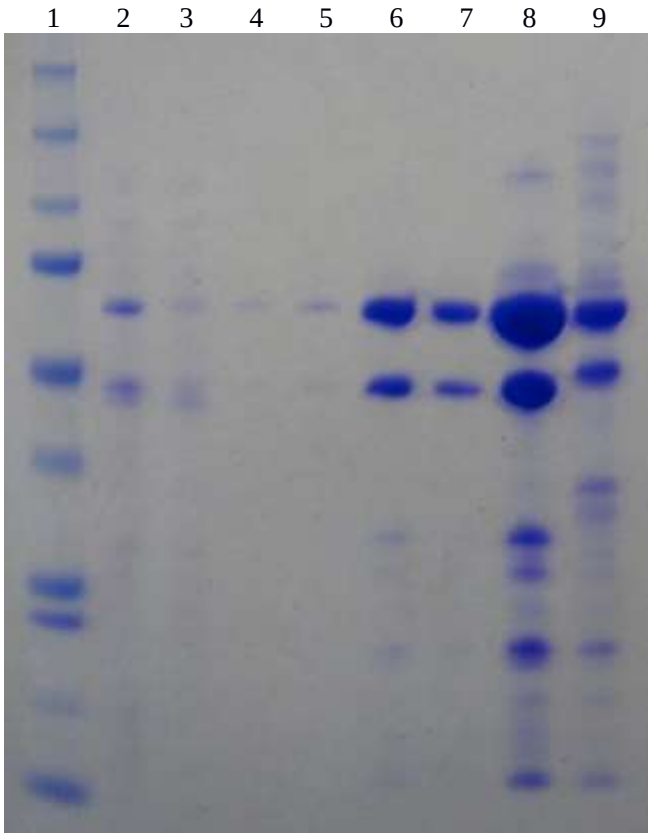
This enzyme was purified and analysed using the ATPase assay as both TRD specificities were known and the DNA cleavage assay showed cutting of all plasmids.

Recombinant S.SauRQ CC72-1 GARA-6-RTGT

MSNTQKKNVPELRFPGFEGEWEEKKLGEVAKIYDGTHQTPKYTNEGIKFLSVENIKTLNSSKYISE
EAFEKEFKIRPEFGDILMTRIGDIGTPNIVSSNEKFAYYVSLALLKTKNLNSYFLKNLILSSSIQN
ELWRKTLHVAFPKKINKNEIGKIKINYPKKQEQQKIGQFFSKLDRQIELEEQKLELLQQQKKGYMQ
KIFSQELRFKDENGEDYSEWEERRFADIFKFHNKLRKPIKENLRVKGSYPYYGATGIIDYVDDFIF
DGNYLLIGEDGANIITRSAPLVYLVNGKFWVNNHAHILSPLNGNIQYLYQVAELVNYEKYNTGTAQ
PKLNIQNLKIINVVISTNLEEQQKIGSFLSKLDRQIDLEEQKLELLQQRKKALLKSMFVPGGSHHH
HHH

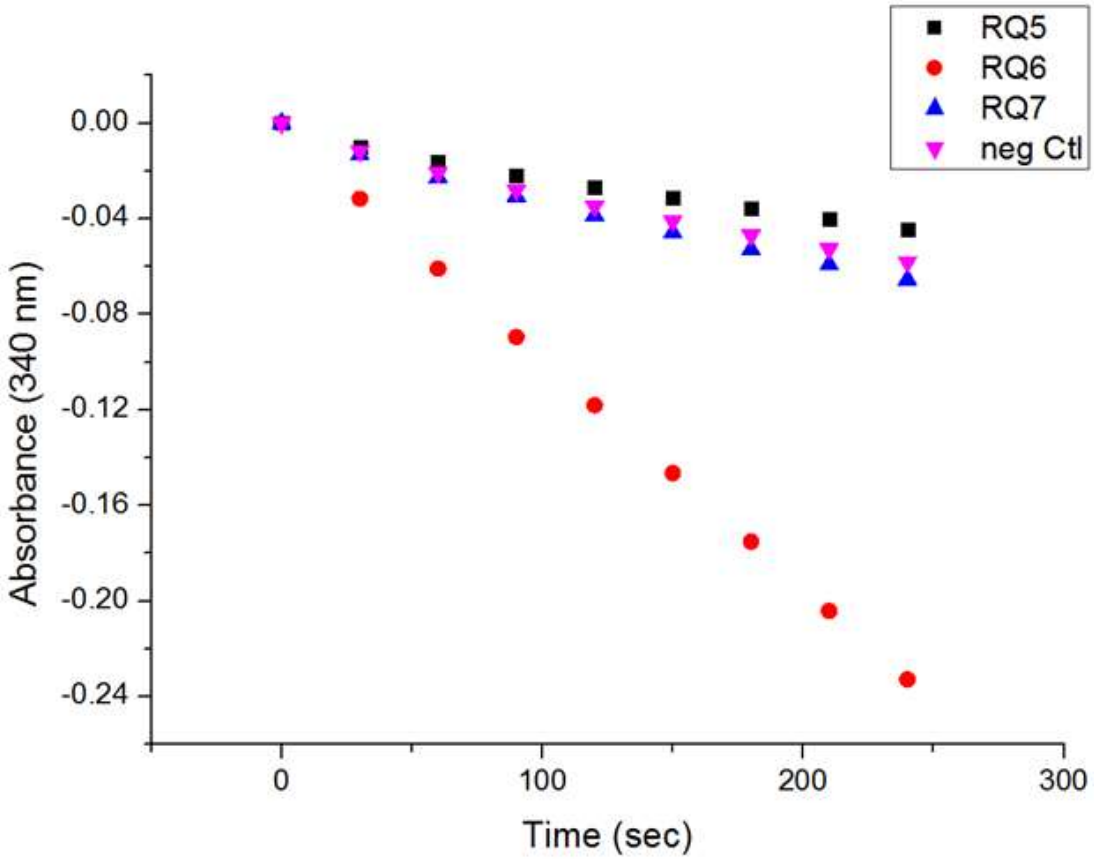
Wild type S.SauRQ

MSNTQKKNVPELRFPGFEGEWEEKKLGEVAKIYDGTHQTPKYTNEGIKFLSVENIKTLNSSKYISE
EAFEKEFKIRPEFGDILMTRIGDIGTPNIVSSNEKFAYYVSLALLKTKNLNSYFLKNLILSSSIQN
ELWRKTLHVAFPKKINKNEIGKIKINYPKKQEQQKIGQFFSKLDRQIELEEQKLELLQQQKKGYMQ
KIFSQELRFKDENGNDYPEWEERRFADIFKFHNKLRKPIKENLRVKGSYPYYGATGIIDYVDDFIF
DGNYLLIGEDGANIITRSAPLVYLVNGKFWVNNHAHILSPLNGNIQYLYQVAELVNYEKYNTGTAQ
PKLNIQNLKIISVVISTNLEEQQKIGSFLSKLDRQIDLEEQKLELLQQRKKALLKSMFV*



1- marker 2- soluble cell extract 3- Nickel column flow through
4- Nickel column wash 1 5- Nickel column wash 2 6- Nickel column eluate
7- eluate after PD10 desalting 8- Final protein after concentration
9- NP purified protein as marker

S.SauRQ
CC72
Recombinant S.SauRQ GARA-6-RTGT



N=6 shows activity.

Oligonucleotide name	DNA sequence (5' to 3')
RQ5for	AGATGATGGAATCAATGCGAGATTCCAGTGTGCCCTATACGATATAA
RQ5rev	TTATATCGTATAGGGCACACTGGAATCTCGCATTGATTCCATCATCT
RQ6for	AGATGATGGAATCAATGCGAGATGTCCAGTGTGCCCTATACGATATAA
RQ6rev	TTATATCGTATAGGGCACACTGGACATCTCGCATTGATTCCATCATCT
RQ7for	AGATGATGGAATCAATGCGAGATGTACCAGTGTGCCCTATACGATATAA
RQ7rev	TTATATCGTATAGGGCACACTGGTACATCTCGCATTGATTCCATCATCT

S.SauJS

This second enzyme from CC72 was purified and analysed using the ATPase assay. There are minor variations between S subunit sequences in CC72-2.

CC72

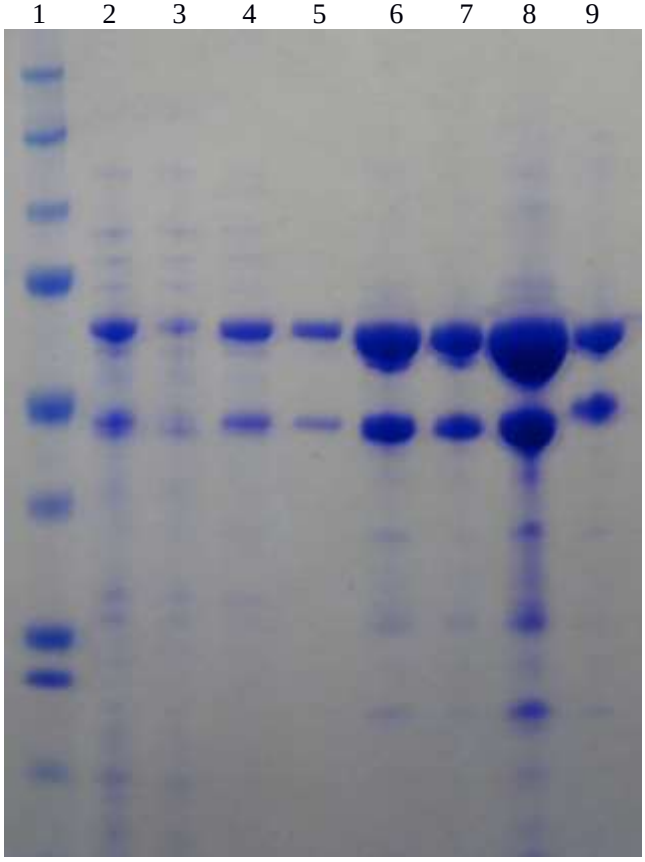
Recombinant S.SauJS CC72-2 GGA-7-TGC

MSNTQKKNVPELRFPGFEGEWEEKKLGDLIKVNSGKDYKHLEKGDIPVYGTGGYMTSVSEPLSEID
AVGIGRKGTINKPYLLEAPFWTVDTLFYCTPKKETDILFILSLFRKINWKVYDESTGVPSLSKQTI
NKINRFVPSNKEQQKIGEFFIKLDRQIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPDW
TNERLGEVTTVTMGQSPKSVNYTDNSNDTVLIQGNADIENGLINPRIYTREVTKLIQKDEIILTVR
APVGKLAMAQINACIGRGVCSIKGDKFLYYFLEWFATQNKWIRFSQGSTFESISGNDIRNIHIKIP
VEDERTKIIKLLNSLDVLNSKTDLKIQNLKQRKQSLLQKIFVPGGSHHHHHH

Wild Type S.SauJS

MSNTQKKNVPELRFPEFEGEWEEKQLGNIIKVNSGKDYKHLDKGDIPVYGTGGYMTSVSEPLSEID
AVGIGRKGTINKPYLLEAPFWTVDTLFYCTPKKETDILFILSLFRKINWKVYDESTGVPSLSKQTI
NKINRFVPTNKEQQKIGKFFSKLDRQIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPDW
TNERLGEVTTVTMGQSPKSVNYTDNSNDTVLIQGNADIENGLINPRIYTREVTKLIQKDEIILTVR
APVGKLAMAQINACIGRGVCSIKGDKFLYYFLEWFATQNKWIRFSQGSTFESISGNDIRNIHIKIP
VEDERTKIIKLLNSLDVLNSKTDLKIQNLKQRKQSLLQKIFV

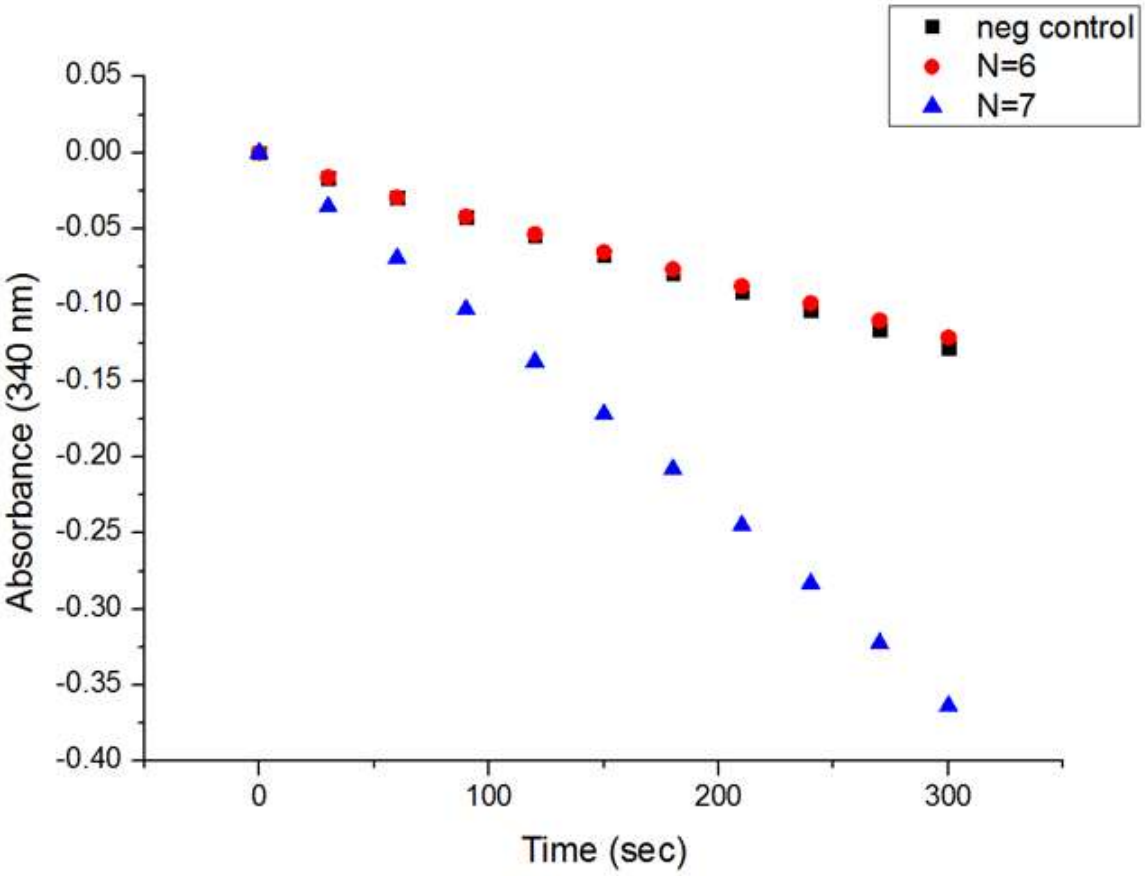
1 2 3 4 5 6 7 8 9



1- marker 2- soluble cell extract 3- Nickel column flow through
4- Nickel column wash 1 5- Nickel column wash 2 6- Nickel column eluate
7- eluate after PD10 desalting 8- final protein after concentration
9- NP purified protein as marker

S.SauJS
CC72
Recombinant S.SauJS CC72-2 GGA-7-TGC
N=7 shows activity.

Oligonucleotide name	DNA sequence (5' to 3')
JS6for	AGATGATGGCATCAATGCGGATTACATTGCGCCCTATACGATATAA
JS6rev	TTATATCGTATAGGGCGCAATGTAATCCGCATTGATGCCATCATCT
JS7for	AGATGATGGCATCAATGCGGATTGACATTGCGCCCTATACGATATAA
JS7rev	TTATATCGTATAGGGCGCAATGTCAATCCGCATTGATGCCATCATCT



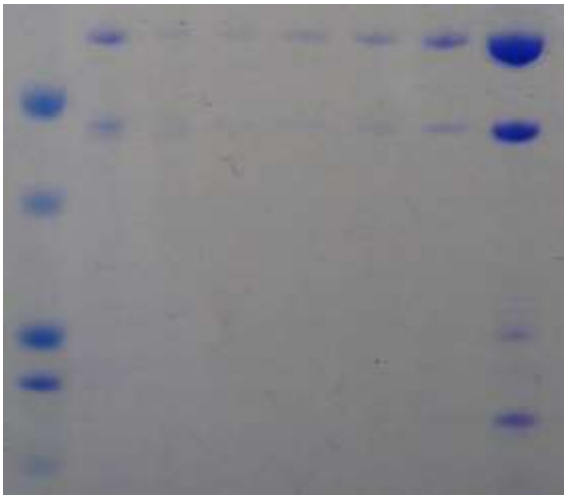
S.SauTU
CC75
Recombinant S.SauTU CC75-1 CAAG-5-RTC

MSNTQKKNVPELRFPGFEGEWEEKELGEIFQIISGSTPLKSNKEFYENGNINWVKTTDLNNSKVTH
SKEKITEYAMKSLKLKLVPKNSVLIAMYGGFNQIGRTGLLKIDATINQAISALLMNHETNPEFIQA
FLNYQVKGWKRYAASSRKDPNITKKDIEQFKVPYVSINEQQKIGEFFSKIDHQIELEEQKLELLQQ
QKKGYMQKIFSQELRFKDENGEDYPDWEVTTIQNITKYTSSKKSSNQYADKDNSKGYPVYDAVQEI
GKDSNYDIEESYISILKDGAGVGRLNLRPGKSSVIGTMGYIQSNNVDIEFLYYRMKVVDFKKYIIG
STIPHLYFKDYSKETLYIPSSIQEQAKIGMFISNLDKLIENKNLKLNCLKQLKQGLLQSMFIPGGS
HHHHHH

Wild type S.SauTU

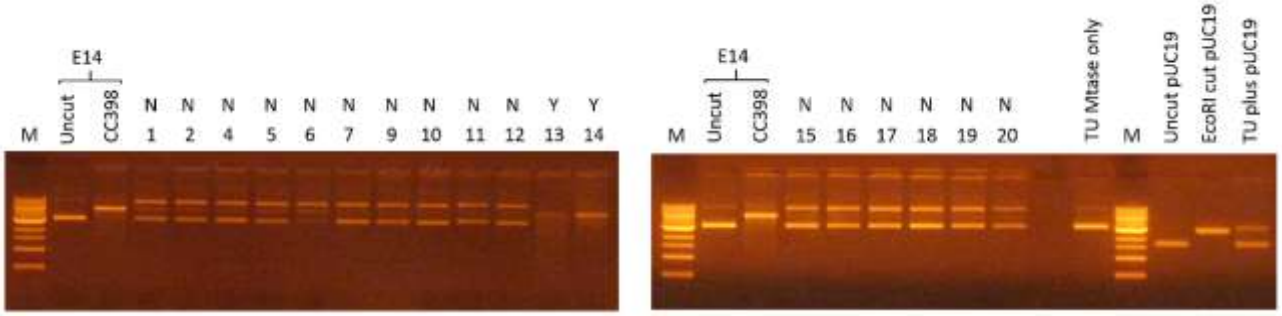
MSNTQTKNVPELRFPGFEGEWEEKELGEIFQIISGSTPLKSNKEFYENGNINWVKTTDLNNSKVTH
SKEKITEYAMKSLKLKLVPKNSVLIAMYGGFNQIGRTGLLKIDATINQAISALLMNHETNPEFIQA
FLNYQVKGWKRYAASSRKDPNITKKDIEQFKVPYVSINEQQKIGEFFSKIDHQIELEEQKLELLQQ
QKKGYMQKIFSQELRFKDENGEDYPDWEVTTIQNITKYTSSKKSSNQYADKDNSKGYPVYDAVQEI
GKDSNYDIEESYISILKDGAGVGRLNLRPGKSSVIGTMGYIQSNNVDIEFLYYRMKVVDFKKYIIG
STIPHLYFKDYSKETLYIPSSIQEQAKIGMFISNLDKLIENKNLKLNCLKQLKQGLLQSMFI

1 2 3 4 5 6 7 8



1- marker 2- soluble cell extract 3- Nickel column flow through 4- Nickel column wash 1 5- Nickel column wash 2 6- Nickel column eluate 7- eluate after conc. and PD10 desalting 8- final protein after concentration

DNA cleavage assay.



S.SauVW
CC75
Recombinant S.SauVW CC75-2 CNGA-7-TTYG

MSNTQKKNVPELRFPGFEGEWEEKELRELRNPKDKYSYTGGPFGSDLKKSDYTTDGIQIIQLQNIG
DGYFYNSNKVFTSNEKAEVLKSCNVFPGDIVIAKMADPIARAAIVPDNNIGKYLMASDGIRLSVDT
VHFNTKFVLECINRKSFRKKVEDNSSGSTRMRIGLSTLGSLTLKTTTLKEQQKIGQFFSKLDRQIE
LEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPDWEEKQLGELSQIVRGASPRPIKDPKWFNK
ESDIGWLRISDVTNQNGKIYHLEQKLSIEGQEKTRVLVTTHLLLSIAASIGKPVMNFVKTGVHDGF
LIFLKPKFNLFFMYYWLEYFKDKWSKYGQPGSQVNLNSEIVKSQTLNMPSNHEQEKVGQFFNRNEK
LIELQQEKIMYIKRCKQVLLQKMFIPGGSHHHHHH

Wild Type S.SauVW

MSNTGKMNVPELRFPGFEGEWEEKELRELRNPKDKYSYTGGPFGSDLKKSDYTTDGIQIIQLQNIG
DGYFYNSNKVFTSNEKAEVLKSCNVFPGDIVIAKMADPIARAAIVPDNNIGKYLMASDGIRLSVDT
VHFNTKFVLECINRKSFRKKVEDNSSGSTRMRIGLSTLGSLTLKTTTLKEQQKIGQFFSKLDRQIV
LEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPDWEEKQLGELSQIVRGASPRPIKDPKWFNK
ESDIGWLRISDVTNQNGKIYHLEQKLSIEGQEKTRVLVTTHLLLSIAASIGKPVMNFVKTGVHDGF
LIFLKPKFNLFFMYYWLEYFKDKWSKYGQPGSQVNLNSEIVKSQTLNMPSNHEQEKVGQFFNRNEK
LIELQQEKIMYIKRCKQVLLQKMFI*

Reports for Job Dryden_V_W_MODs

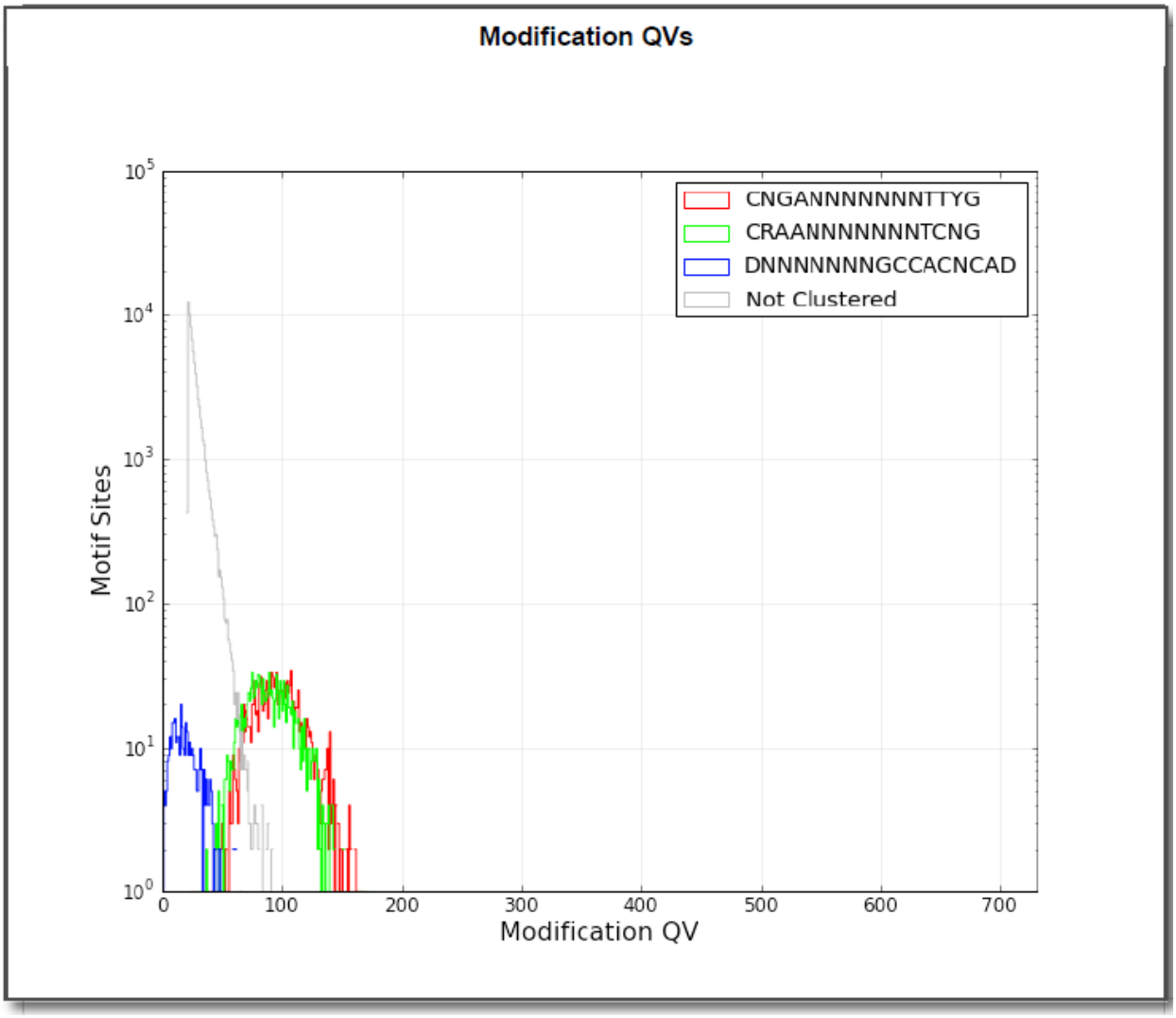
PACIFIC BIOSCIENCES

SMRT Cells: 1 Movies: 1

Motif Summary

Motifs	Modified Position	Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
CNGANNNNNNNTTYG	4	m6A	99.93%	1442	1443	97.87	66.11	CRAANNNNNNNTCNG
CRAANNNNNNNTCNG	4	m6A	99.86%	1441	1443	89.76	63.95	CNGANNNNNNNTTYG
DNNNNNNNGCCACNCAD	9	unknown	19.1%	72	377	38.56	67.49	

S.SauVW
CC75
Recombinant S.SauVW CC75-2 CNGA-7-TTYG

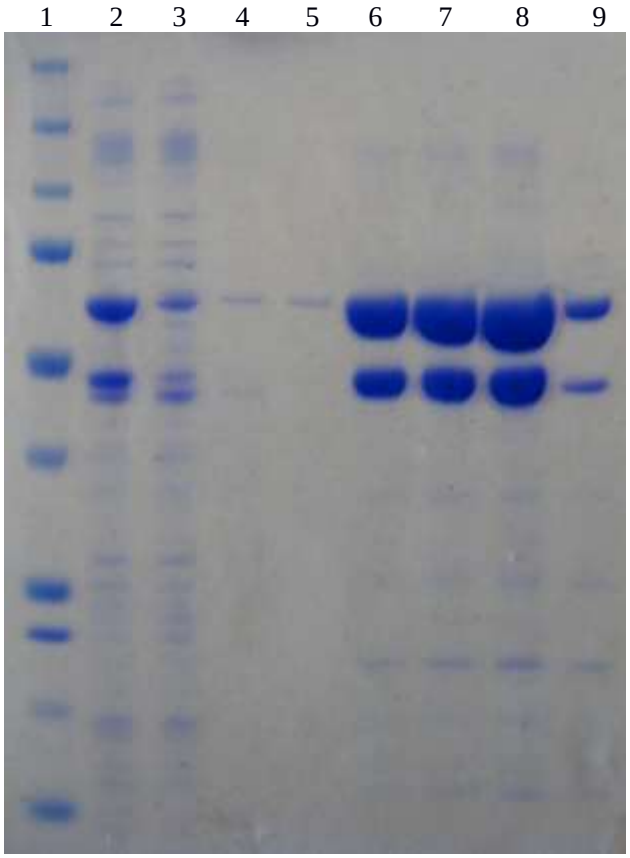


S.SauZW
CC80
Recombinant S.SauZW CC80-2 GAC-6-TTYG

MSNTQKKNVPELRFPGFEGEYSLDIFGNLATNKSEKFNPQNENASIDIELDCIEQNTGRLIKIYNS
KEFSSQKNKFNPQNVLYGKLRPYLNKYYFTKKSGVCSSEIWVLKSTKEDKLLNLFLYYFIQTKRYS
DVASKSAGSKMPRADWGLIENIRVYFPELCEQQKIGQFFSKLDRQIELEEQKLELLQQQKKGYMQK
IFSQELRFKDENGNDYPDWEEKQLGELSQIVRGASPRPIKDPKWFNKESDIGWLRISDVTNQNGKI
YHLEQKLSIEGQEKTRVLVTTHLLLSIAASIGKPVMNFVKTGVHDGFLIFLKPKFNLFFMYYWLEY
FKDKWSKYGQPGSQVNLNSEIVKSQTLNMPSNHEQEKVGQFFNRNEKLIELQQEKIMYIKRCKQVL
LQKMFIPGGSHHHHHH

Wild Type S.SauZW

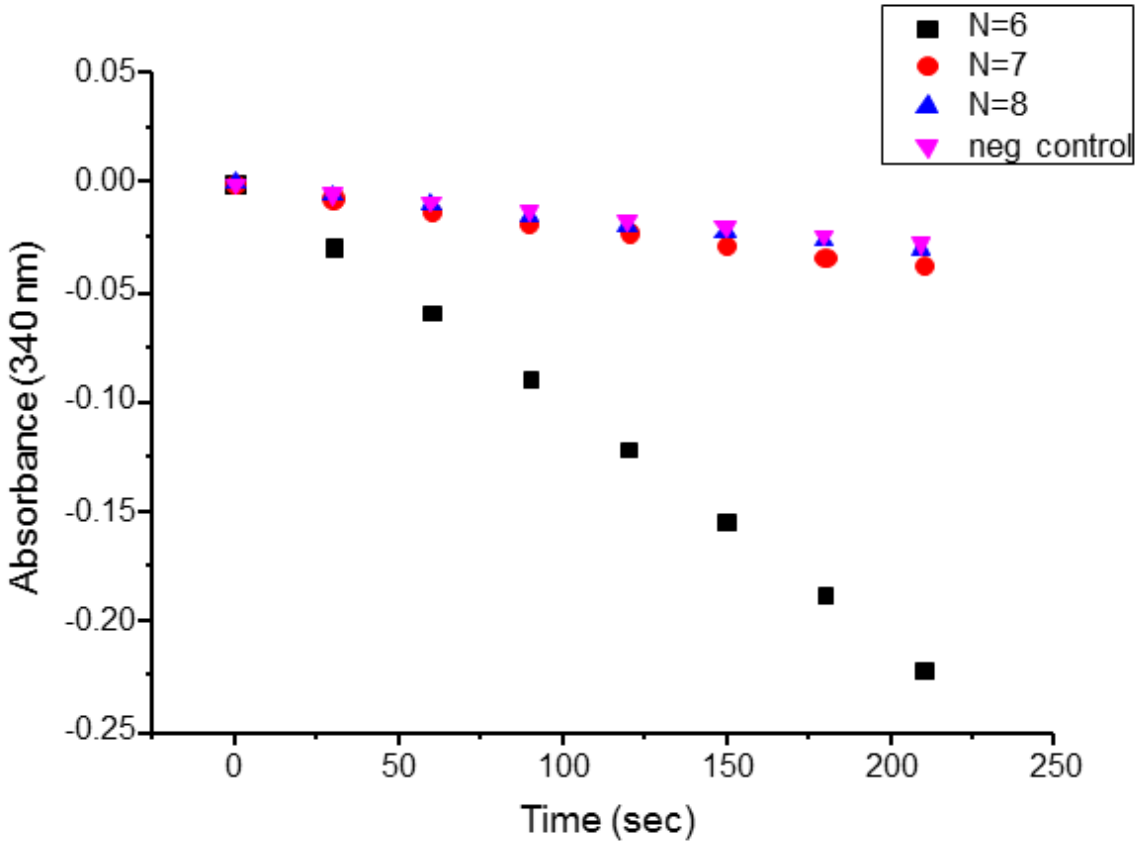
MSNTQTKNVPELRFPGFEGEYSLDIFGNLATNKSEKFNPQNENASIDIELDCIEQNTGRLIKIYNS
KEFSSQKNKFNPQNVLYGKLRPYLNKYYFTKKSGVCSSEIWVLKSTKEDKLLNLFLYYFIQTKRYS
DVASKSAGSKMPRADWGLIENIRVYFPELCEQQKIGQFFSKLDRQIELEEQKLELLQQQKKGYMQK
IFSQELRFKDENGNDYPDWEEKQLGELSQIVRGASPRPIKDPKWFNKESDIGWLRISDVTNQNGKI
YHLEQKLSIEGQEKTRVLVTTHLLLSIAASIGKPVMNFVKTGVHDGFLIFLNPKFNLFFMYYWLEY
FKDKWSKYGQPGSQVNLNTEIVKSQTLNMPSNHEQEKVGQFFNRNEKLIELQQEKIMYLKRRKQVL
LQKMFI*



1- marker 2- soluble cell extract 3- Nickel column flow through
4- Nickel column wash 1 5- Nickel column wash 2 6- Nickel column eluate
7- eluate after conc. and PD10 desalting
8- final protein after concentration 9- CC75-1 purified protein marker

Although purified, this enzyme cut all plasmids in the DNA cleavage assay so the ATPase assay was used as we knew the specificities of the TRDs.

S.SauZW
CC80
Recombinant S.SauZW CC80-2 GAC-6-TTYG
N=6 shows activity.



Oligonucleotide name	DNA sequence (5' to 3')
ZW6for	AGATGATGGAATCAATGCGACTTCCATTTCGGCCCTATACGATATAA
ZW6rev	TTATATCGTATAGGGCCGAAATGGAAGTCGCATTGATTCCATCATCT
ZW7for	AGATGATGGAATCAATGCGACTTCTCATTTCGGCCCTATACGATATAA
ZW7rev	TTATATCGTATAGGGCCGAAATGAGAAGTCGCATTGATTCCATCATCT
ZW8for	AGATGATGGAATCAATGCGACTTCTACATTTCGGCCCTATACGATATAA
ZW8rev	TTATATCGTATAGGGCCGAAATGTAGAAGTCGCATTGATTCCATCATCT

S.SauXf*
ST80
Recombinant S.SauXf* CC80-3 TCTA-6-RTTC

MSNTQKKNVPELRFPGFEGEWEEKQFADFTKINQGLQIAINERKTEYSPELYFYITNEFLRPNSQT
KYFIENPPQSVIANKEDILMTRTGNTGKVVTNVFGAFHNNFFKIKFDKNLYDRLFLVEVLNSSKIQ
NKILSLAGSSTIPDLNHSDFYSISSSYPLLREQQKIGKFFSKLDRQIELEEQKLELLQQQKKGYMQ
KIFSQELRFKDENGEDYPDWKEKKLGDITEQSMYGIGASATRFDSKNIYIRITDIDEKSRKLNYQN
LTTPDELNNKYKLKRNDILFARTGASTGKSYIHKEEKDIYNYYFAGFLIKFKINEQNSPLFIYQFT
LTSKFNKWVKVMSVRSGQPGINSEEYAKLPLVLPNKLEQQKIAKFLDRFDRQIELEKQKIEILQQQ
KKGLLQSMFIPGGSHHHHHH

Wild Type S.SauXf*

MSNTQKKNVPELRFPEFEGEWEEKQFADFTKINQGLQIAINERKTEYSPELYFYITNEFLRPNSQT
KYFIENPPQSVIANKEDILMTRTGNTGKVVTNVFGAFHNNFFKIKFDKNLYDRLFLVEVLNSSKIQ
NKILSLAGSSTIPDLNHSDFYSISSSYPLLREQQKIGKFFSKLDRQIELEEQKLELLQQQKKGYMQ
KIFSQELRFKDENGEDYPDWKEKKLGDITEQSMYGIGASATRFDSKNIYIRITDIDEKSRKLNYQN
LTTPDELNNKYKLKRNDILFARTGASTGKSYIHKEEKDIYNYYFAGFLIKFKINEQNSPLFIYQFT
LTSKFNKWVKVMSVRSGQPGINSEEYAKLPLVLPNKLEQQKIAKFLDRFDRQIELEKQKIEILQQQ
KKGLLQSMFI

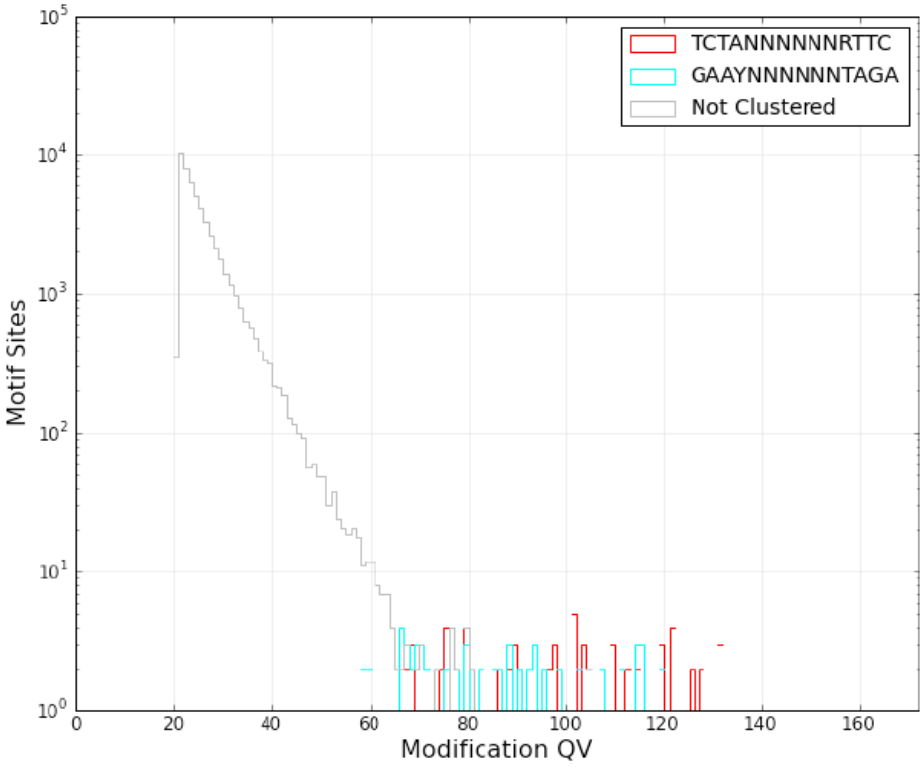
Reports for Job Dryden_X_zeta_MODs

SMRT Cells: 1 Movies: 1

Motif Summary

Motifs	Modified Position	Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
TCTANNNNNNRTTC	4	m6A	100.0%	92	92	96.27	61.85	GAAYNNNNNNTAGA
GAAYNNNNNNTAGA	3	m6A	100.0%	92	92	90.82	60.21	TCTANNNNNNRTTC

Modification QVs

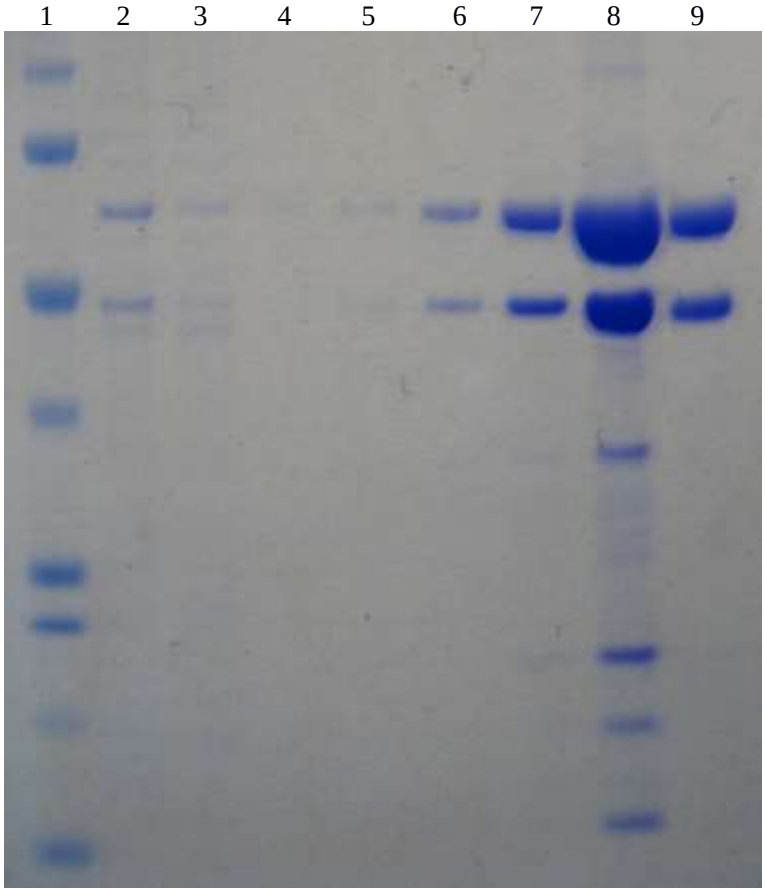


S.Saue*D
CC873
Recombinant S.Saue*D CC873-1 GAG-6-GAT

MSNTQKKNVPELRFPGFEGEWEEKSISSFLKESKIKGSNGSHAKKLTVKLWGKGVVPKKETFKGSD
NTQYYKRKAGQLMYGKLDFLNCAFGIVPDSLNNYESTIDSPSFDFINGDSKFLLERIKLKSFYKKF
GDIANGSRKAKRINQDTFLSLPVFAPKYDEQLRIGEFFSKLDRQIELEEQKLELLQQQKKGYMQKI
FSQELRFKDENSEDYPHWENSKIEKYLKERNERSDKGQMLSVTINSGIIKFSELDRKDNSSKDKSN
YKVVRKNDIAYNSMRMWQGASGRSNYNGIVSPAYTVLYPTQNTSSLFIGYKFKTHRMIHKFKINSQ
GLTSDTWNLKYKQLKNINIDIPVLEEQEKIGDFFKKMDILISKQKIKIEILEKEKQSFLQKMFLPG
GSHHHHHH

Wild Type S.Saue*D

MSNTQKKNVPELRFPGFEGEWEEKSISSFLKESKIKGSNGSHAKKLTVKLWGKGVVPKKETFKGSD
NTQYYKRKAGQLMYGKLDFLNCAFGIVPDSLNNYESTIDSPSFDFINGDSKFLLERIKLKSFYKKF
GDIANGSRKAKRINQDTFLSLPVFAPKYDEQLRIGEFFSKLDRQIELQKQKLELLQQQKKGYMQKI
FSQELRFKDENGEDYPHWENSKIEKYLKERNERSDKGQMLSVTINSGIIKFSELDRKDNSSKNKSN
YKVVRKNDIAYNSMRMWQGASGKSNYNGIVSPAYTVLYPTQNTSSLFIGYKFKTHRMIHKFKINSQ
GLTSDTWNLKYKQLKNINIDIPVLEEQEKIGDFFKKMDILISKQKIKIEILEKEKQSFLQKMFL*



1- marker 2- soluble cell extract 3- Nickel column flow through
4- Nickel column wash 1 5- Nickel column wash 2 6- Nickel column eluate
7- eluate after PD10 desalting and concentration
8- Final concentrated protein 9- CC398-1 purified protein marker

S.Saue*D
CC873-1
Recombinant S.Saue*D GAG-6-GAT

e*D clearly digests pUC19 so the ATPase assay was used as we knew the specificities of both TRDs.

Likely site: GAG-N_x-GAT

GAG-4-GAT 2 sites in pUC19

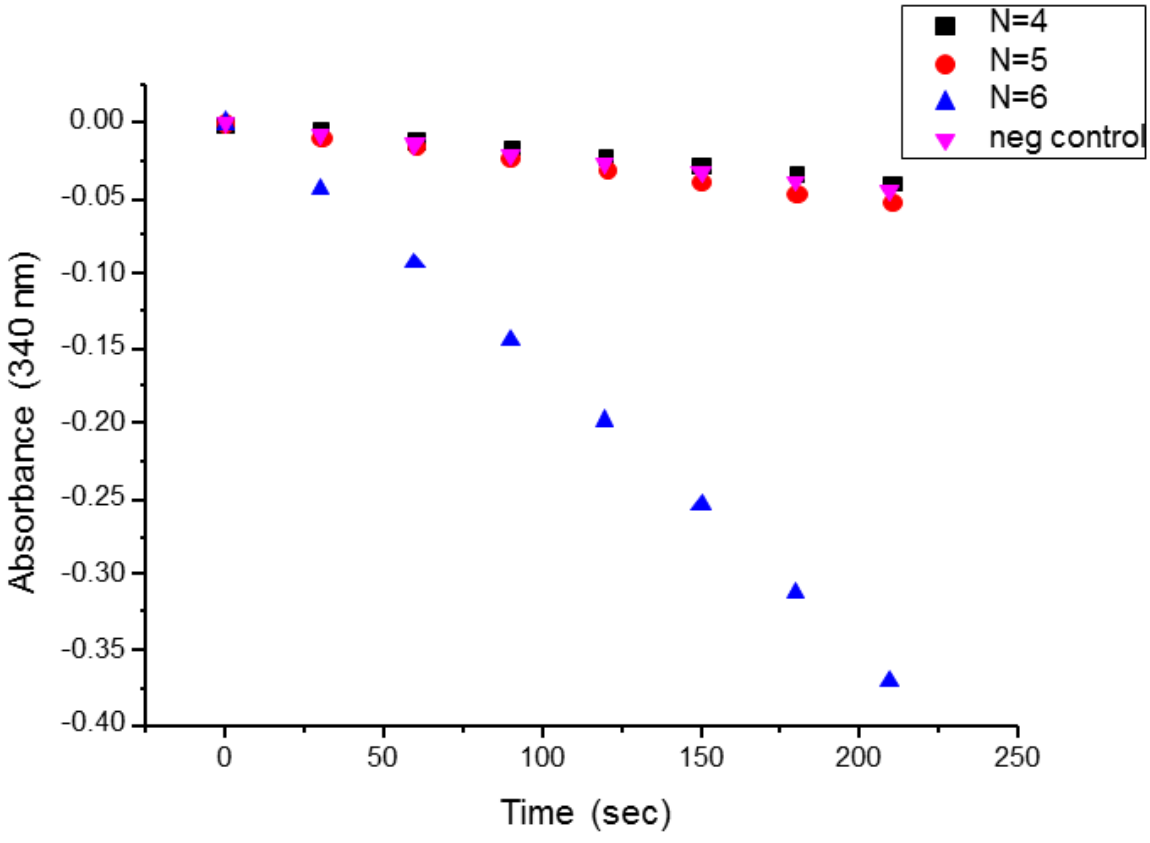
GAG-5-GAT 0 sites in pUC19

GAG-6-GAT 2 sites in pUC19

GAG-7-GAT 0 sites in pUC19

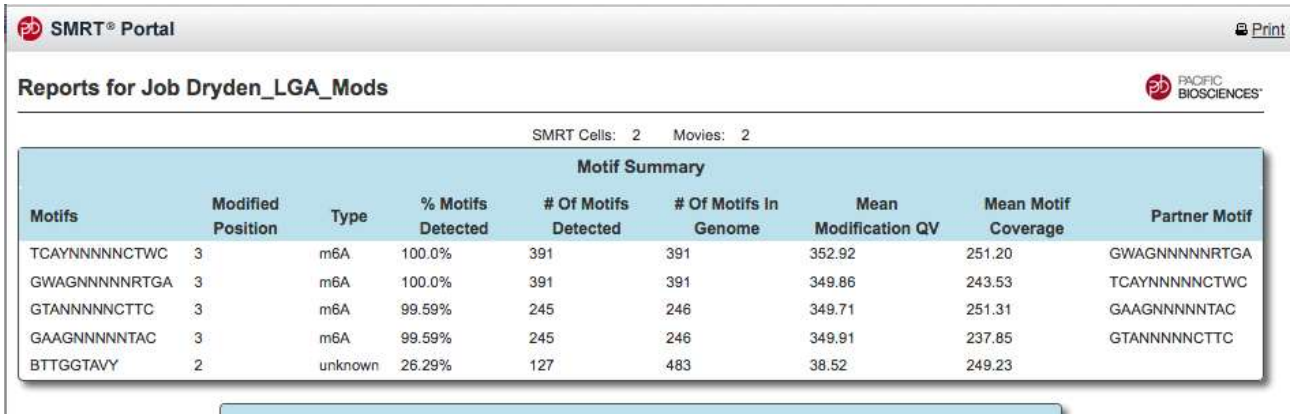
Oligonucleotide name	DNA sequence (5' to 3')
e*D6for	AGATGATGGAATCAATGCGAGTTCCATGATGCCCTATACGATATAA
e*D6rev	TTATATCGTATAGGGCATCATGGAACTCGCATTGATTCCATCATCT
e*D5for	AGATGATGGAATCAATGCGAGTTCCAGATGCCCTATACGATATAA
e*D5rev	TTATATCGTATAGGGCATCTGGAACTCGCATTGATTCCATCATCT
e*D4for	AGATGATGGAATCAATGCGAGTTCAGATGCCCTATACGATATAA
e*D4rev	TTATATCGTATAGGGCATCTGAACTCGCATTGATTCCATCATCT

N=6 shows activity.



SMRT results for S. aureus strains LGA251 and NCTC13435

LGA251



SMRT® Portal

Print

Reports for Job Dryden_LGA_Mods

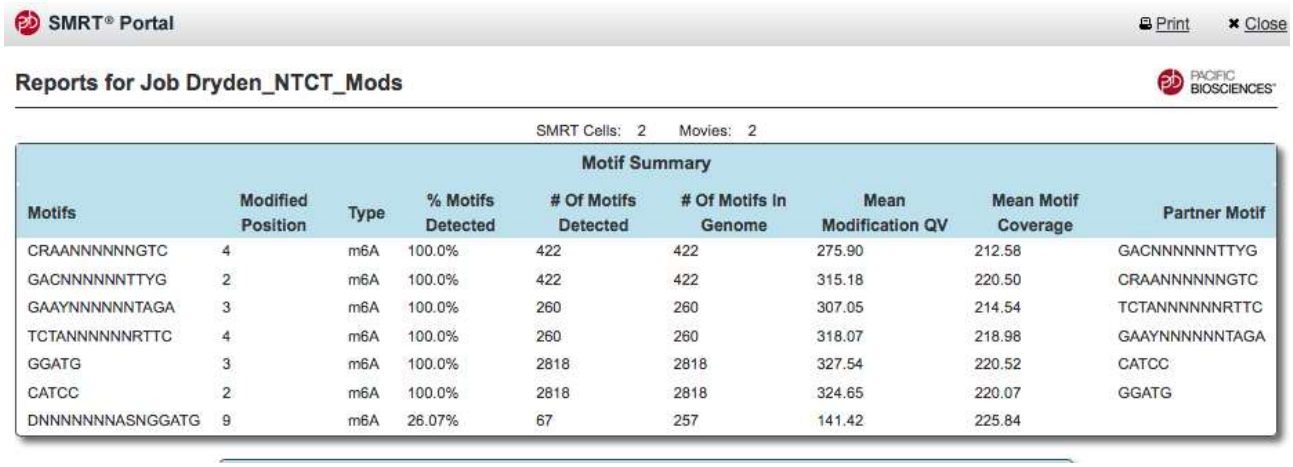
PACIFIC BIOSCIENCES

SMRT Cells: 2 Movies: 2

Motif Summary

Motifs	Modified Position	Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
TCAYNNNNNCTWC	3	m6A	100.0%	391	391	352.92	251.20	GWAGNNNNNRTGA
GWAGNNNNNRTGA	3	m6A	100.0%	391	391	349.86	243.53	TCAYNNNNNCTWC
GTANNNNNCTTC	3	m6A	99.59%	245	246	349.71	251.31	GAAGNNNNNTAC
GAAGNNNNNTAC	3	m6A	99.59%	245	246	349.91	237.85	GTANNNNNCTTC
BTTGGTAVY	2	unknown	26.29%	127	483	38.52	249.23	

NCTC13435



SMRT® Portal

Print Close

Reports for Job Dryden_NTCT_Mods

PACIFIC BIOSCIENCES

SMRT Cells: 2 Movies: 2

Motif Summary

Motifs	Modified Position	Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
CRAANNNNNNGTC	4	m6A	100.0%	422	422	275.90	212.58	GACNNNNNNTTYG
GACNNNNNNTTYG	2	m6A	100.0%	422	422	315.18	220.50	CRAANNNNNNGTC
GAAYNNNNNNTAGA	3	m6A	100.0%	260	260	307.05	214.54	TCTANNNNNNRTTC
TCTANNNNNNRTTC	4	m6A	100.0%	260	260	318.07	218.98	GAAYNNNNNNTAGA
GGATG	3	m6A	100.0%	2818	2818	327.54	220.52	CATCC
CATCC	2	m6A	100.0%	2818	2818	324.65	220.07	GGATG
DNNNNNNNASNGGATG	9	m6A	26.07%	67	257	141.42	225.84	

SUPPLEMENTARY INFORMATION FOR TABLES 5 AND 6.

SUPPLEMENTARY INFORMATION FOR TABLES 5 AND 6.

By combining all TRD 1 with all TRD 2 amino acid sequences and searching sequence databases, we found that some of our "artificial hybrids" described in Table 3 were actually present in real strains of *S. aureus*. We present several examples below.

S.SauAU

A plasmid expressing S.SauAU with the M subunit was prepared but not analysed further. The S.SauAU sequence matches that of the S subunit of the Type I RM system in *S. schweitzeri* FSA084.

>S.SauAU

MSNTQKKNVPELRFPGFEGEWEEKKLGDLTTKIGSGKTPKGGSENYTNKGIPFLRSQNIRNGKLNL
NDLVYISKDIDDEMKNSRTYYGDVLLNITGASIGRTAINSIVEIHANLNQHVCIIRLKKEYYYNFF
GQYLLSRKGKRKIFLAQSGGSREGLNFKEIANLKIFTPTIFEEQQKIGEFISKLDRQIELEEQKLE
LLQQQKKGYMQKIFSQELRFKDENGEDYPDWEVTTIQNITKYTSSKKSSNQYADKDNSKGYPVYDA
VQEIGKDSNYDIEESYISILKDGAGVGRLNLRPGKSSVIGTMGYIQSNNVDIEFLYYRMKVVDFKK
YIIGSTIPHLYFKDYSKETLYIPSSIQEQAKIGMFISNLDKLIENKNLKLNCLKQLKQGLLQSMFI
PGGSHHHHHH

S. schweitzeri FSA084

CLUSTAL O(1.2.1) multiple sequence alignment

```
FSA084       msn-tqkkvpelrfpgfegeweekklgevttkigsgktpkggsenytnkgipflrsqnir
S.SauAU      MSNTQKKNVPELRFPGFEGEWEEKKLGDLTTKIGSGKTPKGGSENYTNKGIPFLRSQNIR
             ***  :*:*******************::******************************

FSA084       ngklnlndlvyiskdiddemknsrtyygdvllnitgasigrtainsivethanlnqhvci
S.SauAU      NGKLNLNDLVYISKDIDDEMKNSRTYYGDVLLNITGASIGRTAINSIVEIHANLNQHVCI
             ********************************************** **********

FSA084       irlkkeyyynffeqyllsrkgkrkiflaqsggsreglnfkeianlkiftstifeeqqkvg
S.SauAU      IRLKKEYYYNFFGQYLLSRKGKRKIFLAQSGGSREGLNFKEIANLKIFTPTIFEEQQKIG
             *********** ********************************** ********:*

FSA084       kffskldrqieleeqklellqqqkkgymqkifsqelrfkdengneypewkvtsiqdvtky
S.SauAU      EFISKLDRQIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGEDYPDWEVTTIQNITKY
             :*:*****************************************::**:*:**:**::***

FSA084       tsskkssnqyadkidskgypvydavreigkdsnydieesyisilkdgagvgrlnlrpeks
S.SauAU      TSSKKSSNQYADKDNSKGYPVYDAVQEIGKDSNYDIEESYISILKDGAGVGRLNLRPGKS
             ************ :**********:****************************** **

FSA084       svigtmgylqannidleflyyrmkivdfkkyiigstiphlyfkdysketiyipssiqeqa
S.SauAU      SVIGTMGYIQSNNVDIEFLYYRMKVVDFKKYIIGSTIPHLYFKDYSKETLYIPSSIQEQA
             ********:*:**:*:********:************************:**********

FSA084       kigkfisnldkmienktrklnclkqlkqgllqgmfi----------
S.SauAU      KIGMFISNLDKLIENKNLKLNCLKQLKQGLLQSMFIPGGSHHHHHH
             *** *******:****. **************.***
```

S.SauJE GGA-6-RTGA
S.SauJE against subspecies 21262, a member of ST49

HsdS sequences from strain 21262.
>EHO91218 This has TRD R + f*
MSNTQKKNVPELRFPGFEGEWEEKKLGEVAKIYDGTHQTPKYTNEGIKFLSVENIKTLNS
SKYISEEAFEKEFKIRPEFGDILMTRIGDIGTPNIVSSNEKFAYYVSLALLKTKNLNSYF
LKNLILSSSIQNELWRKTLHVAFPKKINKNEIGKIKINYPKKQEQQKIGQFFSKLDRQIE
LEEQKLELLQQQKKGYMQKIFSQELRFKDENGEDYPDWKEKKLGDITEQSMYGIGASATR
FDSKNIYIRITDIDEKSRKLNYQNLTTPDELNNKYKLKRNDILFARTGASTGKSYIHKEE
KDIYNYYFAGFLIKFEIDEQNNPLFIYQFTLTSKFNKWVKVMSVRSGQPGINSEEYAKLP
LVLPNKLEQQKIAEFLDRFDQQIELEKQKIEILQQQKKGLLQSMFI

>EHO92010 This has TRD J + E
MSNTQKKNVPELRFPGFEGEWEEKKLEDIIKVNSGKDYKHLDKGDIPVYGTGGYMTSVSE
PLSEIDAVGIGRKGTINKPYLLEAPFWTVDTLFYCTPKKETDILFILSLFRKINWKVYDE
STGVPSLSKQTINKINRFVPTNKEQQKIGKFFSKLDRQIELEEQKIELLQQQKKGYIQKI
FSQELRFKDENGDDYPEWEETTIQEIAQINTGKKDTKDAITNGSYDFYVRSPIVYKINTF
SYEGEAILTVGDGVGVGKVFHYVNGKFDYHQRVYKISDFKNYYGLLLFYYFSQNFLKETK
KYSAKTSVDSVRKDMVANMKVPRPIYIEQEKIGQFIKKVDNKIKIQKQVIELLKQRKKAL
LQKMFI

CLUSTAL O(1.2.1) multiple sequence alignment

```
EHO91218      msntqkknvpelrfpgfegeweekklgevakiydgthqtpkytnegikflsveniktlns
S.SauJE       MSNTQKKNVPELRFPGFEGEWEEKKLGDLIKVNSGKDYKHL---------EKGDIPVYGT
EHO92010      msntqkknvpelrfpgfegeweekkledиikvnsgkdykhl---------dkgdipvygt
              ************************** :: *: .*.. .           .  :* .  :

EHO91218      skyiseeafekefkirpefgdilmtrigdigtpnivssnekfayyvslallk--tknlns
S.SauJE       GGYMTSVS-----EPLSEIDAVGIGRKGTINKPYLLEAP---FWTVDTLFYCTPKKETDI
EHO92010      ggymtsvs-----eplseidavgigrkgtinkpylleap---fwtvdtlfyctpkketdi
              . *::. :     :   *:  : : * * * .* ::.:     : *.  :    .*: :

EHO91218      yflknlilsssiqnelwrktlhvafpkkinkneigkikinypkkqeqqkigqffskldrq
S.SauJE       LF-----ILSLFRKINWKVYDESTGVPSLSKQTINKINRFVPSNKEQQKIGEFFIKLDRQ
EHO92010      lf-----ilslfrkinwkvydestgvpslskqtinkinrfvptnkeqqkigkffskldrq
               *     : * :::  *:   . :   .:.*: * **:   *.::******:** *****

EHO91218      ieleeqklellqqqkkgymqkifsqelrfkdengedypdwkekklgditeqsmygigas-
S.SauJE       IELEEQKLELLQQQKKGYMQKIFSQELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKD
EHO92010      ieleeqkiellqqqkkgyiqkifsqelrfkdengddypeweettiqeiaqintgkkdtkd
              *******:**********:***************.***:*:*..: :*:: .     :.

EHO91218      atrfdskniyiritdideksrklnyqnlttpdelnnkyklkrndilfartgastgksyih
S.SauJE       AITNGSYDFYVRSPIV-YK---------------INTFSYEGEAILTVGDGVGVGKVFHY
EHO92010      aitngsydfyvrspiv-yk---------------intfsyegeailtvgdgvgvgkvfhy
              *    * ::*:*   :  *                *.:. : : ** .  *...** : :

EHO91218      keekdiynyyfagflikfeideqnnplfiyqftltskfnkwvkvmsvrsgqpginseeya
S.SauJE       VNGK--FDYHQRVYKIS-DFKNYYGLLLFYYF--SQNFLKETKKYSAKTSVDSVRKDMIA
EHO92010      vngk--fdyhqrvykis-dfknyyglllfyyf--sqnflketkkysaktsvdsvrkdmva
               : *  ::*:   : *. ::.:    *::* *  :.:* * .*  *.::.  .:..:  *

EHO91218      klplvlpnkleqqkiaefldrfdqqielekqkieilqqqkkgllqsmfi----------
S.SauJE       NMKVPRPIYIEQKKIGQFIKRVDNKTKIQKQVIELLKQRKKSLLQKMFIPGGSHHHHHH
EHO92010      nmkvprpiyieqekigqfikkvdnkikiqkqviellkqrkkallqkmfi----------
              :: :  *  :**:**.:*:.:.*:: :::** **:*:*:**.***.***
```

The above alignment shows that SauJE is identical to the EHO92010 sequence from this strain.

S.SauJE GGA-6-RTGA
Sub species 21262, a member of ST49

```
CLUSTAL O(1.2.1) multiple sequence alignment
TRD R and TRD f* against EHO91218, the second HsdS in this strain.

CC80-3        ------------------------------------------------------------
EHO91218      msntqkknvpelrfpgfegeweekklgevakiydgthqtpkytnegikflsveniktlns
CC72-1        MSNTQKKNVPELRFPGFEGEWEEKKLGEVAKIYDGTHQTPKYTNEGIKFLSVENIKTLNS

CC80-3        ------------------------------------------------------------
EHO91218      skyiseeafekefkirpefgdilmtrigdigtpnivssnekfayyvslallktknlnsyf
CC72-1        SKYISEEAFEKEFKIRPEFGDILMTRIGDIGTPNIVSSNEKFAYYVSLALLKTKNLNSYF

CC80-3        ------------------------------------------------------------
EHO91218      lknlilsssiqnelwrktlhvafpkkinkneigkikinypkkqeqqkigqffskldrqie
CC72-1        LKNLILSSSIQNELWRKTLHVAFPKKINKNEIGKIKINYPKKQEQQKIGQFFSKLDRQIE

CC80-3        ----------------------QELRFKDENGEDYPDWKEKKLGDITEQSMYGIGASATR
EHO91218      leeqklellqqqkkgymqkifsqelrfkdengedypdwkekklgditeqsmygigasatr
CC72-1        LEEQKLELLQQQKKGYMQKIFS--------------------------------------

CC80-3        FDSKNIYIRITDIDEKSRKLNYQNLTTPDELNNKYKLKRNDILFARTGASTGKSYIHKEE
EHO91218      fdskniyiritdideksrklnyqnlttpdelnnkyklkrndilfartgastgksyihkee
CC72-1        ------------------------------------------------------------

CC80-3        KDIYNYYFAGFLIKFKINEQNSPLFIYQFTLTSKFNKWVKVMSVRSGQPGINSEEYAKLP
EHO91218      kdiynyyfagflikfeideqnnplfiyqftltskfnkwvkvmsvrsgqpginseeyaklp
CC72-1        ------------------------------------------------------------

CC80-3        LVLPNKLEQQKIAKFLDRFDRQIELEKQKIEILQQQKKGLLQSMFI
EHO91218      lvlpnkleqqkiaefldrfdqqielekqkieilqqqkkgllqsmfi
CC72-1        ---------------------------------------------
```

S.SauJE GGA-6-RTGA

S.SauJE against ST49 strain "Tager 104"

The ST49 Tager genome has the same TRD combinations as the ST49 strain 21262.

PATRIC db

>fig|1381115.3.peg.1063|VBIStaAur301678_1063| Type I restriction-modification system, specificity subunit S (EC 3.1.21.3) [Staphylococcus aureus subsp. aureus Tager 104 | 1381115.3] This is TRD R+f*

MSNTQKKNVPELRFPGFEGEWEEKKLGEVAKIYDGTHQTPKYTNEGIKFLSVENIKTLNS
SKYISEEAFEKEFKIRPEFGDILMTRIGDIGTPNIVSSNEKFAYYVSLALLKTKNLNSYF
LKNLILSSSIQNELWRKTLHVAFPKKINKNEIGKIKINYPKKQEQQKIGQFFSKLDRQIE
LEEQKLELLQQQKKGYMQKIFSQELRFKDENGEDYPDWKEKKLGDITEQSMYGIGASATR
FDSKNIYIRITDIDEKSRKLNYQNLTTPDELNNKYKLKRNDILFARTGASTGKSYIHKEE
KDIYNYYFAGFLIKFEIDEQNNPLFIYQFTLTSKFNKWVKVMSVRSGQPGINSEEYAKLP
LVLPNKLEQQKIAEFLDRFDQQIELEKQKIEILQQQKKGLLQSMFI

>fig|1381115.3.peg.2628|VBIStaAur301678_2628| Type I restriction-modification system, specificity subunit S (EC 3.1.21.3) [Staphylococcus aureus subsp. aureus Tager 104 | 1381115.3] This is TRD J+E

MSNTQKKNVPELRFPGFEGEWEEKKLEDIIKVNSGKDYKHLDKGDIPVYGTGGYMTSVSE
PLSEIDAVGIGRKGTINKPYLLEAPFWTVDTLFYCTPKKETDILFILSLFRKINWKVYDE
STGVPSLSKQTINKINRFVPTNKEQQKIGKFFSKLDRQIELEEQKIELLQQQKKGYIQKI
FSQELRFKDENGDDYPEWEETTIQEIAQINTGKKDTKDAITNGSYDFYVRSPIVYKINTF
SYEGEAILTVGDGVGVGKVFHYVNGKFDYHQRVYKISDFKNYYGLLLFYYFSQNFLKETK
KYSAKTSVDSVRKDMVANMKVPRPIYIEQEKIGQFIKKVDNKIKIQKQVIELLKQRKKAL
LQKMFI

S.SauJE against ST49 Tager 104 GGA-6-RTGA

```
CLUSTAL O(1.2.1) multiple sequence alignment


S.SauJE                                 MSNTQKKNVPELRFPGFEGEWEEKKLGDLIKVNSGKDYKHLEKGDIPVYGTGGYMTSVSE
fig|1381115.3.peg.2628|VBIStaAur301678_2628|  MSNTQKKNVPELRFPGFEGEWEEKKLEDIIKVNSGKDYKHLDKGDIPVYGTGGYMTSVSE
                                        ************************* *:************:*****************

S.SauJE                                 PLSEIDAVGIGRKGTINKPYLLEAPFWTVDTLFYCTPKKETDILFILSLFRKINWKVYDE
fig|1381115.3.peg.2628|VBIStaAur301678_2628|  PLSEIDAVGIGRKGTINKPYLLEAPFWTVDTLFYCTPKKETDILFILSLFRKINWKVYDE
                                        ************************************************************

S.SauJE                                 STGVPSLSKQTINKINRFVPSNKEQQKIGEFFIKLDRQIELEEQKLELLQQQKKGYMQKI
fig|1381115.3.peg.2628|VBIStaAur301678_2628|  STGVPSLSKQTINKINRFVPTNKEQQKIGKFFSKLDRQIELEEQKIELLQQQKKGYIQKI
                                        ********************:********:** ************:**********:***

S.SauJE                                 FSQELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKDAITNGSYDFYVRSPIVYKINTF
fig|1381115.3.peg.2628|VBIStaAur301678_2628|  FSQELRFKDENGDDYPEWEETTIQEIAQINTGKKDTKDAITNGSYDFYVRSPIVYKINTF
                                        ************.**********:************************************

S.SauJE                                 SYEGEAILTVGDGVGVGKVFHYVNGKFDYHQRVYKISDFKNYYGLLLFYYFSQNFLKETK
fig|1381115.3.peg.2628|VBIStaAur301678_2628|  SYEGEAILTVGDGVGVGKVFHYVNGKFDYHQRVYKISDFKNYYGLLLFYYFSQNFLKETK
                                        ************************************************************

S.SauJE                                 KYSAKTSVDSVRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTKIQKQVIELLKQRKKSL
fig|1381115.3.peg.2628|VBIStaAur301678_2628|  KYSAKTSVDSVRKDMVANMKVPRPIYIEQEKIGQFIKKVDNKIKIQKQVIELLKQRKKAL
                                        ***************:*************:*******:**** ***************:*

S.SauJE                                 LQKMFIPGGSHHHHHH
fig|1381115.3.peg.2628|VBIStaAur301678_2628|  LQKMFI----------
                                        ******
```

S.SauNQ ACC-5-RTGT

This TRD pair was found in strains KPL1845 (ST96) and 21343(ST88).

Subspecies 21343 contains SauNQ and a novel TRD (NOVEL 1) paired with TRD K.

>EHQ67679 THIS IS TRD NOVEL 1 + TRD K

MSNTQKKNVPELRFPGFEGEWEEKKLGEVATFAKGKLGAKKDVSQNGVPVILYGELYTKY
GAIVSKIFSKTDIPENKLKMAKKNDVLIPSSGETAIDIATASCIYLNKGVAVGGDINILT
PQKQDGRFISLSINGINKNELSKYAQGKTVVHLYNNDIKNLKIAFPSEFEEQVRIGNFFS
KLDRQIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPKWEEKKIEDIASQVYGG
GTPNTKIKEFWNGDIPWIQSSDVKVNDLILQQCNKFISKNSIELSSAKLIPANSIAIVTR
VGVGKLCLVEFDYATSQDFLSLSSLKYDKLYSLYSLLYTMKKISANLQGTSIKGITKKEL
LDSIIKIPHNLEEQQKIGDLFYKIDKYISFNKCKIEILKSLKQGLLKKMFI

>EHQ71248 THIS IS TRD N+Q ACC-5-RTGT

MSNTQTKNVPELKFPEFEGEWEEKKLGEFAGKVTKKNVDKKYIETLTNSAELGIISQKDY
FDKEISNIDNIKKYYVVEENDFVYNPRISNYAPFGPVNRNKLGKKGVMSPLYTVFKIQNI
DLNFIEFYFKSSKWYRFMALNGDSGARADRFSIKNRTFMEMPLHIPCMDEQIKIGQFFSK
LDRQIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPEWEERRFADIFKFHNKLR
KPIKENLRVKGSYPYYGATGIIDYVDDFIFDGNYLLIGEDGANIITRSAPLVYLVNGKFW
VNNHAHILSPLNGNIQYLYQVAELVNYEKYNTGTAQPKLNIQNLKIISVVISTNLEEQQK
IGSFLSKLDRQIDLEEQKLELLQQRKKALLKSMFV

SPECIES KPL1845 CONTAINS THREE *Sau1* S SUBUNITS.

>ETD06224 THIS IS TRD N+Q ACC-5-RTGT

MSNTQTKNVPELKFPEFEGEWEEKKLGEFAGKVTKKNVDKKYIETLTNSAELGIISQKDY
FDKEISNIDNIKKYYVVEENDFVYNPRISNYAPFGPVNRNKLGKKGVMSPLYTVFKIQNI
DLNFIEFYFKSSKWYRFMALNGDSGARADRFSIKNRTFMEMPLHIPCMDEQIKIGQFFSK
LDRQIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPEWEERRFADIFKFHNKLR
KPIKENLRVKGSYPYYGATGIIDYVDDFIFDGNYLLIGEDGANIITRSAPLVYLVNGKFW
VNNHAHILSPLNGNIQYLYQVAELVNYEKYNTGTAQPKLNIQNLKIISVVISTNLEEQQK
IGSFLSKLDRQIDLEEQKLELLQQRKKALLKSMFV

>ETD11204 THIS HAS TWO NOVEL TRDS, NOVEL 2 + NOVEL 3.

MTEQINTPELRFPEFKNEWSYDLVSDVVTNKSKKFDPKKEEAKKDIELDSIEQNTGRLLD
TYISNDFTSQKNKFNKGNVLYSKLRPYLNKYYYATIDGVCSSEIWVLNTLNKDVLANKFL
YYFIQTNRFSSVTNKSAGSKMPRADWELVKNIRLYKGSIEEQEKIGYFFSKLDRQIELEE
KKLELLEQQKKGYMQKIFAQELRFKDENGNDYPDWVTKKLGDIGKVAMNKRIYKNETTEN
GEIPFYKIGNFGKNADTFITREKFDEYKEKYPYPNVGDILISASGSIGRTIEYTGEDAYY
QDSNIVWLNHNDEVINKYLKYFYKIVKWSGIEGTTIKRLYNKNILNTKIELPTVEEQYKM
ANFLSKLDKIIDIQIEKIELLKQRKQGLLQKMFV

>ETD09130 THIS HAS A NOVEL TRD (NOVEL 4) PAIRED WITH TRD f*

1MSNTQKKNVPELRFPEFEGEWKDVKFVSIFQEVSNKTSDLAKYPLFSLTVEKGITPKTER
61YKRDFLVKKSDNFKIVEPRDIVYNPMNVTLGAIDLSKYNYDIALSGYYHVMKIINSFNPD
121FISNFLKTEKMIIHYKKIATGSLMEKQRVHFSEFKNIIKKFPTNKEQQKIGDFFSKLDRQ
181IELQVQKLELLQQQKKGYMQKIFSQELRFKDENGEDYPDWKEKKLGDITEQSMYGIGASA
241TRFDSKNIYIRITDIDEKSRKLNYQNLTTPDELNNKYKLKRNDILFARTGASTGKSYIHK
301EEKDIYNYYFAGFLIKFEIDEQNNPLFIYQFTLTSKFNKWVKVMSVRSGQPGINSEEYAK
361LPLVLPNKLEQQKIAEFLDRFDQQIELEKQKIEILQQQKKGLLQSMFI

PROMALS ALIGNMENT OF TRD AMINO ACID SEQUENCES WITH SECONDARY STRUCTURE PREDICTIONS.

"e" means beta strand and "h" means alpha helix in the consensus secondary structure.

```
PROMALS alignment of all first TRDs.

Conservation:                    999876797999999798898665 5 5
NOVEL_4_189_              1  M--TEQINTPELRFPEFKNEWSYDLVSDVVTNKSKKFDPKK------------EEAKKDIELDSIEQNTG   56
Z_GAC_191_                1  MSNTQTKNVPELRFPGFEGEYSLDIFGNLATNKSEKFNPQN------------ENASIDIELDCIEQNTG   58
NOVEL2_194                1  MSNTQKKNVPELRFPEFEGEWKDVKFVSIFQEVSNKTSDLA------------KYPLFSLTVEKGITPKT   58
NOVEL1_199                1  MSNTQKKNVPELRFPGFEGEWEEKKLGEVATFAKGKLGAKK---------DVSQNGVPVILYGELYTKYG   61
R_GARA_192                1  MSNTQKKNVPELRFPGFEGEWEEKKLGEVAKIYDGTHQTPK----------YTNEGIKFLSVENIKTLNS   60
J_GGA_172                 1  MSNTQKKNVPELRFPEFEGEWEERKLGDLIKVNSGKDYKH-----------LDKGDIPVYGTGGYMTS--   57
N_ACC_198                 1  MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKY-----IETLTNSAELGIISQKDYFDKEI   65
O_CAAC_195                1  MSNKQKKNVPELRFPGFEGEWEEKKLGEVGTFTSGGTPLKS-------KSEYWNGDIPWITTGDIHNIKR   63
T_CAAG_199                1  MSNTQTKNVPELRFPGFEGEWEEKELGEIFQIISGSTPLKSN------KEFYENGNINWVKTTDLNNSKV   64
C_GWAG_206                1  MSNTQTKNVPELRFPGFEGEWEEKQVGELLEFKNGLNKGKE----------YFGSGSSIVNFKDVFNNRS   60
M_CAG_203                 1  MSNTQTKNVPELRFPGFEGEWEEKKLEDLGLFQKSYSFSRA---------KEGNGKTKHIHYGDIHSKFK   61
X_TCTA_192                1  MSNTQKKNVPELRFPGFEGEWEEKQFADFTKINQGLQIAINE------RKTEYSPELYFYITNEFLRPNS   64
B_AGG_199                 1  MSNTQKKNVPELRFPGFEGEWEEKQLGDLTDRVIRKNKNLES-----KKPLTISGQLGLIDQTEYFSKSV   65
A_CCAY_203                1  MSNTQKKNVPELRFPGFEGEWEEKQLGDLTTKIGSGKTPKGG------SENYTNKGIPFLRSQNIRNGKL   64
e*_GAG_190                1  MSNTQKKNVPELRFPGFEGEWEEKSISSFLKESKIKGSNGS---------HAKKLTVKLWGKGVVPKKET   61
V_CNGA_210                1  MSNTGKMNVPELRFPGFEGEWEEKELRELRNPKDKYSYTGGPFGSDLKKSDYTTDGIQIIQLQNIGDGYF   70
b*_GGHA_200_              1  MSNTQKKNAPELRFPEFEGEWKEKKLEDTLEFIKDGTHGTH---------ENVNNGPWLLSAKNIKNNKI   61
Consensus_ss:                          eee        eeeeeeeeeeee                    eeeeeeeeee

Conservation:                                                5
NOVEL_4_189_             57  RLLDTY-----ISNDFTSQKNKFNKGNVLYSKLRPY---LNKYYYATI----DGVCSSEIWVLNTLNK-D  113
Z_GAC_191_               59  RLIKIY-----NSKEFSSQKNKFNPQNVLYGKLRPY---LNKYYFTKK----SGVCSSEIWVLKSTKE-D  115
NOVEL2_194               59  ERYKRDFL-----VKKSDNFKIVEPRDIVYNPMNVT---LGAIDLSKYN--YDIALSGYYHVMKIIN---  115
NOVEL1_199               62  AIVSK--IFS-KTDIPENKLKMAKKNDVLIPSSGETAIDIATASCIYLN--KGVAVGGDINILTPQ----  122
R_GARA_192               61  SK-----YIS-EEAFEKEFKIRPEFGDILMTRIGDI----GTPNIVSSN--EKFAYYVSLALLKTK----  114
J_GGA_172                58  ------------------VSEPLSEIDAVGIGRKGTI----NKPYLLEA----PFWTVDTLFYCTPEK---   99
N_ACC_198                66  S-----------NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLG--KKGVMSPLYTVFKIQ----  118
O_CAAC_195               64  ENITN--FIT-EKGLNESSAKLITNEAILIAMYGQGK-TRGMSAILNF----EATTNQACAIYQ-----T  120
T_CAAG_199               65  THSKE--KIT-EYAMKSLKLKLVPKNSVLIAMYGGFN-QIGRTGLLK----IDATINQAISALLMNH---  123
C_GWAG_206               61  INTNNLTGKV-NVNSKELKNYSVEKGDVFFTRTSEVIGEIGYPSVILNDP-ENTVFSGFVLRGRPKSGID  128
M_CAG_203                62  TVLDSD-GNI-PNIIEKAVFELIQKGDIVFADASEDYSDLGKAVMIDFEP-NSLISGLHTHLFRPLN---  125
X_TCTA_192               65  QTKYF---------IENPPQSVIANKEDILMTRTGN----TGKVVTNV----FGAFHNNFFKIKFDKN---  115
B_AGG_199                66  SSK------------NLENYTLIKNGEFAYNKSYSNGYPLGAIKRLTRY--DSGVLSSLYICFSIKS---  118
A_CCAY_203               65  NLNDLV-YIS-KDIDDEMKNSRTYYGDVLLNITGAS---IGRTAINSIVE-THANLNQHVCIIRLKK---  125
e*_GAG_190               62  F-----------KGSDNTQYYKRKAGQLMYGKLDFL---NCAFGIVPDS--LNNYESTIDSPSFDFI---  112
V_CNGA_210               71  YNSNKV-FTS-NEKAEVLKSCNVFPGDIVIAKMADP---IARAAIVPDNNIGKYLMASDGIRLSVDT--V  133
b*_GGHA_200_              62  IISSDDRKISESDYKKIYKNYKLEKGDLLLTIVGTI----GRAAIVKNP--NNIAFQRSVAILKTKA---  122
Consensus_ss:                        e            ee     eeeeee       eeeeee       eeee  eeeee

Conservation:                     9                                      5    5    99 799 9999898998
NOVEL_4_189_             114  VLANKFLYYFIQTNRFSS-VTNKSAG----SKMPRADWELVKNIRLYKGS-IEEQEKIGYFFSKLDRQIE  177
Z_GAC_191_               116  KLLNLFLYYFIQTKRYSD-VASKSAG----SKMPRADWGLIENIRVYFPE-LCEQQKIGQFFSKLDRQIE  179
NOVEL2_194               116  SFNPDFISNFLKTEKMIIHYKKIATGS--LMEKQRVHFSEFKNIIKKFPT-NKEQQKIGDFFSKLDRQIE  182
NOVEL1_199               123  KQDGRFISLSINGI-NKNELSKYAQG----KTVVHLYNNDIKNLKIAFPSEFEEQVRIGNFFSKLDRQIE  187
R_GARA_192               115  NLNSYFLKNLILSSSIQNELWRKTLHV---AFPKKINKNEIGKIKINYPK-KQEQQKIGQFFSKLDRQIE  180
J_GGA_172                100  EADILFILSLFRKINWKL----YDES----TGVPSLSKQTINKINRLVPT-NKEQQKIGEFFSKLDRQIE  160
N_ACC_198                119  NIDLNFIEFYFKSSKWYRFMALNGDSGA-RADRFSIKDRTFMEMPLHIPC-MDEQIKIGQFFSKLDRQIE  186
O_CAAC_195               121  NQNINFVFQYFQK--LYEFLRSLSNE----GSQKNLSLSLLKEITLNYPN-EQEQKKIGDFFSKLDRQIE  183
T_CAAG_199               124  ETNPEFIQAFLNYQ-VKGWKRYAASS----RKDPNITKKDIEQFKVPYVS-INEQQKIGEFFSKIDHQIE  187
C_GWAG_206               129  LINNNFKRYVFFTNSFRKEMITKSSM----TTRALTSGTAINKMKVIYPVSAKEQKKIGDFFSKLDRQIE  194
M_CAG_203                126  NAISNFLIFYTKTLSYKKFIRQQGTG----ISVLGISKKSLLNLNVLIPRSELEQQKIGQFFSKLDRQIE  191
X_TCTA_192               116  LYDRLFLVEVLNSSKIQNKILSLAGS----STIPDLNHSDFYSISSSYPL-LREQQKIGKFFSKLDRQIE  180
B_AGG_199                119  EMSKDFMEAYFDSTHWYREVSGIAVEGARNHGLLNVSVNDFFTILIKYPS-LEEQQKIGKFFSKLDRQIE  187
A_CCAY_203               126  EYYYIFFGQYLLSRKGKRKIFLAQSG----GSREGLNFKEIANLKIFTPTIFEEQQKIGKFFSKLDRQIE  191
e*_GAG_190               113  NGDSKFLLERIKLKSFYKKFGDIANGS---RKAKRINQDTFLSLPVFAPK-YDEQLRIGEFFSKLDRQIE  178
V_CNGA_210               134  HFNTKFVLECINRKSFRKKVEDNSSG----STRMRIGLSTLGSLTLKTTT-LKEQQKIGQFFSKLDRQIV  198
b*_GGHA_200_             123  TYDVGFIFQLFQTKYFKNLLLRKQVV----SAQPGLYLGDIRKIKISITNIIEEQRKIGIFFSKLDRQIE  188
Consensus_ss:                   hhhhhhhhh hhhhhhhhhh              hhhhh  ee     hhhhhhhhhhhhhhhhhhh

Conservation:                    977899999899
NOVEL_4_189_             178  LEEKKLELLEQQ  189
Z_GAC_191_               180  LEEQKLELLQQQ  191
NOVEL2_194               183  LQVQKLELLQQQ  194
NOVEL1_199               188  LEEQKLELLQQQ  199
R_GARA_192               181  LEEQKLELLQQQ  192
J_GGA_172                161  LEEQKLELLQQQ  172
N_ACC_198                187  LEEQKLELLQQQ  198
O_CAAC_195               184  LEEQKLELLQQQ  195
T_CAAG_199               188  LEEQKLELLQQQ  199
C_GWAG_206               195  LEEQKLELLQQQ  206
M_CAG_203                192  LEEQKLELLQQQ  203
X_TCTA_192               181  LEEQKLELLQQQ  192
B_AGG_199                188  LEEQKLELLQQQ  199
A_CCAY_203               192  LEEQKLELLQQQ  203
e*_GAG_190               179  LQKQKLELLQQQ  190
V_CNGA_210               199  LEEQKLELLQQQ  210
b*_GGHA_200_             189  LEEQKLELLQQQ  200
Consensus_ss:                 hhhhhhhhhh
```

```
PROMALS alignment of all second TRDs.
Conservation:       9989799997999999989696797 9    5
NOVEL3_205       1  KKGYMQKIFAQELRFKDENGNDYPDWVTKKLGDIGKVAMNKRIYKNE-------TTENGEIPFYKIGNFG   63
S_GCA_200        1  KKGYMQKIFSQELRFKDENGNDYPDWTNERLGEVTTVTMGQSPKSVN-----YTDNSNDTVLIQGNADIE   65
d*_CYAA_220      1  KKGYMQKIFSQELRFKDENGNDYPEWENVMLQKVLKDKTEGIKRGPFGGALKKDIFVESGYAVYEQRNAI   70
a*_GAA_208       1  KKGYMQKIFSQELRFKDENGNDYPEWENKRIEDIANVNKGFTPSTNN-----NEYWDNNDKNWLSIAGMN   65
E_TCAY_194       1  KKGYMQKIFSQELRFKDENGNDYPEWEETTIKEIAQINXGKKDTKD--------AITNGSYDFYVRSPIV   62
W_CRAA_211       1  KKGYMQKIFSQELRFKDENGNDYPDWEEKQLGELSQIVRGASPRPIK---DPKWFNKESDIGWLRISDVT   67
Q_ACAY_197       1  KKGYMQKIFSQELRFKDENGEDYSEWEERRFADIFKFHNKLRKPIKE------NLRVKGSYPYYGATGII   64
G_ACA_196        1  KKGYMQKIFTQELRFKDENGEEYPEWENKFIKDIFIFENNRRKPITS------SLREKGLYPYYGATGII   64
f*_GAAY_224      1  KKGYMQKIFSQELRFKDENGEDYPDWKEKKLGDITEQSMYGIGASA--------TRFDSKNIYIRITDID   62
L_TTTA_213       1  KKGYMQKIFSQELRFKDENGNDYPNWRTIELKNILENIVDNRGKTPD-------NAPSEKYPLLEVNALG   63
Y_CTA_209        1  KKGYMQKIFSQELRFKDENGNDYPDWEKKKLKEIACVYTGNTPSKKE-----NIYWNKGEYVWVTPTDIN   65
U_GAY_193        1  KKGYMQKIFSQELRFKDENGEDYPDWEVTTIQNITKYTSSKKSSNQY-----ADKDNSKGYPVYDAVQEI   65
I_YTCA_220       1  KKGYMQKIFSQELRFKNENGNDYPDWERIKFFDVIDKVIDFRGRTPK---KLNMEWSDEGYLALSAVNVK   67
K_CGA_212        1  KKGYMQKIFSQELRFKDENGNDYPKWEEKKIEDIASQVYGGGTPNTK-----IKEFWNGDIPWIQSSDVK   65
D_ATC_204        1  KKGYMQKIFSQELRFKDENGEDYPHWENSKIEKYLKERNERSDK--------------GQMLSVTINSGI   56
c*_GAY_209       1  KKGYLQKIFSQELRFKDENGNDYPEWRFARFKDFMYKPINIRPAINI-----------SKSELLTVKLHC   59
P_AGG_214        1  KKGYMQKIFSQELRFKDESGNDYPDWEEKELGEVADRVIRKNKNFES----KKPLTISGQLGLIDQTEYF   66
F_TTAA_216       1  KKGYMQKIFSQELRFKDEEGKDYPDWKSKSIQEIFENKGGTALETE--------FNFDGNYKVISIGSYS   62
H_TAC_206        1  KKCYIQKIFSQELRFKDEEGNYYKGWNKKQLKDVLEFSNKRTINE-------------NEYPVLTSSRQG   57
Consensus_ss:         hhhhhh                eeeeeeheeeee                        eeeeeeeeee

Conservation:                                          5             7
NOVEL3_205      64  KNADTFITR--EKF----DEYKEKYPYPNVGD-ILISASG------SIGRTIEYTG--EDA-YYQDSNIV  117
S_GCA_200       66  NGLINPR-----------IYTREVTKLIQKDE-IILTVRA------PVGKLAMAQIN-----ACIGRGVC  112
d*_CYAA_220     71  YDISNFRYY-INEN----KYKEMQSFSVQPND-IIMSCSG------TIGRLALIPH-NYTK-GIINQALI  126
a*_GAA_208       66  QKYLY-KGN-KGIS----KDAAKNYMKVKNDT-LIMSFKL------TIGKLAIVKAP-----LYTNEAIC  117
E_TCAY_194      63  YK--------------------INTFSYEGEAILTVGDGV------GVGKVFHYVNGK----FDYHQRVY  102
W_CRAA_211      68  NQNGKIYHLEQKLS----IEGQEKTRVLVTTH-LLLSIAA------SIGKPVMNFVK-----TGVHDGFL  121
Q_ACAY_197      65  DYV--------------------DDFIFDGNY-LLIGEDGANIITRSAPLVYLVNGK-----FWVNNHAH  108
G_ACA_196       65  DYV--------------------KDYLFNNEERLLIGEDGAK-WGQFETSSFIANGQ-----YWVNNHAH  108
f*_GAAY_224     63  EKSRKLNYQ-NLTT----PDELNNKYKLKRND-ILFARTGA-----STGKSYIHKEEKDIYNYYFAGFLI  121
L_TTTA_213      64  YYRPAYIKV-SKFVS-ENTYNNWFREHLKEND-ILFSTVG------NTGIVSLMDNYK----AVIAQNIV  120
Y_CTA_209       66  NSKNIYESE-NKLT----QEGYKKARQLPENT-LLVTCIA------SIGKNAILRKQ-----GSCNQQIN  118
U_GAY_193       66  GK--------------------DSNYDIEESY-ISILKDGA-----GVGRLNLRPGK-----SSVIGTMG  104
I_YTCA_220      68  KGYIDFNVE-AKYGNLDLYTRWMRGNELYKGQ-VLFTTEA------PMGNVAQVPDNKG---YILSQRTI  126
K_CGA_212       66  VNDLILQQCNKFISK--NSIELSSAKLIPANS-IAIVTRV------GVGKLCLVEFD-----YATSQDFL  121
D_ATC_204       57  IKFSELDR--KDNS----SKNKSNYKVVRKND-IAYNSMRM-----WQGASGKSNYN-----GIVSPAYT  109
c*_GAY_209      60  KGIEK-ANI-NRVL----KLGATNYYKRFEGQ-FIYGKQNF-----FNGAFDIVPKKFDG--LYSSSDVP  115
P_AGG_214       67  SKSVS-------------SKNLENYTLIKNGE-FAYNKSYSN--GYPLGAIKRLTRYDS---GVLSSLYI  117
F_TTAA_216      63  INSTYNDQN-IRVN----KNKKTEKYILSKGD-LAMVLNDKTKDGKIIGRSIFIDKDNQ---YIYNQRTE  123
H_TAC_206       58  LILQSDYYKDRKTF----AESNIGYFILPKNH-ITYRSRS------DDGIFKFNLNLMIDV-GIISKYYP  115
Consensus_ss:                              ee      eeeee          eeeeee         eeee   ee

Conservation:                  6                                  5            7    98 76  7
NOVEL3_205     118  WLNHND-EVINKYLKYFYKI-----VKWSGIEG----TTIKRLYNKNILNTKIELPT-VEEQYKMANFLS  176
S_GCA_200      113  SIKG------DKFLYYFLEWFATQNKWIRFSQG----STFESISGNDIRNIHIKIPV-EDERTKIIKLLN  171
d*_CYAA_220    127  RFRTNH-KIRSEFFLIFMRSNQMQRKILEANPG---SAITNLVPVKELKLIPFPLPV-KFEQDKISQFIH  191
a*_GAA_208     118  HFIWKVNKINTEFIYYYLNSL---NISTFGVQA----VKGVTLNNDSINSIIVKLPN-EEEQNIIAKFLL  179
E_TCAY_194     103  KISDFK-NYYGLLLFYYFSQN-FLKETKKYSAK----TSVDSVRKDMIANMKVPRPI-YIEQKKIGQFIK  165
W_CRAA_211     122  IFLKP--KFNLFFMYYWLEYF--KDKWSKYGQP----GSQVNLNSEIVKSQTLNMPS-NHEQEKVGQFFN  182
Q_ACAY_197     109  ILSPL--NGNIQYLYQVAEL----VNYEKYNTG----TAQPKLNIQNLKIINVVISTNLEEQQKIGSFLS  168
G_ACA_196      109  VVKSN--DHNLFFMNYYLNF----KELRAFVTG----NAPAKLTHANLCNINLKIPC-LTEQDKVSALLK  167
f*_GAAY_224    122  KFKINE-QNSPLFIYQFTLTSKFNKWVKVMSVR----SGQPGINSEEYAKLPLVLPN-KLEQQKIAKFLD  185
L_TTTA_213     121  GLRVNN-NNLPSFIYYMLSYKGNQKKIKRIQMG----AVQPSVKVSQFKFIKYLVPI-KDEQEKVAKLLI  184
Y_CTA_209      119  AVVPFE-NINIDYLYYISDSL--STFMKSIAGK----TATQIVNKNTFENLEIYLAP-FEEQNKIADLIS  180
U_GAY_193      105  YIQSN--NVDIEFLYYRMKVV----DFKKYIIG----STIPHLYFKDYSKETLYIPSSIQEQAKIGMFIS  164
I_YTCA_220     127  AFNSNE-KITDNFLASLLSSENVYNDLLKLCSG----ATAKGVSQKNLNRLYVTIPHSISEQEEIAEFFR  191
K_CGA_212      122  SLSSLK--YDKLYSLYSL--LYTMKKISANLQG----TSIKGITKKELLDSIIKIPHNLEEQQKIGDLFY  183
D_ATC_204      110  VLYPTQ-NTSSLFIGYKFKTHRMIHKFKINSQGL--TSDTWNLKYKQLKNINIDIPV-LEEQEKIGDFFK  175
c*_GAY_209     116  AFEINTEKIEPNYFISYISRPSFYKSKEKYSTG----TGSKRIHENTVLNFSLHLPC-LNEQLKIASFVC  180
P_AGG_214      118  CFSIKS-EMSKDFMEAYFDSTHWYREVSGIAVEGARNHGLLNISVNDFFTILIKYPS-LEEQRKIGDFFI  185
F_TTAA_216     124  RLIPFA-ENDNKFLWFLMNTDLIRNKIKGMMQG----ATQVYINYSSIKLISIQLPL-LEEQQKIRGFLE  187
H_TAC_206      116  VFKGI--DANQYYLTLHLNYQ-LKKEYIKYATG----TSQLVLSQKDLQNIKTKLPS-YEEQQKIGDFFS  177
Consensus_ss:         eee       hhhhhhhhh hhhhhhhhhh              hhhhh  ee     hhhhhhhhhhhh
Conservation:          5        75  56  97 697 786
NOVEL3_205     177  KLDKIIDIQIEKIELLKQRKQGLLQKMFV----------  205
S_GCA_200      172  SLDVLNSKTDLKIQNLKQRKQSLLQKIFV----------  200
d*_CYAA_220    192  IINRRIEQSEKKIESLKNRKQGFLQKLFV----------  220
a*_GAA_208     180  EVDKTVNNQLVKTKLLKQRKKGLLQRMFV----------  208
E_TCAY_194     166  RVDNKTKIQKQVIELLKQRKKALLQKMFI----------  194
W_CRAA_211     183  RNEKLIELQQEKIMYIKRCKQVLLQKMFI----------  211
Q_ACAY_197     169  KLDRQIDLEEQKLELLQQRKKALLKSMFV----------  197
G_ACA_196      168  SIDNKMNNQMNRIELLKERKKELLQKMFI----------  196
f*_GAAY_224    186  RFDRQIELEKQKIEILQQQKKGLLQSMFI----------  214
L_TTTA_213     185  EIDKLVNKQLIKIELLQQRKKALLKSMFI----------  213
Y_CTA_209      181  SLEELIEKQASKLIKMKSRKQGMLQIMFI----------  209
U_GAY_193      165  NLDKLIENKNLKLNCLKQLKQGLLQSMFI----------  193
I_YTCA_220     192  KINQLVELQKYKIEHTKSQKQVFLQKMFI----------  220
K_CGA_212      184  KIDKYISFNKCKIEMLKSLKQGLLKKMFI----------  212
D_ATC_204      176  KMDILISKQKIKIEILEKEKQSFLQKMFL----------  204
c*_GAY_209     181  FLNRKIELLERKIYLIKKQKQALLQQMFI----------  209
P_AGG_214      186  KLDRQIELEEQKLELLQQRKKALLKSMLI----------  214
F_TTAA_216     188  VLSGITTKQLHKIDQLKERKKAFLQKMFI----------  216
H_TAC_206      178  EIDRLVEKQSSKVGRLKVRKKELLQKMFV----------  206
Consensus_ss:       hhhhhhhhhhhhhhhhhhhhhhhhhhh
```

PROMALS alignment of all TRDs.

```
Conservation:                                 799997576    6 85   5
CC80-3 f*                        1  ----------QELRFKDENGEDYPDWKEKKLGDITEQSMYGIGASA---------TRFDSKNIYIRITDI   51
CC45-1L                          1  ----------QELRFKDENGNDYPNWRTIELKNILENIVDNRGKTP--------DNAPSEKYPLLEVNAL   52
CC97 c*                          1  ----------QELRFKDENGNDYPEWRFARFKDFMYKPINIRPAIN-------ISKSELLTVKLHCK-GI   52
CC22-1I                          1  ----------QELRFKNENGNDYPDWERIKFFDVIDKVIDFRGRTPKK----LNMEWSDEGYLALSAVNV   56
CC873D                           1  ----------QELRFKDENGEDYPHWENSKIEKYLKERNERSDKGQM------LSVTIN--SGIIKFSEL   52
CC5-1D                           1  ----------QELRFKDENGEDYPDWENSKIEKYLKERNERSDKGQM------LSVTIN--SGIIKFSEL   52
CC30-1D                          1  ----------QELRFKDENSEDYPHWENSKIEKYLKERNERSDKGQM------LSVTIN--SGIIKFSEL   52
CC5-2H                           1  ----------QELRFKDEEGNYYKGWNKKQLKDVLEFSNKRTINE-------------NEYPVLTSSRQ   46
CC133-2fromED133 d*              1  ----------QELRFKDENGNDYPEWENVMLQKVLKDKTEGIKRGPFGG-ALKKDIFVESGYAVYEQRNA   59
CC72-2S                          1  ----------QELRFKDENGNDYPDWTNERLGEVTTVTMGQSPKSVN------YTDNSNDTVLIQGNADI   54
CC93-3 a*                        1  ----------QELRFKDENGNDYPEWENKRIEDIANVNKGFTPSTNN------NEYWDNNDKNWLSIAGM   54
CC93-2K                          1  ----------QELRFKDENGNDYPKWEEKKIEDIASQVYGGGTPNTK------IKEFWNGDIPWIQSSDV   54
CC30-2K                          1  ----------QELRFKDENGNDYPNWEEKKIEDIASQVYGGGTPNTK------IKEFWNGDIPWIQSSDV   54
CC80-2W                          1  ----------QELRFKDENGNDYPDWEEKQLGELSQIVRGASPRPIKD----PKWFNKESDIGWLRISDV   56
CC75-2W                          1  ----------QELRFKDENGNDYPDWEEKQLGELSQIVRGASPRPIKD----PKWFNKESDIGWLRISDV   56
CC59Q                            1  ----------QELRFKDENGEDYSEWEERRFADIFKFHNKLRKPIK-------ENLRVKGSYPYYGATGI   53
CC72-1Q                          1  ----------QELRFKDENGNDYPEWEERRFADIFKFHNKLRKPIK-------ENLRVKGSYPYYGATGI   53
CC1-2G                           1  ----------QELRFKDENGEEYPEWENKFIKDIFIFENNRRKPIT-------SSLREKGLYPYYGATGI   53
ST425-1E                         1  ----------QELRFKDENGNDYPEWEETTIKEIAQINTGKKDTKD---------AITNGSYDFYVRSPI   51
CC15TRD2E                        1  ----------QELRFKDENGNDYPEWEETTIKEIAQINXGKKDTKD---------AITNGSYDFYVRSPI   51
CC133_771E                       1  ----------QELRFKDENGDDYPEWEETTIKEIAQINTGKKDTKD---------AITNGSYDFYVRSPI   51
CC398-1E                         1  ----------QELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKD---------AITNGSYDFYVRSPI   51
CC80-1Y                          1  ----------QELRFKDENGNDYPDWEKKKLKEIACVYTGNTPSKKE------NIYWNKGEYVWVTPTDI   54
CC75-1U                          1  ----------QELRFKDENGEDYPDWEVTTIQNITKYTSSKKSSNQY------ADKDNSKGYPVYDAVQE   54
CC1-1F                           1  KKGYMQKIFSQELRFKDEEGKDYPDWKSKSIQEIFENKGGTALETE---------FNFDGNYKVISIGSY   61
CC873 e*                         1  -MSNTQKKNVPELRFPGFE----GEWEEKSISSFLKESKIKGSNGS---------HAKKLTVKLWGKGVV   56
CC80-2Z                          1  -MSNTQTKNVPELRFPGFE----GEYSLDIFGNLATNKSEKFNPQN-------ENASIDIELDCIEQNTG   58
CC80-3XS.Sau11819ORF2227P        1  -MSNTQKKNVPELRFPEFE----GEWEEKQFADFTKINQGLQIAINE------RKTEYSPELYFYITNEF   59
CC80-1X                          1  -MSNTQKKNVPELRFPGFE----GEWEEKQFADFTKINQGLQIAINE------RKTEYSPELYFYITNEF   59
CC75-1T                          1  -MSNTQTKNVPELRFPGFE----GEWEEKELGEIFQIISGSTPLKSN------KEFYENGNINWVKTTDL   59
ST130-1T                         1  -MSNTQKKNVPELRFPGFE----GEWEEKKLGEIFQIISGSTPLKSN------KKFYENGNINWVKTTDL   59
CC93-3M                          1  -MSNTQTKNVPELRFPGFE----GEWEEKKLEDLGLFQKSYSFSRA---------KEGNGKTKHIHYGDI   56
CC133_771-1strain32320Hsd        1  -MSNTQTKNVPELRFPGFE----GEWEEKKLGDLGLFQKSYSFSRA---------KEGNGKTKHIHYGDI   56
CC133-2fromED133J                1  -MSNTQKKNVPELRFPGFE----GEWEEKKLESIIKVNSGKDYKH-----------LDKGDIPVYGTGGY   54
CC72-2J                          1  -MSNTQKKNVPELRFPEFE----GEWEEKQLGNIIKVNSGKDYKH-----------LDKGDIPVYGTGGY   54
CC51TRD1J                        1  -MSNTQTKNVPELRFPGFE----GEWEEKKLEDIIKVNSGKDYKH-----------LDKGDIPVYGTGGY   54
CC30-2strainMRSA252HsdSJ         1  -MSNTQTKNVPELRFPGFE----GEWEEKKLGDLIKVNSGKDYKH-----------LEKGDIPVYGTGGY   54
CC59-1J                          1  -MSNTQKKNVPELRFPEFE----GEWEERKLGDLIKVNSGKDYKH-----------LDKGDIPVYGTGGY   54
CC72-1R                          1  -MSNTQKKNVPELRFPGFE----GEWEEKKLGEVAKIYDGTHQTPK----------YTNEGIKFLSVENI   55
CC15TRD1O                        1  -MSNKQKKNVPELRFPGFE----GEWEEKKLGEVGTFTSGGTPLKS-------KSEYWNGDIPWITTGDI   58
CC398-1strain398HsdSN            1  -MSNTQKKNVPELRFPGFE----GEWEEKKLGEFAGKVTQKNVDKKY-----IETLTNSAELGIISQKDY   60
ST425-1C                         1  -MSNTQTKNVPELRFPGFE----GEWEEKQVGELLEFKNGLNKGKE----------YFGSGSSIVNFKDV   55
CC30-1strainMRSA252HsdSC         1  -MSNTQTKNVPELRFPGFE----GEWEEKKVGELLEFKNGLNKGKE----------YFGSGSSIVNFKDV   55
CC45-1strain3067HsdSC            1  -MSNTQKKNVPELRFPGFE----GEWEEKKVGELLEFKNGLNKGKE----------YFGSGSSIVNFKDV   55
CC97A                            1  -MSNTQKKNVPELRFPGFE----GEWEEKQLGDLTTKIGSGKTPKGG------SENYTNKGIPFLRSQNI   59
CC1-2strainMW2HsdSA              1  -MSNTQTKNVPELRFPGFE----GEWEEKKLGNLTTKIGSGKTPKGG------SENYTNKGIPFLRSQNI   59
CC1-1strainMW2HsdSA              1  -MSNTQKKNVPELRFPGFE----GEWEEKKLGDLTTKIGSGKTPKGG------SENYTNKGIPFLRSQNI   59
CC5-2strainN315HsdSA             1  -MSNTQTKNVPELRFPGFE----GEWEEKKLGNLTTKIGSGKTPKGG------SENYTNKGIPFLRSQNI   59
CC75-2V                          1  -MSNTGKMNVPELRFPGFE----GEWEEKELRELRNPKDKYSYTGGPFGSDLKKSDYTTDGIQIIQLQNI   65
CC22-1strain5096HsdSB            1  -MSNTQKKNVPELRFPGFE----GEWEEKKLGDLTDRVIRKNKNLES-----KKPLTISGQLGLIDQTEY   60
CC51TRD2P                        1  ----------QELRFKDESGNDYPDWEEKELGEVADRVIRKNKNFES-----KKPLTISGQLGLIDQTEY   55
CC5-1strainN315HsdSB             1  -MSNTQKKNVPELRFPGFE----GEWEEKQLGDLTDRVIRKNKNLES-----KKPLTISGQLGLIDQTEY   60
CC93-2 b*                        1  -MSNTQKKNAPELRFPEFE----GEWKEKKLEDTLEFIKDGTHGTH---------ENVNNGPWLLSAKNI   56
Consensus_ss:                                              eeeeeeeeeeee                      eeeeeee e
```

```
Conservation:                                          5             5
CC80-3 f*                    52  DEKSRKLN-YQNLTTP----DELNNKYKLKRNDILFARTGAST-----GKS-YIHKEEKDIYNYYFAGFL  110
CC45-1L                      53  GYYRPAYI-KVSKFVSE-NTYNNWFREHLKENDILFSTVGNT------GIV-SLMDN----YKAVIAQNI  109
CC97 c*                    53  EKANINRV----------LKLGATNYYKRFEGQFIYGKQNFFN-----GAF-DIVPKK--FDGLYSSSDV  104
CC22-1I                      57  KKGYIDFNVEAKYGNLD-LYTRWMRGNELYKGQVLFTTEAPM------GNV-AQVPD---NKGYILSQRT  115
CC873D                       53  DRKDN-------------SSKNKSNYKVVRKNDIAYNSMRMWQ-----GAS-GKSNY-----NGIVSPAY   98
CC5-1D                       53  DRKDN-------------SSKDKSNYKVVRKNDIAYNSMRMWQ-----GAS-GKSNY-----NGIVSPAY   98
CC30-1D                      53  DRKDN-------------SSKDKSNYKVVRKNDIAYNSMRMWQ-----GAS-GRSNY-----NGIVSPAY   98
CC5-2H                       47  GLILQSD---YYKDRKT-FAESNIGYFILPKNHITYRSRSDD------GIFKFNLNLM--IDVGIISKYY  104
CC133-2fromED133 d*          60  IYDISNF----RYYINE-NKYKEMQSFSVQPNDIIMSCSGTI------GRL-ALIPHN--YTKGIINQAL  115
CC72-2S                      55  ENGL----------INP-RIYTREVTKLIQKDEIILTVRAPV------GKL-AMAQI-----NACIGRGV  101
CC93-3 a*                    55  NQKYLYK---GNKGIS---KDAAKNYMKVKNDTLIMSFKLTI------GKL-AIVKA-----PLYTNEAI  106
CC93-2K                      55  KVNDLILQ-QCNKFISK-NSIELSSAKLIPANSIAIVTRVGV------GKL-CLVEF-----DYATSQDF  110
CC30-2K                      55  KVNDLILR-QCNKFISK-NSIELSSAKLIPANSIAIVTRVGV------GKL-CLVEF-----DYATSQDF  110
CC80-2W                      57  TNQNGKIY-HLEQKLS---IEGQEKTRVLVTTHLLLSIAASI------GKP-VMNFV-----KTGVHDGF  110
CC75-2W                      57  TNQNGKIY-HLEQKLS---IEGQEKTRVLVTTHLLLSIAASI------GKP-VMNFV-----KTGVHDGF  110
CC59Q                        54  IDYV--------------------DDFIFDGNYLLIGEDGA-NIITRSAPLVYLVNG-----KFWVNNHA   97
CC72-1Q                      54  IDYV--------------------DDFIFDGNYLLIGEDGA-NIITRSAPLVYLVNG-----KFWVNNHA   97
CC1-2G                       54  IDYVK-------------------DYLFNNEERLLIGEDGA-KWGQFETSS-FIANG-----QYWVNNHA   97
ST425-1E                     52  VYKI--------------------NTFSYEGEAILTVGDGVGV-----GKV-FHYVN----GKFDYHQRV   91
CC15TRD2E                    52  VYKI--------------------NTFSYEGEAILTVGDGVGV-----GKV-FHYVN----GKFDYHQRV   91
CC133_771E                   52  VYKI--------------------NTFSYEGEAILTVGDGVGV-----GKV-FHYVN----GKFDYHQRV   91
CC398-1E                     52  VYKI--------------------NTFSYEGEAILTVGDGVGV-----GKV-FHYVN----GKFDYHQRV   91
CC80-1Y                      55  NNSKNIY--ESENKLT---QEGYKKARQLPENTLLVTCIASI------GKN-AILRK-----QGSCNQQI  107
CC75-1U                      55  IGK--------------------DSNYDIEESYISILKDGAGV-----GRL-NLRPG-----KSSVIGTM   93
CC1-1F                       62  SINSTYN--DQNIRVN---KNKKTEKYILSKGDLAMVLNDKTKDGKIIGRS-IFIDK---DNQYIYNQRT  122
CC873 e*                     57  PKKETF------------KGSDNTQYYKRKAGQLMYGKLDFLN-----CAF-GIVPD---SLNNYESTID  105
CC80-2Z                      59  RLIKIYN------------SKEFSSQKNKFNPQNVLYGKLRPYL-----NKY-YFTKK-----SGVCSSEI  106
CC80-3XS.Sau11819ORF2227P    60  LRPNS--------QTKY-FIENPPQSVIANKEDILMTRTGNT------GKV-VTNVF-----GAFHNNFF  108
CC80-1X                      60  LRPNS--------QTKY-FIENPPQSVIANKEDILMTRTGNT------GKV-VTNVF-----GAFHNNFF  108
CC75-1T                      60  NNSKVTH---SKEKITE-YAMKSLKLKLVPKNSVLIAMYGGFNQI---GRT-GLLKI-----DATINQAI  116
ST130-1T                     60  NNSKVTH---SKEKITE-YAMNSLKLKLVPKNSVLIAMYGGFNQI---GRT-GLLKI-----DATINQAI  116
CC93-3M                      57  HSKFKTV--LDSDGNIP-NIIEKAVFELIQKGDIVFADASEDYSDL--GKA-VMIDFE--PNSLISGLHT  118
CC133_771-1strain32320Hsd    57  HSKFKTV--LDSDGNIP-NIIEKAVFELIQKGDIVFADASEDYSDL--GKA-VMIDFK--PNSLISGLHT  118
CC133-2fromED133J            55  MTS--------------------VSEPLSEIDAVGIGRKGTI------NKP-YLLEA-----PFWTVDTL   92
CC72-2J                      55  MTS--------------------VSEPLSEIDAVGIGRKGTI------NKP-YLLEA-----PFWTVDTL   92
CC51TRD1J                    55  MTS--------------------VSEPLSEIDAVGIGRKGTI------NKP-YLLEA-----PFWTVDTL   92
CC30-2strainMRSA252HsdSJ     55  MTS--------------------VSEPLSEIDAVGIGRKGTI------NKP-YLLEA-----PFWTVDTL   92
CC59-1J                      55  MTS--------------------VSEPLSEIDAVGIGRKGTI------NKP-YLLEA-----PFWTVDTL   92
CC72-1R                      56  KTLNSS------KYISE-EAFEKEFKIRPEFGDILMTRIGDI------GTP-NIVSS---NEKFAYYVSL  108
CC15TRD1O                    59  HNIKREN---ITNFITE-KGLNESSAKLITNEAILIAMYGQGKTR---GMS-AILNF-----EATTNQAC  115
CC398-1strain398HsdSN        61  FDKEIS------------NIDNIKKYYVVEENDFVYNPRMSNYAPF--GPV-NRNKL---GKKGVMSPLY  112
ST425-1C                     56  FNNRSINT-NNLTGKVN-VNSKELKNYSVEKGDVFFTRTSEVIGEI--GYP-SVILND--PENTVFSGFV  118
CC30-1strainMRSA252HsdSC     56  FNNRSINT-NNLTGKVN-VNSKELKNYSVEKGDVFFTRTSEVIGEI--GYP-SVILND--PENTVFSGFV  118
CC45-1strain3067HsdSC        56  FNNRSINT-NNLTGKVN-VNSKELKNYSVEKGDVFFTRTSEVIGEI--GYP-SVILND--PENTVFSGFV  118
CC97A                        60  RNGKLNL--NDLVYISK-DIDDEMKNSRTYYGDVLLNITGASI-----GRT-AINSIV--ETHANLNQHV  118
CC1-2strainMW2HsdSA          60  RNGKLNL--NDLVYISK-DIDDEMKNSRTYYGDVLLNITGASI-----GRT-AINSIV--ETHANLNQHV  118
CC1-1strainMW2HsdSA          60  RNGKLNL--NDLVYISK-DIDDEMKNSRTYYGDVLLNITGASI-----GRT-AINSIV--EIHANLNQHV  118
CC5-2strainN315HsdSA         60  RNGKLNL--NDLVYISK-DIDDEMKNSRTYYGDVLLNITGASI-----GRT-AINSIV--ETHANLNQHV  118
CC75-2V                      66  GDGYFYN--SNKVFTSN-EKAEVLKSCNVFPGDIVIAKMADPI-----ARA-AIVPDN-NIGKYLMASDG  125
CC22-1strain5096HsdSB        61  FSKSVS-------------SKNLENYTLIKNGEFAYNKSYSNGYPL--GAI-KRLTR---YDSGVLSSLY  111
CC51TRD2P                    56  FSKSVS-------------SKNLENYTLIKNGEFAYNKSYSNGYPL--GAI-KRLTR---YDSGVLSSLY  106
CC5-1strainN315HsdSB         61  FSKSVS-------------SKNLENYTLIKNGEFAYNKSYSNGYPL--GAI-KRLTR---YDSGVLSSLY  111
CC93-2 b*                    57  KNNKIIIS-SDDRKISESDYKKIYKNYKLEKGDLLLTIVGTI------GRA-AIVKN---PNNIAFQRSV  115
Consensus_ss:                              e            ee     eeeeee         eee eee        eeee  e
```

```
Conservation:                                              7                                            6    98 775
CC80-3 f*                         111  IKFKINE---QNSPLFIYQFTLTSKFNKWVKVMS----VRSGQPGINSEEYAKLPLVLPN-KLEQQKIAK  172
CC45-1L                           110  VGLRVNN---NNLPSFIYYMLSYKGNQKKIKRIQ----MGAVQPSVKVSQFKFIKYLVPI-KDEQEKVAK  171
CC97 c*                           105  PAFEINT--EKIEPNYFISYISRPSFYKSKEKYS----TGTGSKRIHENTVLNFSLHLPC-LNEQLKIAS  167
CC22-1I                           116  IAFNSNE---KITDNFLASLLSSENVYNDLLKLC----SGATAKGVSQKNLNRLYVTIPHSISEQEEIAE  178
CC873D                             99  TVLYPTQ---NTSSLFIGYKFKTHRMIHKFKINSQ--GLTSDTWNLKYKQLKNINIDIPV-LEEQEKIGD  162
CC5-1D                             99  TVLYPTQ---NTSSLFIGYKFKTHRMIHKFKINSQ--GLTSDTWNLKYKQLKNINIDIPV-LEEQEKIGD  162
CC30-1D                            99  TVLYPTQ---NTSSLFIGYKFKTHRMIHKFKINSQ--GLTSDTWNLKYKQLKNINIDIPV-LEEQEKIGD  162
CC5-2H                            105  PVFKGI----DANQYYLTLHLNYQ-LKKEYIKYA----TGTSQLVLSQKDLQNIKTKLPS-YEEQQKIGD  164
CC133-2fromED133 d*               116  IRFRTNH---KIRSEFFLIFMRSNQMQRKILEAN---PGSAITNLVPVKELKLIPFPLPV-KFEQDKISQ  178
CC72-2S                           102  CSIKGD--------KFLYYFLEWFATQNKWIRFS----QGSTFESISGNDIRNIHIKIPV-EDERTKIIK  158
CC93-3 a*                         107  CHFIWKV--NKINTEFIYYYLNS---LNISTFGV----QAVKGVTLNNDSINSIIVKLPN-EEEQNIIAK  166
CC93-2K                           111  LSLSSL----KYDKLYSLYSLLY--TMKKISANL----QGTSIKGITKKELLDSIIKIPHNLEEQQKIGD  170
CC30-2K                           111  LSLSSL----KYDKLYSLYSLLY--TMKKISANL----QGTSIKGITKKELLDSIIKIPHNLEEQQKIGD  170
CC80-2W                           111  LIFLNP----KFNLFFMYYWLEY--FKDKWSKYG----QPGSQVNLNTEIVKSQTLNMPS-NHEQEKVGQ  169
CC75-2W                           111  LIFLKP----KFNLFFMYYWLEY--FKDKWSKYG----QPGSQVNLNSEIVKSQTLNMPS-NHEQEKVGQ  169
CC59Q                              98  HILSPL----NGNIQYLYQVAEL----VNYEKYN----TGTAQPKLNIQNLKIINVVISTNLEEQQKIGS  155
CC72-1Q                            98  HILSPL----NGNIQYLYQVAEL----VNYEKYN----TGTAQPKLNIQNLKIISVVISX----------  145
CC1-2G                             98  HVVKSN----DHNLFFMNYYLNF----KELRAFV----TGNAPAKLTHANLCNINLKIPC-LTEQDKVSA  154
ST425-1E                           92  YKISDFK---NYYGLLLFYYFSQ-NFLKETKKYS----AKTSVDSVRKDMVANMKVPRPI-YIEQEKIGQ  152
CC15TRD2E                          92  YKISDFK---NYYGLLLFYYFSQ-NFLKETKKYS----AKTSVDSVRKDMIANMKVPRPI-YIEQKKIGQ  152
CC133_771E                         92  YKISDFK---NYYGLLLFYYFSQ-NFLKETKKYS----AKTSVDSVRKDMVANMKVPRPI-YIEQEKIGQ  152
CC398-1E                           92  YKISDFK---NYYGLLLFYYFSQ-NFLKETKKYS----AKTSVDSVRKDMIANMKVPRPI-YIEQKKIGQ  152
CC80-1Y                           108  NAVVPFE---NINIDYLYYISDS--LSTFMKSIA----GKTATQIVNKNTFENLEIYLAP-FEEQNKIAD  167
CC75-1U                            94  GYIQSN----NVDIEFLYYRMKV----VDFKKYI----IGSTIPHLYFKDYSKETLYIPSSIQEQAKIGM  151
CC1-1F                            123  ERLIPFA---ENDNKFLWFLMNTDLIRNKIKGMM----QGATQVYINYSSIKLISIQLPL-LEEQQKIRG  184
CC873 e*                          106  SPSFDFI---NGDSKFLLERIKLKSFYKKFGDIA---NGSRKAKRINQDTFLSLPVFAPK-YDEQLRIGE  168
CC80-2Z                           107  WVLKSTKE-DKLLNLFLYYFIQTKRYS-DVASKS----AGSKMPRADWGLIENIRVYFPE-LCEQQKIGQ  169
CC80-3XS.Sau11819ORF2227P         109  KIKFDKN---LYDRLFLVEVLNSSKIQNKILSLA----GSSTIPDLNHSDFYSISSSYPL-LREQQKIGK  170
CC80-1X                           109  KIKFDKN---LYDRLFLVEVLNSSKIQNKILSLA----GSSTIPDLNHSDFYSISSSYPL-LREQQKIGK  170
CC75-1T                           117  SALLMNH---ETNPEFIQAFLNYQV-KGWKRYAA----SSRKDPNITKKDIEQFKVPYVS-INEQQKIGE  177
ST130-1T                          117  SALLMNH---ETNPEFIQAYLNYQV-KGWKRYAA----SSRKDPNITKKDIEQFKVPYVS-INEQQKIGE  177
CC93-3M                           119  HLFRPLN---NAISNFLIFYTKTLSYKKFIRQQG----TGISVLGISKKSLLNLNVLIPRSELEQQKIGQ  181
CC133_771-1strain32320Hsd         119  HLFRPLN---NAISNFLIFYTKTLSYKKFIRQQG----TGISVLGISKKSLLNLNVLIPRSELEQQKVGK  181
CC133-2fromED133J                  93  FYCTPKK---ETDILFILSLFRKIN----WKVYD----ESTGVPSLSKQTINKINRFVPT-NKEQQKIGK  150
CC72-2J                            93  FYCTPKK---ETDILFILSLFRKIN----WKVYD----ESTGVPSLSKQTINKINRFVPT-NKEQQKIGK  150
CC51TRD1J                          93  FYCTPKK---ETDILFILSLFRKIN----WKVYD----ESTGVPSLSKQTINKINRFVPT-NKEQQKIGE  150
CC30-2strainMRSA252HsdSJ           93  FYCTPKK---ETDILFILSLFRKIN----WKVYD----ESTGVPSLSKQTINKINRFVPS-NKEQQKIGE  150
CC59-1J                            93  FYCTPEK---EADILFILSLFRKIN----WKLYD----ESTGVPSLSKQTINKINRLVPT-NKEQQKIGE  150
CC72-1R                           109  ALLKTK----NLNSYFLKNLILSSSIQNELWRKT---LHVAFPKKINKNEIGKIKINYPK-KQEQQKIGQ  170
CC15TRD1O                         116  AIYQTN-----QNINFVFQYFQK--LYEFLRSLS----NEGSQKNLSLSLLKEITLNYPN-EQEQKKIGD  173
CC398-1strain398HsdSN             113  TVFKIQ----NIDLNFIEFYFKSSKWYRFMALNGD-SGARADRFSIKDRTFMEMPLHIPC-MDEQIKIGQ  176
ST425-1C                          119  LRGRPKSGIDLINNNFKRYVFFTNSFRKEMITKS----SMTTRALTSGTAINKMKVIYPVSAKEQKKIGD  184
CC30-1strainMRSA252HsdSC          119  LRGRPKSGIDLINNNFKRYVFFTNSFRKEMITKS----SMTTRALTSGSAINKMKVIYPVSAKEQRKIGD  184
CC45-1strain3067HsdSC             119  LRGRPKSGIDLINNNFKRYVFFTNSFRKEMITKS----SMTTRALTSGTAINRMKVIYPVSAKEQKKIGD  184
CC97A                             119  CIIRLKK---EYYYIFFGQYLLSRKGKRKIFLAQ----SGGSREGLNFKEIANLKIFTPTIFEEQQKIGK  181
CC1-2strainMW2HsdSA               119  CIIRLKK---EYYYNFFGQYLLSRKGKRKIFLAQ----SGGSREGLNFKEIANLKIFTPTIFEEQQKIGK  181
CC1-1strainMW2HsdSA               119  CIIRLKK---EYYYNFFGQYLLSRKGKRKIFLAQ----SGGSREGLNFKEIANLKIFTPTIFEEQQKIGK  181
CC5-2strainN315HsdSA              119  CIIRLKK---EYYYNFFGQYLLSRKGKRKIFLAQ----SGGSREGLNFKEIANLKIFTPTIFEEQQKIGQ  181
CC75-2V                           126  IRLSVDT--VHFNTKFVLECINRKSFRKKVEDNS----SGSTRMRIGLSTLGSLTLKTTT-LKEQQKIGQ  188
CC22-1strain5096HsdSB             112  ICFSIKS---EMSKDFMEAYFDSTHWYREVSGIAVEGARNHGLLNVSVNDFFTILIKYPS-LEEQQKIGK  177
CC51TRD2P                         107  ICFSIKS---EMSKDFMEAYFDSTHWYREVSGIAVEGARNHGLLNISVNDFFTILIKYPS-LEEQRKIGD  172
CC5-1strainN315HsdSB              112  ICFSIKS---EMSKDFMEAYFDSTHWYREVSGIAVEGARNHGLLNVSVNDFFTILIKYPS-LEEQQKIGK  177
CC93-2 b*                         116  AILKTKA---TYDVGFIFQLFQTKYFKNLLLRKQ----VVSAQPGLYLGDIRKIKISITNIIEEQRKIGI  178
Consensus_ss:                          eeeee        hhhhhhhh  hhhhhhhhhh              hhhhh  ee     hhhhhhhhh
```

```
Conservation:                        75   6  5     76  66 597 568568
CC80-3 f*                      173   FLDRFDRQIELEKQKIEILQQQKKGLLQSMFI  204
CC45-1L                        172   LLIEIDKLVNKQLIKIELLQQRKKALLKSMFI  203
CC97 c*                        168   FVCFLNRKIELLERKIYLIKKQKQALLQQMFI  199
CC22-1I                        179   FFRKINQLVELQKYKIEHTKSQKQVFLQKMFI  210
CC873D                         163   FFKKMDILISKQKIKIEILEKEKQSFLQKMFL  194
CC5-1D                         163   FFKKMDILISKQKMKIEILEKEKQSFLQKMFL  194
CC30-1D                        163   FFKKMDILISKQKIKIEILEKEKQSFLQKMFL  194
CC5-2H                         165   FFSEIDRLVEKQSSKVGRLKVRKKELLQKMFV  196
CC133-2fromED133 d*            179   FIHIINRRIEQSEKKIESLKNRKQGFLQKLFV  210
CC72-2S                        159   LLNSLDVLNSKTDLKIQNLKQRKQSLLQKIFV  190
CC93-3 a*                      167   FLLEVDKTVNNQLVKTKLLKQRKKGLLQRMFV  198
CC93-2K                        171   LFYKIDKYISFNKCKIEMLKSLKQGLLKKMFI  202
CC30-2K                        171   LFYKIDKYISFNKCKIEILKSLKQGLLQKIFI  202
CC80-2W                        170   FFNRNEKLIELQQEKIMYLKRRKQVLLQKMFI  201
CC75-2W                        170   FFNRNEKLIELQQEKIMYIKRCKQVLLQKMFI  201
CC59Q                          156   FLSKLDRQIDLEEQKLELLQQRKKALLKSMFV  187
CC72-1Q                              --------------------------------
CC1-2G                         155   LLKSIDNKMNNQMNRIELLKERKKELLQKMFI  186
ST425-1E                       153   FIKKVDNKIKIQKQVIELLKQRKKALLQKMFI  184
CC15TRD2E                      153   FIKRVDNKTKIQKQVIELLKQRKKALLQKMFI  184
CC133_771E                     153   FIKKVDNKIKIQKQVIELLKQRKKALLQKMFI  184
CC398-1E                       153   FIKRVDNKTKIQKQVIELLKQRKKSLLQKMFI  184
CC80-1Y                        168   LISSLEELIEKQASKLIKMKSRKQGMLQIMFI  199
CC75-1U                        152   FISNLDKLIENKNLKLNCLKQLKQGLLQSMFI  183
CC1-1F                         185   FLEVLSGITTKQLHKIDQLKERKKAFLQKMFI  216
CC873 e*                       169   FFSKLDRQIELQKQKLELLQQQKKGYMQKIFS  200
CC80-2Z                        170   FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS  201
CC80-3XS.Sau11819ORF2227P      171   FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS  202
CC80-1X                        171   FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS  202
CC75-1T                        178   FFSKIDHQIELEEQKLELLQQQKKGYMQKIFS  209
ST130-1T                       178   FFSKLDRQIELEEQKLELLQQQKK--------  201
CC93-3M                        182   FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS  213
CC133_771-1strain32320Hsd      182   FFSKLDRQIELEEQKIELLQQQKKGYIQKIFS  213
CC133-2fromED133J              151   FFSKLDRQIELQEQKLELLQQQKKGYMQKIFS  182
CC72-2J                        151   FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS  182
CC51TRD1J                      151   FFSKLDRQIELEEQKLELFQQQKKGYMQKIFS  182
CC30-2strainMRSA252HsdSJ       151   FFIKLDRQIELEEQKLELLQQQKKGYMQKIFS  182
CC59-1J                        151   FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS  182
CC72-1R                        171   FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS  202
CC15TRD1O                      174   FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS  205
CC398-1strain398HsdSN          177   FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS  208
ST425-1C                       185   FFSKLDRQIELEEQKLELLQQQKKGYMQKIFT  216
CC30-1strainMRSA252HsdSC       185   FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS  216
CC45-1strain3067HsdSC          185   FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS  216
CC97A                          182   FFSKLDRQIELEEQKLELLQQQKKGYLQKIFS  213
CC1-2strainMW2HsdSA            182   FFSKLDRQIELEEQKLELLQQQKKGYMQKIFT  213
CC1-1strainMW2HsdSA            182   FISKLDRQIELEEQKLELLQQQKKGYMQKIFS  213
CC5-2strainN315HsdSA           182   FFSKLDQQIELEEQKLELLQQQKKCYIQKIFS  213
CC75-2V                        189   FFSKLDRQIVLEEQKLELLQQQKKGYMQKIFS  220
CC22-1strain5096HsdSB          178   FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS  209
CC51TRD2P                      173   FFIKLDRQIELEEQKLELLQQRKKALLKSMLI  204
CC5-1strainN315HsdSB           178   FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS  209
CC93-2 b*                      179   FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS  210
Consensus_ss:                        hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
```