

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

SCHOLARONE™
Manuscripts

1
2 1 **DNA target recognition domains in the Type I restriction and modification**
3 2 **systems of *Staphylococcus aureus*.**

4 Laurie P. Cooper¹⁺, Gareth A. Roberts¹⁺, John H. White¹⁺, Yvette Luyten²⁺, Edward K.M. Bower¹,
5 Richard D. Morgan², Richard J. Roberts², Jodi A. Lindsay³, David T.F. Dryden^{4*}

6
7 1. EaStCHEM School of Chemistry, University of Edinburgh, The King's Buildings, Edinburgh EH9 3FJ,
8 UK.

9 2. New England Biolabs, 240 County Road, Ipswich, MA 01938-2723, USA.

10 3. Institute of Infection and Immunity, St George's, University of London, Cranmer Terrace, London,
11 SW17 0RE, UK.

12 4. Department of Biosciences, Durham University, Stockton Road, Durham, DH1 3LE, UK.

13
14 + Joint first authors

15 * Author for correspondence:

16 david.t.dryden@durham.ac.uk, Tel. +44 (0)191 3341200

17
18
19 Keywords: DNA restriction, DNA modification, epigenetics, *Staphylococcus aureus*, Type I restriction
20 system, endonuclease, methyltransferase.

21
22
23 Abbreviations: Restriction and Modification, RM; Horizontal Gene Transfer, HGT; hsd, host specificity
24 for DNA; Methyltransferase, MTase; N6-methyl adenine, m6A; N4-methyl cytosine, m4C; C5-methyl
25 cytosine, 5mC; Clonal Complex, CC; Sequence Type, ST; Target Recognition Domain, TRD; Single
26 Molecule Real Time, SMRT; Enhanced Green Fluorescent Protein, EGFP; Polymerase Chain Reaction,
27 PCR; ATP hydrolysis, ATPase; Lysogeny Broth, LB; S-adenosyl-L-methionine, SAM.

28
29
30 Abstract

31 *Staphylococcus aureus* displays a clonal population structure in which horizontal gene transfer
32 between different lineages is extremely rare. This is due, in part, to the presence of a Type I DNA
33 restriction and modification (RM) system given the generic name of *Sau1*, which maintains different
34 patterns of methylation on specific target sequences on the genomes of different lineages. We have
35 determined the target sequences recognised by the *Sau1* Type I RM systems present in a wide range
36 of the most prevalent *S. aureus* lineages and assigned the sequences recognised to particular target
37 recognition domains within the RM enzymes. We used a range of biochemical assays on purified
38 enzymes and single molecule real-time sequencing on genomic DNA to determine these target
39 sequences and their patterns of methylation. Knowledge of the main target sequences for *Sau1* will
40 facilitate the synthesis of new vectors for transformation of the most prevalent lineages of this
41 “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 R₂M₂S₁ 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 M₂S₁ 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 *hsdM*/*hsdS* 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

1
2
3 95 informal one not following the usual conventions (34), we retain it as it is established in the
4 96 literature. However, it is important to note that some strains of *S. aureus* show additional Type I RM
5 97 systems, which show limited amino acid sequence identity to the HsdR, HsdM and HsdS of *Sau1*
6 98 ([Figure 1a](#)). For instance, Monk *et al.* (35) identified an active Type I RM system, SauJKDIII, in *S.*
7 99 *aureus* JKD6159 which showed low sequence identity to members of the *Sau1* family. This is clearly
8 100 a member of a new and different Type I RM family whose subunits will be unable to interact with the
9 101 *Sau1* HsdM and HsdR (DTFD, JAL and MTG Holden, manuscript in preparation).
10 102

11 103 The *Sau1* Type I RM systems are so effective because they show great variability in the target
12 104 sequences recognised thus preventing HGT between CC but allowing HGT between strains within a
13 105 CC (31,35,36). This variability in target sequences is due to the modular construction of the Type I
14 106 RM systems ([Figure 1b](#)). The S subunit contains two target recognition domains (TRD) each of which
15 107 recognises one half of a bipartite target, for example the first Type I RM system in CC1, given the
16 108 generic name CC1-1, recognises CCAYNNNNNTTAA (adenine methylation sites are underlined)
17 109 (35,36). Swapping TRDs between S subunits generates new targets, for example the second Type I
18 110 RM enzyme in CC1, termed CC1-2, couples the first TRD of CC1-1 with a different second TRD to
19 111 recognise CCAYNNNNNNTGT. This swapping is easy because the DNA for S subunits contain
20 112 conserved sequences bounding each TRD. Most *S. aureus* strains have two copies of *hsdS*, two of
21 113 *hsdM* and one of *hsdR*. Thus there are often four TRDs in each CC, which define the restriction
22 114 barrier against HGT. Some Type I RM enzymes have half-size HsdS incorporating only a single TRD. It
23 115 has been shown that these products are often able to dimerise and recognise symmetric target
24 116 sequences (37-39). We have been able to recapitulate these results on "half-HsdS" enzymes by
25 117 manipulating the CC398-1 *S. aureus* system (EKM Bower and DTFD, unpublished results).
26 118
27 119

28 120 Previously we have identified the target sequences recognised by several common community-
29 121 associated, hospital-associated and livestock-associated) MRSA clonal complexes (31,36) and
30 122 recently several more have been identified (3,35,40). Monk *et al.* (35) and Jones *et al.* (40) have used
31 123 this information to prepare DNA methylated by the MTase M₂S₁ component enzymes to aid the
32 124 transformation of *S. aureus* strains that are usually resistant to transformation.
33 125
34 126

35 127 The identification of further targets recognised by the S subunits of *Sau1* Type I RM systems would in
36 128 principle allow more CC to be transformed for genetic analysis. In addition, further understanding of
37 129 the structural requirements for TRDs to recognise different specific DNA sequences is of intense
38 130 interest as the Type I RM systems are very widespread in bacteria and archaea (1,4) and exert a
39 131 considerable pressure on HGT and the evolution of prokaryotes. For instance, the use of multiple
40 132 TRDs being exchanged between strains has been observed in *Helicobacter* (41), *Mycoplasma* (42,43),
41 133 *Streptococci* (44,45) and *Bacteroides* (46).
42 134
43 135

44 136 Here we identify many further TRDs and their targets using both biochemical and PacBio single-
45 137 molecule real-time (SMRT) sequencing methods to define the barriers to HGT in a wide range of *S.*
46 138 *aureus* CC of global importance.
47 139
48 140

49 138 Materials and Methods

50 139 Nomenclature for expression plasmids encoding new MTases.

51 140 As each Type I S subunit contains two TRDs and we propose to determine the targets recognised by
52 141 each TRD, we have given each TRD a single letter code, Table 1, and refer to the plasmids as
53 142 pSauTRD1-TRD2, e.g. pSauBI expresses an S subunit containing TRD B and TRD I and the M subunit. If
54 143 the TRD combination is the same as that found in a known clonal complex, then that CC is also given
55 144 in brackets. The MTase would be called M.SauBI in this example and the S subunit S.SauBI and is
56 145 from CC22. All sequences are given in the supplementary information.
57 146
58 147
59 148
60 149

1
2
3
4 146
5 147 **Preparation of M.SauBI (CC22-1), M.SauCD (CC30-1), M.SauJK (CC30-2) and M.SauCL (CC45-1).**
6 These four MTases were prepared as EGFP-His tag fusions as described in Roberts *et al.* (31). pSauBI-
7 EGFP (CC22-1, genomic DNA from MRSA5906), pSauCD-EGFP (CC30-1, genomic DNA from MRSA252),
8 pSauJK-EGFP (CC30-2, genomic DNA from MRSA252) and pSauCL-EGFP (CC45-1, genomic DNA from
9 strain 70642) were all constructed by the polymerase chain reaction (PCR) with their *hsdS* fused to
10 DNA encoding EGFP and a His-tag, with the following locus-specific oligonucleotides priming from
11 the 3' end of the genes encoding the S subunits:
12

13 154 CC22-1 BI BS 5'GATCGAATTCCGGATCCAATAAACATCTTTGTAAAAACAC3'
14 155 CC30-1 CD BS 5'GATCGAATTCCGGATCCTAAGAACATCTTTGTAAAAGG3'
15 156 CC30-2 JK BS 5'GATCGAATTCCGGATCCTATAAAATTTTGAAAGTAATCCTTG3'
16 157 CC45-1 CL R167K BS 5'GATCGAATTCCGGATCCAATAAACATCGATTTAAGTAAGGC3'

17 The sequence for CC45-1 introduced a single mutation R167K in the first TRD in the S subunit but
18 since this change is found in other *S. aureus* isolates containing this TRD, the change is presumed to
19 be completely neutral.
20

21 162 **A new vector for MTase expression: pJF118his.**

22 Although we had not experienced problems in examining the fusion proteins of S subunits and EGFP
23 in biochemical work, we decided to construct a vector encoding *hsdS* with only a C-terminal His-tag.
24 Vector pJF118his was made by PCR of the plasmid encoding the MTase CC5-1-EGFP constructed in
25 Roberts *et al.* (31) with these two oligonucleotides:

26 pJFMShisTS 5'AGCTTCGAGAGGATCCCATCATCATCATCATCATTAAGAACATTAGCTTGGCTGTTTGGCGG3'
27 and pJFMSEGFPHisBS 5'GAGTGAATCCCCGGGGATCCGTCGACC3'.

28 The resulting PCR product was cut with BamHI and unimolecular religation gave pJF118his into
29 which the *hsdMS* operon could be ligated as BamHI fragments and from which all subsequent MTase
30 clones were descended.
31

32 173 **Construction of an MTase plasmid to allow TRD swaps: pSaudeltaXmal.**

33 A PCR-based strategy was devised to allow free pairwise assortment of desired TRDs in HsdS. Many,
34 but not all of the HsdS subunits, including that encoded by the Type I system in CC398 (36), have a
35 predicted proline-glycine sequence near the N-terminus. This dipeptide can be encoded by CCCGGG,
36 which would be a target site for SmaI or XmaI. Oligonucleotides were designed which would
37 introduce this motif in the N-terminus (a replacement with no amino acid changes) and at the C
38 terminus (an insertion of two amino acids) of the S subunit of the CC398 system (36), by a two stage
39 PCR fusion. Thus, primary PCR products were generated by reactions primed by: PromoterJF
40 5'GCTTCTGGCGTCAGGCAGCC3' with 398SmaIOligoBS
41 5'CCCATTGCGCTTCAAACCCGGGGAACTCAACTCTGGCAC3' and 398SmaIOligoTS
42 5'GTGCCAGAGTTGAGATTCCCCGGGTTGAAGGCCAATGGG3' with 398SmaIBamHI
43 5'GATCGATCGGATCCCCGGGAATAAACATCTTTGAAGTAATGAC3'.

44 The purified PCR products were then fused in a secondary PCR reaction primed by PromoterJF with
45 398SmaIBamHI. The product was then cut with BamHI, and ligated into the BamHI site of pJF118his
46 as pSauNE-XmaI. This mutated form of the CC398-1 MTase, could assemble the complete restriction
47 enzyme that proved to be active in endonucleolytic cleavage (36). This indicated that insertion of a
48 proline and glycine towards the C-terminus did not affect the function of the enzyme. Subsequently,
49 on reanalysing the DNA sequence, a single PCR mutation was discovered within the XmaI fragment.
50 This caused a mutation A50S but this clearly did not affect the specificity or function of the S subunit
51 in our assays. Digestion of pSauNE-XmaI with XmaI followed by intramolecular religation of the
52 vector fragment generates pSaudeltaXmal, into which any pairwise combination of TRDS with XmaI
53 cohesive ends may be inserted.
54

1
2
3 196 **Construction of MTases M.SauNI, M.SauND, M.SauNK, M.SauNL, M.SauBE, M.SauJE and M.SauCE**
4 197 **(ST425-1) containing hybrid S subunits.**

5 198 The DNA for each TRD of these S subunits was fused to the DNA for the reciprocal TRD of S.SauNE
6 199 (CC398-1). This was achieved by creating primary PCRs with a short area of homology, which then
7 200 allowed base pairing of single strands of each PCR, in a secondary PCR. For example, S.SauBE TRD B
8 201 was generated from an appropriate plasmid template by PCR with oligonucleotides,
9 202 TRD1FOR398SmaI OligoTS 5'GTGCCAGAGTTGAGATTCCCCGGGTTGAAGGCGAATGGG3' paired with
10 203 TRD1nearuniversal 5'GTTCTTCTAATTCAATTGT3'. TRD E was similarly generated by PCR from
11 204 plasmid template with oligonucleotides TRD2nearuniversal 5'ACAAATTGAATTAGAAGAAC3' and
12 205 398SmaIBamHI 5'GATCGATCGGATCCCCGGGAATAAACATCTTGAAAGTAATGAC3'. The final insert
13 206 was then generated by PCR with the two gel-purified primary oligonucleotides and
14 207 TRD1FOR398SmaI OligoTS 5'GTGCCAGAGTTGAGATTCCCCGGGTTGAAGGCGAATGGG3' and
15 208 398SmaIBamHI 5'GATCGATCGGATCCCCGGGAATAAACATCTTGAAAGTAATGAC3'. S.SauCL was the
16 209 only subunit for which we could not use the central universal oligonucleotides for PCR and required
17 210 specific substitutes: TRDLFOR/CC45-1
18 211 5'ACAAATTGAATTAGAAGAACAAAAACTGAAATTACTTCAACACAG3' and TRDC/CC45-1
19 212 5'GTTCTTCTAATTCAATTGTGATCGATCGAGTTGCTGAAGAAG3'. Each C-terminus is unique and where
20 213 TRD2 was not TRD E, a specific oligonucleotide was employed: TRDIREV/CC22-1c-termsmal
21 214 5'GATCGATCGGATCCCCGGGAATAAACATCTTGAAAGAAC3', TRDDREV/CC30-1c-termsmal
22 215 5'GATCGATCGGATCCCCGGGAATAAACATCTTGAAAGGATTG3', TRDKREV/CC30-2c-termsmal
23 216 5'GATCGATCGGATCCCCGGGTATAAAATTGGTGAAGTAATCCTTG3' and TRDLREV/CC45-1c-termsmal
24 217 5'GATCGATCGGATCCCCGGGAATAAACATCGATTAAAGTAAGGC3'. Each pure secondary PCR product
25 218 was cut with XmaI and ligated into the XmaI site of pSaudeltaXmaI.
26
27
28
29 220 **Construction of further MTases with further combinations of TRDs using synthetic genes.**
30 221 Additional *hsdS* sequences were obtained as synthetic genes from GeneArt (ThermoFisher Scientific)
31 222 with sequences optimised for expression in *E. coli* (Supplementary information). All the first TRDs
32 223 begin with 5'CCCGGGTTGAAGGCGAATGGGAG3', except that for CC80-2 which begins with
33 224 5'CCCGGGTTGAAGGCGAATATTCT3'. All the first TRDs end with
34 225 5'CAAATTGAATTAGAAGAACAGAACAGAAG3'. All the second TRDs begin with
35 226 3'CAAATTGAATTAGAAGAACAGAACAG5' and have a universal reverse oligonucleotide, Trd2unirev
36 227 5'GATCGATCGGATCCCCGGG3'. These conserved sequences were used to create oligonucleotides to
37 228 prime PCR reactions. Each pure secondary PCR product was cut with XmaI and ligated into the XmaI
38 229 site of pSaudeltaXmaI. The orientation of the fragments was determined by PCR.
39
40
41 231 **Expression and purification of MTases.**
42 232 These new MTases and the R subunit of CC5 were expressed in *E. coli* BL21(DE3) and purified via
43 233 HisTrap chromatography, size exclusion chromatography, diethylaminoethyl (DEAE) anion exchange
44 234 chromatography and, if necessary, Heparin HiTrap chromatography (GE Healthcare, Uppsala,
45 235 Sweden) as described previously (31).

46
47 237 **Nuclease and ATPase assays.**

48 238 Purified MTases were mixed with the CC5 R subunit and used in assays for ATP hydrolysis (ATPase)
49 239 activity (coupled enzyme assay following a change in absorbance of NADH) and DNA cleavage
50 240 activity (plasmid cutting assay with analysis via agarose gel electrophoresis) as previously described
51 241 (31,36).

52
53 243 **Preparation of genomic DNA for SMRT sequencing.**

54 244 The expression plasmids harbouring the various MTases were used to transform a non-methylating
55 245 (*dam*⁻ *dcm*⁻) strain of *E. coli* ER2796 (30). Single colonies from the transformation plate of Lysogeny
56 246 Broth (LB) agar medium supplemented with 10 µg/ml kanamycin, 10 µg/ml tetracycline as well as

1
2
3 247 100 µg/ml carbenicillin, which acted as a selection marker for the expression construct, were picked
4 and used to inoculate 5 mL of LB containing the same cocktail of antibiotics. The cultures were
5 incubated overnight with shaking at 37°C and 1 mL aliquots of the overnight culture were then
6 pelleted by centrifugation (6000 g, 6 min, 4°C). The culture medium was carefully removed and the
7 cell pellets stored at -20°C until required. Genomic DNA was prepared from each cell pellet using the
8 Wizard Genomic DNA purification kit (Promega, Madison, WI) according to the manufacturer's
9 instructions. The quality of the genomic DNA preparations was initially assessed by agarose gel
10 electrophoresis and from the shape of the absorbance profile from 240 to 340 nm. Genomic DNA
11 from *S. aureus* strains LGA251 (a kind gift from Mark Holmes) and NCTC13435 (a kind gift from
12 Angela Kearns) was prepared by using the PurElut Bacterial Genomic Kit (EdgeBio, Gaithersburg, MD
13 20877, USA). The DNA library for SMRT sequencing was prepared and subsequently analysed as
14 described in Anton *et al.* (30).
15
16 259
17
18 260 **Methylation of plasmids using M.EcoGII.**
19 M.EcoGII was kindly supplied by Dr. Iain Murray (New England Biolabs) and used to modify plasmids
20 E2, E5, E10, E11 and E12 previously described (31) and plasmid pCN36 (47). 0.45 µg DNA was
21 methylated using 2.0 U of M.EcoGII for 100 min at 37°C in a 50 µl volume. The reaction was in
22 1xNEB4 buffer (50 mM potassium acetate, 20 mM Tris acetate, 10 mM Mg acetate, 1 mM DTT (pH
23 7.9@25°C) supplemented with 320 µM S-adenosyl-L-methionine (SAM). As a negative control, DNA
24 was incubated in the same buffer without M.EcoGII. The DNA samples were then supplemented with
25 ATP (20 µM) and additional SAM (160 µM) and then digested with a Type I enzyme (CC5-1, CC5-2,
26 CC30-1, CC45-1 or the NY TRD hybrid) for 14 min at 37°C. As a control, methylated and
27 unmethylated DNA was digested with EcoRI.
28
29 271
30
31 272 **Results and Discussion**
32
33 273 **Assigning TRDs to target sequences.**
34 Each TRD was given a one letter code (A to Z and a* to f*), Table 1. There were 14 TRD1 examples
35 and 18 TRD2 examples in our survey and these are found in 17 different CC or ST groups. Table 1 lists
36 the target specificity and site of methylation for each TRD in our survey. These data were obtained
37 by pairing TRDs and determining the complete target for each TRD pair as described in the next
38 section and in full in the supplementary information. Of interest are the TRD pairs B and P and U and
39 c*. These pairs recognise the same DNA sequence namely AGG and GAY respectively. Amino acid
40 sequence comparisons of B with P and U with c* are shown in Figure 2.
41
42 282 TRD B and TRD P are virtually identical throughout the TRD region even though TRD B is the first TRD
43 in the HsdS subunit and TRD P is the second TRD in the HsdS subunit, Figure 2a. While the high level
44 of sequence identity is expected for Type I systems in the same family, the high level of identity
45 between TRDs found in the first or second position in the HsdS subunit is more unusual. However,
46 such a situation has previously been observed in comparisons of the Type I systems in *Salmonella*
47 *blegdam* and *E. coli* R124 (48).
48
49 289 In contrast, TRDs U and c* are both examples of the second TRD in the HsdS subunit recognising 5'-
50 GAY-3' but the level of identity between them is much lower (~36%) (Figure 2b). This level of identity
51 between TRDs recognising the same target is expected if the TRDs are from different Type I RM
52 families so the low level of identity observed here is unusual. Despite this low level of sequence
53 identity, the predicted secondary structure elements are the same as expected from the early work
54 of Sturrock and Dryden (49). In fact, all of the TRDs in the *Sau1* family of RM systems align well when
55 secondary structure elements are taken into consideration (50) and they will have the same protein
56 fold (Supplementary information: PROMALS alignments). Therefore, it should in future be possible
57
58
59
60

1
2
3 297 to predict the precise amino acid to nucleotide contacts involved in sequence recognition as was
4 298 done for the Type IIG TRDs (51,52).
5
6
7
8
9 302 **Determination of complete target sequences recognised by pairs of TRDs.**
10 303 Tables 2, 3 and 4 show the TRD combinations investigated in this work and those investigated
11 304 previously by ourselves and others along with their combined target sequences, methylation
12 305 specificity and the methods used to determine these parameters. The full experimental data are
13 306 given in the supplementary information. Many of the TRDs were investigated in more than one
14 307 MTase and in more than one assay thus our set of data represents a self-consistent set. DNA
15 308 cleavage and ATP hydrolysis assays were performed on purified MTases mixed with purified R
16 309 subunit while SMRT data were collected from *E. coli* genomic DNA isolated after the hosts were
17 310 transformed with a plasmid expressing the MTase or directly from *S. aureus* genomic DNA. The
18 311 adenines targeted for methylation were determined easily by SMRT sequencing but for systems not
19 312 examined in this manner, it was assumed if there was a single adenine in the site recognised by the
20 313 TRD that this was the target for methylation.
21
22
23 315 Table 2 contains systems from a range of CC investigated previously as well as several examined in
24 316 this study. It is important to note that in our work those systems containing M.SauMRSII plus S.
25 317 SauMRSII, M.Sau133ORF1794P plus S.Sau133ORF1794P and M.SauMRSI plus S.SauMRSI are paired
26 318 with the HsdR (SauN315ORF189P) from the N315 strain of CC5 in DNA cleavage and ATPase assays.
27 319 Those shown in Tables 3 and 4 are studied as HsdS paired with the HsdM (M.SauSTORF499P) from
28 320 strain S0385 of CC398 and the HsdR (SauN315ORF189P) from the N315 strain of CC5 (if used in DNA
29 321 cleavage or ATPase assays). Therefore, these HsdS are not examined in the context of their natural
30 322 genome, but since they are all from the *Sau1* family of Type I RM systems and the HsdM and HsdR of
31 323 these RM systems are essentially identical in all of the strains, it is reasonable to assume that the
32 324 target specificities identified are those that would be recognised in their natural host.
33
34
35 326 Identifying the complete target recognised by a member of the *Sau1* Type I RM family when both
36 327 TRDs have unknown targets is difficult and ambiguous as either orientation may be correct. Hence,
37 328 we combined TRDs with unknown targets with TRD E or TRD N to make a protein recognising a
38 329 hybrid sequence in which one half of the target was already known (Table 3). A variety of methods
39 330 were used to determine the target associated with each hybrid including DNA cleavage and ATP
40 331 hydrolysis assays when the hybrid enzyme could be expressed and purified from *E. coli* and SMRT
41 332 sequencing when the expression and purification levels were low, for example, the SauJK enzyme
42 333 corresponding to the second Type I RM enzyme in CC30 did not express in *E. coli* despite its
43 334 expression in *S. aureus* by Monk *et al.* (35). The ambiguity in assignment of targets in CC93 in Monk
44 335 *et al.* (35) is resolved because the TRDs M and b* occur in more than one HsdS in our survey.
45
46
47 337 The DNA sequences for further pairs of TRDs found in a wide range of CC and ST groups were then
48 338 inserted after the *hsdM* of CC398-1 in our expression vector and examined to ascertain the spacer
49 339 sequence in the natural system (Table 4).
50
51 341 Genomic DNA from *S. aureus* strains NCTC13435 and LGA251 was prepared and examined using
52 342 SMRT sequencing as these strains contain two TRD pairs, XY and e*f* respectively, which we could
53 343 not express in *E. coli*. While SMRT signatures for the other Type I HsdS in these strains were very
54 344 clear (Supplementary information) and in agreement with our results from *E. coli* (Table 4) and those
55 345 of Monk *et al.* (35), these TRD pairs still showed no methylation activity even in their normal host.
56 346 Thus these TRDs pairs are not active.
57
58
59
60

1
2
3 348 **Analysis of spacer sequence length in *S. aureus* Type I RM systems.**
4 349 It is apparent that the number of base pairs separating the adenines targeted for methylation and
5 350 the number of base pairs in the non-specific spacer between the sequences recognised by the TRDs
6 351 is not constant, with the former varying between 7 and 9 base pairs and the latter varying between 5
7 352 and 7 base pairs. This variation makes it very difficult to predict a Type I RM recognition sequence if
8 353 one knows only the targets recognised by the two TRDs as the length of the spacer in the target is
9 354 not recognised in any obvious manner by the TRDs. An example of this is the CC80-1 enzyme (Table
10 355 4) containing TRDs X and Y of known specificity. Since the enzyme did not methylate DNA *in vivo* for
11 356 the SMRT analysis, the spacer and hence the complete target for CC80-1 remain unknown until the
12 357 enzyme is purified and analysed biochemically. While it has been observed that insertions of
13 358 multiples of four amino acids into the alpha helical spacers separating the TRDs can increase the
14 359 length of the spacer in the target sequence in a predictable manner (65-67), it is clear from the
15 360 structure of HsdS subunits (Figure 1b) that the junction between the TRDs and the alpha helical
16 361 spacers in the conserved region is going to be of crucial importance for determining the fine details
17 362 of the length of the spacer in the target sequence as was found for some Type IIB RM enzymes
18 363 which contain a subunit equivalent to HsdS (68). Perhaps even single amino acid insertions or
19 364 deletions will serve to rotate the TRD with respect to the rest of the subunit and thereby change the
20 365 length of the spacer. Further progress in understanding the correlation between amino acid
21 366 sequence and the length of the target spacer would be greatly aided by an accurate atomic structure
22 367 of a Type I enzyme with DNA as the current models (12,13) lack sufficient resolution to be
23 368 informative on this point.
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3 399 determine the target sequences of the *S. aureus* Type I enzymes and have been previously described
4 400 (31). These plasmids were then mixed with various purified *S. aureus* Type I restriction enzymes or,
5 401 as a control, the EcoRI restriction enzyme. After one hour of methylation by M.EcoGII, the plasmids
6 402 were completely resistant to digestion by EcoRI and by the *S. aureus* restriction enzymes (Figure 3).
7 403 Furthermore, the shuttle vector pCN36 (47) was also protected from digestion by these same
8 404 enzymes (data not shown). Subsequent experiments using the methylated pCN36 to transform *S.*
9 405 *aureus* were unfortunately entirely unsuccessful (unpublished results by JAL using strains HO5096
10 406 (CC22), JE2 (CC8) and RN4220 (CC8, *hsdR*⁺). The reason for the failure of transformation with the
11 407 highly-methylated pCN36 when it should be resistant to all *Sau1* RM systems is not clear. This result
12 408 may imply a further unrecognised barrier to transformation of *S. aureus* or some aspect of the
13 409 physical properties of highly methylated DNA. Nevertheless, the method using MTases with very
14 410 short target recognition sequences may be of use for transformation of other bacterial species.
15 411
16 412
17 413

18 CONCLUSIONS

19 414 In conclusion, we have determined the target recognition sequences of a considerable number of
20 415 TRDs and HsdS specificity subunits of the Type I RM systems in *S. aureus*. This was achieved using a
21 416 combination of gene synthesis, endonuclease activity, ATP hydrolysis activity and single molecule
22 417 real-time genome sequencing. The systems analysed cover a large proportion of the known
23 418 sequence types and clonal complexes of *S. aureus* and delineate more clearly the barrier to
24 419 horizontal gene transfer within the *S. aureus* population.
25 420

26 421 The data obtained here will allow the construction of new *E. coli* strains for preparing methylated
27 422 shuttle vectors (35) and MTase reagents for *in vitro* methylation of DNA (40) to assist transformation
28 423 of further *S. aureus* strains. However, these approaches are time consuming and it is worth noting
29 424 that the common shuttle vector used for transformation of *S. aureus*, pCN36 (47), contains a target
30 425 site for almost every TRD pair investigated in this paper. This means that pCN36 is inevitably a poor
31 426 vector for transformation of *S. aureus*. The construction of new shuttle vectors completely lacking
32 427 *Sau1* targets via DNA synthesis, coupled with careful analysis of the fragments to be ligated into the
33 428 vector so that they also lack targets, may be an effective way forward to improve transformation of *S.*
34 429 *aureus* now that so many target specificities have been determined. Obviously, the avoidance of the
35 430 sequence AN₆₋₉T, although difficult to achieve without altering protein coding sequences in a vector,
36 431 would be a general method to negate the effect of the Type I RM systems in *S. aureus* and other
37 432 prokaryotes.
38 433
39 434

40 435 Lastly, the determination of so many recognition sequences of Type I RM systems in different
41 436 lineages of *S. aureus*, in effect a "Rosetta Stone", means that now the population structure of *S.*
42 437 *aureus* can be investigated from an epigenetic/evolutionary perspective (4) as performed previously
43 438 with, for example, *H. pylori* (71) and *S. pneumoniae* (72).
44 439
45 440

46 Acknowledgements

47 441 DTFD thanks the Institute of Advanced Study, Durham University for providing a fellowship from
48 442 January to April 2016 and an excellent environment for writing this paper. We thank Dr Iain Murray,
49 443 New England Biolabs for supplying M.EcoGII, Mark Holmes for donating strain LGA251 and Angela
50 444 Kearns for donating strain NCTC13435.
51 445
52 446

53 Source of funding

54 447 This work was supported by Biotechnology and Biological Sciences Research Council grant
55 448 BB/K005804/1 to DTFD and the Wellcome Trust grants GR080463MA to D.T.F.D and 090288/Z/09/ZA
56 449 to D.T.F.D. and J.A.L.
57
58
59
60

1
2
3 450
4 451
5 452
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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.

482 REFERENCES

- 483 1) Oliveira, P.H., Touchon, M. and Rocha, E.P. (2014) The interplay of restriction-modification
484 systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.* **42**,
485 10618-10631.
- 486 2) Roberts, R.J., Vincze, T., Posfai J. and Macelis, D. (2015) REBASE - a database for DNA
487 restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* **43**, D298-
488 D299.
- 489 3) Blow, M.J., Clark, T.A., Daum, C.G., Deutschbauer, A.M., Fomenkov, A., Fries, R., Froula, J.,
490 Kang, D.D., Malmstrom, R.R., Morgan, R.D., Posfai, J., Singh, K., Visel, A., Wetmore, K., Zhao,
491 Z., Rubin, E.M., Korlach, J., Pennacchio, L.A. and Roberts, R.J. (2016) The Epigenomic
492 Landscape of Prokaryotes. *PLoS Genet.* **12**, e1005854.
- 493 4) Oliveira, P.H., Touchon, M. and Rocha, E.P. (2016) Regulation of genetic flux between
494 bacteria by restriction-modification systems. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5658-5663.
- 495 5) Loenen, W.A.M., Dryden, D.T.F., Raleigh, E.A., Wilson, G.G. and Murray, N.E. (2014)
496 Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res.*
497 **42**, 3-19.
- 498 6) Loenen, W.A.M., Dryden, D.T.F., Raleigh, E.A. and Wilson, G.G. (2014) Type I restriction
499 enzymes and their relatives. *Nucleic Acids Res.* **42**, 20-44.
- 500 7) Pingoud, A., Wilson, G.G. and Wende, W. (2014) Type II restriction endonucleases--a
501 historical perspective and more. *Nucleic Acids Res.* **42**, 7489-7527.
- 502 8) Rao, D.N., Dryden, D.T.F. and Bheemanaik, S. (2014) Type III restriction-modification
503 enzymes: a historical perspective. *Nucleic Acids Res.* **42**, 45-55.
- 504 9) King, G. and Murray, N.E. (1995) Restriction alleviation and modification enhancement by
505 the Rac prophage of Escherichia coli K-12. *Mol. Microbiol.* **16**, 769-777.
- 506 10) Murray, N.E., Batten, P.L. and Murray, K. (1973) Restriction of bacteriophage lambda by
507 Escherichia coli K. *J. Mol. Biol.* **81**, 395-407.
- 508 11) Webb, J.L., King, G., Ternent, D., Titheradge, A.J.B. and Murray, N.E. (1996) Restriction by
509 EcoKI is enhanced by co-operative interactions between target sequences and is dependent
510 on DEAD box motifs. *EMBO J.* **15**, 2003-2009.
- 511 12) Kennaway, C.K., Taylor, J.E., Song, C.F., Potrzebowski, W., Nicholson, W., White, J.H.,
512 Swiderska, A., Obarska-Kosinska, A., Callow, P., Cooper, L.P., Roberts, G.A., Artero, J.B.,
513 Bujnicki, J.M., Trinick, J., Kneale, G.G. and Dryden, D.T.F. (2012) Structure and operation of
514 the DNA-translocating Type I DNA restriction enzymes. *Genes Dev.* **26**, 92-104.
- 515 13) Kennaway, C.K., Obarska-Kosinska, A., White, J.H., Tuszynska, I., Cooper, L.P., Bujnicki, J.M.,
516 Trinick, J. and Dryden, D.T.F. (2009) The structure of M.EcoKI Type I DNA methyltransferase
517 with a DNA mimic antirestriction protein. *Nucleic Acids Res.* **37**, 762-770
- 518 14) Morgan, R.D., Luyten, Y.A., Johnson, S.A., Clough, E.M., Clark, T.A. and Roberts, R.J. (2016)
519 Novel m4C modification in type I restriction-modification systems. *Nucleic Acids Res.* **44**,
520 9413-9425.
- 521 15) Loenen, W.A.M. and Raleigh, E.A. (2014) The other face of restriction: modification-
522 dependent enzymes. *Nucleic Acids Res.* **42**, 56-69.
- 523 16) Waldron, D.E. and Lindsay, J.A. (2006) Sau1: a novel lineage-specific Type I Restriction-
524 Modification system that blocks horizontal gene transfer into *Staphylococcus aureus*, and
525 between *S. aureus* isolates of different lineages. *J. Bacteriol.*, **188**, 5578-5585.
- 526 17) Lindsay, J.A. (2010) Genomic variation and evolution of *Staphylococcus aureus*. *Intl J. Med.*
527 *Microbiol.*, **300**, 98-103.
- 528 18) Monk, I.R., Shah, I.M., Xu, M., Tan, M.W. and Foster, T.J. (2012) Transforming the
529 untransformable: application of direct transformation to manipulate genetically
530 *Staphylococcus aureus* and *Staphylococcus epidermidis*. *Mbio.*, **3**, doi:pii: e00277-11.
- 531 19) Lindsay, J.A. (2014) *Staphylococcus aureus* genomics and the impact of horizontal gene
532 transfer. *Intl. J. Med. Microbiol.* **304**, 103-109.

- 1
2
3 533 20) Feil, E.J., Cooper, J.E., Grundmann, H., Robinson, D.A., Enright, M.C., Berendt, T., Peacock,
4 534 S.J., Smith, J.M., Murphy, M., Spratt, B.G., Moore, C.E. and Day, N.P. (2003) How clonal is
5 535 *Staphylococcus aureus?* *J. Bacteriol.* **185**, 3307-3316.
6 536 21) Sung, J.M., Lloyd, D.H. and Lindsay, J.A. (2008) *Staphylococcus aureus* host specificity:
7 537 comparative genomics of human versus animal isolates by multi-strain microarray. *Microbiol.*
8 538 **154**, 1949-1959.
9 539 22) Monecke, S., Coombs, G., Shore, A.C., Coleman, D.C., Akpaka, P., Borg, M., Chow, H., Ip, M.,
10 540 Jatzwauk, L., Jonas, D., Kadlec, K., Kearns, A., Laurent, F., O'Brien, F.G., Pearson, J., Ruppelt,
11 541 A., Schwarz, S., Scicluna, E., Slickers, P., Tan, H.L., Weber, S. and Ehricht, R.A. (2011) A field
12 542 guide to pandemic, epidemic and sporadic clones of methicillin-resistant *Staphylococcus*
13 543 *aureus*. *PLoS One.* **6**, e17936.
14 544 23) Méric, G., Miragaia, M., de Been, M., Yahara, K., Pascoe, B., Mageiros, L., Mikhail, J., Harris,
15 545 L.G., Wilkinson, T.S., Rolo, J., Lamble, S., Bray, J.E., Jolley, K.A., Hanage, W.P., Bowden, R.,
16 546 Maiden, M.C., Mack, D., de Lencastre, H., Feil, E.J., Corander, J. and Sheppard, S.K. (2015)
17 547 Ecological Overlap and Horizontal Gene Transfer in *Staphylococcus aureus* and
18 548 *Staphylococcus epidermidis*. *Genome Biol. Evol.* **7**, 1313-1328.
19 549 24) McCarthy, A.J. and Lindsay, J.A. (2010) Genetic variation in *Staphylococcus aureus* surface
20 550 and immune evasion genes is lineage associated: implications for vaccine design and host-
21 551 pathogen interactions. *BMC Microbiol.* **10**, 173-187.
22 552 25) McCarthy, A.J., Lindsay, J.A. (2012) The distribution of plasmids that carry virulence and
23 553 resistance genes in *Staphylococcus aureus* is lineage associated. *BMC Microbiol.* **12**, 104-111.
24 554 26) McCarthy, A.J., Witney, A.A. and Lindsay, J.A. (2012) *Staphylococcus aureus* temperate
25 555 bacteriophage: carriage and horizontal gene transfer (HGT) is lineage associated. *Front. Cell.*
26 556 *Infect. Microbiol.* **2**:6. doi: 10.3389/fcimb.2012.00006.
27 557 27) McCarthy, A.J., Loeffler, A., Witney, A.A., Gould, K.A., Lloyd, D.H. and Lindsay, J.A. (2014)
28 558 Extensive horizontal gene transfer during *Staphylococcus aureus* co-colonization *in vivo*.
29 559 *Genome Biol. Evol.* **6**, 2697-2708.
30 560 28) Corvaglia, A.R., François, P., Hernandez, D., Perron, K., Linder, P. and Schrenzel, J. (2010) A
31 561 Type III-like restriction endonuclease functions as a major barrier to horizontal gene transfer
32 562 in clinical *Staphylococcus aureus* strains. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 11954-11958.
33 563 29) Xu, S.Y., Corvaglia, A.R., Chan, S.H., Zheng, Y. and Linder, P. (2011) A type IV modification-
34 564 dependent restriction enzyme SauUSI from *Staphylococcus aureus* subsp. *aureus* USA300.
35 565 *Nucleic Acids Res.* **39**, 5597-5610.
36 566 30) Anton, B.P., Mongodin, E.F., Agrawal, S., Fomenkov, A., Byrd, D.R., Roberts, R.J. and Raleigh,
37 567 E.A. (2015) Complete Genome Sequence of ER2796, a DNA Methyltransferase-Deficient
38 568 Strain of *Escherichia coli* K-12. *PLoS One.* **10**, e0127446.
39 569 31) Roberts, G.A., Houston, P.J., White, J.H., Chen, K., Stephanou, A.S., Cooper, L.P., Dryden,
40 570 D.T.F. and Lindsay, J.A. (2013) Impact of target site distribution for Type I restriction
41 571 enzymes on the evolution of methicillin-resistant *Staphylococcus aureus* (MRSA) populations.
42 572 *Nucleic Acids Res.* **41**, 7472-7484.
43 573 32) Barcus, V.A., Titheradge, A.J.B. and Murray, N.E. (1995) The diversity of alleles at the *hsd*
44 574 locus in natural populations of *Escherichia coli*. *Genetics.* **140**, 1187-1197.
45 575 33) Titheradge, A.J.B., King, J., Ryu, J. and Murray, N.E. (2001) Families of restriction enzymes: an
46 576 analysis prompted by molecular and genetic data for type I restriction and modification
47 577 systems. *Nucleic Acids Res.* **29**, 4195-4205.
48 578 34) Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., Blumenthal, R.M.,
49 579 Degtyarev, S.Kh., Dryden, D.T.F., Dybvig, K., Firman, K., Gromova, E.S., Gumpert, R.I., Halford,
50 580 S.E., Hattman, S., Heitman, J., Hornby, D.P., Janulaitis, A., Jeltsch, A., Josephsen, J., Kiss, A.,
51 581 Klaenhammer, T.R., Kobayashi, I., Kong, H., Krüger, D.H., Lacks, S., Marinus, M.G., Miyahara,
52 582 M., Morgan, R.D., Murray, N.E., Nagaraja, V., Piekarowicz, A., Pingoud, A., Raleigh, E., Rao,
53 583 D.N., Reich, N., Repin, V.E., Selker, E.U., Shaw, P.C., Stein, D.C., Stoddard, B.L., Szybalski, W.,

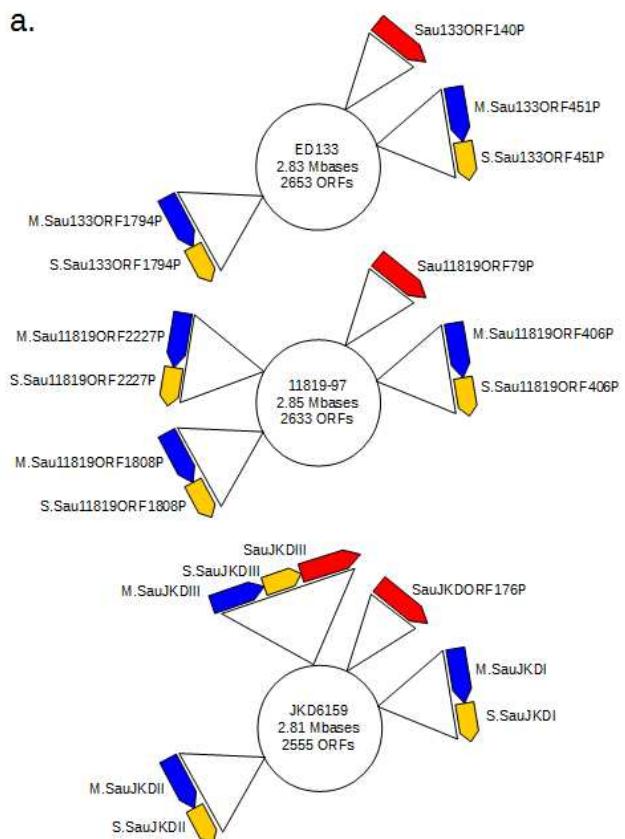
- 1
2
3 Trautner, T.A., Van Etten, J.L., Vitor, J.M., Wilson, G.G. and Xu, S.Y. (2003) A nomenclature
4 for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes.
5 *Nucleic Acids Res.* **31**, 1805-1812.
6 584
7 585
8 586
9 587
10 588
11 589
12 590
13 591
14 592
15 593
16 594
17 595
18 596
19 597
20 598
21 599
22 600
23 601
24 602
25 603
26 604
27 605
28 606
29 607
30 608
31 609
32 610
33 611
34 612
35 613
36 614
37 615
38 616
39 617
40 618
41 619
42 620
43 621
44 622
45 623
46 624
47 625
48 626
49 627
50 628
51 629
52 630
53 631
54 632
55 633
56 634
57 Morgan, R.D. and Luyten, Y.A. (2009) Rational engineering of type II restriction endonuclease
58
59
60
- 584 Monk, I.R., Tree, J.J., Howden, B.P., Stinear, T.P. and Foster, T.J. (2015) Complete Bypass of
585 Restriction Systems for Major *Staphylococcus aureus* lineages. *MBio*. **6**, e00308-15
586
587 35) Chen, K., Stephanou, A.S., Roberts, G.A., White, J.H., Cooper, L.P., Houston, P.J., Lindsay, J.A.
588 and Dryden, D.T.F. (2016) The Type I Restriction Enzymes as Barriers to Horizontal Gene
589 Transfer: Determination of the DNA Target Sequences Recognised by Livestock-Associated
590 Methicillin-Resistant *Staphylococcus aureus* Clonal Complexes 133/ST771 and 398. *Adv. Exp.*
591 *Med. Biol.* **915**, 81-97.
592
593 37) Abadjieva, A., Patel, J., Webb, M., Zinkevich, V. and Firman, K. (1993) A deletion mutant of
594 the type IC restriction endonuclease EcoR1241 expressing a novel DNA specificity. *Nucleic*
595 *Acids Res.* **21**, 4435-4443.
596
597 38) Meister, J., MacWilliams, M., Hübner, P., Jütte, H., Skrzypek, E., Piekarowicz, A. and Bickle,
598 T.A. (1993) Macroevolution by transposition: drastic modification of DNA recognition by a
599 type I restriction enzyme following Tn5 transposition. *EMBO J.* **12**, 4585-4591.
600
601 39) MacWilliams, M.P. and Bickle, T.A. (1996) Generation of new DNA binding specificity by
602 truncation of the type IC EcoDXXI hsdS gene. *EMBO J.* **15**, 4775-4783.
603
604 40) Jones, M.J., Donegan, N.P., Mikheyeva, I.V. and Cheung, A.L. (2015) Improving
605 transformation of *Staphylococcus aureus* belonging to the CC1, CC5 and CC8 clonal
606 complexes. *PLoS One*. **10**, e0119487.
607
608 41) Furuta, Y., Kawai, M., Uchiyama, I. and Kobayashi, I. (2011) Domain movement within a gene:
609 a novel evolutionary mechanism for protein diversification. *PLoS One*. **6**, e18819.
610
611 42) Dybvig, K., Sitaraman, R. and French, C.T. (1998) A family of phase-variable restriction
612 enzymes with differing specificities generated by high-frequency gene rearrangements. *Proc.*
613 *Natl. Acad. Sci. U.S.A.* **95**, 13923-13928.
614
615 43) Xiao, L., Ptacek, T., Osborne, J.D., Crabb, D.M., Simmons, W.L., Lefkowitz, E.J., Waites, K.B.,
616 Atkinson, T.P. and Dybvig, K. (2015) Comparative genome analysis of *Mycoplasma*
617 *pneumoniae*. *BMC Genomics*. **16**, 610. doi: 10.1186/s12864-015-1801-0.
618
619 44) Li, J., Li, J.W., Feng, Z., Wang, J., An, H., Liu, Y., Wang, Y., Wang, K., Zhang, X., Miao, Z., Liang,
620 W., Sebra, R., Wang, G., Wang, W.C. and Zhang JR. (2016) Epigenetic Switch Driven by DNA
621 Inversions Dictates Phase Variation in *Streptococcus pneumoniae*. *PLoS Pathog.* **12**,
622 e1005762.
623
624 45) Willemse, N. and Schultsz, C. (2016) Distribution of Type I Restriction–Modification Systems
625 in *Streptococcus suis*: An Outlook. *Pathogens*, **5**, 62; doi:10.3390/pathogens5040062.
626
627 46) Cerdeño-Tárraga, A.M., Patrick, S., Crossman, L.C., Blakely, G., Abratt, V., Lennard, N., Poxton,
628 I., Duerden, B., Harris, B., Quail, M.A., Barron, A., Clark, L., Corton, C., Doggett, J., Holden,
629 M.T.G., Larke, N., Line, A., Lord, A., Norbertczak, H., Ormond, D., Price, C., Rabbinowitsch, E.,
630 Woodward, J., Barrell, B. and Parkhill, J. (2005) Extensive DNA inversions in the *B. fragilis*
631 genome control variable gene expression. *Science* **307**, 1463-1465.
632
633 47) Charpentier E, Anton AI, Barry P, Alfonso B, Fang Y, Novick RP. Novel cassette-based shuttle
634 vector system for gram-positive bacteria. *Appl Environ Microbiol*. 2004 Oct;70(10):6076-85.
635
636 48) Thorpe, P.H., Ternent, D. and Murray, N.E. (1997) The specificity of sty SKI, a type I
637 restriction enzyme, implies a structure with rotational symmetry. *Nucleic Acids Res.* **25**,
638 1694-1700.
639
640 49) Sturrock, S.S. and Dryden, D.T.F. (1997) A prediction of the amino acids and structures
641 involved in DNA recognition by type I DNA restriction and modification enzymes. *Nucleic*
642 *Acids Res.* **25**, 3408-3414.
643
644 50) Pei, J., Kim, B.H., Tang, M. and Grishin, N.V. (2007) PROMALS web server for accurate
645 multiple protein sequence alignments. *Nucleic Acids Res.* **35**, W649-652.
646
647 51) Morgan, R.D. and Luyten, Y.A. (2009) Rational engineering of type II restriction endonuclease

- 1
2
3 635 DNA binding and cleavage specificity. *Nucleic Acids Res.* **37**, 5222-5233.
4 636 52) Callahan, S.J., Luyten, Y.A., Gupta, Y.K., Wilson, G.G., Roberts, R.J., Morgan, R.D. and
5 637 Aggarwal, A.K. (2016) Structure of Type IIL Restriction-Modification Enzyme Mmel in
6 638 Complex with DNA Has Implications for Engineering New Specificities. *PLoS Biol.* **14**,
7 639 e1002442.
8 640 53) Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K., Oguchi, A., Nagai, Y., Iwama, N.,
9 641 Asano, K., Naimi, T., Kuroda, H., Cui, L., Yamamoto, K. and Hiramatsu, K. (2002) Genome and
10 642 virulence determinants of high virulence community-acquired MRSA. *Lancet* **359**, 1819-1827.
11 643 54) Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki,
12 644 K., Nagai, Y., Lian, J., Ito, T., Kanamori, M., Matsumaru, H., Maruyama, A., Murakami, H.,
13 645 Hosoyama, A., Mizutani-Ui, Y., Takahashi, N.K., Sawano, T., Inoue, R., Kaito, C., Sekimizu, K.,
14 646 Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J., Kanehisa, M., Yamashita, A., Oshima, K.,
15 647 Furuya, K., Yoshino, C., Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H. and Hiramatsu, K.
16 648 (2001) Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. *Lancet* **357**,
17 649 1225-1240.
18 650 55) Holden, M.T., Feil, E.J., Lindsay, J.A., Peacock, S.J., Day, N.P., Enright, M.C., Foster, T.J.,
19 651 Moore, C.E., Hurst, L., Atkin, R., Barron, A., Bason, N., Bentley, S.D., Chillingworth, C.,
20 652 Chillingworth, T., Churcher, C., Clark, L., Corton, C., Cronin, A., Doggett, J., Dowd, L., Feltwell,
21 653 T., Hance, Z., Harris, B., Hauser, H., Holroyd, S., Jagels, K., James, K.D., Lennard, N., Line, A.,
22 654 Mayes, R., Moule, S., Mungall, K., Ormond, D., Quail, M.A., Rabbinowitsch, E., Rutherford, K.,
23 655 Sanders, M., Sharp, S., Simmonds, M., Stevens, K., Whitehead, S., Barrell, B.G., Spratt, B.G.
24 656 and Parkhill, J. (2004) Complete genomes of two clinical *Staphylococcus aureus* strains:
25 657 evidence for the rapid evolution of virulence and drug resistance. *Proc. Natl. Acad. Sci. USA*
26 658 **101**, 9786-9791.
27 659 56) Chua, K., Seemann, T., Harrison, P.F., Davies, J.K., Coutts, S.J., Chen, H., Haring, V., Moore, R.,
28 660 Howden, B.P. and Stinear, T.P. (2010) Complete genome sequence of *Staphylococcus aureus*
29 661 strain JKD6159, a unique Australian clone of ST93-IV community methicillin-resistant
30 662 *Staphylococcus aureus*. *J. Bacteriol.* **192**, 5556-5557.
31 663 57) Guinane, C.M., Ben Zakour, N.L., Tormo-Mas, M.A., Weinert, L.A., Lowder, B.V., Cartwright,
32 664 R.A., Smyth, D.S., Smyth, C.J., Lindsay, J.A., Gould, K.A., Witney, A., Hinds, J., Bollback, J.P.,
33 665 Rambaut, A., Penadés, J.R. and Fitzgerald, J.R. (2010) Evolutionary genomics of
34 666 *Staphylococcus aureus* reveals insights into the origin and molecular basis of ruminant host
35 667 adaptation. *Genome Biol Evol.* **2**, 454-466.
36 668 58) Sung, J.M., Lloyd, D.H. and Lindsay, J.A. (2008) *Staphylococcus aureus* host specificity:
37 669 comparative genomics of human versus animal isolates by multi-strain microarray.
38 670 *Microbiology* **154**, 1949-1959.
39 671 59) Schijfelen, M.J., Boel, C.H., van Strijp, J.A. and Fluit, A.C. (2010) Whole genome analysis of a
40 672 livestock-associated methicillin-resistant *Staphylococcus aureus* ST398 isolate from a case of
41 673 human endocarditis. *BMC Genomics* **11**, 376-386.
42 674 60) Holden, M.T., Hsu, L.Y., Kurt, K., Weinert, L.A., Mather, A.E., Harris, S.R., Strommenger, B.,
43 675 Layer, F., Witte, W., de Lencastre, H., Skov, R., Westh, H., Zemlickova, H., Coombs, G., Kearns,
44 676 A.M., Hill, R.L., Edgeworth, J., Gould, I., Gant, V., Cooke, J., Edwards, G.F., McAdam, P.R.,
45 677 Templeton, K.E., McCann, A., Zhou, Z., Castillo-Ramirez, S., Feil, E.J., Hudson, L.O., Enright,
46 678 M.C., Balloux, F., Aanensen, D.M., Spratt, B.G., Fitzgerald, J.R., Parkhill, J., Achtman, M.,
47 679 Bentley, S.D. and Nubel, U. (2013) A genomic portrait of the emergence, evolution, and
48 680 global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Res.* **23**,
49 681 653-664.
50 682 61) Garcia-Alvarez, L., Holden, M.T., Lindsay, H., Webb, C.R., Brown, D.F., Curran, M.D., Walpole,
51 683 E., Brooks, K., Pickard, D.J., Teale, C., Parkhill, J., Bentley, S.D., Edwards, G.F., Girvan, E.K.,
52 684 Kearns, A.M., Pichon, B., Hill, R.L., Larsen, A.R., Skov, R.L., Peacock, S.J., Maskell, D.J. and
53 685 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., Urbaneck, 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., Rasmussen, 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.



b.

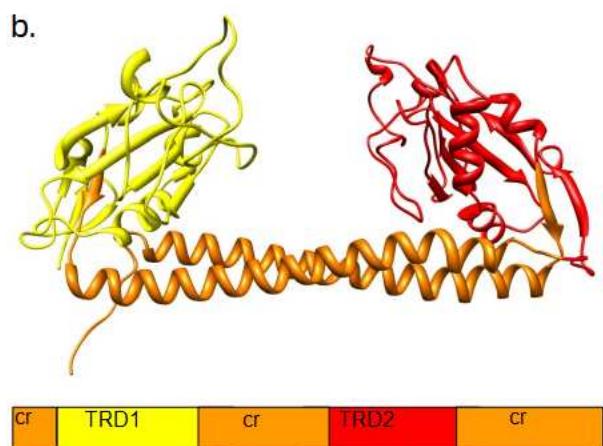


Figure 2.

Figure 3.

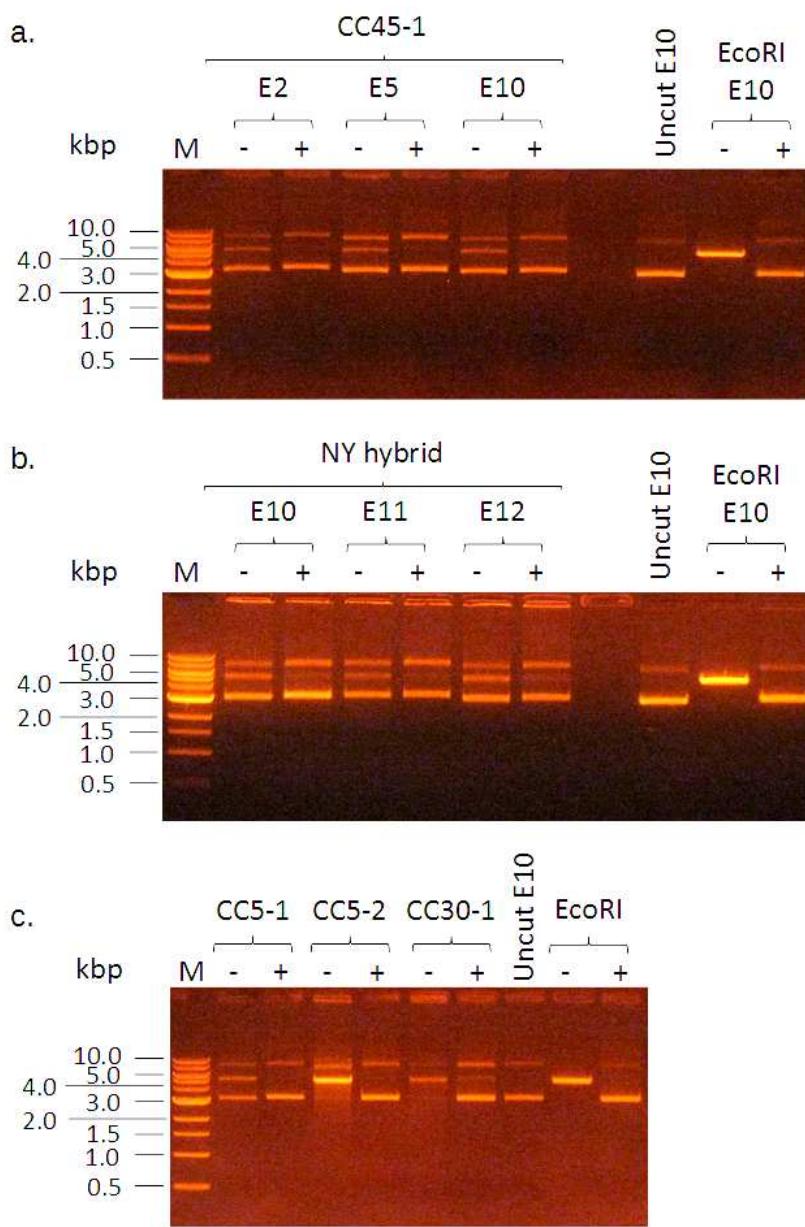


Table 1. TRD targets shown from 5' to 3'. The underlined A indicates the site of adenine methylation. TRD pair B and P, and TRD pair U and c* recognise the same DNA sequence and are highlighted in red. In the targets, Y= C or T; R= A or G; W= A or T; H = A, T or C; N = A, G, T or C.

TRD1 code letter and target	TRD2 code letter and target
A CC <u>A</u> Y	D <u>A</u> TC
B <u>AGG</u>	E TC <u>A</u> Y
C GW <u>A</u> G	F TT <u>AA</u>
J G <u>GA</u>	G AC <u>A</u>
M C <u>A</u> G	H T <u>AC</u>
N <u>AC</u> C	I YT <u>CA</u>
O CA <u>AC</u>	K CG <u>A</u>
R G <u>AR</u> A	L TTT <u>A</u>
T C <u>A</u> AG	P <u>AGG</u>
V CNG <u>A</u>	Q AC <u>A</u> Y
X T <u>CT</u> A	S G <u>CA</u>
Z G <u>AC</u>	U <u>G</u> AY
b*	GG <u>H</u> A
e*	G <u>A</u> G
	W C <u>RA</u> A
	Y C <u>T</u> A
	a* G <u>AA</u>
	c* <u>G</u> AY
	d* CY <u>AA</u>
	f* GA <u>AY</u>

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	CC <u>A</u> Y-5-T <u>A</u> A	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	CC <u>A</u> Y-6-T <u>G</u> T	AG	CC1-2 (CC8-2)	g, s, a	
N315 (54)	CC5	S.SauN315II	CC <u>A</u> Y-6-G <u>T</u> A	AH	CC5-2	g, s, a	s (CC8-1 and CC8-2 in strain NRS384 are from ref. 35)
		S.SauN315I	<u>A</u> GG-5-GAT	BD	CC5-1 (CC8-1)	g, s, a	
MRSA252 (55)	CC30	S.SauMRSII	GW <u>A</u> G-5-GAT	CD	CC30-1	g, s	s (35) g, s (this work)
		S.SauMRSI	G <u>G</u> A-7-T <u>C</u> G	JK	CC30-2	s	
JKD6159 (56)	CC93	S.SauJKDIII	GA <u>A</u> G-5-TAC or complement	Not a <i>Sau1</i> 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	GG <u>H</u> A-7-T <u>C</u> G	b*K	CC93-2	s	
ED133 (57)	CC133	S.SauJKDI	<u>C</u> A <u>G</u> -6-T <u>C</u> T	Ma*	CC93-1	s	g (36) s (this work)
		S.Sau133ORF451P	<u>C</u> A <u>G</u> -5-R <u>T</u> G <u>A</u>	ME	CC133-1	g	
32320 (58)	CC133	S.Sau133ORF1794P	<u>G</u> GA-7-T <u>T</u> R <u>G</u>	Jd*	CC133-2	s	g (36)
		S.Sau32320ORFAP	<u>C</u> A <u>G</u> -5-R <u>T</u> G <u>A</u>	ME	CC133-1	g	
S0385 (59)	CC398	S.SauSTORF499P	<u>A</u> CC-5-R <u>T</u> G <u>A</u>	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
<u>A</u> GG-5- <u>T</u> G	BE	a	<u>A</u> CC-6- <u>T</u> TC	Na*	s
GGA-6- <u>T</u> G	JE	g, s	<u>A</u> CC-6- <u>T</u> TC	Nc*	s
<u>A</u> CC-6- <u>T</u> GAR	NI	g	<u>A</u> CC-6- <u>T</u> TRG	Nd*	g, s
<u>A</u> CC-6- <u>T</u> CG	NK	g	GARA-6- <u>T</u> G	RE	s
<u>A</u> CC-6- <u>T</u> AAA	NL	g	CA <u>A</u> G-5- <u>T</u> G	TE	s
<u>A</u> CC-5-C <u>C</u> T	NP	s	CNGA-6- <u>T</u> G	VE	s
<u>A</u> CC-5-R <u>T</u> GT	NQ	g, s	TCT <u>A</u> -6- <u>T</u> G	XE	g, s
<u>A</u> CC-6- <u>T</u> GC	NS	s	GAC-5- <u>R</u> G	ZE	a
<u>A</u> CC-5-R <u>T</u> C	NU	g, s	G <u>A</u> C-6- <u>T</u> GC	ZS	a
<u>A</u> CC-6- <u>T</u> YG	NW	g, s	GG <u>A</u> A-6- <u>T</u> G	b*E	s
<u>A</u> CC-6- <u>T</u> AG	NY	g, s	G <u>A</u> G-6- <u>T</u> G	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	CC <u>A</u> Y-6- <u>T</u> TC	Ac*	CC97-1	s
HO5096 (60) LGA251 (61)	CC22 ST425	S.Sau5096I	<u>A</u> GG-6- <u>T</u> GAR	BI	CC22-1	g, s
		S.Sau251I	GW <u>A</u> G-5- <u>T</u> GA	CE	ST425-1	g, s*
		S.Sau251ORF16900P	<u>G</u> AG-?- <u>T</u> TC	e*f*	ST425-2	Not expressed, no signature with s*.
		S.Sau251II	GA <u>A</u> G-5-TAC or complement	Not a <i>Sau1</i> system	Same as CC93-3	s*
Isolate 3 (19)	CC51	S.Sau3ORFAP	GG <u>A</u> -6-CCT	JP	CC51-1	s
Isolate 3067 (19)	CC45	S.Sau347I	GW <u>A</u> G-6- <u>T</u> AAA	CL	CC45-1	g
Isolate 3150 (19)	CC15	S.SauL315ORFAP	CA <u>A</u> C-5- <u>T</u> GA	OE	CC15-1	s
SA40 (62)	CC59	S.SauSA40ORF370P	GG <u>A</u> -6- <u>T</u> TGT	JQ	CC59-1	a
CN1 (63) MSHR1132 (64) NCTC13435 NCBI Biosample identifier: SAMEA2479566	CC72	S.SauCN1ORF415P	GA <u>R</u> A-6- <u>T</u> TGT	RQ	CC72-1	a
		S.SauCN1ORF1757P	GG <u>A</u> -7- <u>T</u> GC	JS	CC72-2	a
		S.Sau1132ORF3780P	CA <u>A</u> G-5- <u>T</u> TC	TU	CC75-1	g
	CC75	S.Sau1132ORF16570P	CNG <u>A</u> -7- <u>T</u> YG	VW	CC75-2	s
		S.Sau13435ORF394P	TCT <u>A</u> -?- <u>T</u> AG	XY	ST80-1	Not expressed, no signature with s or s*.
		S.Sau13435ORF1751P	GAC-6- <u>T</u> YG	ZW	ST80-2	a, s*
	ST80	S.Sau13435ORF2165P	TCT <u>A</u> -6- <u>T</u> TC	Xf*	ST80-3	s, s*
32326 (58)		S.Sau32326ORFAP	GA <u>G</u> -6-GAI	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	<i>S. schweitzeri</i> 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	<i>S. argenteus</i> 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 MSNTQKKNVPELRFPGFEGEWEEKKLGEVATFAKGKLGAKKDVSQNGVPVILYGELYTKYGAIVSKIFS KTDI PENKLKMAKKNDVLIPSSGETAIDIATASCIYLNKGVA VGGDINILTPQKQDGRFISLSIN GINKNELSKYAQGKTVVHLYNNNDIKNLKIAFPSEFEEQVRIGNFFSKLDRQIELEEQQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPKWEEKKIEDIASQVYGGGT PNTKIKEFWNGDIPWIQSSDVKVNLD ILQQCNKFISKNSIELSSAKLIPANSIAIVTRVGVGKLCIVEFDYATSQDFSLSSLKYDKLYSLYTMKKISANLQGTSIKGITKELLDSTIKIPHNLLEEQQKIGDLFYKIDKYISFNKCKIEILKSLK QGLLKKMFI
Species KPL1845. Bioproject accession: PRJNA169473
> S.Sau1845ORF1619P TRD NOVEL 2 + NOVEL 3 MTEQINTPELRFPEFKNEWSYDLVDVVTNKSKKFDPKKEAKKDI ELD SIEQNTGRLLDTYISNDFTSQKNKFNKG NVLYSKLRPYLNKYYATIDGVCSSEIWVLNTLNKDV LANKFLYYFI QTNRFSSVTN KSAGSKMPRADWEVKNIRLYKG SIEEQE KIGYFFSKLDRQIELEEKKLELLQQQKKGYMQKIFFAQELRFKDENGNDYPDWTKKLGDIGKVAMNKRIYKNETTENGEIPFYKIGNFGKNADTFITREKFDEYK EKYPYPNVGDILISASGSIGRTIEYTGEDAYYQDSNIVWLHNHDEVINKYLKYFYKIVKWSGIEGTTIKRLYNKNILNTKIELPTVEEQYKMANFLSKLDKIIDIQIEKIELLKQRKQGLLQKMFV
> S.Sau1845ORF2199P TRD NOVEL 4 + TRD f* MSNTQKKNVPELRFPEFE GEWVKDFVVSIFQEVSNKTSIDLAKYPLFSILTVEKGITPKTERYKRDFLVKKSDNFKIVEPRDIVNPMNVTLGAI DLSK NYDIALSGYYHVMKIINSFNPDFISNFLKTEKMIIH YKKIATGSLMEQQRVHFSEFKNIKKFPTNKEQQKIGDFFSKLDRQIELQVQKLELLQQQKKGYMQKIFSQELRFKDENGEDYPDWKEKLG DITEQSMY GIGASATRFDSKNIYIRITDIDEKS RKLNYQNLT TPDELNNKYKLKRNDILFARTGASTGKSYIHKEEKDIYNNYFAGFLIKFEIDEQNNPLFIYQFTLTSKFNWKVMSVRSGQPGINSEYYAKLPLVLPN KLEQQKIAEFLDRFDQQIELEKQKIEILOQQOKGL LQSMFI

1
2 **SUPPLEMENTARY INFORMATION**3
4 DNA target recognition domains in the Type I
5 restriction/modification systems of *Staphylococcus aureus*.6
7 Laurie P. Cooper, Gareth A. Roberts, John H. White, Yvette Luyten,
8 Edward K.M. Bower, Richard D. Morgan, Richard J. Roberts, Jodi A.
9 Lindsay, David T.F. Dryden.10
11 Pages 2 and 3: SUPPLEMENTARY INFORMATION FOR TABLE 1.12
13 Pages 4 to 9: Supplementary information for MATERIALS AND METHODS
14 SECTION "Construction of further MTases with further combinations
15 of TRDs using synthetic genes."16
17 Pages 10 to 16: SUPPLEMENTARY INFORMATION FOR TABLE 2.18
19 Pages 17 to 57: SUPPLEMENTARY INFORMATION FOR TABLE 3.20
21 Pages 58 to 85: SUPPLEMENTARY INFORMATION FOR TABLE 422
23 Pages 86 to 91: SUPPLEMENTARY INFORMATION FOR TABLES 5 AND 6.24
25 Pages 92 to 97: PROMALS ALIGNMENT OF TRD AMINO ACID SEQUENCES WITH
26 SECONDARY STRUCTURE PREDICTIONS.

1
2 **SUPPLEMENTARY INFORMATION FOR TABLE 1.**
3

4 The amino acid sequences in FASTA format of the first TRD with its
 5 letter code, DNA target (5' to 3') and methylation site underlined.
 6 The TRDs labelled as NOVEL 1, NOVEL 2 and NOVEL 4 were found once
 7 all of the other TRDs had been analysed but are included in this
 8 list for completeness.

9 The TRD sequences are flanked by the conserved regions so to
 10 obtain the amino acid sequence of any HsdS subunit simply paste
 11 the sequence for the second TRD directly on to the end of the
 12 sequence for the first TRD.
 13

14 >A CCAY
 15 MSNTQKKNVP~~E~~LRFPGFEGEWEEKQLGDLTTKIGSGKTPKGGS~~E~~N~~T~~NKGIPFLRSQNIRNGKLNLDLVYISKDIDDEMKN~~S~~R~~T~~YYGDVLLNITGASIG
 16 RTAINSIVETHANLNQHVC~~I~~IRLKKEYYYIFFGQYLLSRKGKR~~K~~IFLAQSGGS~~R~~ELNFKEIANL~~K~~IFTPTIFEEQQKIGKFFSKLDRQIELEEQKLELL
 17 QQQ
 18 >B AGG
 19 MSNTQKKNVP~~E~~LRFPGFEGEWEEKQLGDLTDRVIRKNKNLESKKPLTISGQLGLIDQTEYFSKS~~V~~SSKNLEN~~T~~LIKNGEFAYNKSYSNGYPLGAIKRLT
 20 RYDGVVLSSLYICFSIKSEM~~S~~KDFMEAYFDSTHWYREVSGIAVEGARNHG~~L~~LN~~V~~SVNDFFTILIKYPS~~L~~EEQQKIGKFFSKLDRQIELEEQKLELLQQQ
 21 >C GWAG
 22 MSNTQTKNVP~~E~~LRFPGFEGEWEEKQVGELLEFKNGLNKGKEYFGSGSSIVNFKDVFNNRSINTNNLTGKVNVNSKELKN~~S~~V~~E~~KGDVFFFTRSEVIGEIG
 23 YPSV~~L~~NDPENTVFSGFVLRGRPKSG~~I~~D~~L~~NNNFKRYVFTNSR~~K~~E~~M~~ITKSSMT~~T~~RAL~~T~~SGTA~~I~~NKM~~V~~Y~~P~~VSAKEQKKIGDFFSKLDRQIELEEQK~~L~~
 24 ELLQQQ
 25 >J GGA
 26 MSNTQTKNVP~~E~~LRFPGFEGEWEERKL~~G~~DLIKVN~~S~~GKDYK~~H~~LDKG~~D~~IPVYGTGGYMTS~~V~~SEPL~~S~~E~~I~~DAVG~~I~~GR~~K~~GTINKPYL~~E~~APFWTVDTLFYCTPEKE
 27 ADILFILS~~L~~FR~~K~~INW~~K~~L~~Y~~DESTGVPSLSKQT~~I~~NK~~I~~RLV~~P~~TN~~K~~EQQKIGEFFSKLDRQIELEEQKLELLQQQ
 28 >M CAG
 29 MSNTQTKNVP~~E~~LRFPGFEGEWEEK~~K~~LG~~E~~FAGKV~~T~~Q~~N~~V~~D~~KK~~Y~~IE~~T~~LT~~N~~AEL~~G~~I~~S~~Q~~K~~DYFD~~K~~E~~I~~S~~N~~ID~~N~~I~~K~~Y~~V~~V~~E~~ENDFVYNPRMSNYAPFGPVNR~~N~~
 30 KLGKKGVMS~~P~~LYTVFKI~~Q~~N~~I~~DN~~N~~IEFYFKSSK~~W~~YRFMALNGD~~S~~ARADRF~~S~~IK~~D~~RTFM~~E~~PL~~H~~IPCM~~D~~EQIKIGQFFSKLDRQIELEEQKLELLQQQ
 31
 32 >O CAAC
 33 MSNKQKKNVP~~E~~LRFPGFEGEWEEK~~K~~LG~~E~~EVGTFTSGG~~T~~PL~~K~~SK~~E~~Y~~W~~NGD~~I~~PWIT~~T~~GD~~I~~HNIKREN~~I~~NTNFITEKGLNESSAKL~~I~~ITNEA~~I~~LIAMYQGKTRG
 34 MS~~A~~IL~~N~~FEATT~~N~~QAC~~A~~Y~~Q~~T~~N~~Q~~N~~INFVFQYFQ~~K~~LYE~~F~~LR~~S~~LS~~E~~GSQ~~K~~N~~L~~S~~L~~KE~~I~~TLN~~P~~N~~E~~Q~~B~~Q~~K~~IGDFFSKLDRQIELEEQKLELLQQQ
 35 >R GARA
 36 MSNTQKKNVP~~E~~LRFPGFEGEWEEK~~K~~LG~~E~~EVAK~~I~~YDG~~T~~HQ~~T~~PK~~Y~~NEG~~I~~KFL~~S~~VEN~~I~~KT~~L~~N~~S~~SK~~Y~~I~~S~~EEA~~E~~KE~~F~~K~~I~~R~~P~~E~~F~~G~~D~~IL~~M~~TR~~I~~G~~D~~ITGP~~N~~IV~~S~~NE
 37 KFA~~Y~~VS~~L~~ALL~~K~~T~~K~~N~~L~~NS~~Y~~FL~~K~~N~~L~~LS~~S~~SI~~Q~~N~~E~~WL~~R~~K~~T~~L~~H~~V~~A~~FP~~K~~K~~I~~N~~K~~NE~~I~~G~~K~~I~~N~~Y~~P~~KK~~Q~~EQ~~Q~~K~~I~~GQFFSKLDRQIELEEQKLELLQQQ
 38 >T CAAG
 39 MSNTQTKNVP~~E~~LRFPGFEGEWEEK~~K~~LG~~E~~EV~~K~~GF~~I~~Q~~F~~II~~S~~G~~T~~PL~~K~~SN~~E~~F~~Y~~ENG~~N~~IN~~W~~V~~K~~TT~~D~~LN~~N~~SK~~V~~TH~~S~~KE~~K~~ITE~~Y~~AM~~K~~SL~~K~~L~~K~~LP~~K~~NS~~V~~L~~I~~AMYGGF~~N~~Q~~I~~
 40 GRTG~~L~~L~~K~~I~~D~~AT~~I~~Q~~A~~IS~~A~~LL~~M~~HET~~N~~PE~~F~~I~~Q~~AF~~L~~NY~~Q~~V~~K~~GW~~K~~R~~Y~~A~~S~~SR~~K~~D~~P~~N~~I~~KK~~D~~IE~~Q~~FK~~V~~P~~V~~S~~I~~NE~~Q~~Q~~K~~IGEFFSK~~I~~D~~H~~Q~~I~~E~~E~~QKLELLQQQ
 41 >V CNGA
 42 MSNTGKMNV~~P~~ELRFPGFEGEWEEK~~K~~REL~~R~~NPKD~~K~~Y~~S~~YT~~G~~PF~~G~~SDL~~K~~KS~~D~~Y~~T~~D~~G~~I~~Q~~I~~I~~Q~~L~~Q~~N~~I~~G~~D~~G~~Y~~F~~Y~~N~~SN~~K~~V~~F~~T~~S~~N~~E~~KA~~E~~VL~~K~~SC~~N~~V~~F~~PG~~D~~IV~~I~~A~~K~~
 43 MADPI~~A~~RA~~A~~IV~~P~~D~~N~~NI~~G~~Y~~K~~LMAS~~D~~G~~I~~R~~L~~S~~V~~DT~~H~~F~~N~~TF~~V~~LC~~I~~NR~~K~~FR~~K~~V~~E~~D~~N~~SS~~G~~STR~~M~~R~~I~~G~~L~~ST~~L~~GT~~T~~TL~~K~~EQ~~Q~~K~~I~~GQFFSKLDRQIV~~L~~
 44 EQKLELLQQQ
 45 >X TCTA
 46 MSNTQKKNVP~~E~~LRFPGFEGEWEEK~~K~~Q~~F~~AD~~F~~TK~~I~~Q~~G~~Q~~I~~AI~~N~~ER~~K~~TE~~Y~~SP~~E~~LY~~F~~Y~~I~~IT~~N~~E~~F~~LP~~R~~NS~~Q~~T~~K~~Y~~F~~I~~E~~N~~P~~P~~Q~~S~~V~~IAN~~K~~E~~D~~IL~~M~~TR~~G~~NT~~G~~V~~V~~TN~~V~~
 47 GAFHNN~~F~~FK~~I~~K~~F~~DK~~N~~LY~~D~~RL~~F~~L~~V~~E~~V~~LN~~S~~SK~~I~~Q~~N~~K~~I~~LS~~L~~AG~~S~~ST~~I~~PD~~L~~HS~~D~~F~~S~~I~~S~~SSY~~P~~LL~~R~~EQ~~Q~~K~~I~~GKFFSKLDRQIELEEQKLELLQQQ
 48 >Z GAC
 49 MSNTQTKNVP~~E~~LRFPGFEGEY~~S~~LD~~I~~FG~~N~~LA~~T~~N~~K~~SE~~K~~F~~N~~P~~Q~~N~~E~~NA~~S~~DI~~E~~LD~~C~~IE~~Q~~NT~~G~~R~~L~~I~~K~~I~~Y~~N~~S~~KE~~F~~S~~Q~~KN~~K~~F~~N~~P~~Q~~N~~V~~LY~~G~~K~~L~~RP~~Y~~LN~~K~~Y~~Y~~FT~~K~~KG
 50 VCS~~E~~I~~W~~L~~K~~ST~~K~~ED~~K~~LL~~N~~FL~~Y~~Y~~F~~IQ~~T~~K~~R~~Y~~S~~DS~~V~~ASK~~S~~AG~~S~~K~~M~~PR~~A~~DW~~G~~L~~I~~N~~R~~V~~Y~~F~~P~~CE~~Q~~Q~~K~~IGQFFSKLDRQIELEEQKLELLQQQ
 51 >b* GGHA
 52 MSNTQKKN~~A~~PE~~L~~RF~~P~~FE~~E~~GEWEKK~~L~~DT~~L~~EF~~I~~K~~D~~G~~T~~H~~G~~THE~~V~~N~~N~~G~~P~~W~~L~~LS~~A~~KN~~I~~KK~~N~~K~~I~~I~~I~~S~~S~~DDR~~K~~I~~S~~E~~D~~Y~~K~~Y~~K~~Y~~N~~K~~L~~E~~K~~G~~D~~LL~~L~~IT~~V~~GT~~I~~GR~~A~~
 53 AIV~~K~~N~~P~~NN~~I~~A~~F~~Q~~R~~S~~V~~AI~~L~~K~~T~~K~~A~~TY~~D~~V~~G~~F~~I~~Q~~F~~LT~~K~~Y~~F~~K~~N~~LL~~R~~Q~~V~~V~~S~~A~~Q~~P~~G~~LY~~G~~LD~~I~~R~~K~~K~~I~~S~~I~~T~~N~~I~~E~~EE~~Q~~R~~K~~I~~G~~FF~~S~~KLDRQIELEEQKLELLQQQ
 54 >e* GAG
 55 MSNTQKKNVP~~E~~LRFPGFEGEWEEK~~S~~SI~~S~~FL~~K~~ES~~K~~I~~K~~GS~~N~~G~~H~~AK~~K~~LT~~V~~KL~~W~~KG~~V~~V~~P~~KK~~E~~TF~~K~~GS~~D~~NT~~Q~~YY~~K~~R~~K~~AG~~Q~~QL~~M~~Y~~G~~K~~L~~DF~~L~~NC~~A~~FG~~I~~VP~~D~~SL~~N~~NY
 56 ESTIDSP~~S~~F~~D~~FI~~G~~DS~~K~~FL~~L~~ER~~I~~KL~~K~~FS~~F~~Y~~K~~K~~G~~DI~~A~~NG~~S~~R~~K~~A~~R~~IN~~Q~~DT~~F~~LS~~L~~SP~~V~~F~~A~~PK~~Y~~DE~~Q~~LR~~I~~G~~E~~FF~~S~~KLDRQIELEEQKLELLQQQ
 57 >NOVEL 1
 58 MSNTQKKNVP~~E~~LRFPGFEGEWEEK~~K~~LG~~E~~EV~~A~~T~~F~~AK~~G~~KL~~G~~A~~K~~D~~V~~S~~Q~~NG~~V~~P~~V~~I~~L~~Y~~G~~E~~L~~Y~~T~~K~~G~~A~~I~~V~~S~~K~~I~~F~~S~~K~~T~~D~~I~~P~~E~~N~~K~~MA~~K~~N~~D~~V~~L~~I~~P~~S~~S~~GET~~A~~DI~~I~~AT~~A~~
 59 ASCIYLN~~K~~GV~~A~~V~~G~~GD~~I~~NL~~T~~P~~Q~~K~~D~~GR~~F~~I~~S~~L~~S~~ING~~I~~NC~~N~~LS~~K~~Y~~A~~Q~~G~~K~~T~~V~~V~~H~~Y~~NN~~D~~I~~K~~N~~K~~I~~A~~F~~P~~SE~~E~~EQ~~V~~R~~I~~G~~N~~FF~~S~~KLDRQIELEEQKLELLQQQ
 60 >NOVEL 2
 61 MSNTQKKNVP~~E~~LRF~~P~~FE~~E~~GEW~~K~~DK~~V~~K~~F~~V~~S~~I~~F~~Q~~E~~V~~S~~N~~K~~T~~S~~DL~~A~~Y~~K~~PL~~F~~SL~~T~~VE~~K~~GT~~P~~K~~T~~ERY~~K~~RD~~F~~LV~~K~~KS~~D~~N~~F~~K~~I~~VE~~P~~R~~D~~IV~~V~~N~~P~~M~~N~~V~~T~~L~~G~~A~~I~~D~~L~~SK~~Y~~NY
 62 DIAL~~G~~Y~~H~~V~~M~~K~~I~~I~~N~~S~~F~~N~~P~~D~~F~~I~~S~~N~~F~~LN~~K~~TE~~K~~M~~I~~I~~Y~~KK~~I~~AT~~G~~SL~~M~~E~~K~~Q~~R~~V~~H~~F~~S~~E~~F~~K~~N~~I~~I~~KK~~F~~PT~~N~~K~~E~~QQ~~K~~IGDFFSKLDRQIELEEQKLELLQQQ
 63 >NOVEL 4
 64 MTEQINTPE~~L~~RF~~P~~FE~~E~~GEW~~K~~DK~~V~~K~~F~~V~~S~~I~~F~~Q~~E~~V~~S~~N~~K~~T~~S~~DL~~A~~Y~~K~~PL~~F~~SL~~T~~VE~~K~~GT~~P~~K~~T~~ERY~~K~~RD~~F~~LV~~K~~KS~~D~~N~~F~~K~~I~~VE~~P~~R~~D~~IV~~V~~N~~P~~M~~N~~V~~T~~L~~G~~A~~I~~D~~L~~SK~~Y~~NY
 65 SSEI~~W~~V~~L~~N~~K~~D~~V~~L~~A~~N~~K~~F~~Y~~Y~~F~~IQ~~T~~NR~~F~~S~~S~~VT~~N~~K~~S~~AG~~S~~K~~M~~PR~~A~~DW~~G~~L~~I~~N~~R~~V~~Y~~F~~P~~CE~~Q~~Q~~K~~IGY~~F~~FF~~S~~KLDRQIELEEKK~~L~~LE~~Q~~QQ

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.

```

1 >D ATC
2 KKGYMQKIFSQELRFKDENGDYPHWENSKIEYKLKERNERSDKGQMLSVTINSGIIIKFSELDRKDNSSKNKSNYKVVRKNDIAYNSMRMWQGASGKSN
3 NGIVSPAYTVLYPTQNTSSLFIGYKFKTHRMIHKFKINSQGLTSDTWNLKYQLNINDIPVLEEQEKIGDFFKMDILISKQKKIELEKQSFLQ
4 KMF
5 >E TCAY
6 KKGYMQKIFSQELRFKDENGDYNPEWEETTKEIAQINXGKKDTKDAITNGSYDFYVRSPIVYKINTFSYEEGEAILTVGDGVGKVFHYVNGKFDYHQ
7 VYKISDFKNYYGLLLFYSQNFLKETKKSAKTSVRKDMIANMKVPRIYIEQKIGQFIKRVDNKTIQKVIELLKQRKALLQKMFI
8 >F TTAA
9 KKGYMQKIFSQELRFKDEEGKDYPDWKSKSIQEIFENKGGTAELETEFNFDGNYKVISINTYNDQNIRVNKNKTEKYILSKGDLAMVLNDKTKD
10 GKIIGRSIFIDKNQYIYNQRTELIPFAEDNKFLWFLMNTDLIRNKIGMMQGATQVINYSISKILIQPLLEEQQKIRGFLEVLSGITKOLHKI
11 DQLKERRKAFLQKMFI
12 >G ACA
13 KKGYMQKIFTQELRFKDENGEEEYPEWENKFIKDIFIFENNRRKPITSSLREKGLPYYGATGIIDYVKDYLFNNEERLLIGEDGAKWQFETSFIANGQ
14 YWVVNHAHVVKSNDHNLFMNYLNFKELRAFVTGNAPAKLTHANCNINLKIPCLEQDKVSALLKSIDNKMNQMRIELLKELLQKMFI
15 >H TAC
16 KKCYCIFKISQELRFKDEEGNYYKGWNKQLKDVLEFSNKRTINEEPVLTSRQGLILQDSDYYKDRKTFAESNIGYFILPKNHITYRSRSDDGIKFKN
17 LNLIMDVGISKYPVFKGIDANQYYLTHLNYQLKEYIKYATGTSQLVSQKDLQNIKTKLPSYEQQKIGDFSEIDRLVEKQSSKVGRLKVRKKELLQKMFI
18 >I YTCA
19 KKGYMQKIFSQELRFKNENGDYPDWERIKFVIDKVIDFRGRTPKKLNMEWSDEGYLASAVNVKGYIDFNVEAKYGNLDLYTRWMRGNELYKQVQVLNDKTKDI
20 >K CGA
21 KKGYMQKIFSQELRFKDENGDYPWERIKFVIDKVIDFRGRTPKKLNMEWSDEGYLASAVNVKGYIDFNVEAKYGNLDLYTRWMRGNELYKQVQVLNDKTKDI
22 >L TTA
23 KKGYMQKIFSQELRFKDENGDYPWERIKFVIDKVIDFRGRTPKKLNMEWSDEGYLASAVNVKGYIDFNVEAKYGNLDLYTRWMRGNELYKQVQVLNDKTKDI
24 >P AGG
25 KKGYMQKIFSQELRFKDENGDYPWERIKFVIDKVIDFRGRTPKKLNMEWSDEGYLASAVNVKGYIDFNVEAKYGNLDLYTRWMRGNELYKQVQVLNDKTKDI
26 >S GCA
27 KKGYMQKIFSQELRFKDENGDYPWERIKFVIDKVIDFRGRTPKKLNMEWSDEGYLASAVNVKGYIDFNVEAKYGNLDLYTRWMRGNELYKQVQVLNDKTKDI
28 >Q ACY
29 KKGYMQKIFSQELRFKDENGDYPWERIKFVIDKVIDFRGRTPKKLNMEWSDEGYLASAVNVKGYIDFNVEAKYGNLDLYTRWMRGNELYKQVQVLNDKTKDI
30 >R CAA
31 KKGYMQKIFSQELRFKDENGDYPWERIKFVIDKVIDFRGRTPKKLNMEWSDEGYLASAVNVKGYIDFNVEAKYGNLDLYTRWMRGNELYKQVQVLNDKTKDI
32 >G CAA
33 KKGYMQKIFSQELRFKDENGDYPWERIKFVIDKVIDFRGRTPKKLNMEWSDEGYLASAVNVKGYIDFNVEAKYGNLDLYTRWMRGNELYKQVQVLNDKTKDI
34 >Y CTA
35 KKGYMQKIFSQELRFKDENGDYPWERIKFVIDKVIDFRGRTPKKLNMEWSDEGYLASAVNVKGYIDFNVEAKYGNLDLYTRWMRGNELYKQVQVLNDKTKDI
36 >A GAA
37 KKGYMQKIFSQELRFKDENGDYPWERIKFVIDKVIDFRGRTPKKLNMEWSDEGYLASAVNVKGYIDFNVEAKYGNLDLYTRWMRGNELYKQVQVLNDKTKDI
38 >C GAT
39 KKGYMQKIFSQELRFKDENGDYPWERIKFVIDKVIDFRGRTPKKLNMEWSDEGYLASAVNVKGYIDFNVEAKYGNLDLYTRWMRGNELYKQVQVLNDKTKDI
40 >U GAY
41 KKGYMQKIFSQELRFKDENGDYPWERIKFVIDKVIDFRGRTPKKLNMEWSDEGYLASAVNVKGYIDFNVEAKYGNLDLYTRWMRGNELYKQVQVLNDKTKDI
42 >W CRAA
43 KKGYMQKIFSQELRFKDENGDYPWERIKFVIDKVIDFRGRTPKKLNMEWSDEGYLASAVNVKGYIDFNVEAKYGNLDLYTRWMRGNELYKQVQVLNDKTKDI
44 >X CKQLI
45 >Y CAA
46 KKGYMQKIFSQELRFKDENGDYPWERIKFVIDKVIDFRGRTPKKLNMEWSDEGYLASAVNVKGYIDFNVEAKYGNLDLYTRWMRGNELYKQVQVLNDKTKDI
47 >a* GAA
48 KKGYMQKIFSQELRFKDENGDYPWERIKFVIDKVIDFRGRTPKKLNMEWSDEGYLASAVNVKGYIDFNVEAKYGNLDLYTRWMRGNELYKQVQVLNDKTKDI
49 >c* GAY
50 KKGYMQKIFSQELRFKDENGDYPWERIKFVIDKVIDFRGRTPKKLNMEWSDEGYLASAVNVKGYIDFNVEAKYGNLDLYTRWMRGNELYKQVQVLNDKTKDI
51 >d* CYAA
52 KKGYMQKIFSQELRFKDENGDYPWERIKFVIDKVIDFRGRTPKKLNMEWSDEGYLASAVNVKGYIDFNVEAKYGNLDLYTRWMRGNELYKQVQVLNDKTKDI
53 >e* GAA
54 KKGYMQKIFSQELRFKDENGDYPWERIKFVIDKVIDFRGRTPKKLNMEWSDEGYLASAVNVKGYIDFNVEAKYGNLDLYTRWMRGNELYKQVQVLNDKTKDI
55 >f* GAA
56 KKGYMQKIFSQELRFKDENGDYPWERIKFVIDKVIDFRGRTPKKLNMEWSDEGYLASAVNVKGYIDFNVEAKYGNLDLYTRWMRGNELYKQVQVLNDKTKDI
57 >g* GAA
58 KKGYMQKIFSQELRFKDENGDYPWERIKFVIDKVIDFRGRTPKKLNMEWSDEGYLASAVNVKGYIDFNVEAKYGNLDLYTRWMRGNELYKQVQVLNDKTKDI
59 >h* GAA
60 KKGYMQKIFSQELRFKDE
```

1
 2 **SUPPLEMENTARY INFORMATION FOR MATERIALS AND METHODS SECTION**
 3 **"CONSTRUCTION OF FURTHER MTASES WITH FURTHER COMBINATIONS OF TRDS**
 4 **USING SYNTHETIC GENES."**

5 The DNA sequence is followed by the predicted amino acid sequence
 6 for each TRD. Some synthetic sequences encoded two TRDs.
 7

8 **CC15 TRD O**

9 CCCGGTTGAAGGCGAATGGGAGGAAAAAAACTGGGTGAAGTGGCACCTTACCAGCGGTGGC
 10 ACTCCGCTAAAAGCAAAGCGAATATTGAATGGTATTCGGTGGATTACACAGGCATATT
 11 CATAACATTAAACGCGAAAACATCACCAACTTATCACCAGAAAAGGCCTGAATGAAAGCAGCGCA
 12 AAACTGATTACCAATGAAGCAATTCTGATTGCCATGTATGGTCAGGGTAAACCCGTGGTATGAGC
 13 GCCATTCTGAATTTGAAGCAACCACCAATCAGGCCTGTCAATTATCAGACAAACCAAGAACATC
 14 AACTCGTGTCCAGTATTCCAGAAACTGTATGAATTCTCGTAGCCTGAGCAATGAAGGTAGC
 15 CAGAAAAATCTGAGCCTGAGCTGCTGAAAGAAATTACCCCTGAATTATCCGAACGAGCAAGAACAG
 16 AAAAAAAATCGGCATTCTTCAGCAAACCTGGATCGTCAAATTGAATTAGAAGAACAGAACAG
 17 CC15 TRD O

18 PGFEGEWEEKKLGEVGTFTSGGTPLSKSEYWNIDIPWITTGDIHNIKRENITNFITEKGLNESSA
 19 KLITNEAILIAMYGQGKTRGMSAILNFEATTNQACAIYQTQNQINFVFQYFQKLYEFLRSLSNEGS
 20 QKNLSSLKEITLNYPNEQEKKIGDFFSKLDRQIELEEQK

21 **CC51 TRD P**

22 CAAATTGAATTAGAAGAACAGAACAGCTGAACTGTTCAGCAGCAGAAAAAGGCTATATGCAGAAA
 23 ATCTTAGCCAAGAGCTGCGCTTAAAGATGAAAGCGGTAAATGATTATCCGGATTGGGAAGAAAAAA
 24 GAACTGGGTGAAGTTGCAGATCGTGTGATTGTAACAAAACAAACTTGAAGCAGAAAAACCGCTG
 25 ACCATTAGCGGTCACTGGGCTGATTGATCAGACAGAAATATTCAAGAAAAGCTATAGCAATGGT
 26 AACCTGGAAAACATACCCCTGATTAAAACGGCGAGTCGCCTATAACAAAAGCTATAGCAATGGT
 27 TATCCGCTGGGTGCAATTAAACGTCTGACCCGTATGATAGCGGTGTTCTGAGCAGCCTGTATATT
 28 TGCTTAGCATCAAAGCGAGATGAGCAAAGATTTCATGGAAGCCTATTTGATAGCACCCATTGG
 29 TATCGTGAAGTTAGCGGTATTGCAGTTGAAGGTGCACGTAATCATGGTCTGCTGAATATTAGCGTG
 30 AACGATTTTTCAACCACCTGATCAAATATCCGAGCCTGGAAGAACAGCGTAAACCGTGAATTTC
 31 TTCATTAAACTGGATGCCAGATTGAGCTGGAAGAACAAAAACTGGAACTGCTGCAACAGCGCAA
 32 AAAGCACTGCTGAAAGTATGCTGATCCCCGGGGATCCGATCGATC
 33 CC51 TRD P

34 QIELEEQKLELFQQQKKGYMQKIFSQELRFKDESGNDYPD WEEKELGEVADRVIRKNKNFESKKPL
 35 TISQLGLIDQTEYFSKSVSSKNLENYTLIKNGEFAYNKSYSNGYPLGAIKRLTRYDSGVLSLYI
 36 CFSIKSEMSKDFMEAYFDSTHWYREVSGIAVEGARNHGLLNISVNDFTILIKYPSLEEQRKIGDF
 37 FIKLDRQIELEEQKLELLQQRKKALLKSMLI

38 **CC72-1 TRD R + CC59-1 TRD Q**

39 CCCGGTTGAAGGCGAATGGGAGGAAAAAAACTGGGTGAAGTGGCAGAAATCTATGATGGCACC
 40 CATCAGACCCCGAAATATACCAATGAAGGTATCAAATTCTGAGCGTGGAAAACATCAAAACCTG
 41 AATAGCAGCAAATACATTAGCGAAGAACGCTTCGAGAAAGAATTCAAATCGTCCGAATTGGC
 42 GATATTCTGATGACCGTATTGGTATTCGGACCCCGAATATTGTTAGCAGCAATGAAAATTC
 43 GCCTACTATGTTAGCCTGGCACTGCTGAAACAAAAATCTGAACAGCTACTCCTGAAAAACCTG
 44 ATTCTGAGCAGCAGCATTGAGAACTGTGGCTAAACCCCTGCATGTTGCAATTCCGAAAAAAA
 45 ATCAACAAAACGAGATCGGCAAATCAAATCAACTACCCGAAAAAACAGAACAGCAGAAAATC
 46 GGTCAGTTTCAGCAAACCTGGATCGCAAATTGAATTAGAACAGAACAGCTGGAACGCTGCAA
 47 CAGCAGAAAAAGGTTATATGCAGAAATCTCAGCCAAGAGCTGCGCTTAAAGATGAAAATGGT
 48 GAAGATTATAGCGAGTGGGAAGAACGTCGTTTGCCGATATTCAAATTTCACAACAAACTGCGC
 49 AAACCGATCAAAGAAAATCTCGTGTAAAGGCAGCTATCCGTATTATGGTCAACCGGCATTATT
 50 GATTATGTGGATGATTATCTCGATGGCAACTATCTGCTGATTGGCGAAGATGGTCAAACATT
 51 ATTACCCGTAGCGCACCGCTGGTTATCTGGTTAATGGTAAATTGGTGAACAACCAGGCCAT
 52 ATTCTGAGTCCGCTGAATGGTAATATTCAAGTATCTGTATCAGGTTGCCGAACGGTGAACATGAA

1 AAATACAATACCGGCACCGCACAGCGAAACTGAACATTGAGAATCTGAAAATTATCAACGTGGTG
 2 ATCAGCACCAATCTGGAAGAACAGCAAAAAATTGGTAGCTCCTGAGCAAACGTGAAAGCATGTT
 3 GACCTGGAAGAACAAAAACTGGAACACTGCTGCAACAAACGTAAAAAGCACTGCTGAAAGCATGTT
 4 GTGCCCGGGGATCCGATCGATC
 5

CC59-1 TRD Q

6
 7 QIELEEQKLELLQQQKKGYMQKIFSQERLFKDENGEDYSEWEERRFADIFKFHNKLKP
 8 IKENLRV KGSYPYYGATGIIDYVDDFIFDGNYLLIGEDGANIITRSAPLVLYVNGKF
 9 WVNNHAHILSPLNGNI QYLYQVAELVNYEKYNTGTAQPKLNIQNLKIINV
 10 VISTNLEEQQKIGSFLSKLDRQIDLEEQKLE LQQRKKALLKSMFV
 11
 12 **CC72-1 TRD R**
 13 PGFEGEWEEKKLGEVAKIYDGTHQTPKYTN
 14 EGIKFLSVENIKTLNSSKYI
 15 SEEAFEKEFKIRPEFG DILMTRIGDIGTPNIVSSNEKFAYYVSLALLKTKNLNSYFLKNL
 16 LISSSIQNELWRKTLHVAFPKK INKNEIGKIKINYPKKQEQQKIGQFFSKLDRQIELEEQK
 17

CC75-1 TRD T and TRD U

18 CCCGGTTTGAAGGCGAATGGGAGGAAAAAGAACTGGGCGAAATCTTCAGATTATTAGCGGTAGC
 19 ACACCGCTGAAAGCAACAAAGAATTATGAGAACGGCAACATCAACTGGGTTAAACCACCGAT
 20 CTGAATAATAGCAAAGTGACCATAGCAAAGAAAAATCACCAGTATGCAATGAAAAGCCTGAAA
 21 CTGAAACTGGTGCCGAAAATAGCCTCTGATTGCAATGTATGGTGGCTTAATCAGATTGGCGT
 22 ACCGGTCTGCTGAAATTGATGCAACCATTAAATCAGGCAATTAGCGCACTGCTGATGAATCATGAA
 23 ACCAACCCGGAATTTATTCAAGGCCTTCTGAATTATCAGGTAAAGGTTGAAACGTTATGCAGCA
 24 AGCAGCCGTAAAGATCCGAATATCACCAAAAAAGATATCGAACAGTTCAAAGTGCCTACGTGAGC
 25 ATTAATGAACAGCAGAAAATTGGCGAGTTTTAGCAAAATCGATCATCAAATTGAATTAGAAGAA
 26 CAGAAGCTGGAACTGCTGCAACAGCAGAAAAAGGTTATATGCAGAAAATCTTCAGCCAAGAGCTG
 27 CGCTTAAAGATGAAAATGGTGAAGATTATCCGGATTGGGAGTTACCAACATTAGAACATTAC
 28 AAATACACCAGCAGAAAAAGCAGCAATCAGTATGCCGATAAAGACAACAGCAAAGGTTATCCG
 29 GTTATGATGCCGTTCAAGAAATTGGCAAAGATAGCAACTATGACATCGAAGAGAGCTATATCAGC
 30 ATTCTGAAAGATGGTCCGGTGGTCTGAATCTCGTCCGGTAAAAGCAGCGTTATTGGC
 31 ACCATGGGTTATATTCAAGAGCAACACGTGGATATCGAGTTCTGTATTATCGTATGAAAGTGGT
 32 GACTCAAAAATACATTATCGTAGCACCATTCCGACCTGTATTCAAAGATTATAGCAAAGAA
 33 ACCCTGTACATTCCGAGCAGCATTCAAGAACAGGCAAAATTGGTATGTTCATCAGAACCTGGAT
 34 AAACTGATCGAGAACAAAACCTGAAACTGAACACTGTCTGAAACAACAGGATTGCTACAA
 35 TCTATGTTATTCCGGGGGATCCGATCGATC
 36
 37 **CC75-1 TRD T**
 38 PGFEGEWEEKELGEIFQIISGSTPLKS
 39 NKEFYENGNIWVKTDLNNSKVTHS
 40 SKEKITEYAMKSLK LKLVPKNSVLIAMYGGFNQIGRTGL
 41 KIDATINQAISALLMNHETNPEFIQAF
 42 LNYQVKGWKRYAA SSRKDPNITKKDIEQFKV
 43 PYVSINEQQKIGEFFSKIDHQIELEEQK
 44
 45 **CC75-1 TRD U**
 46 QIELEEQKLELLQQQKKGYMQKIFSQERLFKDENGEDYPDWEVTTIONITKY
 47 TSSKKSSNQYADKD NSKGYPVYDAVQEIGKDSNYDIEESYISIL
 48 DGAGVGRNL
 49 RPGKSSV
 50 IGMYIQSN
 51 NVDIEFLY YRMKV
 52 VDFKKYIIGSTIPHL
 53 FDYSKETLYIP
 54 PSSIQEAKIGMF
 55 ISNLDK
 56 LIENKNL
 57 KLNCL
 58 KQL
 59 QG
 60 LQSMF
 61 I

CC75-2 TRD V

62 CCCGGTTTGAAGGCGAATGGGAGGAAAAAGAACTGCGTGA
 63 ACTGCGCAATCCGAAAGATAAATAC AGCTATACCGGTGGTCCGTTGGTAGCGATCTG
 64 AAAAAGCGATTACCGATCTGAGGCTATTCTATAACAGCAACAAAGTGT
 65 TTACCA
 66 GCAAC GAAAAAGCCGAAGTTCTGAAAAGCTGT
 67 AATGTTCCGGTGA
 68 ATTGTGATTGCA
 69 AAAAGCCGATTGCACGTGCC
 70 GAATTGTTCCGGATA
 71 ATAACATTGGTAA
 72 ATACCTGATGGCCAG
 73 TGATGGT
 74 GGTATT
 75 CGTCTGAGCGTT
 76 GATACCGTT
 77 CATTGTTAACAC
 78 CAAATT
 79 TGCTGGA
 80 ATGCGATCAACCGT
 81 AAAAGCTTC
 82 GTAAAAAGCTGAG
 83 GAGGATA
 84 ATAGCAGCG
 85 TAGCACC
 86 CGTATGCGT
 87 ATTGGTCTGAG
 88 CCGTACCC
 89 GAGAAACAGCAG
 90 AAAATTGGTCAG
 91 TTTC

1 AGCAAACGGATCGTCAAATTGAATTAGAAGAACAGAACAG
 2

3 **CC75-2 TRD V**

4 PGFEGEWEEKELRELRNPKDKYSYTGGPGSDLKSDYTTDGIQIIQLQNIQDGYFYNSNKVFTSN
 5 EKAEVLKSCNVFPGDIVAKMADPIARAAIVPDNNIGKYLMASDGIRLSVDTVFNTKFVLECINR
 6 KSFRKKVEDNSSGSTRMRIGLSTLGSLLKTTLKEQQKIGQFFSKLDRQIELEEQK
 7

8 **CC75-2 TRD W**

9 CAAATTGAATTAGAAGAACAGAACAGCTGGAACACTGCTGCAACAGCAGAAAAAAGGTTATATGCAGAAA
 10 ATCTTCAGCCAAGAGCTGCGCTTAAAGATGAAAATGGTAACGATTATCCGGATTGGGAAGAAAAA
 11 CAGCTGGGTGAACTGAGCCAGATTGTTGGTGCAAGTCCCGTCCGATTAAAGATCCGAAATGG
 12 TTTAACAAAGAAAGCGATATTGGTTGGCTGCGCATTAGTGTACCAATCAGAATGGCAAATC
 13 TATCATCTGGAACAGAAACTGAGCATCGAAGGTCAAGAAAAACCCGTGTCTGGTTACCACCCAT
 14 CTGCTGCTGAGCATTGCAAGCATTGGTAAACCGTTATGAACCTTGAAACCCGTGTGCAT
 15 GATGGCTTCTGATTTCTGAAACCGAAATTCAACCTGTTATGTACTATTGGCTGGAATAT
 16 TTCAAAGATAATGGTCAAATATGGTCAGCCTGGTAGCCAGGTTAATCTGAATAGCGAAATTGTT
 17 AAAAGCCAGACCCCTGAATATGCCGAGCAATCATGAACAAGAAAAGTGGCCAGTTTTAACCGC
 18 AACGAAAAACTGATTGAACTGCAAGAGAAAATCATGTATATCAAACGTTGCAAACAGGTGCTG
 19 CTGCAAAATGTTATTCCCCGGGGATCCGATCGATC
 20

21 **CC75-2 TRD W**

22 QIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPDWEWKLKEIACVYTGNTPSKKENIY
 23 FNKESDIGWLRISDVNTQNQNGKIYHLEQKLSIEGQEKRVLVTTHLLSIAASIGKPVMNFVKTGVH
 24 DGFLIFLKPKNLFFMYYWLEYFKDKWSKYGQPGSQVNLNSEIVKSQTLNMPSONHEQEKGQFFNR
 25 NEKLIELQQEKIMYIKRCKQVLLQKMF
 26

27 **CC80-1 TRD X and TRD Y**

28 CCCGGGTTGAAGGCGAATGGGAGGAAAAACAGTTGCCGACTTCACCAAAATTAACCAGGGCTG
 29 CAGATTGCCATTAATGAACGTAACCGAATATAGCCCTGAGCTGTATTTCTATATCACCAACGAA
 30 TTTCTGCGTCCGAATAGCCAGACCAAAATATTCATTGAAAATCCGCCTCAGAGCGTGTGATTGCCAAC
 31 AAAGAAGATATTCTGATGACCCGACCGTAATACCGGCAAAGTTGTTACCAATGTTGGTGC
 32 TTCCACAACAACCTTTCAAATTCGATAAAACCTGTATGATCGCCTGTTCTGGTTGAA
 33 GTTCTGAACAGCAGCAAATCCAGAACAAAATTCTGAGCCTGGCAGGTAGCAGCACCATTCCGGAT
 34 CTGAATCATAGCATTCTATAGCATTAGCAGCTATCCGCTGCGCGAACAGCAAAATT
 35 GGCAAATTCTTAGCAAACCTGGATCGTCAAATTGAATTAGAACAGAACAGCTGAACTGCTGAA
 36 CAGCAGAAAAAAGGTTATATGCAGAAAATCTCAGCCAAGAGCTGCGTTAAAGATGAAAATGGT
 37 AACGATTATCCGGATTGGGAGAAAAAAACTGAAAGAAATTGCCCTGCGTGTATACCGTAATACC
 38 CCGAGCAAAAGAACATCTATTGGAACAAAGGCAGTATGTTGGTTACCCGACCGATATT
 39 AACAAACAGCAAAACATTATGAAAGCGAAAACAAACTGACCCAAGAACAGGCTACAAAAAGCACGT
 40 CAGCTGCCGAAAATACCCTGCTGGTTACCTGTATTGCAAGCATTGGTAAAATGCCATTCTGCGT
 41 AAACAGGGTAGCTGTAATCAGCAGATTATGCAGTTGCGTTGAGAACATCAACATCGATTAT
 42 CTGTATTATATCAGCGATAGCCTGAGCACCTTCATGAAAAGCATTGCGAGTAAAACCGCAACCCAG
 43 ATTGTGAACAAAAACACCTTGAAAACCTGGAAATTACCTGGCACCTTGAGGAACAGAACAAA
 44 ATTGCAGATCTGATTAGCAGCTGGAAAGAACTGATTGAAAACAGGCAAGCAAACAGTATCAAAATG
 45 AAAAGCCGCAAACAGGGCATGCTGCAGATTATGTTATTCCCCGGGGATCCGATCGATC
 46

47 **CC80-1 TRD X**

48 PGFEGEWEEKQFADFTKINQGLQIAINERKTEYSPELYFYITNEFLRPNQTKYFIENPPQSVIAN
 49 KEDILMTRTGNTGVVTVNFGAFHNNFKIKFDKNLYDRLFLVEVLNSSKIQNKILSLAGSSTIPD
 50 LNHSDFYSISSSYPLLREQQKIGKFFSKLDRQIELEEQK
 51

52 **CC80-1 TRD Y**

53 QIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPDWEKKLKEIACVYTGNTPSKKENIY
 54 NKGEYVWVTPTDINNSKNIYESENKLTQEGYKKARQLPENTLLVTCIASIGKNAILRKQGSCNQQI
 55 NAVVPFENINIDYLYYISDSLSTFMKSIAGKTATQIVNKNTFENLEIYLAPFEEQNKIADLISSLE
 56 ELIEKQASKLIKMKSRKQGMLQIMFI
 57

1
2 **CC80-2 TRD Z + CC72-2 TRD S**
3

4 CCCGGGTTGAAGGCGAATATTCTCTGGATATTGGTAATCTGCCACCAACAAAAGCGAAAAAA
 5 TTCAATCCGAGAATGAAAAGCCAGCATTGATATTGAACACTGGATTGCATTGAACAGAACATACCGGT
 6 CGTCTGATCAAATCTATAACAGCAAAGAATTAGCAGCCAGAAAAACAAATTAAACCGCAGAAC
 7 GTGCTGTATGGTAAACTCGTCCGTATCTGAACAAATATTACTCACCAAAAAAGTGGTGTGC
 8 AGCAGCGAAATTGGGTTCTGAAAAGCACCAAAAGAAGATAAACTGCTGAACCTGTTCTGTACTAT
 9 TTCATTACGACCAACGCTATAGTGTGCAAGCAGGAGTAGCAAAATGCCTCGTCA
 10 GATTGGGGTCTGATTGAAAATATTCTGTGTATTTCCGAACTGTGCGAACAGCAGAAAATTGGT
 11 CAGTTTTAGCAAACGGTCAAATTGAATTAGAAGAACAGAAGCTGGAACACTGCTGCAACAG
 12 CAGAAAAAAGGTTATATGCAGAAAATCTCAGCCAAGAGCTGCGCTTAAAGATGAAAATGGTAAC
 13 GATTATCCGGACTGGACCAATGAACGCTGGGTGAAGTTACCAACCGTTACCATGGTCAGAGCCG
 14 AAAAGCGTGAATTATACCGATAATAGCAATGACACCCTCTGATTCAAGGTAATGCCGATATTGAA
 15 AACGGTCTGATTAATCCCGTATCTACCCGTGAAGTGACCAAACACTGATTCAAGAAGATGAGATT
 16 ATTCTGACCGTTCTGCACCAGGTTGGTAAACTGGCAATGGCACAGATTAATGCATGTATTGGTCGT
 17 GGTGTTGCAGCATTAAAGGCATAAATTCTGTATTATTCCTGAAATGGTCGCCACCCAGAAT
 18 AAATGGATTCTGTTAGCCAGGGTAGCACCTTGAAAGCATTAGCGGTAATGATATTGCAACATC
 19 CATATCAAAATCCGGTTGAAGATGAACGCACCAAAATTATCAAACACTGCTGAATAGCCTGGATGTG
 20 CTGAATTCAAAACCGATCTGAAAATCAGAATCTGAAACAGCGTAAACAGAGCCTGCTGCAAAAAA
 21 ATCTTGTGCCGGGGATCCGATCGATC
 22
 23 **CC80-2 TRD Z**
 24 PGFEGEYSLDIFGNLATNKSEKFNPQNENASIDIELDCIEQNTGRLIKIYNKEFSSQKNFKNPQN
 25 VLYGKLRLPYLNKYYFTKKSGVCSSEIWVLKSTKEDKLLNLFLYYFIQTKRYSDVASKSAGSKMPRA
 26 DWGLIENIRVYFPELCEQQKIGQFFSKLDRQIELEEQK
 27
 28 **CC72-2 TRD S**
 29 QIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPDWTNERLGEVTVTMGQSPKSVDNYTDN
 30 SNDTVLIQGNADIENGLINPRIYTREVTKLIQKDEIILTVRAPVGKLAMAQINACIGRGVCSIKGD
 31 KFLYYFLEWFATQNWKIRFSQGSTFESISGNDIRNIHIKIPVEDERTKIIKLLNSLDVLNSKTDLK
 32 IQNLKQRKQSLLQKIFV
 33
 34
 35 **CC93-2 TRD b***
 36 CCCGGGTTGAAGGCGAATGGGAGGAGAAAAACTGGAAGATAACCTGGAATTCAATTAAAGATGGC
 37 ACCCATGGTACACATGAAAATGTTAATAATGGTCCGTGGCTGCTGAGCGCCAAAACATTAAAAAC
 38 AACAAAATCATCATCAGCAGCGACGATCGCAAATTAGCGAAAGCGATTACAACAAATCTACAAA
 39 AACTATAAAACTGGAAAAAGGCGATCTGCTGCTGACCATTGGCACCATTGGCGTGCAGCAATT
 40 GTTAAAATCCGAACAATATTGCCTTCAGCGTAGCGTTGCAATCCTGAAACACAAAGCAACCTAT
 41 GATGTGGGCTTATCTTCAGCTGTTCCAGACCAAATACTTTAAACACTGCTGCTGCGTAAACAG
 42 GTTGTAGCGCACAGCCTGGCTGTATCTGGGTGATATTGTAACAAATCAGCATTACCAAC
 43 ATCATCGAAGAACAGCGAAAATCGGTATCTTTCAGCAAACGGATCGTCAAATTGAATTAGAA
 44 GAACAGAAG
 45
 46 **CC93-2 TRD b***
 47 PGFEGEWEEKKLEDITLEFIKDGTGTHENVNNGPWLLSAKNKNKIIISSDDRKISESDYKKIYK
 48 NYKLEKGDLLLTIVGTIGRAAIVKNPNNAFQRSVAILKTATYDVGIFQLFQTKYFKNLLLRQK
 49 VVSAQPGLYLDIRKIKISITNIIIEQRKIGIFFSKLDRQIELEEQK
 50
 51
 52
 53 **C93-3 TRD a***
 54 CAAATTGAATTAGAACAGAACAGCTGGAACACTGCTGCAACAGCAGAAAAAGGTTATATGCAGAAA
 55 ATCTTCAGCCAAGAGCTGCGCTTAAAGATGAAAATGGTAACGATTATCCGAATGGGAAAACAAA
 56 CGCATTGAAGATATTGCCAATGTGAACAAAGGTTTACCCGAGCACCAACAATAACGAATATTGG
 57 GATAACAACGATAAAAACGGCTGAGCATTGCAGGCATGAATCAGAAATATCTGTATAAAGGCAAC
 58 AAAGGCATCAGCAAAGATGCAGCAAAACATATGAAAGTGAACAAACGACACCCCTGATCATGTCC
 59 TTTAAACTGACCATTGGTAAACTGGCGATTGTTAAAGCACCCTGTATACCAATGAAGCCATTG
 60 CATTAACTCTGGAAAGTGAACAAACCCGAGTTCATCTACTATTACCTGAACAGCCTGAAC
 ATTAGCACCTTGGTGTTCAGGCAGTTAAAGGTGTTACCCCTGAATAACGATAGCATCACAGCATT

1
 2 ATTGTGAAACTGCCGAATGAAGAGGAACAGAACATTATCGCAAAATTCTGCTGGAAGTGGACAAA
 3 ACCGTTAATAATCAGCTGGTAAAACCAAACGTGAAACACGTAAAAAAGGCCTGCTGCAGCGT
 4 ATGTTGTTCCGGGGATCCGATCGATC
 5 **CC93-3 TRD a***
 6 QIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPEWENKRIEDIANVNKGFTPSTNNNEYW
 7 DNNDKNWLSIAGMNQYLYKGNKGISKDAKNYMVKNDTLIMSKLTIGKLAIVKAPLYTNEAIC
 8 HFIWKVNKINTEFIYYYLNSLNISTFGVQAVKGVTNNDSINSIIVKLPNEEQNIIAKFLLEVDK
 9 TVNNQLVTKLLKQRKKGLLQRMFV
 10
 11

12 **CC873 TRD e* + CC97 TRD c***
 13

14 CCCGGTTGAAGGCGAATGGGAGGAAAATCGATCAGCAGCTTCTGAAAGAAAGCAAAATCAA
 15 GGTAGCAATGGTAGCCATGCAAAAAACTGACCCTTAAACTGTGGGTAAGGTGTGTTCCGAAA
 16 AAAGAACGTTAAAGGCAGCGATAACACCCAGTATTACAAACGTAAAGCAGGTAGCTGATGTAT
 17 GGCAAACGGATTCTGAATTGCGCTTGGTATTGTTCCGGATAGCCTGAATAACTATGAAAGC
 18 ACCATTGATAGCCCGAGCTTGATTTCATTAATGGCGATAGCAAATTCTGCTGGAACGCATTAAA
 19 CTGAAAAGCTCTACAAAAATTGGCGATATTGCAAATGGCAGCCGTAAAGCAAAACGTATTAAT
 20 CAGGATACCTTCTGAGCCTGCCGGTTTGACCGAAATATGATGAACAGCTGCGTATTGGTGA
 21 TTTTCAGTAAACTGGATCGTCAAATTGAATTAGAACAGAACAGCTGGAACGTGCTGCAACAC
 22 CAGGAAAGGTTATCTGAGAAAATCTTAGCCAAGAGCTGCGCTTAAAGATGAAAACGGTAATGAT
 23 TATCCGGAATGGCGTTTGCCCCGTTCAAAGATTATGACAAACCGATTAATATCCGTCGGCA
 24 ATCAACATTAGCAAAGCGAACTGCTGACCGTTAAACTGCATTGCAAAGGTATTGAAAAGCCA
 25 ATTAACCGTGTGCTGAAACTGGGTGCAACCAATTATTACAAACGTTTGAAAGGCCAGTTATCTAT
 26 GGCAAACAGAACCTTTAACGGTGCCTTGATATCGTGCCTTAAAGGTTATCTGATGGCTGTATAGC
 27 AGCAGTGATGTTCCGGCATTTGAAATCAATACCGAGAAAATTGAGCCAACTACTTCATCAGCTAT
 28 ATTAGCCGTCCGAGCTCTATAAAAGCAAAGAGAAATATAGCACCGGACCGGTAGCAAACGTATT
 29 CATGAAAATACCGTGCACCTTAGCCTGCATCTGCCGTGTGAATGAACAGCTGAAAATTGCA
 30 AGCTTGTGTGCTTCTGACCGTAAATTGAACGTGCTGGAACGCCAAATCTATCTGATCAAAAAA
 31 CAGAACAGGCCCTGTCAGCAAATGTTATTCCCGGGGGATCCGATCGATC
 32
 33
 34
 35 **CC873 TRD e***
 36 PGFEGEWEEKSISSFLKESKIKGSNGSHAKKLTVKLWKGVVPKETFKGSDNTQYYKRKAGOLMY
 37 GKDFLNCAFIVPDSLNNYESTIDSPSFDFINGDSKFLERIKLKSFYKFGDIANGSRKAKRIN
 38 QDTFLSLPVFAPKYDEQLRIGEFFSKLDRQIELEEQK
 39
 40 **CC97 TRD c***
 41 QIELEEQKLELLQQQKKGYLQKIFSQELRFKDENGNDYPEWRFARFKDFMYKPINIRPAINISKSE
 42 LLTVKLHCKGIEKANINRVLKLGATNYYKRFEQFIYGKQNFFNGAFDIVPKFDGLYSSSDVPAF
 43 EINTEKIEPNYFISYISRPSFYKSKEKYSTGTGSKRIHENVTLNFSLHLPCLNEQLKIASFCFLN
 44 RKIELLERKIYLIKKQKQALLQQMFI
 45
 46
 47 **CC133-2 from ED133 TRD d***

48 CAAATTGAATTAGAACAGAACAGCTGGAACTGCTGCAACAGCAGAAAAAGGTTATATGAGAAA
 49 ATCTTCAGCCAAGAGCTGCGTTAAAGATGAAAATGGTAACGATTATCCGAATGGAAAATGTG
 50 ATGCTGCAGAAAGTTCTGAAAGATAAAACCGAAGGTATTAAACGTGGTCCGTTGGTGTGCACTG
 51 AAAAGATATTTGTGGAAAGCGGCATGCCGTTATGAACAGCGTAATGCCATTATGATATC
 52 AGCAACTTCCGCTACTATATCAACGAGAACAAATACAAAGAGATGCAGAGCTTAGCGTTAGCCG
 53 AATGATATTATCATGAGCTGTAGCGGCACCATTGGTCGTGGCACTGATTCCGCATAACTATACC
 54 AAAGGTATTATCAACCAGGCCCTGATTGTTCTGTAACATCATAAAATCCGCAGCGAATTCTTT
 55 CTGATCTTATGCGTAGCAATCAGATGCAGCGTAAATTCTGGAAGCAAATCCGGTAGCGCAATT
 56 ACCAATCTGGTCCGGTTAAAGAACTGAAACTGATCCCCTGCGCCGGTTAAATTGAACAG
 57 GATAAAATGCCAGGTTATCACATTATTAACCGTCGTATTGAACAGAGCGAGAAAAAAATCGAA
 58 AGCCTGAAAAATCGCAAACAGGGTTCTGCAGAAACTGTTGTTCCCGGGATCCGATCGATC
 59
 60 **CC133-2 from ED133 TRD d***
 QIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPEWENVMLQKVLKDTEGIKRGPFGGAL
 KKDIFVESGYAVYEQRNAIYDISNFRYYINENKYKEMQSFSVQPNDIIMSCSGTIGRLALIPHNYT

1
2 KGIINQALIRFRTNHKIRSEFFLIFMRSNQMQRKILEANPGSAITNLVPVKELKLIPFPLPVKFEQ
3 DKISQFIHIINRRIEQSEKKIESLKNRKQGFLQKLFV
4
5

6 **ST80-3 TRD X + TRD f***
7

8 CCCGGGTTGAAGGCGAATGGGAGGAAAAACAGTTGCCGATTTACCAAAATTAACCAGGGTCTG
9 CAGATTGCCATTAATGAACGTAACCGAATAGCCCTGAGCTGTATTCATATCACCAACGAA
10 TTTCTGCGTCCGAATAGCCAGACCAAATATTCATTGAAAATCCGCCTCAGAGCGTGATTGCCAAC
11 AAAGAAGATATTCTGATGACCCGACCGTAATACCGGAAAGTTGTTACCAATGTTGGTGC
12 TTCCACAACAACCTTTCAAATCAAATTGATAAAACCTGTATGATCGCCTGTTCTGGTGA
13 GTTCTGAACACGAGCAAATCCAGAACAAAATTCTGAGCCTGGCAGGTAGCAGCACCATTCCGGAT
14 CTGAATCATAGCGATTCTATAGCATTAGCAGCTATCCGCTGTCGCGAACAGCAAAATT
15 GGCAAATTCTTAGCAAACGGATCGCCAGATTGAACTGGAAGAACAGAAACTGGAACGTGCTGCAA
16 CAGCAGAAAAAAGGCTATATGCAGAAATCTTAGCCAAGAGCTGCGCTTAAAGATGAAAACGGT
17 GAAGATTATCCGGATTGGAAAGAAAAAAACTGGCGATATTACCGAGCAGAGCATGTATGGTATT
18 GGTGCAAGCGAACCGTTGATAGCAAAATATCTATATCCGCATCACCGACATCGATGAAAAAA
19 GCCGTAACACTGAATTATCAGAACCTGACCAACACGGATGAACTGAAACAATAAAACTGAAAC
20 GCAACGACATCCTGTTGCACGTACCGGTGCAAGTACCGGTAAAGCTATATTCAAAAGAAGAGA
21 AAGACATCTACAACTACTTTGCGGGTTCTGATCAAATTCAAAATTACGAACAGAACAGTC
22 CGCTGTTCATCTACAGTTACCCCTGACCAAGCAAATTCAACAAATGGTTAAAGTTATGAGCGTGC
23 GTAGCGGTGAGCTGGTATTAATAGCGAAGAATATGCAAAACTGCCGCTGTTCTGCCGAATAAAC
24 TGGAACAACAAAAATCGCGAAATTCTGGATGTTGATCGTCAGATCGAGCTGGAAAAACAAA
25 AAATTGAAATTCTGCAGCAACAAAAAAGGCCTGCTGCAGAGTATGTTATTCCGGGGATCCG
26 ATCGATC
27
28

29 **ST80-3 TRD X**
30

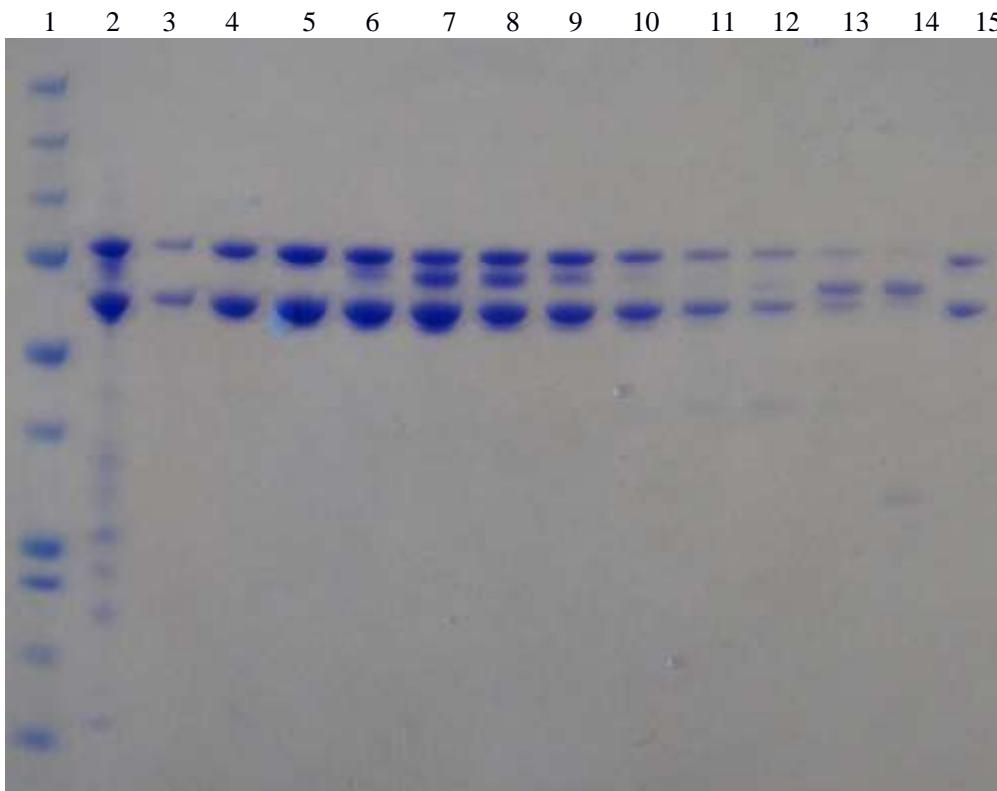
31 MSNTQKKNVPELRFPGEWEEEKQFADFTKINQGLQIAINERKTEYSPELYFYITNEFLRPNQ
32 KYFIENPPQSVIANKEDEILMTRGNTGVVTNVFGAFHNNFFKIKFDKNLYDRLFLVEVLNSSKIQ
33 NKILSLAGSSTIPDLNHSDFYSISSLSPPLLREQQKIGKFFSKLDR
34

35 **ST80-3 TRD f***
36

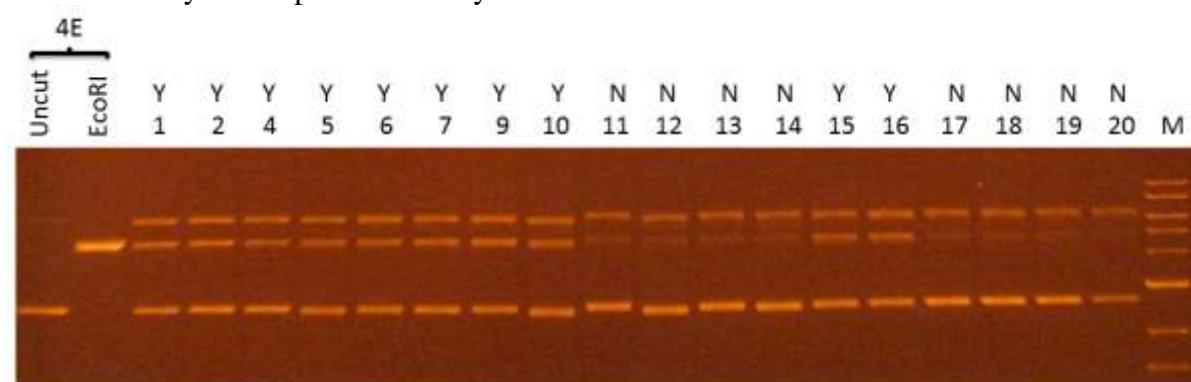
37 QIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGEDYPDWKEKKLDITEQSMYGIGASATRFDS
38 KNIYIRITDIDEKSRLNYQNLTPDELNNKYKLKRNDILFARTGASTGKSYIHKEEKDIYNYYFA
39 GFLIKFKINEQNSPLFIYQFTLTSKFNWKVVKVMSVRSGQPGINSEEEYAKLPLVLPNKLEQQKIAKF
40 LDRFDRQIELEKQKIEILQQQKKGLLQSMFIPGGSHHHHH
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2 **SUPPLEMENTARY INFORMATION FOR TABLE 2.**
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
 2 **S.SauCD-EGFP**
 3 CC30-1 **GWAG-5-GAT**
 4 **This MTase was expressed and purified as a fusion with EGFP.**
 5 MSNTQTKNVPELRFPGEFEWEEKVGELLEFKNGLNKGKEYFGSGSSIVNFKDVNNRSLNTNNL
 6 TGKVNVNSKELKNYSVEKGDVFFTRTSEVIGEIGYPSVIWNDPENTVFSGFVLGRPKSGIDLINN
 7 NFKRYVFFTNSFRKEMITKSSMTTRALTSGSAINMKVYIPVSAKEQRKIGDFFSKLDRQIELEEQ
 8 KLELLQQQKKGYMQKIFSQELRFKDENSEDYPHWENSKIEKYLKERNERSDKGQMLSVTINSGIK
 9 FSELDRKD NSSKDKSNYKVVRKNDIAYNSMRMWQGASGRSNYNGIVSPAYTVLYPTQNTSSLFIGY
 10 KFKTHRMIHKFKINSQGLTS DTWNLK YKQLKNINIDIPVLEEQE KIGDFFKKMDILISKQKIKIEI
 11 LEKEKQSFLQKMF LGSMSVSKGEELFTGVVPILVELDGDVN GHKF SVSGE GEGDATY GKLTLKFICT
 12 TGKLPVPWP TLVTTLYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFE
 13 GTTLVNRIELKGIDF KEDGNILGHKLEYN YN SHNVYIMADKQKNGIKVNFKIRHNIE DGSVQLADH
 14 YQQNTPIGDGPVLLPDNHYLSTQ SALSKDPNEKRDHMVLLEFVTAA GITLGMD ELYKHHHHHH
 15



44 1- marker 2- Nickel column eluate 3-14 Fractions from gel filtration column
 45 15- CC5-1 Purified protein marker
 46 Nuclease assay on the plasmid library.



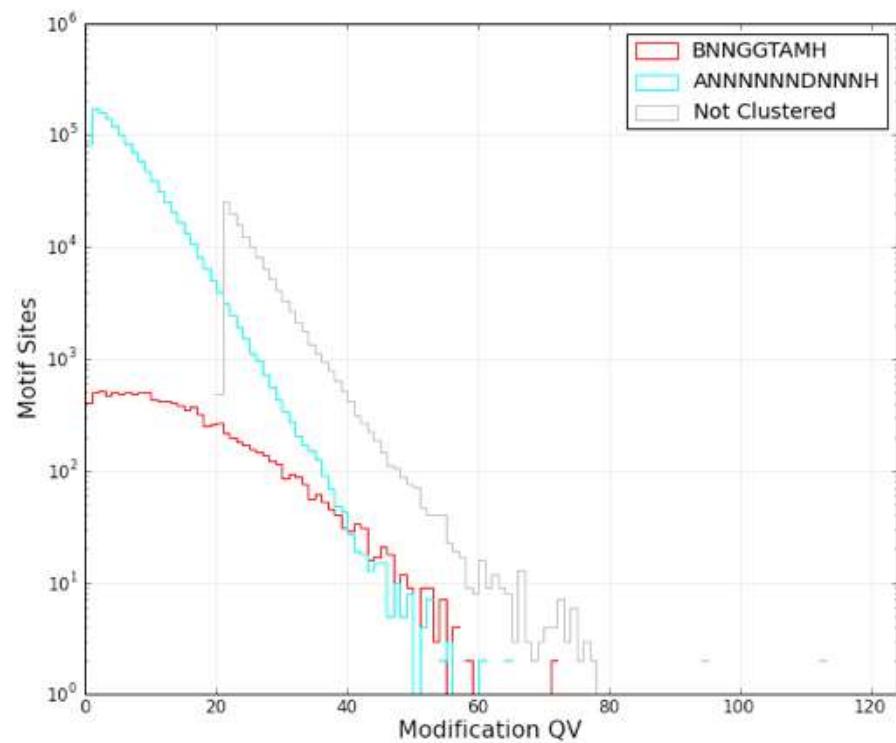
1
 2 **S.SauCD-EGFP**
 3 CC30-1 **GWAG-5-GAT**
 4

5 **SMRT did not work for the CC30-1 system when looking for**
 6 **methylation of genomic DNA from *E. coli*.**
 7

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	
<i>Not Clustered</i>	0		0.19	14583	7880091	36.1	107.4	

14 **Modification QV Histogram By Motif**
 15

16 **Modification QV Histogram**
 17



S . SauJK-EGFP

CC30-2 GGA-7-TCG

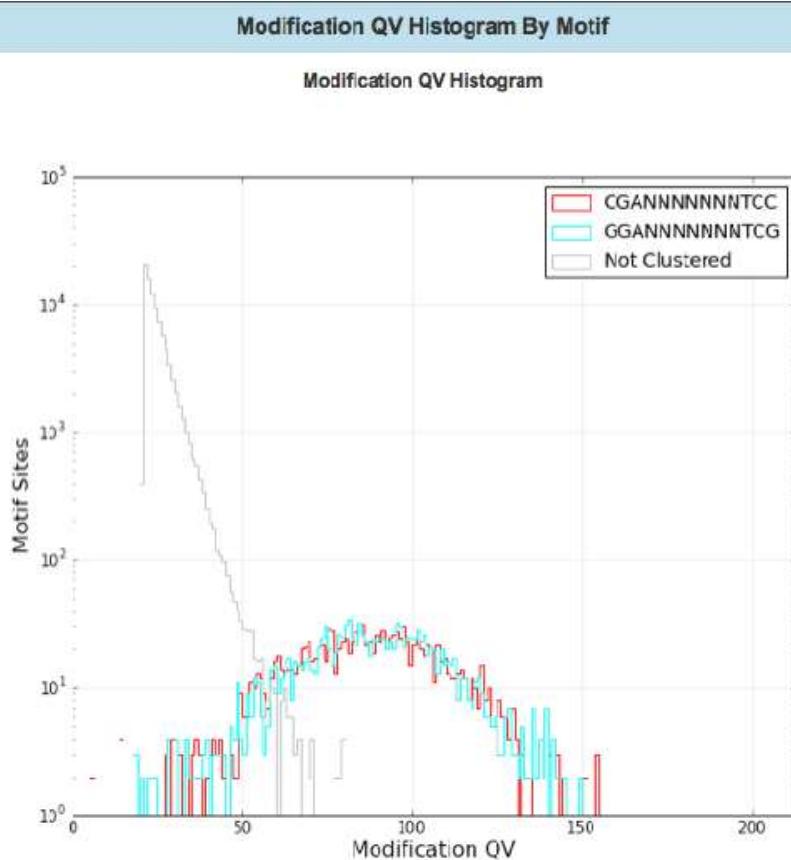
This MTase was a fusion with EGFP.

MSNTQKKNVPRLFPEFEGERWEERKLGDLIKVNNSGKDYKHLKDGDIPVYGTGGYMTSVSEPLSEID
AVGIGRKGTINKPYLLEAPFWTVDTLFYCTPEKEADILFILSLFRKINWKLYDESTGVPSLSKQTI
NKNRVLPTNKEQQKIGEFFSKLDRQIELEEQQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPKW
EEKKIEDIASQVYGGGTPNTKIKEFWNGDIPWIQSSDVKVNDLILQQCNKFISKNSIELSSAKLIP
ANSIAIVTRVGVKLCIVEFDYATSQDFLSLSSLKYDKLYSLYSLLYTMKKISANLQGTSIKGITK
KELLDIPIKPHNLEEQQKIGDLFYKIDKYISFNCKIEMLKSLKQGLLKKMFIGSMVSKGEELFT
GVVPILVLDGDVNGHKFSVSGEGEGLATYGKLTLKFICTTGKLPVPWPTLVTTLTGYVQCFSRYP
DHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVRIELKGIDFKEDGNILGHKL
EYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSAL
SKDPNEKRDHMVLLEFVTAAAGITLGMDELYKHHHHHH

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
<i>Not Clustered</i>	0		0.09	8260	9123294	35.7	87.7	



1
2 S.SauJd*

3 CC133-2 from ED133 GGA-7-TTRG

4 This enzyme was studied using the SMRT assay. There are minor
5 variations in S subunit sequence in CC133-2.

6 Recombinant S.SauJd* CC133-2

7 MSNTQKKNVPELRFPGEFEWEEKKLGDLIKVNNSGKDYKHLKGDI PVYGTGGYMTSVSEPLSEID
8 AVGIGRKGTINKPYLLEAPFWTVDLFYCTPKKETDILFILSLFRKINWKVYDESTGVPSLSKQTI
9 NKINRFVPSNKEQQKIGEFFIKLDRQIEEQQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPEW
10 ENVMLQKVLDKTEGIKRGPFGGALKKDIFVESGYAVYEQRNAIYDISNFRYYINENKYKEMQSFS
11 VQPNDIIMSCSGTIGRLALIPHNYTKGIINQALIRFRTNHKIRSEFFLIFMRSNQMQRKILEANPG
12 SAITNLVPVKELKLIPPLPVKFEQDKISQFIHIINRRIEQSEKKIESLKNRKQGFLQKLFVPAGS
13 HHHHHH
14

15 Wild type S.SauJd*

16 MSNTQKKNVPELRFPGEFEWEEKKLESIIKVNSGKDYKHLKGDI PVYGTGGYMTSVSEPLSEID
17 AVGIGRKGTINKPYLLEAPFWTVDLFYCTPKKETDILFILSLFRKINWKVYDESTGVPSLSKQTI
18 NKINRFVPTNKEQQKIGKFFSKLDRQIELQEQQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPEW
19 ENVMLQKVLDKTEGIKRGPFGGALKKDIFVESGYAVYEQRNAIYDISNFRYYINENKYKEMQSFS
20 VQPNDIIMSCSGTIGRLALIPHNYTKGIINQALIRFRTNHKIRSEFFLIFMRSNQMQRKILEANPG
21 SAITNLVPVKELKLIPPLPVKFEQDKISQFIHIINRRIEQSEKKIESLKNRKQGFLQKLFV*

22 Reports for Job Dryden_J_delta_MODs

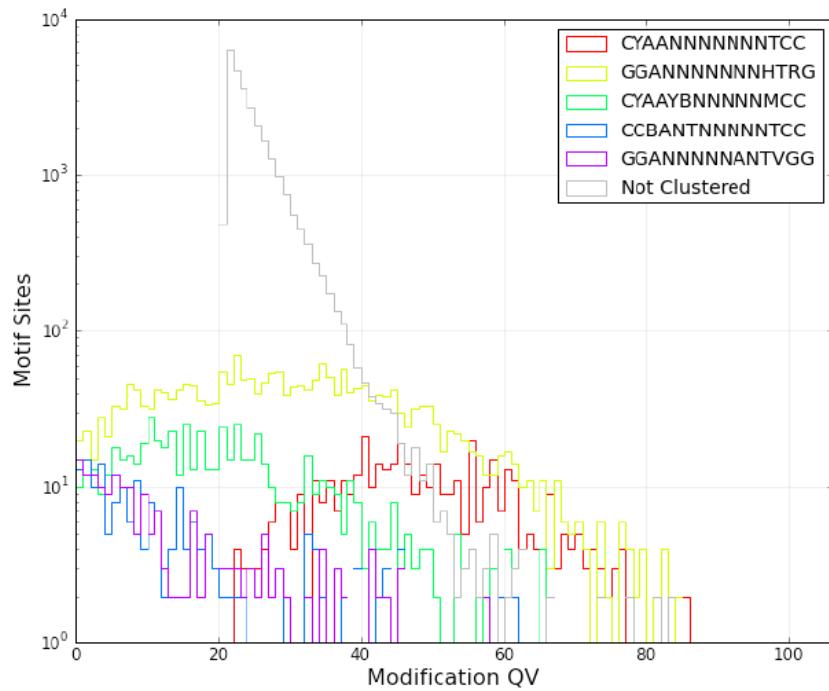
23 Pacific Biosciences

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

24 Motif Summary

25 SMRT Cells: 1 Movies: 1

26 Modification QVs



1
2 **S.SauNE**
3 **CC398-1 ACC-5-RTGA**

4
5 The clone obtained contained a single amino acid substitution A50S
6 which did not affect activity. The enzyme was expressed using
7 plasmid pSauNE-XmaI.

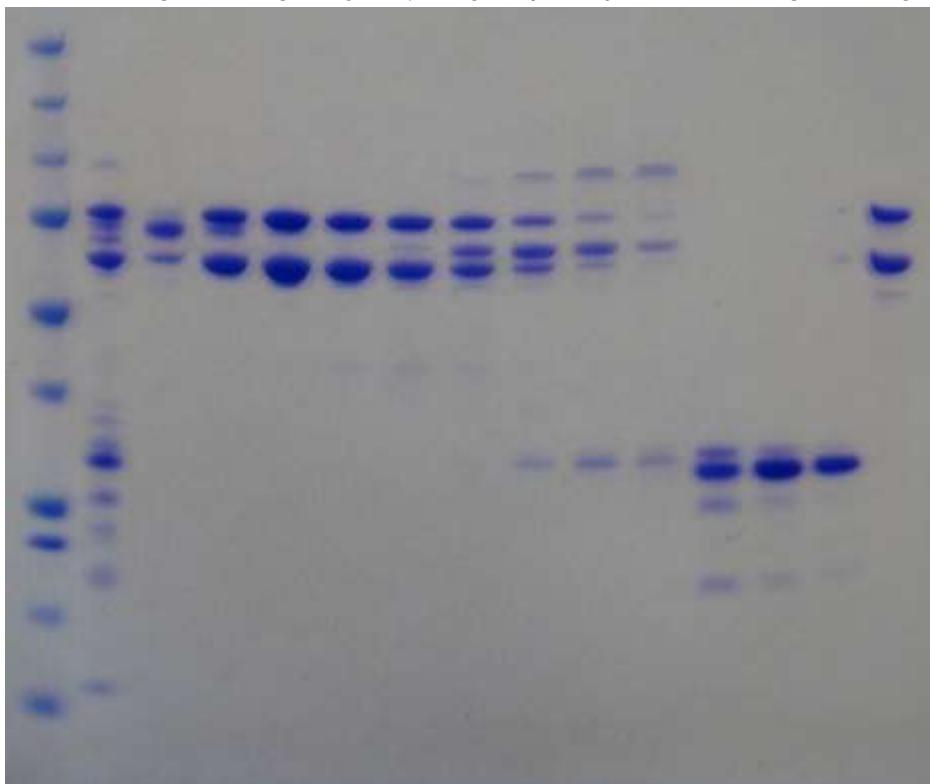
8 **S.SauNEXmaI "Expected" sequence**

9
10 MSNTQKKNVPELRFPGEFEWEEKKLGEFAGKVTQKNVDKKYIETLTNS**A**ELGIISQKDYFDKEIS
11 NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGGKGVMSPLYTVFKIQNIIDLNFIEFYFKSS
12 KWYRFMALNGDSGARADRFSIKDRTFMEMPLHPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
13 KKGYMQKIFSQELRFKDENGKDYPEWEETTIKEIAQINTGKDKDAITNGSYDFYVRSPIVYKIN
14 TFSYEGEAILTVGDGVGVGKVFHYVNGKFDYHQRVYKISDFKNYYGLLFYYFSQNFLKETKKYSA
15 KTSVDSVRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTQKQVIELLKQRKKSLLQKMFIPGGS
16 HHHHHH*

17 **S.SauNEXmaI "Actual" sequence**

18 MSNTQKKNVPELRFPGEFEWEEKKLGEFAGKVTQKNVDKKYIETLTNS**S**ELGIISQKDYFDKEIS
19 NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGGKGVMSPLYTVFKIQNIIDLNFIEFYFKSS
20 KWYRFMALNGDSGARADRFSIKDRTFMEMPLHPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
21 KKGYMQKIFSQELRFKDENGKDYPEWEETTIKEIAQINTGKDKDAITNGSYDFYVRSPIVYKIN
22 TFSYEGEAILTVGDGVGVGKVFHYVNGKFDYHQRVYKISDFKNYYGLLFYYFSQNFLKETKKYSA
23 KTSVDSVRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTQKQVIELLKQRKKSLLQKMFIPGGS
24 HHHHHH*

25 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15



53 1- marker 2- Nickel column eluate 3-14 Fractions from gel filtration column
54 15- CC5-1 purified protein marker

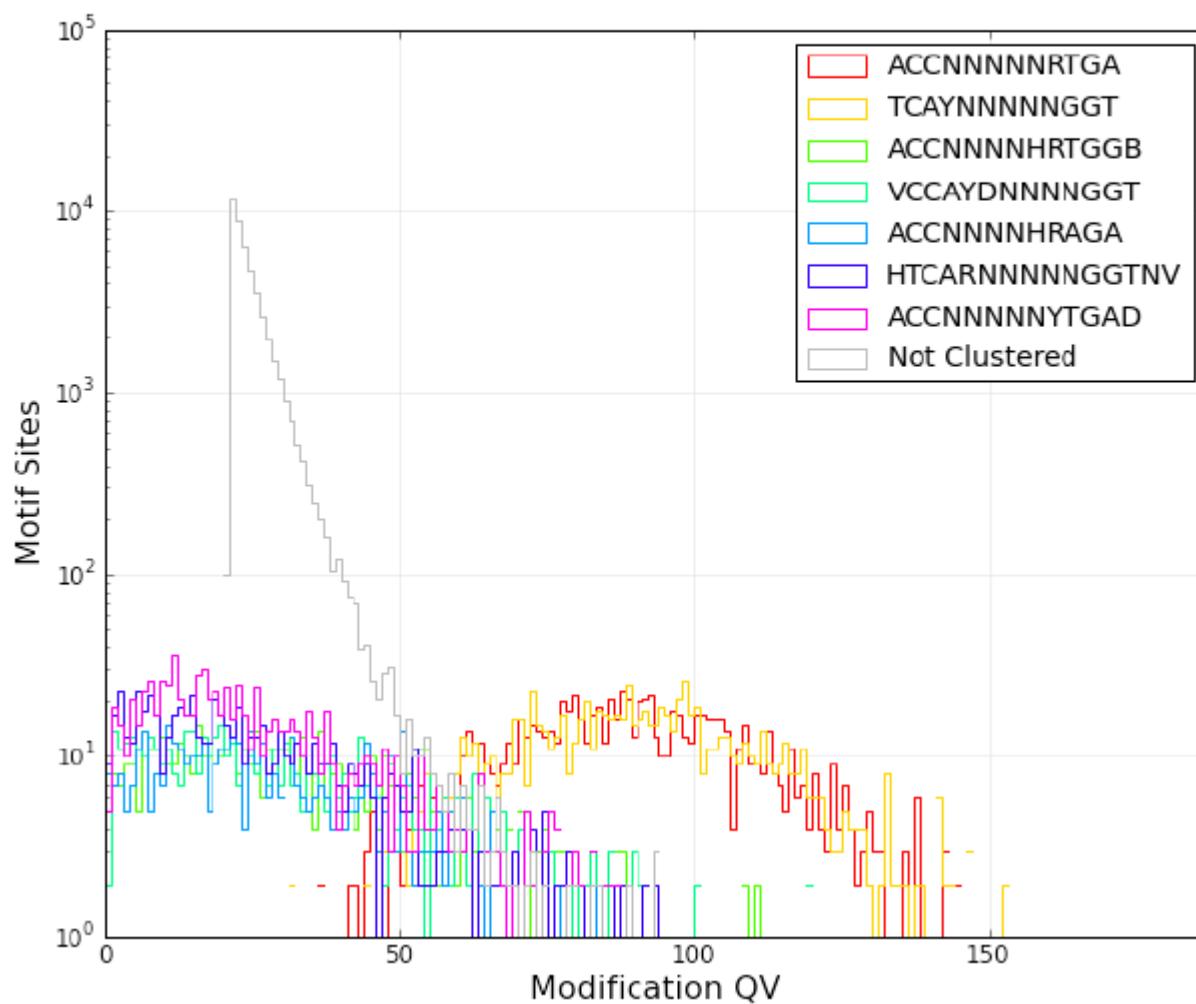
1
 2 S . SauNE
 3 CC398-1 ACC-5-RTGA
 4

5 Reports for Job Ed_1_Dryden_MODs

PACIFIC BIOSCIENCES

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	VCCAYDNNNNNGGT
VCCAYDNNNNNGGT	4	m6A	45.38%	269	593	54.62	61.85	ACCNnnnHRTGGB
ACCNnnnHRAGA	1	m6A	41.75%	200	479	48.38	61.76	
HTCARNnnnnNGTNV	4	m6A	36.31%	264	727	51.22	62.33	
ACCNnnnNYTGAD	1	m6A	34.9%	320	917	49.93	60.88	

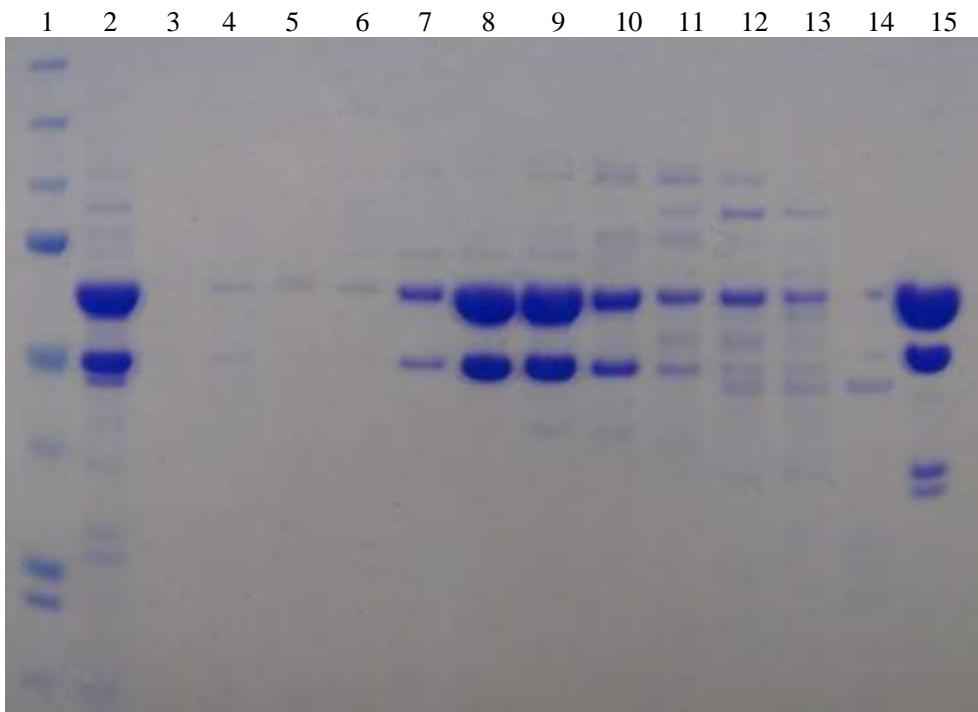
Modification QVs



1
2 **SUPPLEMENTARY INFORMATION FOR TABLE 3.**
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
 2 S.SauBE AGG-5-RTGA
 3 This MTase was purified but cut all the plasmids in the nuclease
 4 assay. Therefore once the targets for each TRD had been determined
 5 from other MTases, we used the ATPase assay to verify the length
 6 of the non-specific spacer.
 7
 8

9 MSNTQKKNVPELRFPGEFEWEEEKKLGDLTDRVIRKNKNLESKKPLTISGQLGLIDQTEYFSKSVS
 10 SKNLENYTLIKNGEFAYNKSYSNGYPLGAIKRLTRYDSGVLSLYICFSIKSEMSKDFMEAYFDST
 11 HWYREVSGIAVEGARNHGLLNVSNDFTILIKYPSLEEQQKIGKFFSKLDRQIELEEQKLELLQQ
 12 QKKGYMQKIFSQELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKDAITNGSYDFYVRSPIVYKI
 13 NTFSYEGERAILTVDGVGVGVKFHYVNGKFDYHQRVYKISDFKNEYGLLLFFYYFSQNFLKETKKYS
 14 AKTSVDSVRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTQKQVIELLKQRKKSLLQKMFIPGG
 15 SHHHHH
 16



40 1- marker 2- Nickel column eluate 3-14 Fractions from gel filtration column
 41 15- CC398-1 purified protein marker
 42

43 **Oligonucleotides for checking BE target site using ATPase assay.**

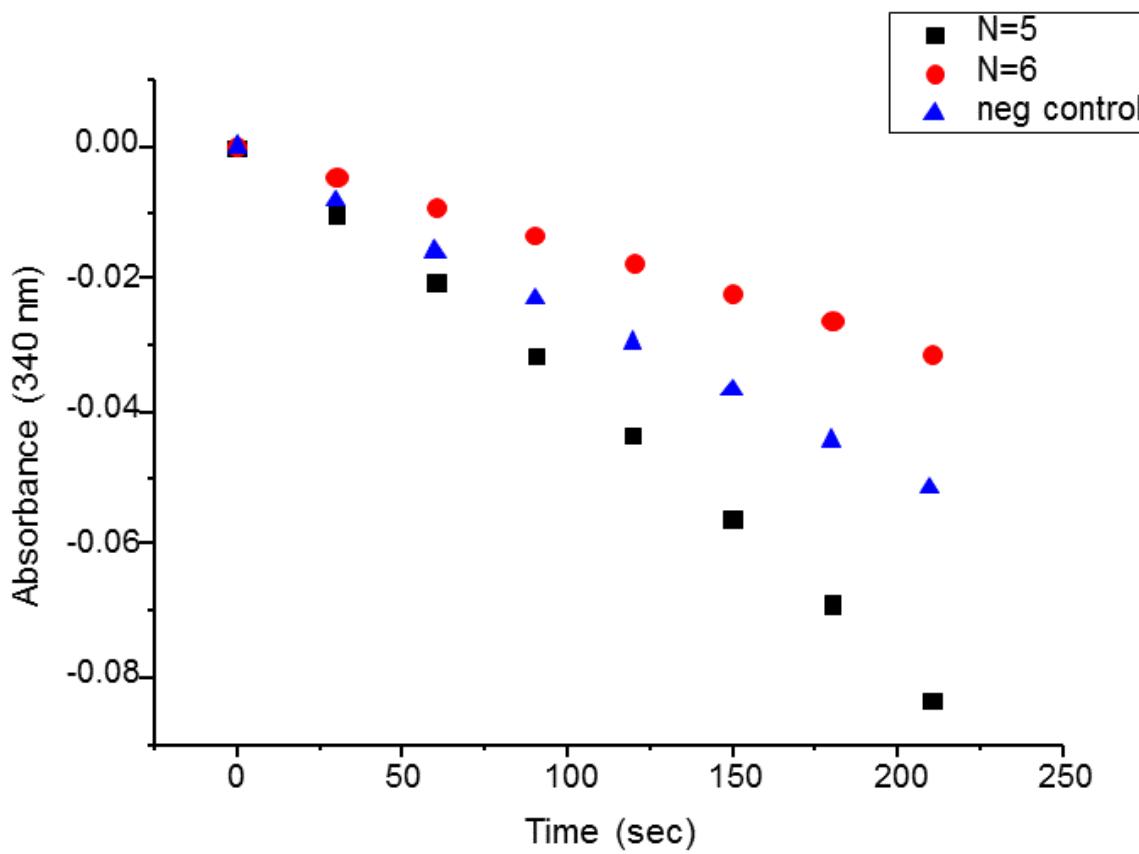
44 Underlined refers to methylated bases.

45 5'-AGG-N-RTGA-3'

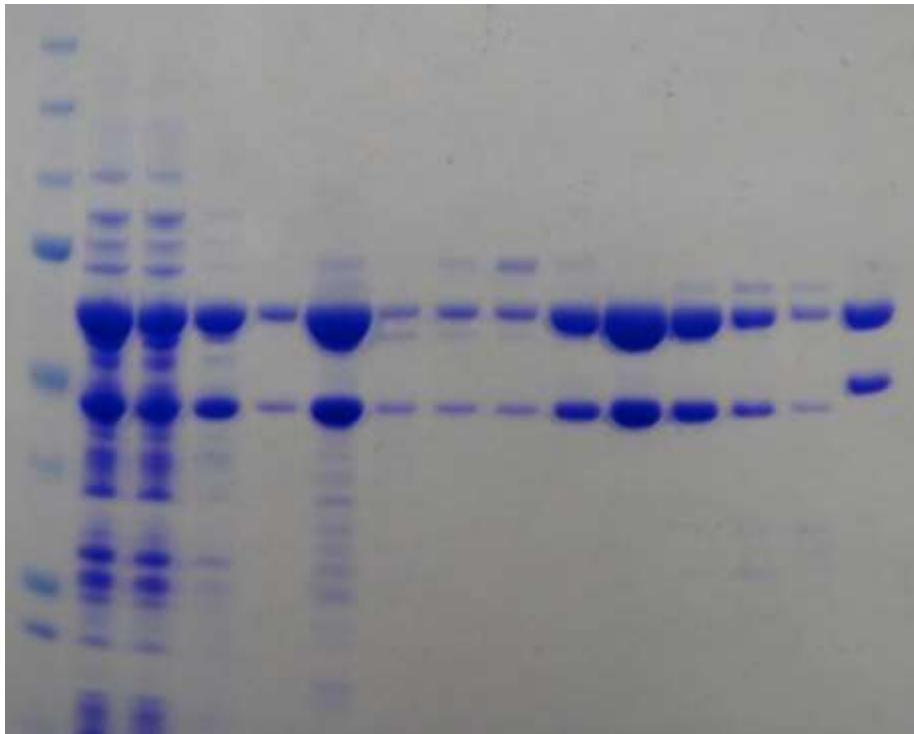
46 3'-TCC-N-YACT -5'

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

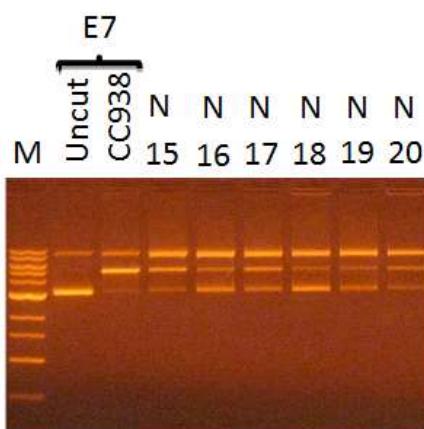
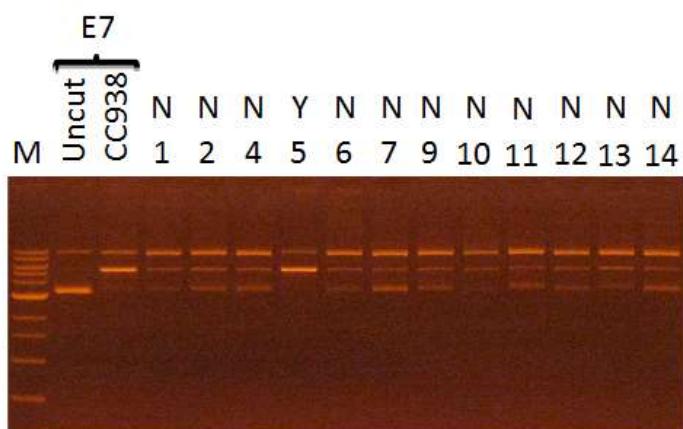
Oligonucleotide name	DNA sequence (5' to 3')
BE5for	AGATGATGGAATCAATGCAGGTTCCAGTGAGCCCTATACGATATAA
BE5rev	TTATATCGTATAGGGCTCACTGGAACCTGCATTGATTCCATCATCT
BE6for	AGATGATGGAATCAATGCAGGTTCACAGTGAGCCCTATACGATATAA
BE6rev	TTATATCGTATAGGGCTCACTGTGAACCTGCATTGATTCCATCATCT

1
2 S . SauBE AGG-5-RTGA3
4 N=5 gives the most activity therefore we conclude from the ATPase
5 assay that the site for the BE TRD combination is AGG-5-RTGA.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
 2 S.SauJE GGA-6-RTGA
 3 This MTase was used in both nuclease and SMRT assays. The TRD pair
 4 JE occurs in other ST groups namely ST49 and ST50.
 5
 6 MSNTQKKNVPELRFPGEFEWEEKKLGDLIKVNNSGKDYKHLERGDIPVYGTGGYMTSVSEPLSEID
 7 AVGIGRKGTINKPYLLEAPFWTVDTLFYCTPKKETDILFILSLFRKINWKVYDESTGVPSLSKQTI
 8 NKINRFVPSNKEQQKIGEFFIKLDRQIELEEQQKLELLQQQKKGYMQKIFSQELRFKDENGKDYPEW
 9 EETTIKEIAQINTGKDKTDKDAITNGSYDFYVRSPIVYKINTFSYEGERAILTVDGTVGVGVKFHYVN
 10 GKFDYHQRVYKISDFKNYYGLLFYYFSQNFLKETKKYSAKTSVDSVRKDMIANMKVPRPIYIEQK
 11 KIGQFIKRVDNKTKIQKQVIELLKQRKKSLLQKMFIPGGSHHHHH
 12 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15



37 1- marker 2- soluble cell extract 3- Nickel column flow through 4- Nickel column wash 1
 38 5- Nickel column wash 2 6- Nickel column eluate 7-14 Fractions from gel filtration column
 39 15- CC398-1 purified protein marker
 40 Possible site: GGANNNNNNRTGA Note that the background
 41 linearisation may be due to the enzyme displaying star activity
 42 against a similar site (i.e., a single GGAN7RTGA site is found in
 43 pUC19) to the real site (GGAN6RTGA). Repeated digests generate an
 44 identical pattern of digestion.
 45

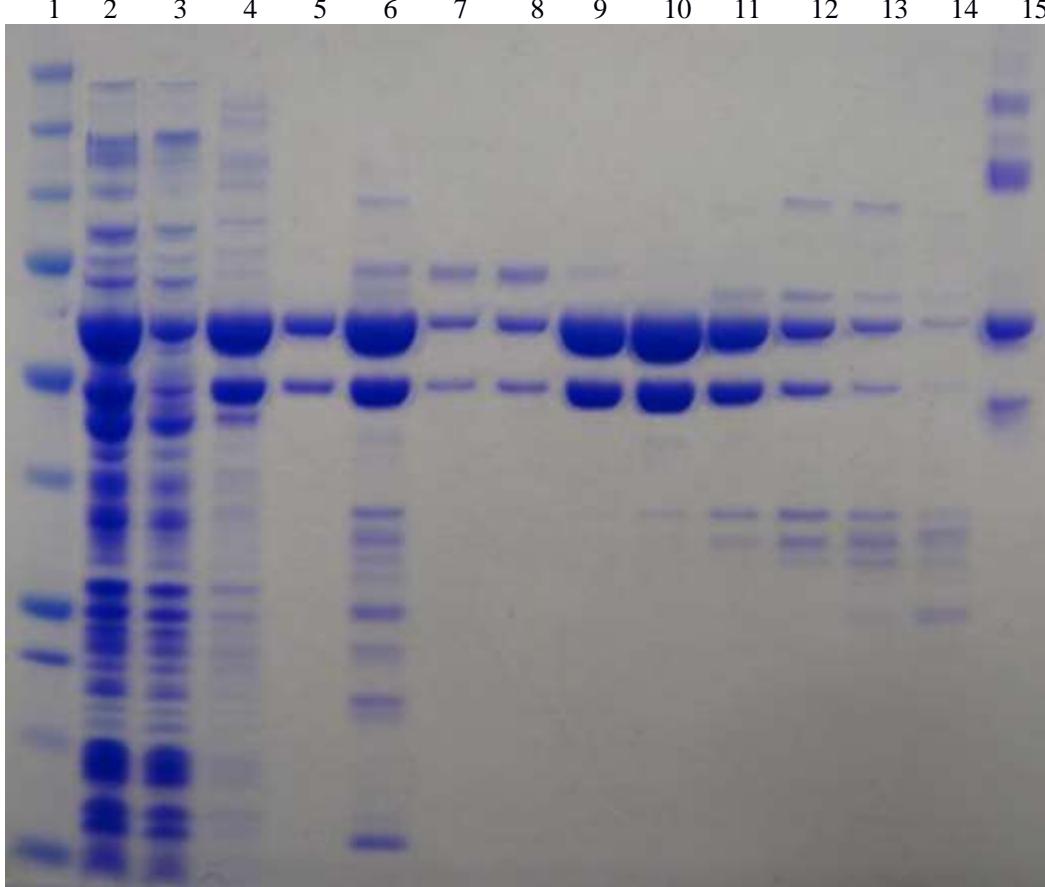


1
2 **S . SauJE GGA-6-RTGA**
3 SMRT data showed only the N=6 spacer giving modification.
4
5

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		
TCAYNNNNNNNTCC	3	m6A	31.44	305	970	38.3	17.1	GGANNNNNNRTGA		
GGANNNNNNRTGA	3	m6A	24.43	237	970	38.3	17.5	TCAYNNNNNNNTCC		
Not Clustered	0		0.00	324	9,124,268	34.2	15.8			

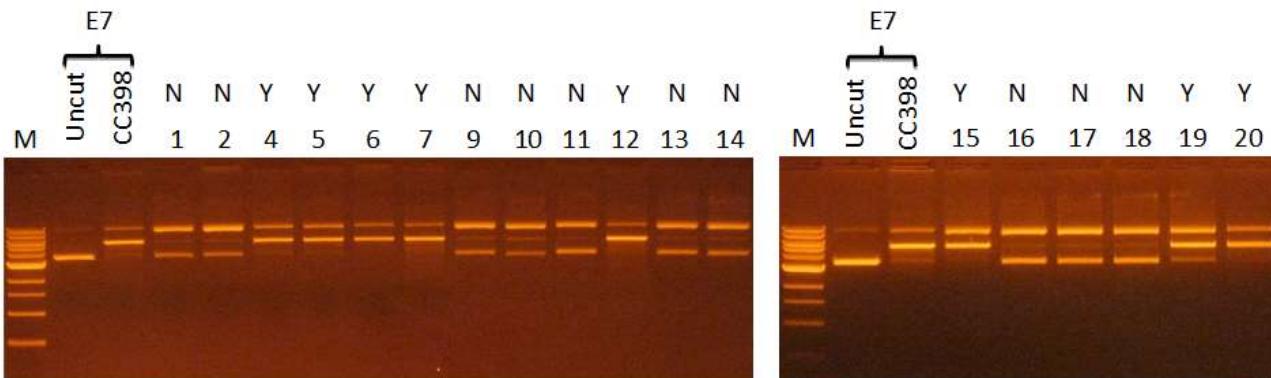
1
2 **S. SauNI ACC-6-TGAR**
3

4 MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
 5 NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNLGKKGVMSPLYTVFKIQNIIDLNFIEFYFKSS
 6 KWYRFMALNGDGSARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
 7 KKGYMQKIFSQELRFKNENGNDYPDWERIKFFDVIDKVIDFRGRTPKKLNMEWSDEGYLALSANV
 8 KKGYIDFNVEAKYGNLDLYTRWMRGNELYKGQVLFTTEAPMGNAQVPDNKGYILSQRTIAFNSNE
 9 KITDNFLASLLSSENVYNDLLKLCGATAKGVSQKNLNRLYVTIPHISSEQEEIAEFFRKINQLVE
 10 LQKYKIEHTKSQKQVFLQKMFIPGGSHHHHHH
 11 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
 12



40 1- marker 2- soluble cell extract 3- Nickel column flow through 4- Nickel column wash 1
 41 5- Nickel column wash 2 6- Nickel column eluate 7-14 Fractions from gel filtration column
 42 15- CC398-1 purified protein marker

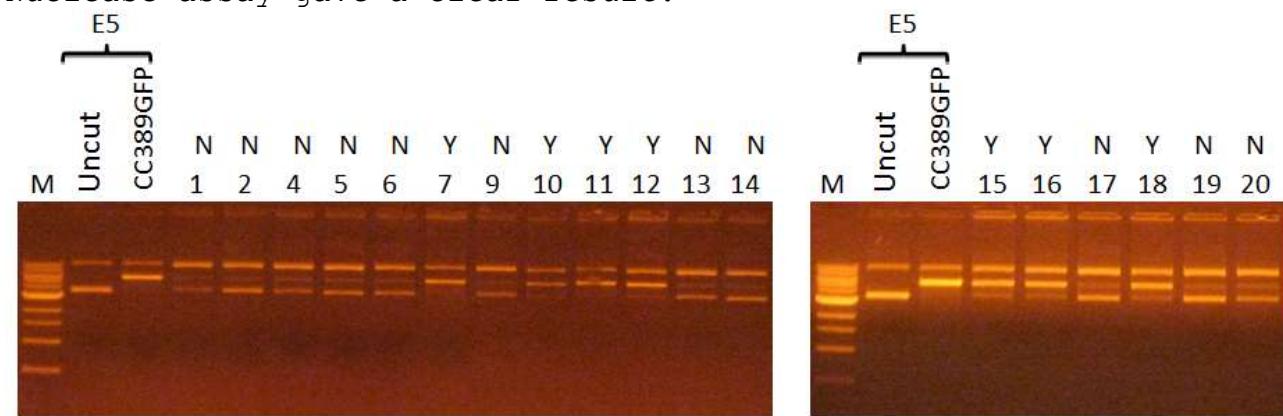
43 Nuclease assay on the plasmid library gave a clear result.



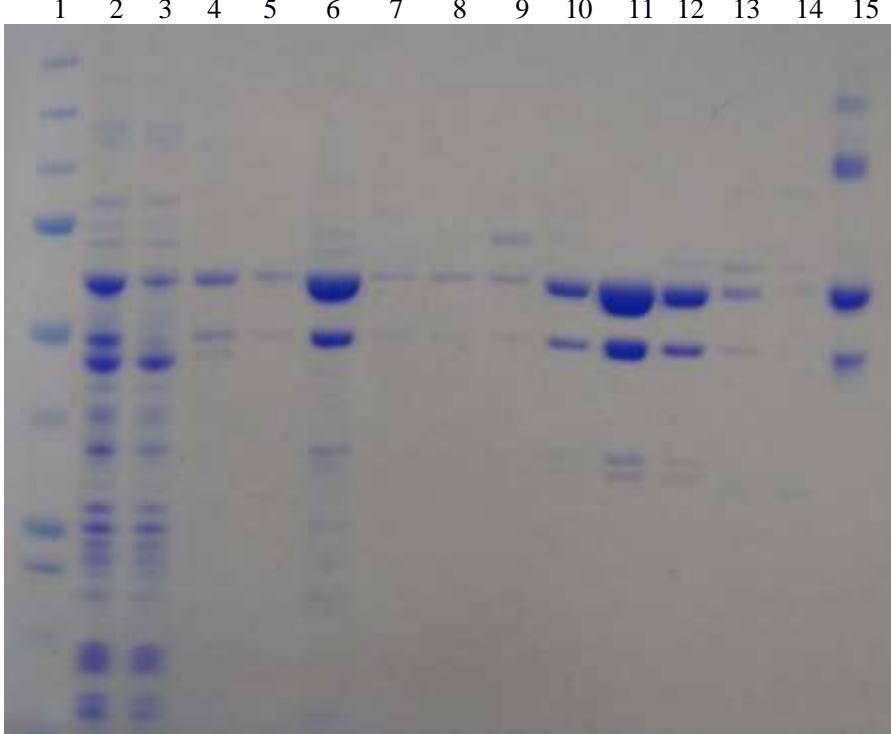
1
 2 **S . SauNK ACC-6-TCG**
 3 MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
 4 NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGGKGVMSPLYTVFKIQNIDLNFIEFYFKSS
 5 KWYRFMALNGDSGARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
 6 KKGYMQKIFSQELRFKDENGNDYPNWEEKKIEDIASQVYGGGTPTKIKEFWNGDIPWIQSSDVKV
 7 NDLILRQCNKFISKNSIELSSAKLIPANSIAIVTRVGVGKLCLVFDYATSQDFLSLSSLKYDKLY
 8 SLYSLLYTMKKISANLQGTSIKGITKELLDI_IKIPHNEEQQQIGDLFYKIDKYISFNKCKIEI
 9 LKSLKQGLLQKIFIPGGSHHHHHH
 10
 11 1 2 3 4 5 6 7 8
 12



40 1- marker 2- soluble cell extract 3- Nickel column flow through
 41 4- Nickel column wash 1 5- Nickel column wash 2 6- Nickel column eluate
 42 7- Eluate after PD10 desalting 8- CC398-1 purified protein marker
 43 Nuclease assay gave a clear result.
 44

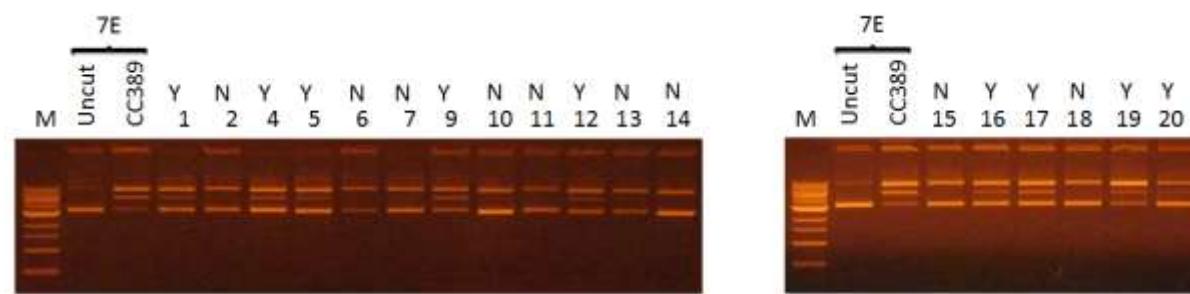


1
 2 **S . SauNL ACC-6-TAAA**
 3 MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
 4 NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGGKGVMSPLYTVFKIQNIIDLNFIEFYFKSS
 5 KWYRFMALNGDGSARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
 6 KKGYMQKIFSQELRFKDENGNDYPNWRTIELKNILENIVDNRGKTPDNAPSEKYPLLEVNALGYYR
 7 PAYIKVSFKFVSENTYNNWFREHLKENDILFSTVGNTGIVSLMDNYKAVIAQNIVGLRVNNNNLPSF
 8 IYYMLSYSKGNQKKIKRIQMGAQPSVKVSQFKFIKYLVPIKDEQEKVAKLLIEIDKLVNKQLIKIE
 9 LLQQRKKALLKSMFIPGGSHHHHHH
 10
 11 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15



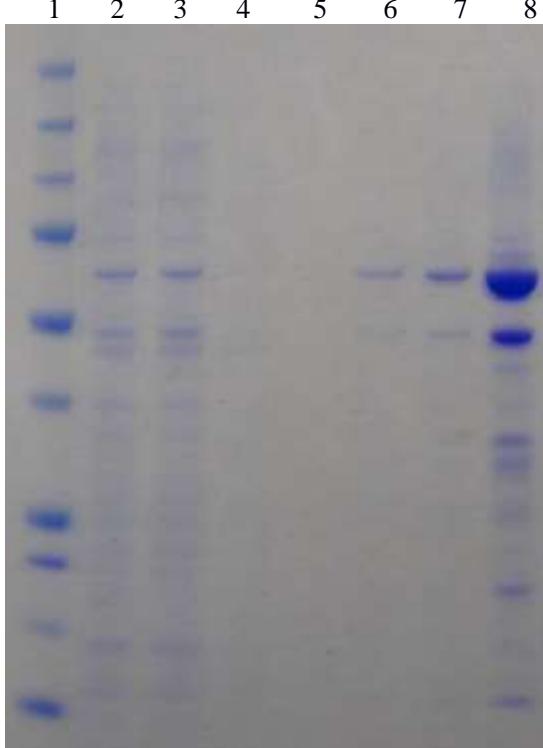
35 1- marker 2- soluble cell extract 3- Nickel column flow through 4- Nickel column wash 1
 36 5- Nickel column wash 2 6- Nickel column eluate 7-14 Fractions from gel filtration column
 37 15- CC398-1 purified protein marker
 38

39 Nuclease assay gave a clear result.



1
2 **S. SauNP ACC-5-CCT**

3 MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
4 NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGGKVMSPLYTVFKIQNIIDLNFIEFYFKSS
5 KWYRFMALNGDSGARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELFQQQ
6 KKGYMQKIFSQELRFKDESGNDYPDWEEKELGEVADRVIRKKNFESKKPLTISGQLGLIDQTEYF
7 SKSVSSKNLENYTLIKNGEFAYNKSYSNGYPLGAIKRLTRYDSGVLLSILYICFSIKSEMSKDFMEA
8 YFDSTHWYREVSGIAVEGARNHGLLNISVNDFFTILIKYPSLEEQRKIGDFFIKLDRQIELEEQKL
9 ELLQQRKKALLKSMLIPGGSHHHHHH
10
11 1 2 3 4 5 6 7 8
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35



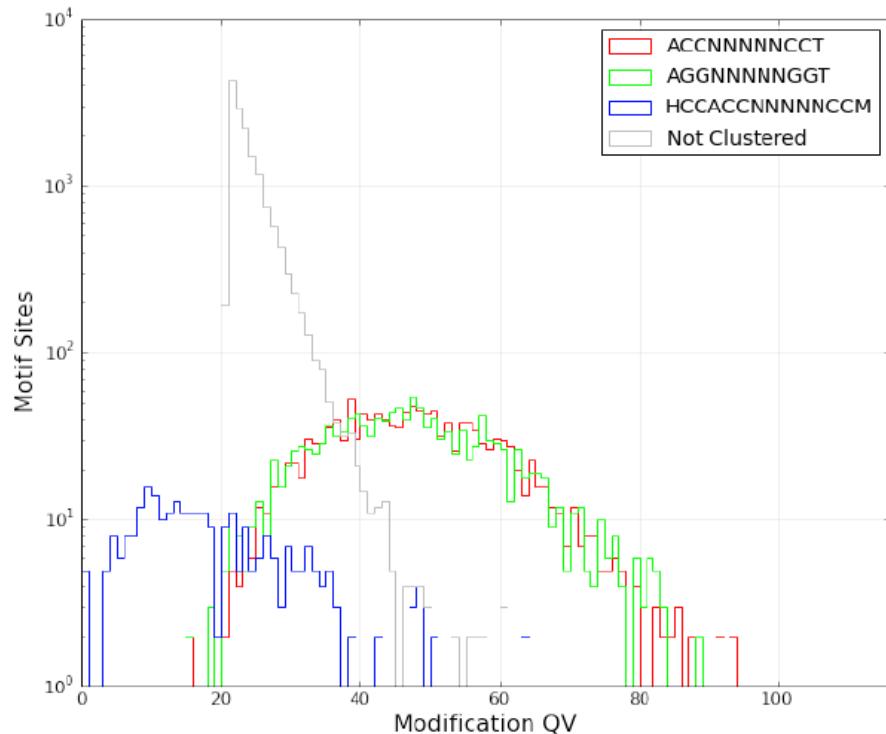
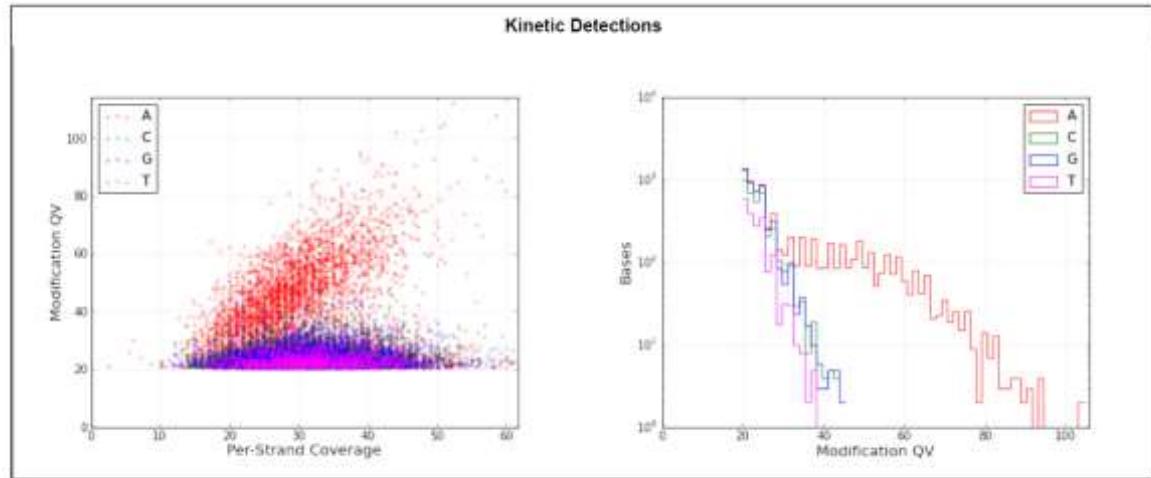
36 1- marker, 2- soluble cell extract, 3- Nickel column flow through, 4- Nickel column wash 1, 5-
37 Nickel column wash 2, 6- Nickel column eluate, 7- eluate after conc. and PD10 desalting, 8- Final
38 concentrated protein
39

40 Although purified this MTase was only assayed via SMRT.
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2 S . SauNP ACC-5-CCT
3

Motifs

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

10 Modification QV Histogram By Motif
11
12
13
1440 Kinetic Detections
41
4255 Motifs
56

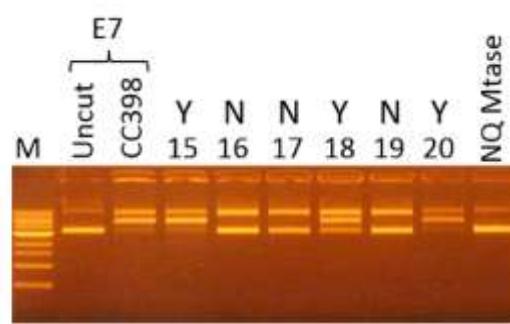
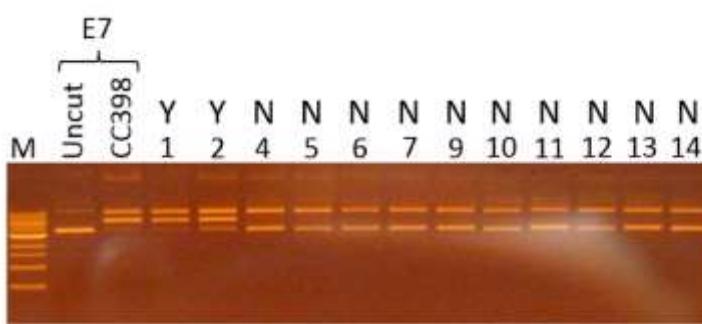
Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACCNINNNNOCCT	1	m6A	91.03	1320	1450	49.8	29.6	AGGNNNNNGGT
AGGNNNNNGGT	1	m6A	89.79	1302	1450	50.0	29.7	ACCNINNNNCCCT
HCCACCNNNNNCCM	4	m6A	17.39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

1
2 **S. SauNQ ACC-5-RTGT**
3

4 MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
5 NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGKKGVMSPLYTVFKIQNIIDLNFIEFYFKSS
6 KWYRFMALNGDSGARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
7 KKGYMQKIFSQELRFKDENGEDYSEWEERRFADIFKFHNKLRKPIKENLRVKGSYPYYGATGIIDY
8 VDDFI FDGYNYLLIGEDGANIITRSAPLVYLVNGKFWVNNAHILSPLNGNIQYLYQVAELVNYEKY
9 NTGTAQP KLN IQNLKIINVVI STN LEE QQKIGSFLSKLDRQIDLEEQKLELLQQRKKALLKSMFVP
10 GGSHHHHHH



39 1- marker 2- soluble cell extract 3- Nickel column flow through
40 4- Nickel column wash 5- Nickel column eluate 6- eluate after conc. and PD10 desalting
41 7- Final concentrated protein
42 DNA cleavage assay and SMRT assay agreed.
43

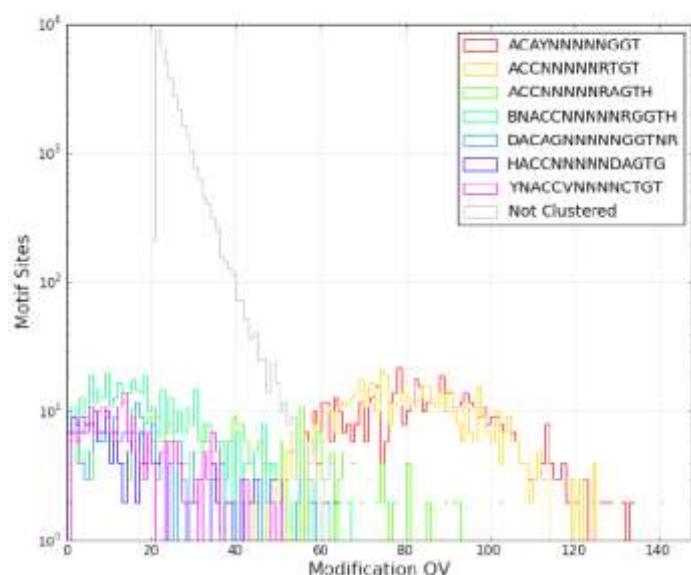


1
2 S . SauNQ ACC-5-RTGT
3
4

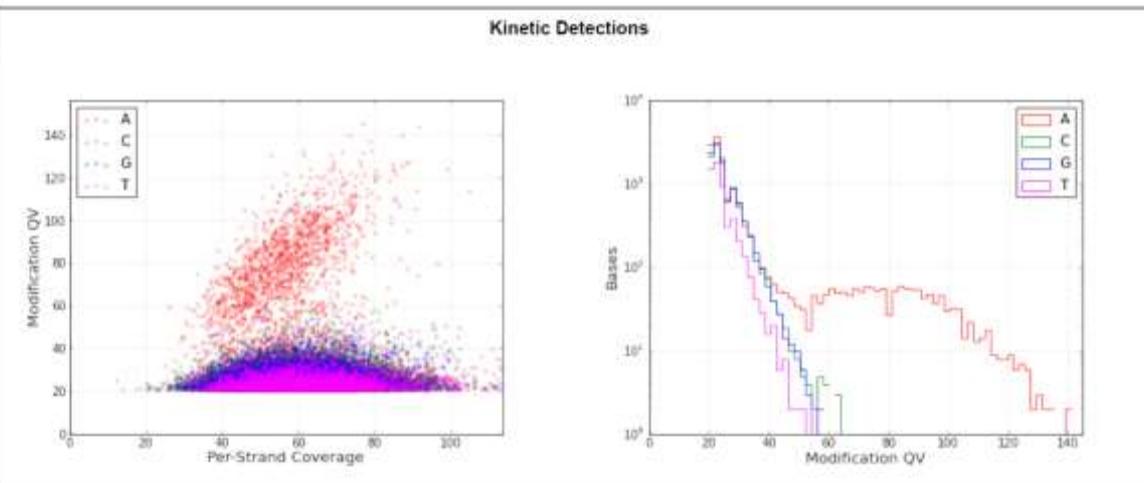
Motifs

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	ACCN>NNNRTGT
ACCN>NNNRTGT	1	m6A	99.85	654	655	80.7	55.5	ACAYNNNNNGGT
ACCN>NNNRRAGTH	1	m6A	55.56	215	387	54.3	56.5	
BNACCN>NNNRRGGTH	3	m6A	23.74	118	497	45.8	57.6	
DACAGNNNNNGGTNR	4	m6A	21.65	50	231	42.8	55.5	
HACCN>NNNNDAGTG	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



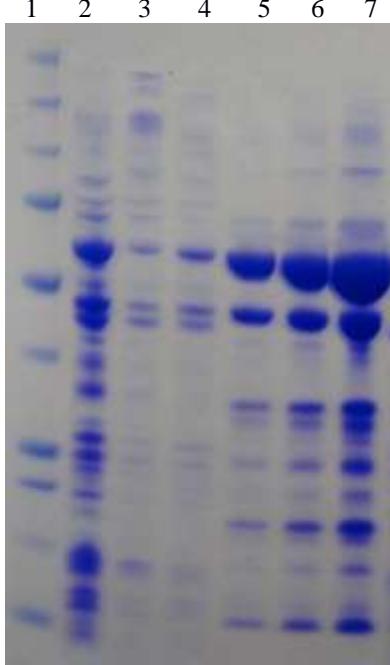
Kinetic Detections



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	ACCN>NNNRTGT
ACCN>NNNRTGT	1	m6A	99.85	654	655	80.7	55.5	ACAYNNNNNGGT
ACCN>NNNRRAGTH	1	m6A	55.56	215	387	54.3	56.5	
BNACCN>NNNRRGGTH	3	m6A	23.74	118	497	45.8	57.6	
DACAGNNNNNGGTNR	4	m6A	21.65	50	231	42.8	55.5	
HACCN>NNNNDAGTG	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	

1
2 **S . SauNS ACC-6-TGC**
3 MSNTQKKNVPELRFPGEFEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
4 NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGKKGVMSPLYTVFKIQNIIDLNFIEFYFKSS
5 KWYRFMALNGDGSARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
6 KKGYMQKIFSQELRFKDENGNDYPDWTNERLGEVTTVTMGQSPKSVDNYTDNSNDTVLIQGNADIEN
7 GLINPRIYTREVTKLIQKDEIILTVAAPVGKLAMAQINACIGRGVCSIKGDKFLYYFLEWFATQNQ
8 WIRFSQGSTFESISGNDIRNIHIKIPVEDERTKIIKLLNSLDVLNSKTDLKIQNLKQRKQSLLQKI
9 FVPGGSHHHHHH
10
11 1 2 3 4 5 6 7
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32



33 1- marker 2- soluble cell extract

34 3- Nickel column flow through 4- Nickel column wash

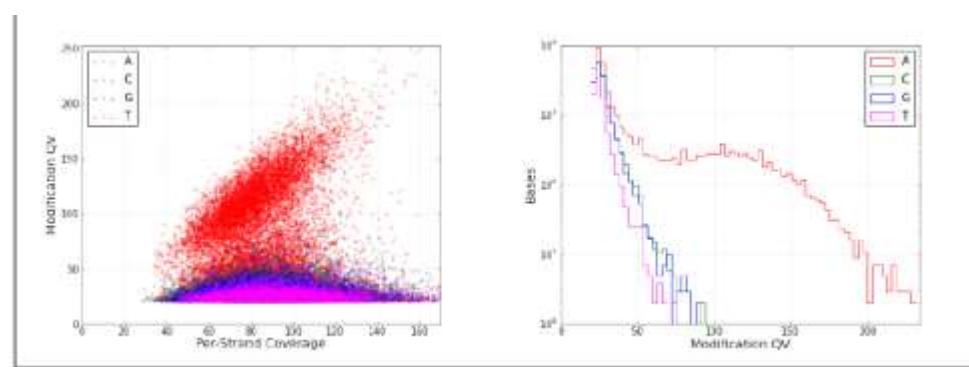
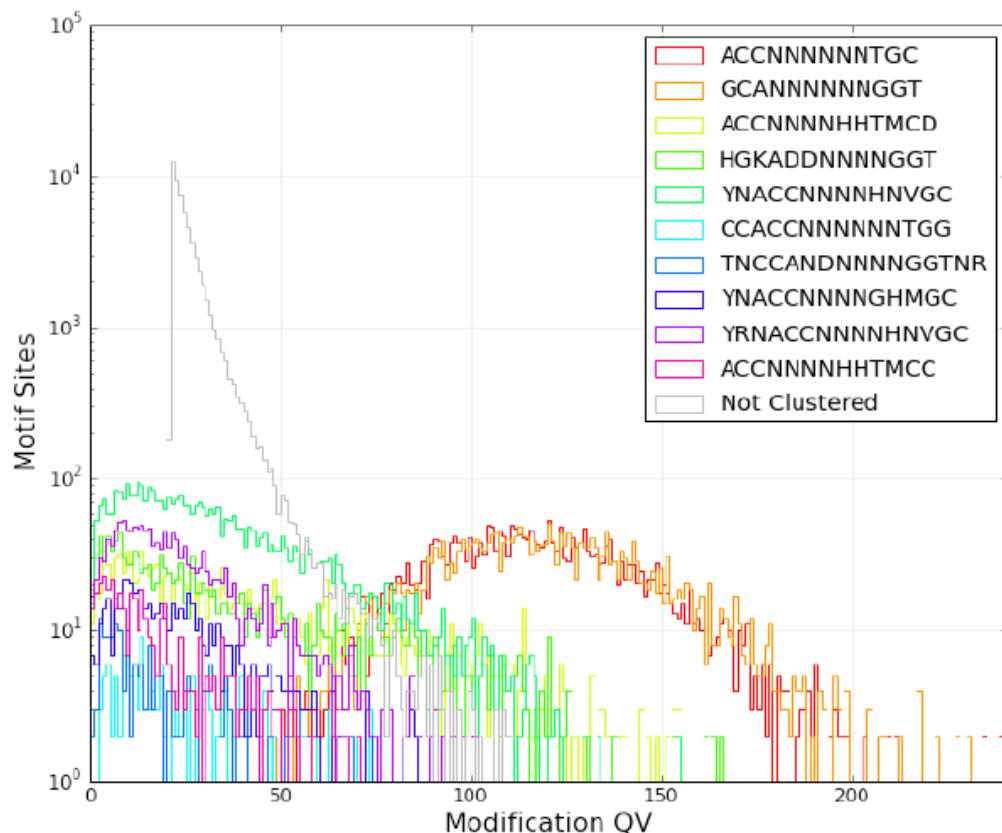
35 5- Nickel column eluate 6- eluate after conc. and PD10 desalting

36 7- final protein after concentration

37 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
ACCN>NNNNNTGC	1	m6A	100.00	2938	2938	118.8	81.7	GCANNNNNNNGT
GCANNNNNNNGGT	3	m6A	99.90	2935	2938	120.7	83.9	ACCN>NNNNNTGC
ACCN>NNNNHHTMCD	1	m6A	57.03	925	1622	71.1	83.7	HGKADDNNNNNGT
HGKADDNNNNNGT	4	m6A	48.83	792	1622	68.8	86.4	ACCN>NNNNHHTMCD
YNACCN>NNNNHVGC	3	m6A	46.49	1925	4141	57.7	84.6	
CCACCN>NNNNNTGG	3	m6A	39.15	74	189	55.7	85.9	
TNCCANDNNNNNGGTNR	5	m6A	31.60	73	231	53.5	82.5	
YNACCN>NNNNGHMGC	3	m6A	31.35	195	622	53.2	87.4	
YRNACCN>NNNNHVGC	4	m6A	28.65	465	1623	48.9	86.3	
ACCN>NNNNHHTMCC	1	m6A	27.58	131	475	58.1	84.3	
Not Clustered	0		0.09	8284	9100925	38.6	92.5	

1
2 S . SauNS ACC-6-TGC
34
5 Modification QV Histogram By Motif
6
7
8

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACONNNNNNTGC	1	m6A	100.00	2998	2998	118.8	81.7	GCANNNNNNGGT
GCANNNNNNGGT	3	m6A	99.90	2995	2998	120.7	83.9	ACONNNNNNTGC
ACONNNNHHTMCD	1	m6A	57.03	925	1622	71.1	83.7	HOKADDNNNNNGGT
HOKADDNNNNNGGT	4	m6A	48.83	792	1622	68.8	86.4	ACONNNNHHTMCD
YNACCNNNNNNVGC	3	m6A	46.48	1928	4141	57.7	84.6	
CCACCNNNNNNTGG	3	m6A	39.15	74	189	65.7	85.9	
TNCCANDNNNNNGTNR	5	m6A	31.60	73	231	53.5	82.5	
YNACCNNNNGHMGC	3	m6A	31.35	195	622	53.2	87.4	
YRNACCNNNHHNVGC	4	m6A	28.65	465	1623	48.9	86.3	
ACONNNNHHTMCC	1	m6A	27.88	131	475	58.1	84.3	
Not Clustered	0		0.09	8204	9100925	38.6	92.5	

1
2 **S . SauNU ACC-5-RTC**
3 MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
4 NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGGKGVMSPLYTVFKIQNIDLNFI
5 EKFSS KKYRFMALNGDGSARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEE
6 QKLELLQQQ KKGYMQKIFSQELRFKDENGEDYPDWEVTTIQNITKYTSSKKSSNQYADKDNSKGYPVY
7 DAVQEIG KDSNYDIEESYISILKDAGVGRLNLRPGKSSVIGTMGYIQSNNVDIEFLYYRMKV
8 VDFKKYIIGS TIPHLYFKDYSKETLYIPSSIQEQA
9 KIGMFISNL
10 DKL
11 LIEN
12 KNL
13 KLN
14 CLK
15 QGLL
16 QSMF
17 I
18 PGGSH
19 HHHHH



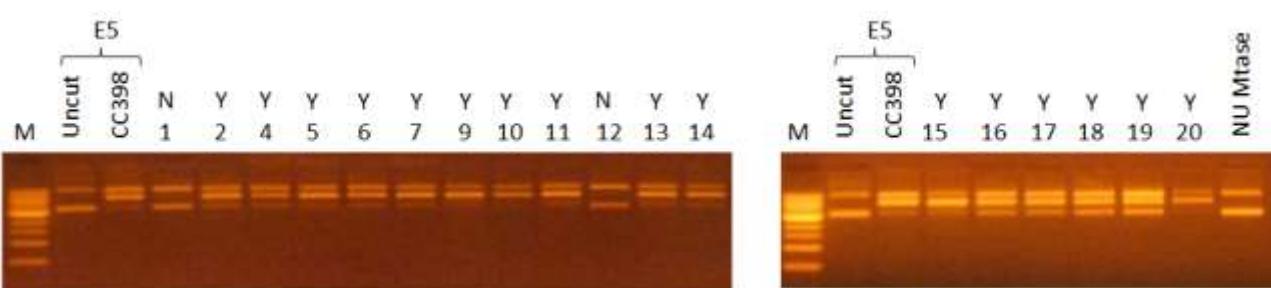
39 1- marker 2- soluble cell extract

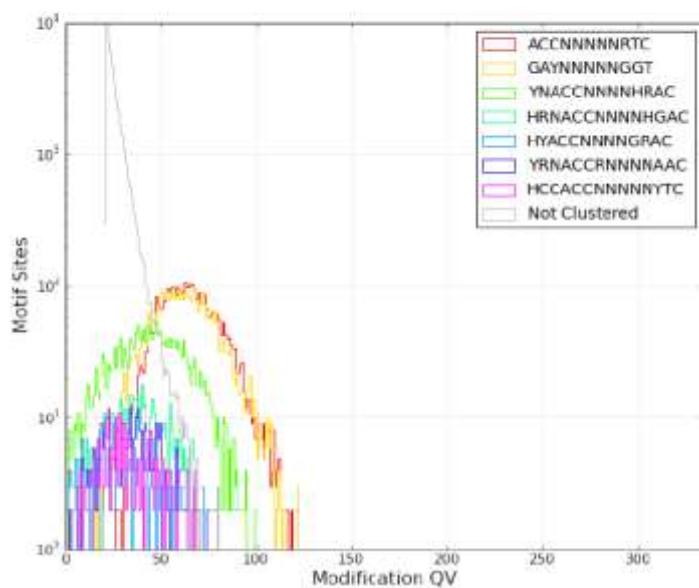
40 3- Nickel column flow through 4- Nickel column wash

41 5- Nickel column eluate 6- eluate after PD10 desalting

42 7- final protein after concentration

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

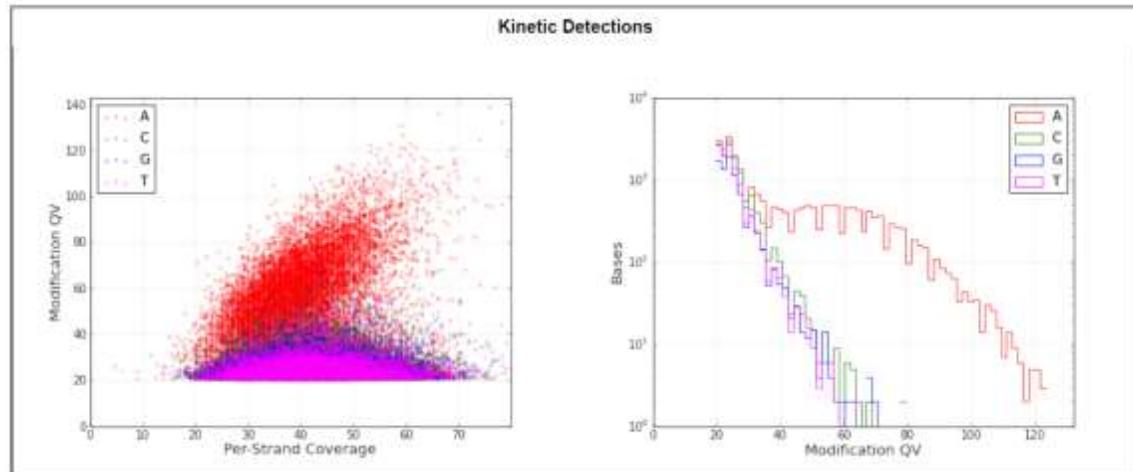


1
2 S . SauNU ACC-5-RTC3
4 Modification QV Histogram By Motif

Motifs

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

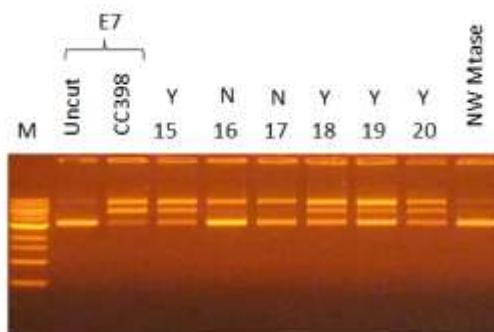
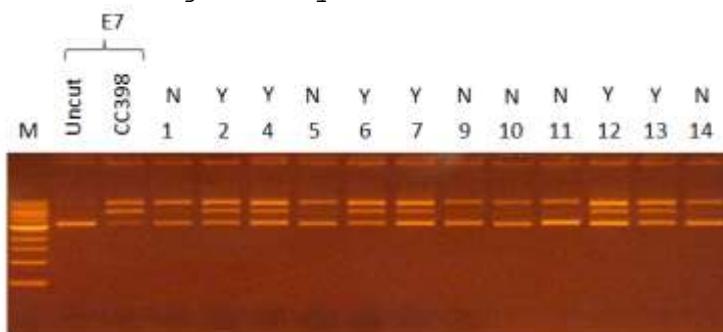
1
2 **S . SauNW ACC-6-TTYG**
3

4 MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
 5 NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGGKVMSPLYTVFKIQNIDLNFIEFYFKSS
 6 KWYRFMALNGDGSARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
 7 KKGYMQKIFSQELRFKDENGNDYPDWEEKQLGELSQIVRGASPRPIKDPKWFNKESDIGWLRISDV
 8 TNQNGKIYHLEQKLSIEGQEKTTRVLVTTHLLSIAASIGKPVMNFVKTGVDGFLIFLKPKFNLFF
 9 MYYWLEYFKDKWSKYQPGSQVNLNSEIVKSQTLNMPSNHEQEKGQFFNRNEKLIELQQEKIMYI
 10 KRCKQVLLQKMFIPGGSHHHHH
 11 1 2 3 4 5 6 7



- 36 1- marker 2- soluble cell extract
 37 3- Nickel column flow through 4- Nickel column wash
 38 5- Nickel column eluate 6- eluate after PD10 desalting
 39 7- final protein after concentration

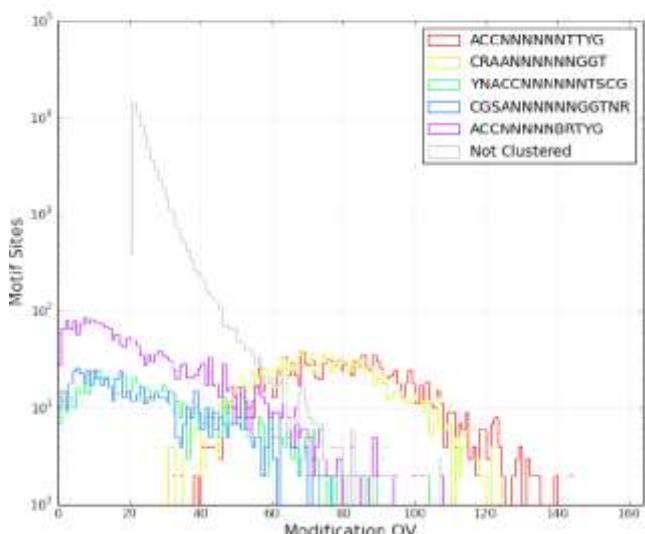
44 DNA cleavage assay.



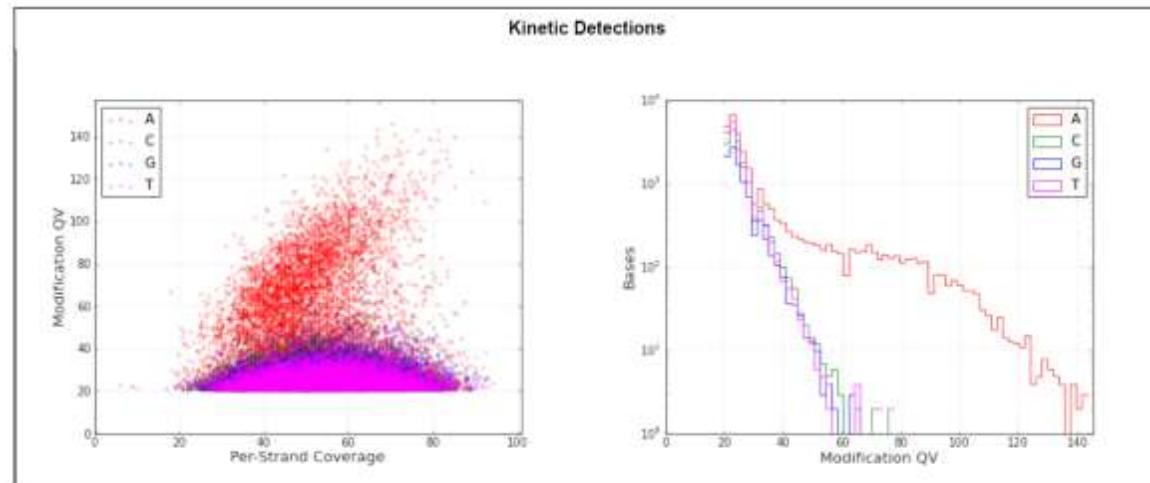
1
2 S . SauNW ACC-6-TTYG
3
4

Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACCN>NNNNNTTYG	1	m6A	99.86	1461	1463	80.7	49.6	CRAANNNNNNNGGT
CRAANNNNNNNGGT	4	m6A	99.59	1457	1463	74.7	48.6	ACCN>NNNNNTTYG
YNACCN>NNNNNTSCG	3	m6A	39.52	313	792	52.0	52.7	CGSANNNNNNNGGTNR
CGSANNNNNNNGGTNR	4	m6A	35.23	279	792	50.1	52.3	YNACCN>NNNNNTSCG
ACCN>NNNNBRTYG	1	m6A	28.16	680	2415	49.3	51.3	
Not Clustered	0		0.08	6917	9110401	37.6	55.8	

13
14 Modification QV Histogram By Motif
15

37 Kinetic Detections



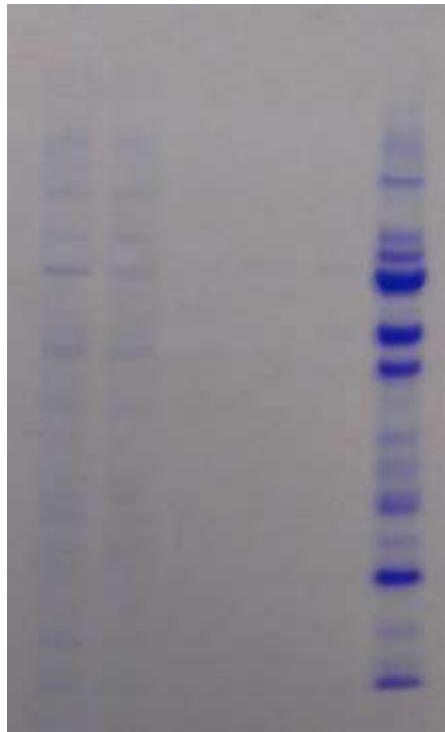
52 Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACCN>NNNNNTTYG	1	m6A	99.86	1461	1463	80.7	49.6	CRAANNNNNNNGGT
CRAANNNNNNNGGT	4	m6A	99.59	1457	1463	74.7	48.6	ACCN>NNNNNTTYG
YNACCN>NNNNNTSCG	3	m6A	39.52	313	792	52.0	52.7	CGSANNNNNNNGGTNR
CGSANNNNNNNGGTNR	4	m6A	35.23	279	792	50.1	52.3	YNACCN>NNNNNTSCG
ACCN>NNNNBRTYG	1	m6A	28.16	680	2415	49.3	51.3	
Not Clustered	0		0.08	6917	9110401	37.6	55.8	

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35**S . SauNY ACC-6-TAG**

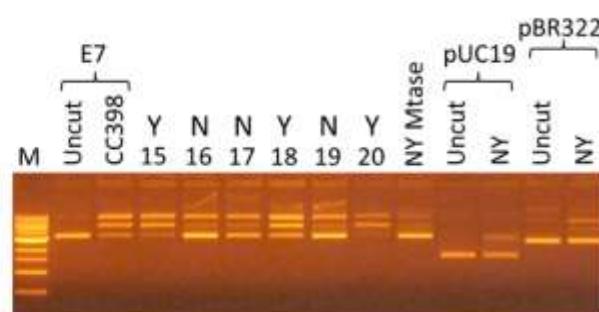
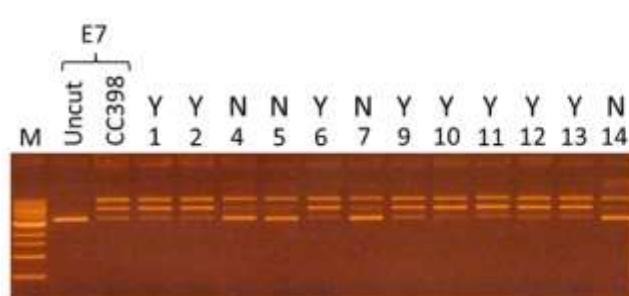
MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
 NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGGKGVMSPLYTVFKIQNIIDLNFIEFYFKSS
 KWYRFMALNGDGSARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
 KKGYMQKIFSQELRFKDENGNDYPDWEKKKLKEIACVYTGNTPSKKENIYWNKGEYVVWVTPTDINN
 SKNIYESENKLTQEGLYKKARQLPENTLLVTCIASIGKNAILRKQGSCNQQINAVVPFENINIDLY
 YISDSLSTFMKSIAGKTATQIVNKNTFENLEIYLAPFEEQNKIADLISSLEELIEKQASKLIKMK
 RKQGMLQIMFIPGGSHHHHHH

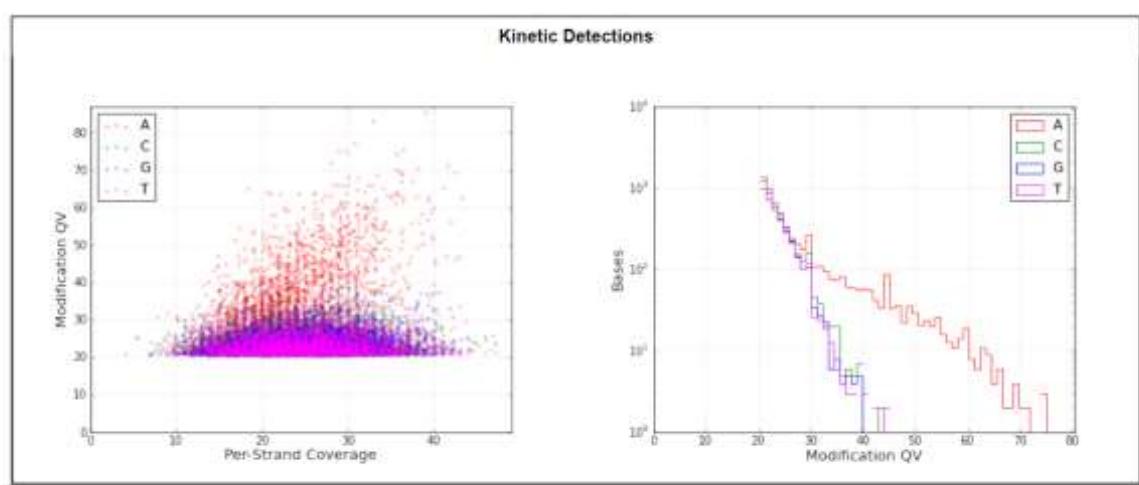
1 2 3 4 5 6



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 of eluate

DNA cleavage assay.

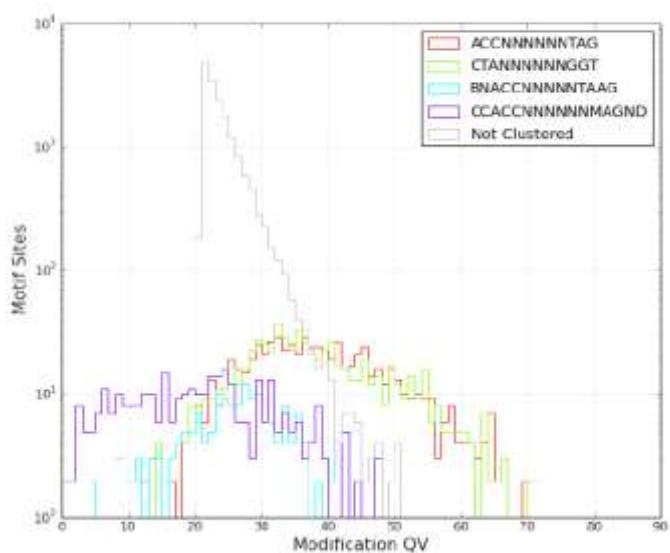


1
2 S . SauNY ACC-6-TAG
34 Kinetic Detections
519 Motifs
20

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACCN>NNNNNTAG	1	m6A	75.92	539	710	43.3	24.7	CTANNNNNNNGGT
CTANNNNNNNGGT	3	m6A	72.39	514	710	42.9	24.7	ACCN>NNNNNTAG
BNACCN>NNNNNTAAG	3	m6A	34.00	68	200	39.3	25.0	
CCACCN>NNNNNMAGND	3	m6A	23.68	85	359	38.3	26.1	
Not Clustered	0		0.01	622	9115347	34.4	27.6	

26 Motifs
27

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACCN>NNNNNTAG	1	m6A	75.92	539	710	43.3	24.7	CTANNNNNNNGGT
CTANNNNNNNGGT	3	m6A	72.39	514	710	42.9	24.7	ACCN>NNNNNTAG
BNACCN>NNNNNTAAG	3	m6A	34.00	68	200	39.3	25.0	
CCACCN>NNNNNMAGND	3	m6A	23.68	85	359	38.3	26.1	
Not Clustered	0		0.01	622	9115347	34.4	27.6	

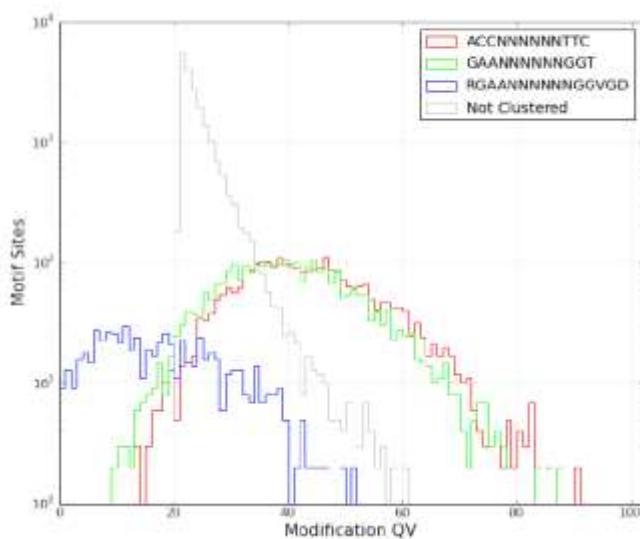
36 Modification QV Histogram By Motif
37

1
2 S . SauNa* ACC-6-TTC
3

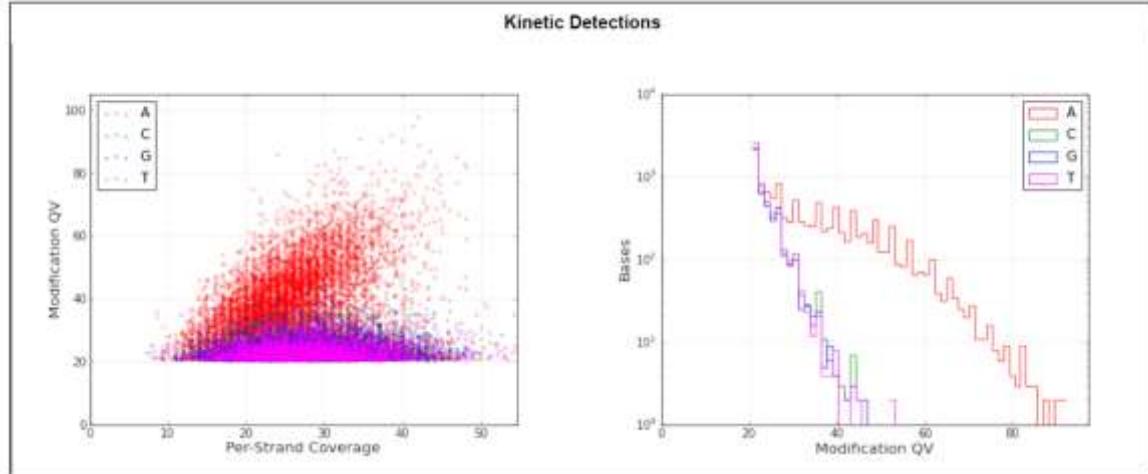
4 Motifs

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

11 MSNTQKKNVPELRFPGEFEWEEKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
12 NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGGKGVMSPLYTVFKIQNIIDLNFIEFYFKSS
13 KWYRFMALNGDSGARADRFSIKDRTFMEMPLHIPC MDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
14 KKGYMQKIFSQELRFKDENGNDYPEWENKRIEDIANVNKGFTPSTNNNEYWDNNDKNWLSIAGMNQ
15 KYLYKGNKGISKDAAKNYMKVKNDTLIMSFKLTIKGKLAIVKAPLYTNEAICHFIWKVNKINTEFIY
16 YYLNSLNISTFGVQAVKGVTNNDSINSIIVKLPNEEEQNIIAKFLLEVDTVNNQLVTKLKLQR
17 KKGLLQRMFVPGGSHHHHHH
18

19 Modification QV Histogram By Motif
20

Kinetic Detections

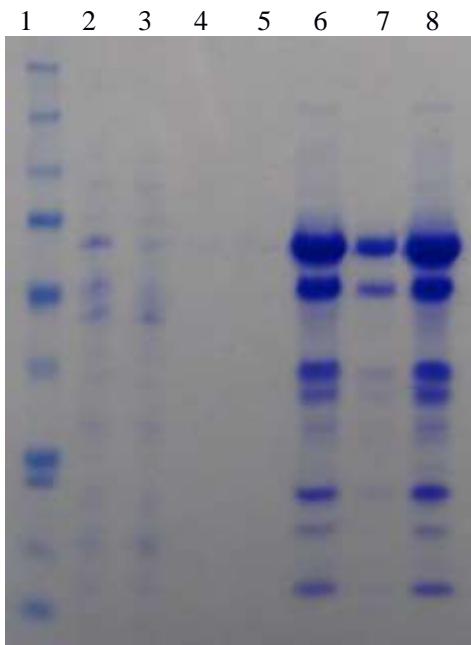


Motifs

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31**S.SauNc* ACC-6-RTC**

MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
 NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGKKGVMSPLYTVFKIQNIIDLNFIEFYFKSS
 KWYRFMALNGDGSARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
 KKGYLQKIFSQELRFKDENGNDYPEWRFARFKDFMYKPINIRPAINISKSELLTVKLHCKGIEKAN
 INRVLKLGATNYYKRFEQFIYGKQNFNGAFDIVPKFDGLYSSSDVPAFEINTEKIEPNYFISY
 ISRPSFYKSKEKYSTGTGSKRIHENTVLNFSLHLPCLNQLKIASFVCFLNRKIELLERKIYLIKK
 QKQALLQQMFIPGGSHHHHHH



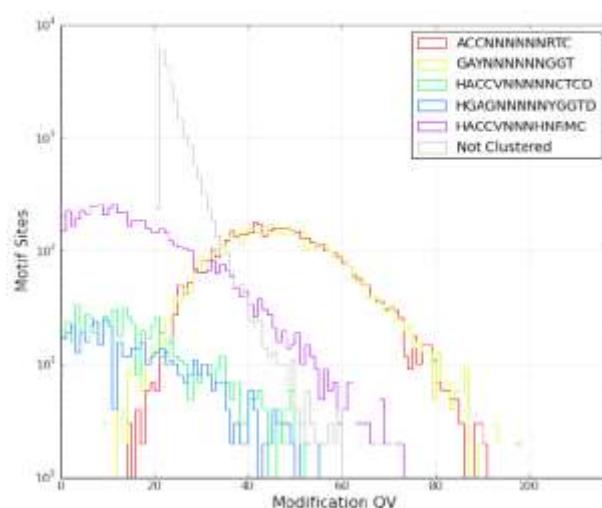
32 1- marker 2- soluble cell extract
 33 3- Nickel column flow through 4- Nickel column wash 1
 34 5- Nickel column wash 2 6- Nickel column eluate
 35 7- eluate after PD10 desalting
 36 8- final protein after concentration
 37 Although purified, this MTase was only assayed by SMRT.
 38
 39

40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60**Motifs**

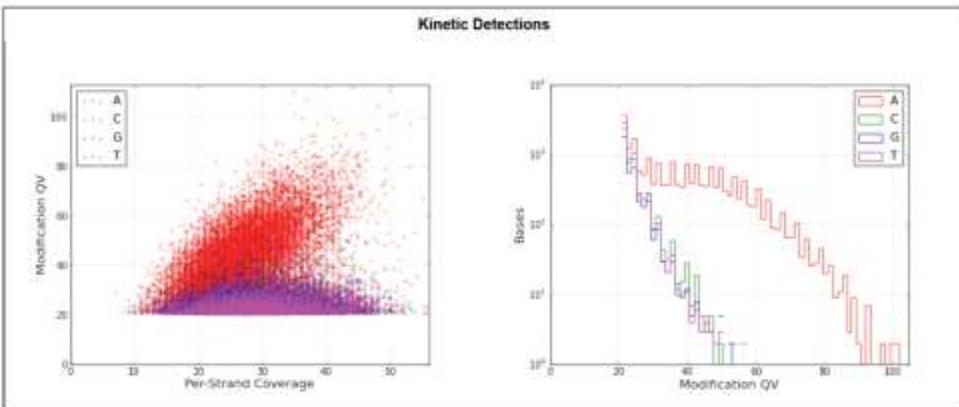
Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACCN>NNNNRTC	1	m6A	90.98	4680	5144	48.9	27.4	GAYNNNNNNNGGT
GAYNNNNNNNGGT	2	m6A	90.18	4639	5144	50.0	27.8	ACCN>NNNNRTC
HACCVNNNNNCTCD	2	m6A	16.64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

1
2 S . SauNc* ACC-6-RTC
3
4

Modification QV Histogram By Motif



Kinetic Detections

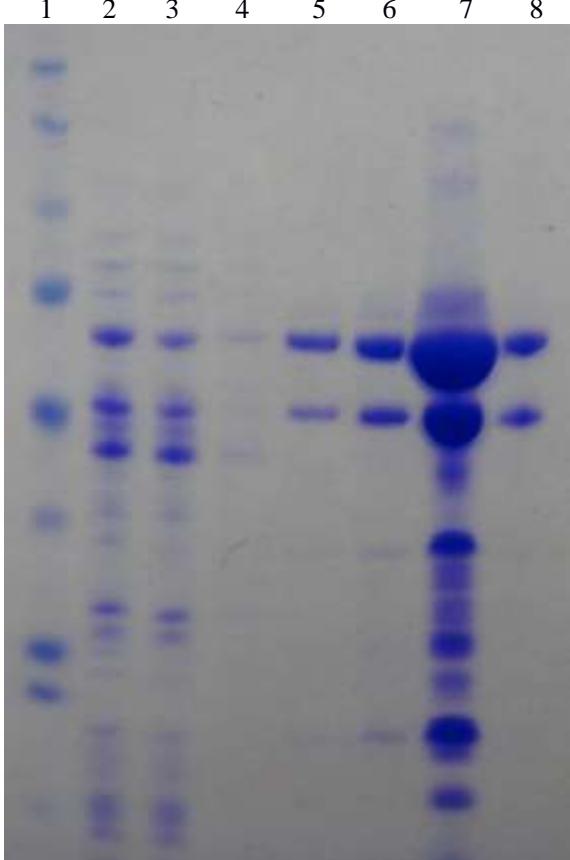


Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
ACNNNNNNRTC	1	m6A	90.98	4680	5144	48.9	27.4	GAYNNNNNNNGGT
GAYNNNNNNNGGT	2	m6A	90.18	4639	5144	50.0	27.8	ACNNNNNNRTC
HACCVNNNNNCTCD	2	m6A	16.64	117	703	40.7	30.2	
HGAGNNNNNNYGGTD	3	m6A	16.69	86	518	41.5	29.5	
HACCVNNNNHNFMC	2	m6A	14.94	935	6205	40.2	29.1	
Not Clustered	0		0.01	1163	909562	35.6	31.2	

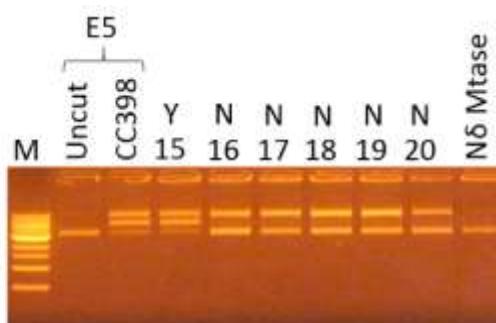
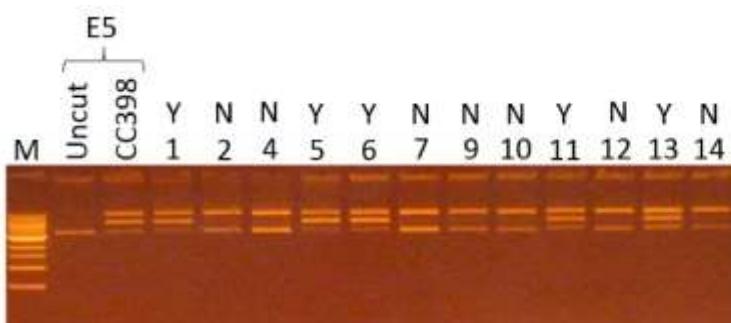
1 **S.SauNd* ACC-6-TTRG**

2
3 MSNTQKKNVPELRFPGFEGEWEEKKLGEFAGKVTQKNVDKKYIETLTNSAELGIISQKDYFDKEIS
4 NIDNIKKYYVVEENDFVYNPRMSNYAPFGPVNRNKLGGKVMSPLYTVFKIQNIIDLNFIEFYFKSS
5 KWYRFMALNGDSGARADRFSIKDRTFMEMPLHIPCMDEQIKIGQFFSKLDRQIELEEQKLELLQQQ
6 KKGYMQKIFSQELRFKDENGNDYPEWENVMLQKVLKDTEGIKRGPFGGALKKDIFVESGYAVYEQ
7 RNAIYDISNFRYYINENKYKEMQSFSVQPNDIIMSCSGTIGRLALIPHNYTKGIINQALIRFRTNH
8 KIRSEFFLIFMRSNQMQRKILEANPGSAITNLVPVKELKLIPFPLPVKFEQDKISQFIHIINRRIE
9 QSEKKIESLKNRKQGFLQKLFVPGGSHHHHHH
10
11 1 2 3 4 5 6 7 8



39 1- marker 2- soluble cell extract 3- Nickel column flow through
40 4- Nickel column wash 5- Nickel column eluate
41 6- eluate after conc. and PD10 desalting
42 7- final protein after concentration 8- NP purified protein marker
43
44

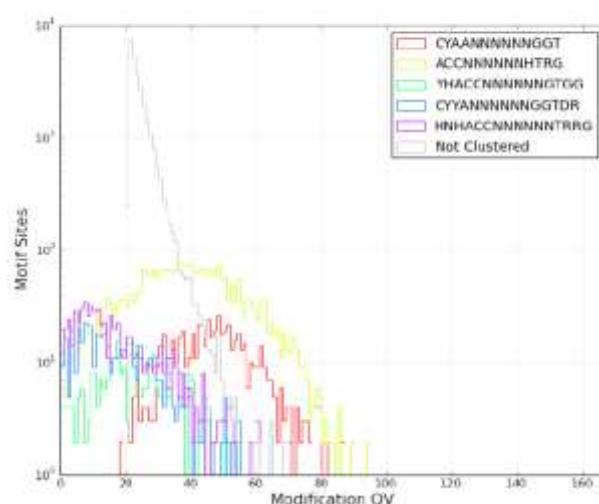
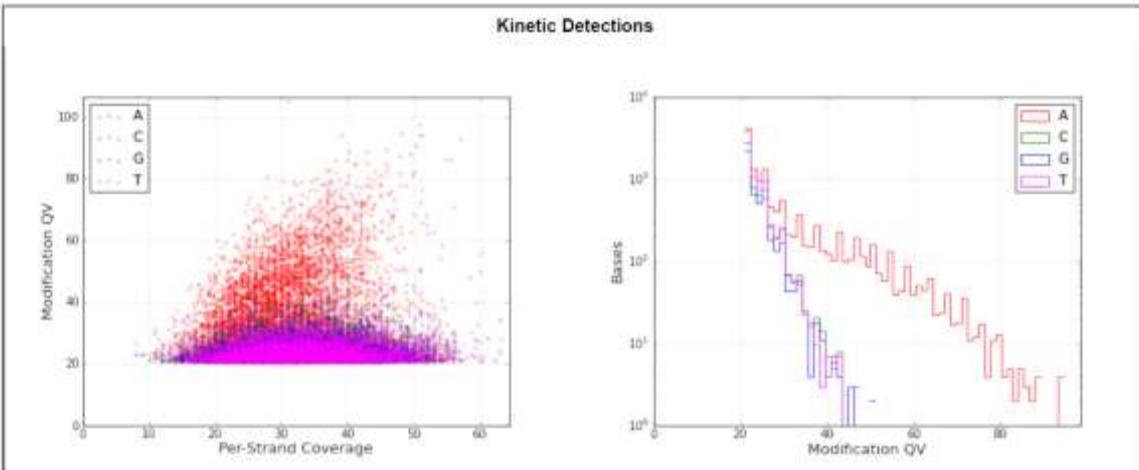
45 DNA cleavage assay.
46



50 Site determined to be ACC-6-TTRG or ACC-6-YTRG. Note that the
51 underlined site was determined by SMRT and is accepted since if Y
52 is a cytosine, then it can't be methylated.
53
54
55
56
57
58
59
60

1
2 S . SauNd* ACC-6-TTRG
34 Motifs
5

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
CYAANNNNNNNGGT	4	m6A	89.98	557	619	50.0	29.9	
ACCN>NNNNHTRG	1	m6A	67.06	2013	3002	47.5	31.0	
YHACCN>NNNNNGTGG	3	m6A	30.03	88	293	40.5	34.0	
CYYANNNNNNGTDR	4	m6A	20.24	102	504	42.0	32.9	
HNHACCN>NNNNNTRRG	4	m6A	17.94	127	708	41.5	32.6	
Not Clustered	0		0.02	1435	9112200	35.8	37.4	

13
14 Modification QV Histogram By Motif
1535 Kinetic Detections
3650 Motifs
51

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
CYAANNNNNNNGGT	4	m6A	89.98	557	619	50.0	29.9	
ACCN>NNNNHTRG	1	m6A	67.06	2013	3002	47.5	31.0	
YHACCN>NNNNNGTGG	3	m6A	30.03	88	293	40.5	34.0	
CYYANNNNNNGTDR	4	m6A	20.24	102	504	42.0	32.9	
HNHACCN>NNNNNTRRG	4	m6A	17.94	127	708	41.5	32.6	
Not Clustered	0		0.02	1435	9112200	35.8	37.4	

1
 2 **S . SauRE GARA-6-RTGA**
 3 MSNTQKKNVPELRFPGEFEGEWEEKKLGEVAKIYDGTHQTPKYTNEGIKFLSVENIKTLNSSKYISE
 4 EAFEKEFKIRPEFGDILMTRIGDIGTPNIVSSNEKFAYYVSLALLTKNLNSYFLKNLILSSION
 5 ELWRKTLHVAFPKKINKNEIGKIKINYPKQEQQQKIGQFFSKLDRQIEEQKLELLQQQKKGYMQ
 6 KIFSQELRFKDENGKDYPEWEETTIKEIAQINTGKDTKDAITNGSYDFYVRSPIVYKINTFSYEG
 7 EAILTVGDGVGVGKVFHVNNGKFDYHQRVYKISDFKNYYGLLFYYFSQNFLKETKKYSAKTSVDS
 8 VRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTQKQVIELLKQRKKSLLQKMFIPGGSHHHHH
 9

10 1 2 3 4 5 6



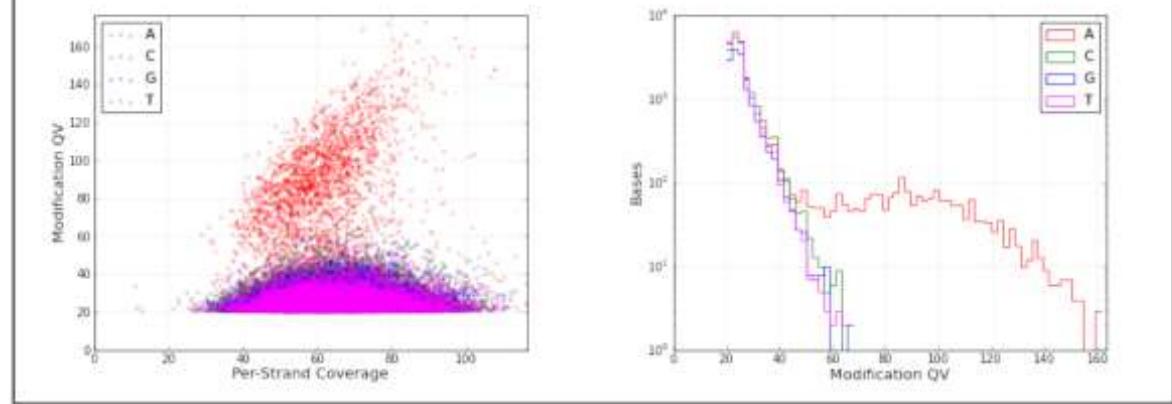
30 1- soluble cell extract, 2- Nickel column flow through, 3-Nickel column wash 1, 4- Nickel column
 31 wash 2, 5- Nickel column eluate, 6- Final protein after PD10 desalting and concentration

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

Motif	Motifs							
	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
TCAYNNNNNNNTYC	3	m6A	91.86	792	862	104.6	62.5	GARANNNNNNRTGA
GARANNNNNNRTGA	4	m6A	79.58	686	862	96.7	61.7	TCAYNNNNNNNTYC
CCACDNNNNNNTYC	3	m6A	15.87	30	189	90.4	65.5	
TCAGNNNTNNNTNCNB	3	m6A	15.03	29	193	91.0	67.7	
Not Clustered	0		0.00	129	9115220	85.4	67.9	

1
2 S . SauRE GARA-6-RTGA
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

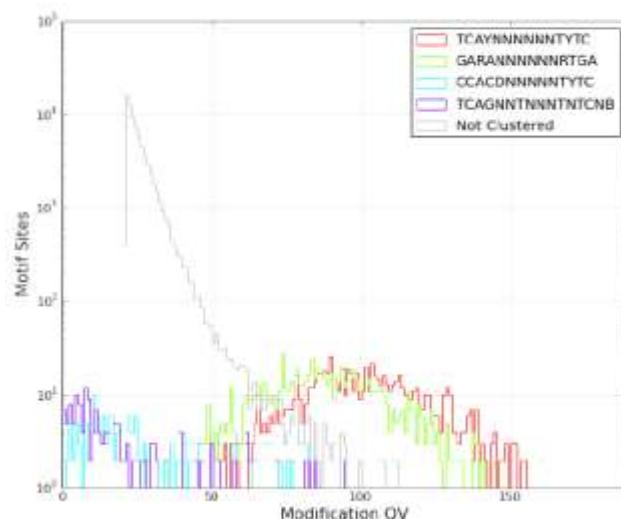
Kinetic Detections



Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
TCAYNNNNNNNTYTC	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	
TCAGNNNTNNNTNTCNB	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



1
 2 **S. SauTE CAAG-5-RTGA**
 3 MSNTQKKNVPELRFPGEFEWEEKELGEIFQIISGSTPLKSNEFYENGNIHWVKTTDLNNSKVTH
 4 SKEKITEYAMKSLKLKVPKNSVLIAMYGGFNQIGRTGLLKIDATINQAISALLMNHETNPEFIQA
 5 FLNYQVKGWKRYAASSRKDPNITKKDIEQFKVPYVSINEQQKIGEFFSKIDHQIELEEQKLELLQQ
 6 QKKGYMQKIFSQELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKDAITNGSYDFYVRSPIVYKI
 7 NTFSYEGEAILTVGDGVGVGVFHYVNGKFDYHQRVYKISDFKNYYGLLFYYFSQNFLKETKKYS
 8 AKTSVDSVRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTQKQVIELLKQRKKSLLQKMFIPGG
 9 SHHHHHH
 10

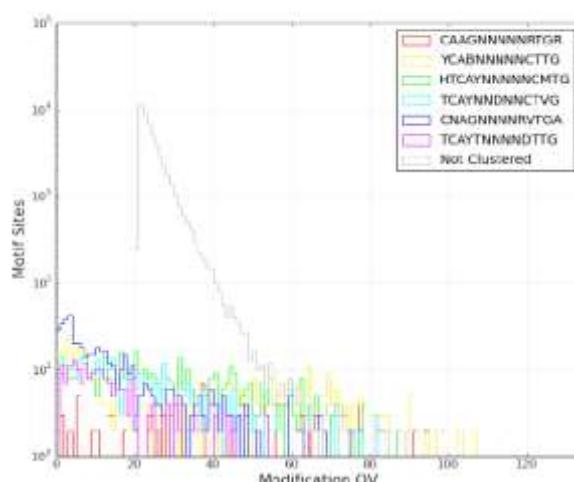


29 1- soluble cell extract 2- Nickel column flow through
 30 3-Nickel column wash 1 4- Nickel column wash 2
 31 5- Nickel column eluate 6- eluate after concentrating and PD10 step
 32 7- Final concentrated protein

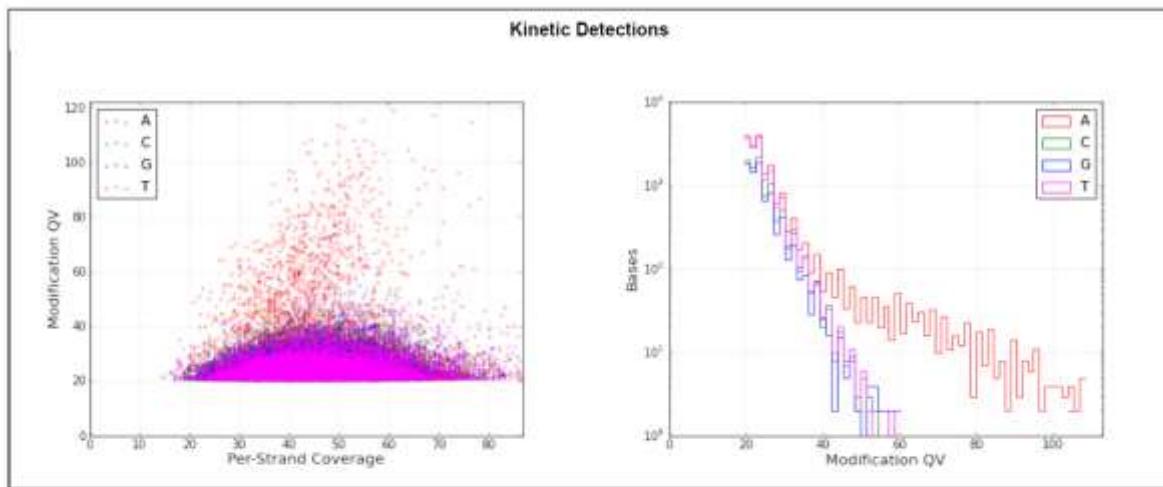
33 This MTase did not purify well and was only analysed by SMRT
 34 sequencing. The degeneracy in the target determined by SMRT
 35 sequencing can be removed using results from other systems.
 36

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	
TCAYNNNNNCTVG	3	m6A	31.07	119	383	51.0	44.3	
CNAGNNNNNRVTGA	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	

1
2 S . SauTE CAAG-5-RTGA3
4 Modification QV Histogram By Motif

23 Kinetic Detections



38 Motifs

39
40
41
42
43
44
45
46
47

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
CAAGNNNNNRG	3	m6A	80.15	214	287	63.8	42.7	
YCABNNNNCTG	3	m6A	55.08	260	472	65.1	43.1	
HTCAVNNTTCMTG	4	m6A	47.41	238	502	51.7	44.4	
TCAYNNNDNCTVG	3	m6A	31.07	119	383	51.0	44.3	
CNAGNNNNRVTGA	3	m6A	28.62	170	584	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	

48
49
50
51
52
53
54
55
56
57
58
59
60

1
 2 **S . SauVE CNGA-6-RTGA**
 3 MSNTQKKNVPELRFPGFEGEWEEKELRELNPDKYSYTGGPGSDLKKSYTTDGIQIIQLQNIG
 4 DGYFYNSNKVFTSNEKAEVLKSCNVFPGDIVIAKMAPIARAAIVPDNNIGKYLMASDGIRLSVDT
 5 VHFNTKFVLECINRKSFRKKVEDNSSGSTRMRIGLSTLGSLLKTTTLKEQQKIGQFFSKLDRQIE
 6 LEEQKLELLQQQKKGYMQKIFSQELRFKDENGKDYPEWEETTIKEIAQINTGKDKTDKDAITNGSYD
 7 FYVRSPIVYKINTFSYEGEAILTVDGVGVGVFHYVNGKFDYHQRVYKISDFKNYYGLLFYYFS
 8 QNFLKETKKYSAKTSVDSVRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTQKQVIELLKQRKK
 9 SLLQKMFIPGGSHHHHHH
 10
 11 1 2 3 4 5 6 7



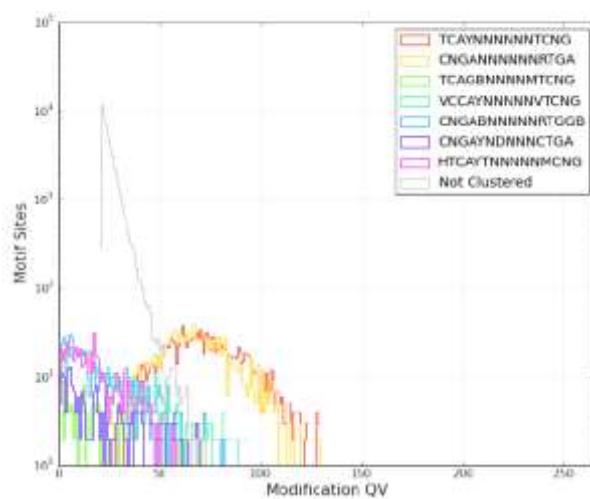
28 1- marker, 2- soluble cell extract, 3- Nickel column flow through, 4- Nickel column wash, 5- Nickel
 29 column eluate, 6- eluate after conc. and PD10 desalting, 7- Final protein after concentration
 30 Although the MTase was purified, it was only analysed via SMRT
 31 sequencing .
 32

Motifs

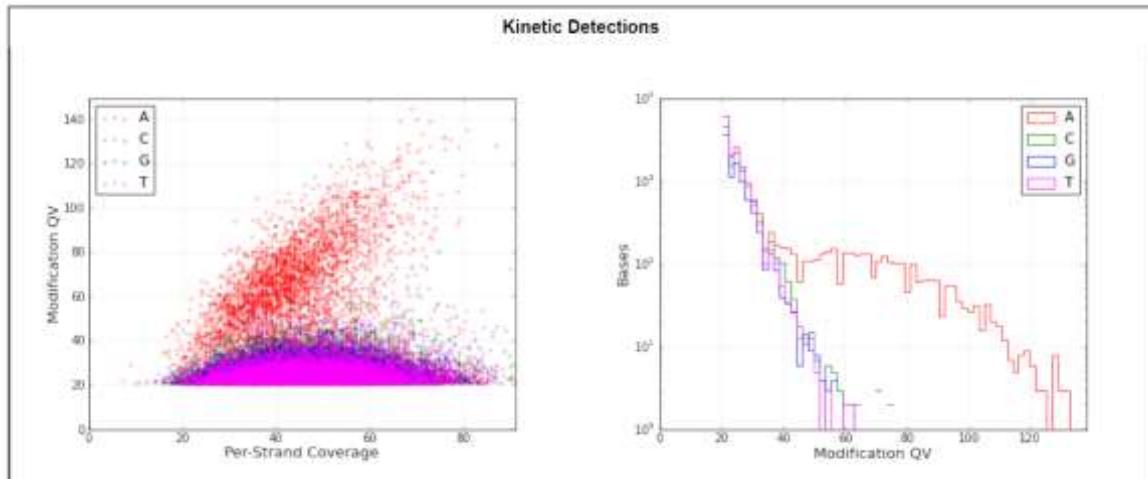
Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
TCAYNNNNNNTCNG	3	m6A	99.34	1364	1363	73.0	43.2	CNGANNNNNNRTGA
CNGANNNNNNRTGA	4	m6A	98.09	1337	1363	71.4	43.4	TCAYNNNNNNTCNG
TCAGBNNNNNMTCNG	3	m6A	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
CNGAYNDNNNNCTGA	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	

1
2 S . SauVE CNGA-6-RTGA
3
4

Modification QV Histogram By Motif



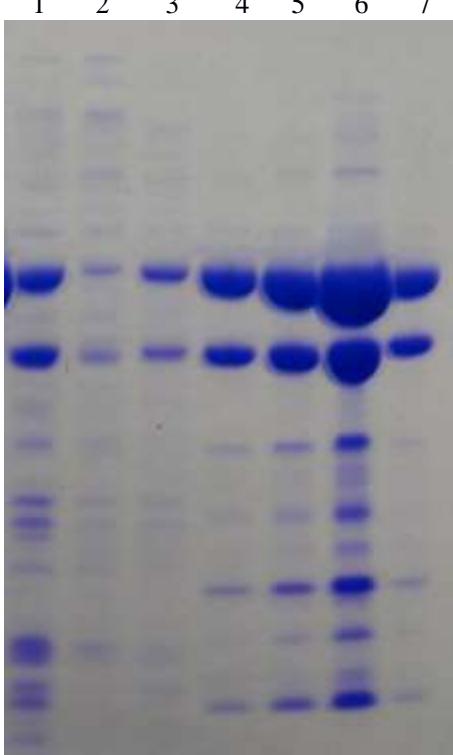
Kinetic Detections



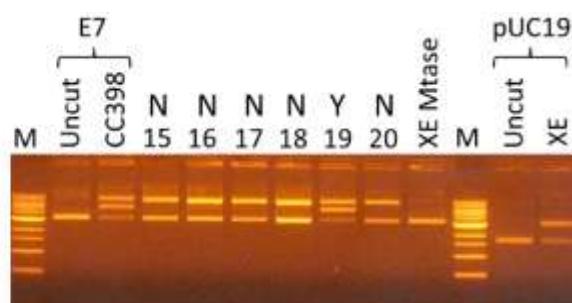
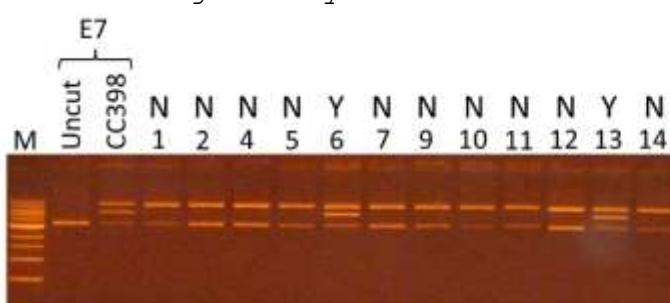
Motifs

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
TCAYNNNNNTCNG	3	m6A	99.34	1354	1363	73.0	43.2	CNGANNNNNRTGA
CNGANNNNNRTGA	4	m6A	98.09	1337	1363	71.4	43.4	TCAYNNNNNTCNG
TCAGBNNNNMTCNG	3	m6A	41.58	79	190	53.7	43.2	
VCCAYNNNNVTTCNG	4	m6A	33.12	211	637	50.8	44.7	CNGABNNNNRTGGB
CNGABNNNNRTGGB	4	m6A	29.98	191	637	50.2	45.8	VCCAYNNNNVTTCNG
CNGAYNDNNNCTGA	4	m6A	31.90	74	232	53.0	44.9	
HTCAYTNNNNMCNG	4	m6A	23.31	131	562	44.6	45.5	
Not Clustered	0		0.05	4345	9112342	36.9	48.9	

1
2 **S . SauXE TCTA-6-RTGA**
3 MSNTQKKNVPELRFPGEFEGEWEEKQFADFTKINQGLQIAINERKTEYSPELYFYITNEFLRPNSQT
4 KYFIENPPQSVIANKEDEILMTRGNTGVVTNVFGAFHNNFFKIKFDKNLYDRLFLVLEVLNSSKIQ
5 NKILSLAGSSTIPDLNHSDFYSISSSYPLLREQQKIGKFFSKLDRQIEEQKLELLQQQKKGYMQ
6 KIFSQELRFKDENGKDYPEWEETTIKEIAQINTGKDTKDAITNGSYDFYVRSPIVYKINTFSYEG
7 EAILTVGDGVGVGVKFHVNGKFDYHQRVYKISDFKNYYGLLFYYFSQNFLKETKKYSAKTSVDS
8 VRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTQKQVIELLKQRKKSLLQKMFIPGGSHHHHHH
9
10 1 2 3 4 5 6 7



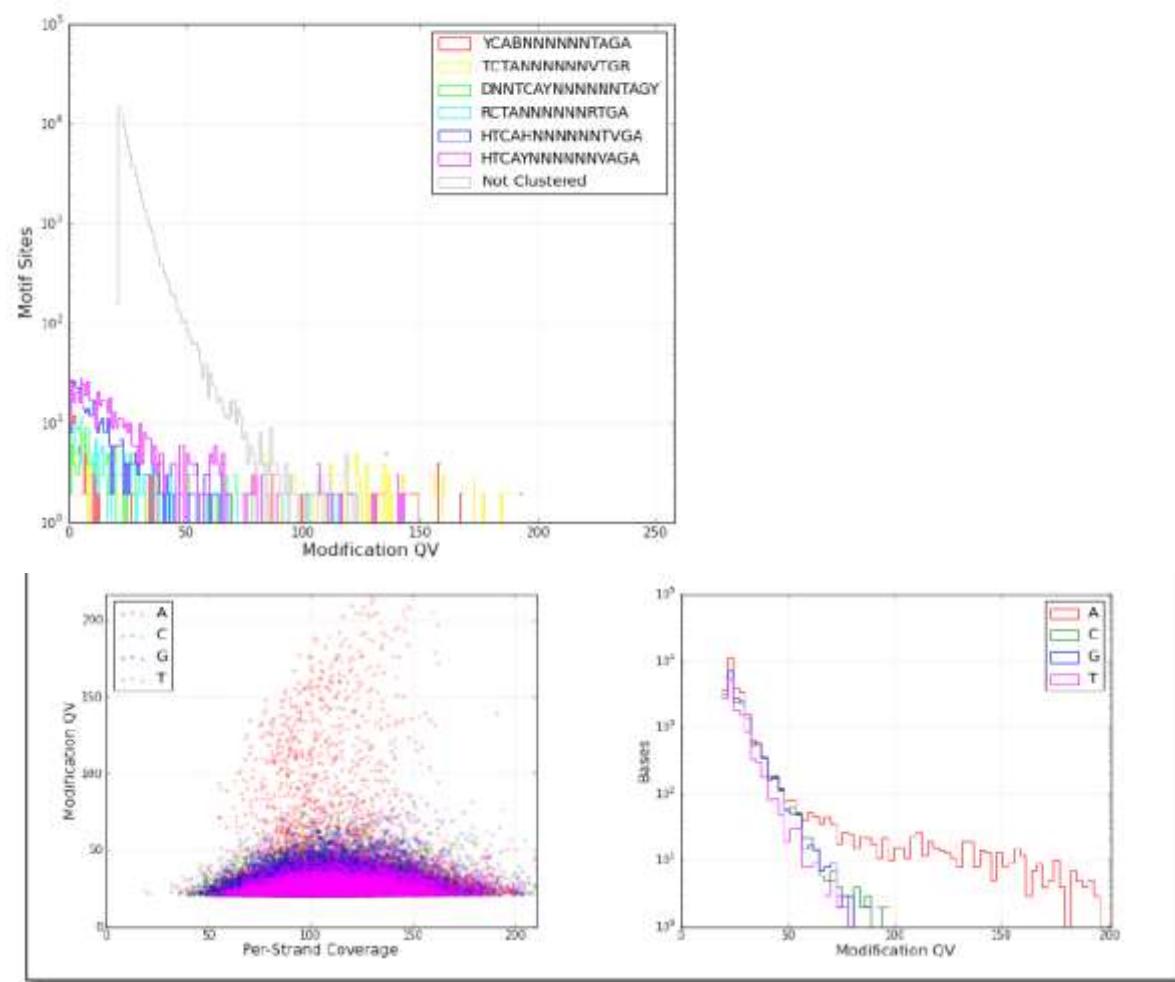
35 1- soluble cell extract 2- Nickel column flow through
36 3- Nickel column wash 4- Nickel column eluate
37 5- eluate after conc. and PD10 desalting
38 6- final protein after concentration
39 7- CC398-1 purified protein marker
40 DNA cleavage assay.



1
2 **S.SauXE TCTA-6-RTGA**
34 The degeneracy in the target determined by SMRT sequencing can be
5 resolved by reference to targets from other systems.
6

Motifs

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

16
17 Modification QV Histogram By Motif
18
19

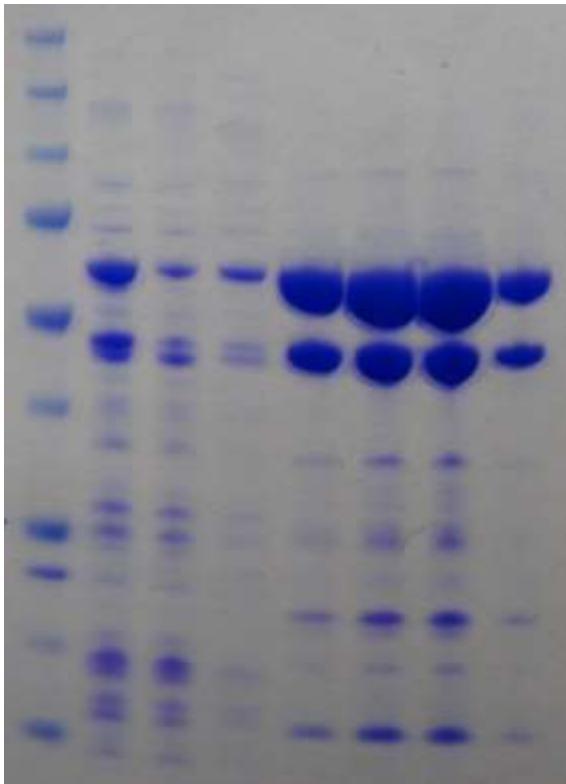
Motifs

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

1
2 **S.SauZE GAC-5-RTGA**
3

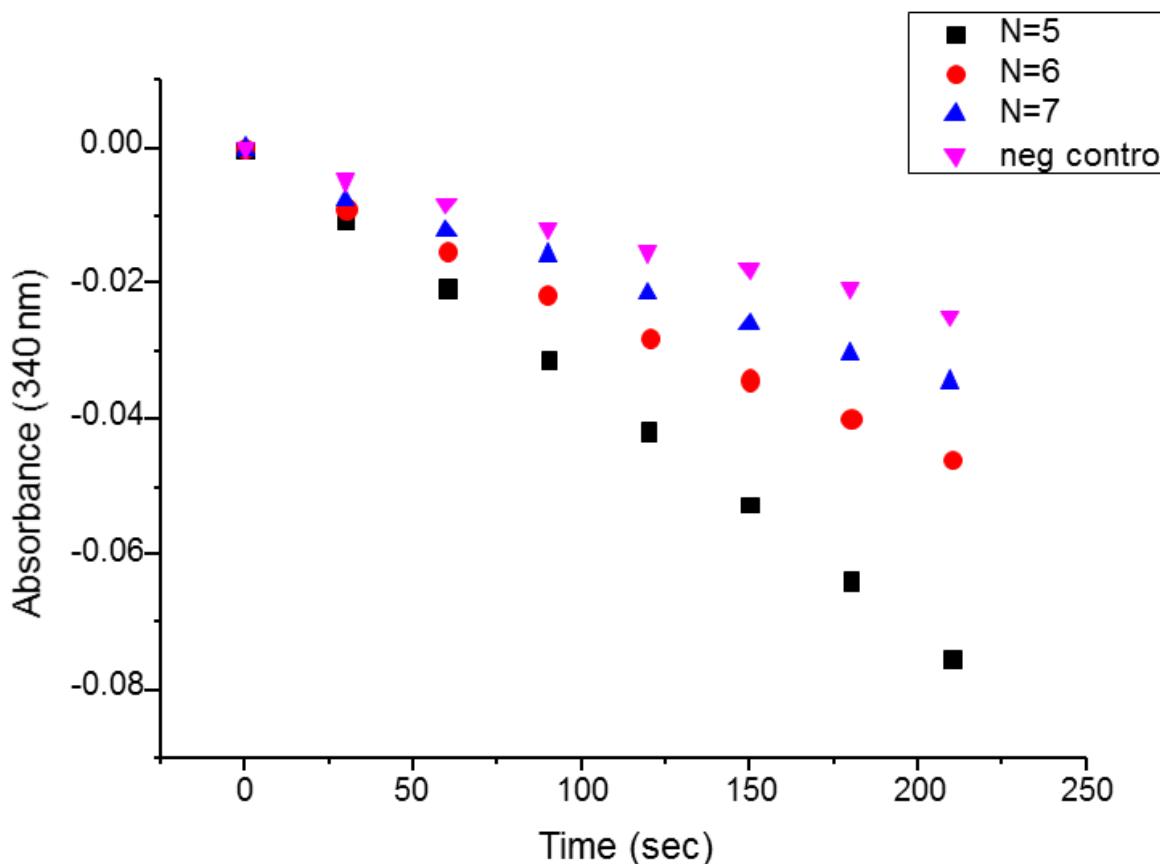
4 MSNTQKKNVPELRFPGEFEYEYSLDIFGNLATNKSEKFPQNENASIDIELDCIEQNTGRLIKIYNS
 5 KEFSSQKNFNPQNVLYGKLRYLNKYYFTKKSGVCSSEIWVLKSTKEDKLLNLFLYYFIQTKRYS
 6 DVASKSAGSKMPRADWGLIENIRVYFPELCEQQKIGQFFSKLDRQIELEEQKLELLQQQKKGYMQK
 7 IFSQELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKDAITNGSYDFYVRSPIVYKINTFSYEGER
 8 AILTVGDGVGVGKVFHYVNGKFDYHQRVYKISDFKNYYGLLLFSQNFKETKKYSAKTSVDSV
 9 RKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTQKQVIELLQRMFIPGGSHHHHH*

10 1 2 3 4 5 6 7 8

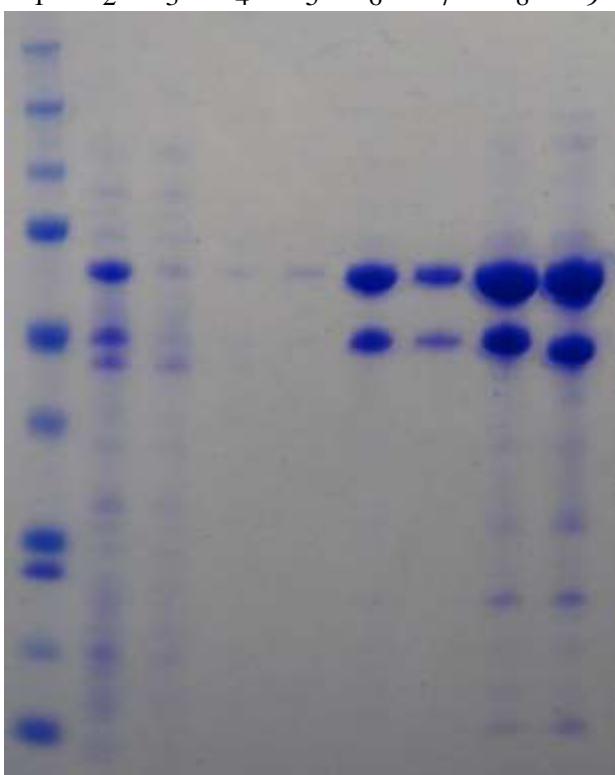
36 1- marker 2- soluble cell extract 3- Nickel column flow through
 37 4- Nickel column wash 5- Nickel column eluate 6- eluate after conc. step and PD10 desalting
 38 7- final concentrated protein 8- CC398-1 purified protein marker
 3940
41 DNA cleavage assay showed cutting of all plasmids so the ATPase
 42 assay was used given that we knew the individual TRD specificities.
 43

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

1
2 S . SauZE GAC-5-RTGA
3 N=5 gives the clearest signal.
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



1
 2 **S . SauZS GAC-6-TGC**
 3 MSNTQKKNVPELRFPGEFEYEYSLDIFGNLATNKSEKFPQNENASIDIELDCIEQNTGRLIKIYN
 4 KEFSSQKNKFNPQNVLYGKLRYLNKYYFTKKSGVCSSEIWVLKSTKEDKLLNLFLYYFIQTKRYS
 5 DVASKSAGSKMPRADWGLIENIRVYFPELCEQQKIGQFFSKLDRQIELEEQKLELLQQQKKGYMQK
 6 IFSQELRFKDENGNDYPDWTNERLGEVTTVTMGQSPKSVNYTDNSNDTVLIQGNADIENGLINPRI
 7 YTREVTKLIQKDEIILTVRAPVGKLAMAQINACIGRGVCSIKGDKFLYYFLEWFATQNKWIRFSQG
 8 STFESISGNDIRNIHIKIPVEDERTKIIKLLNSLDVLNSKTDLKIQNLKQRKQSLLQKIFVPGGSH
 9 HHHHH*
 10
 11 1 2 3 4 5 6 7 8 9

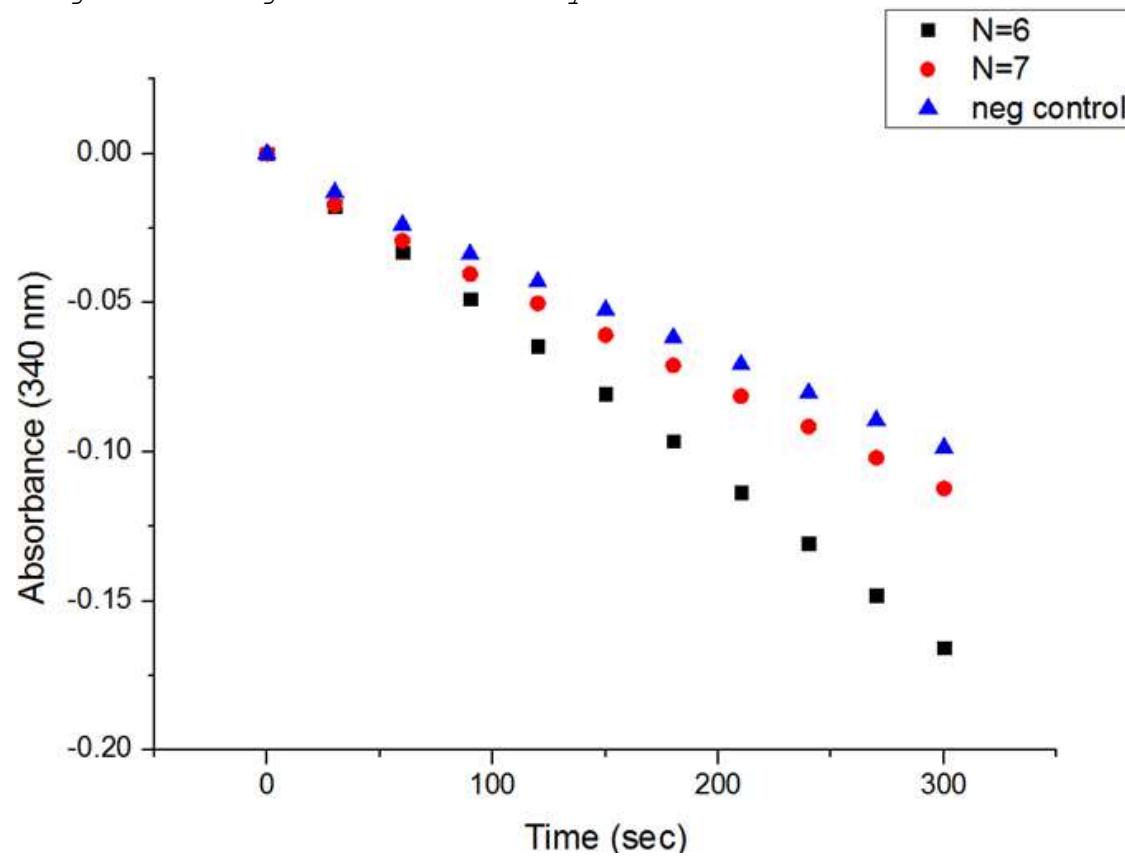


37 1- marker 2- soluble cell extract 3- Nickel column flow through
 38 4- Nickel column wash 1 5- Nickel column wash 2 6- Nickel column eluate
 39 7- eluate after PD10 desalting 8- final protein after concentration
 40 9- NQ purified protein marker

41 The DNA cleavage assay showed cutting of all plasmids so the
 42 ATPase assay was used since we knew the TRD specificities.
 43
 44
 45
 46

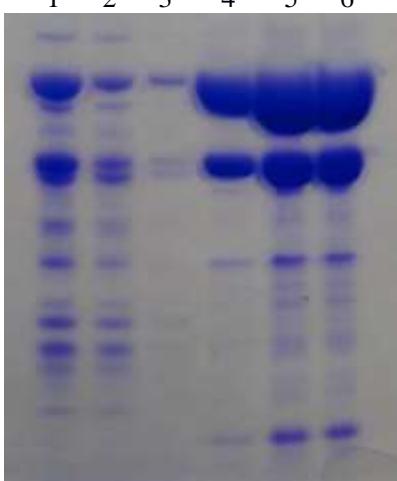
Oligonucleotide name	DNA sequence (5' to 3')
ZS6for	AGATGATGGAATCAATGCGACTTCCATTGCGCCCTATACTGATATAA
ZS6rev	TTATATCGTATAGGGCGCAATGGAAGTCGCATTGATTCCATCATCT
ZS7for	AGATGATGGAATCAATGCGACTTCACATTGCGCCCTATACTGATATAA
ZS7rev	TTATATCGTATAGGGCGCAATGTGAAGTCGCATTGATTCCATCATCT

1
2 S . SauZS GAC-6-TGC
3 N=6 gives the greatest activity.
4
5
6
7
8



1 **S.Saub*E GGHA-6-RTGA**

2
 3 MSNTQKKNVPELRFPGEFEGEWEEKKLEDTLEFIKGTHGTHENVNNGPWLLSAKNIKNNKIISSD
 4 DRKISESDYKKIYKNYKLEKGDLLLTIVGTIGRAAIVKNPNNIAFQRSVAILKTKATYDVGIFQL
 5 FQTKYFKNLLLRKQVSAQPGLYLGDIRKIKISITNIEEQRKIGIFFSKLDRQIELEEQKLELLQ
 6 QQKKGYMQKIFSQELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKDAITNGSYDFYVRSPIVYK
 7 INTFSYEGEAILTVGDGVGVGVFHYVNGKFDYHQRVYKISDFKNYYGLLFYYFSQNFLKETKKY
 8 SAKTSVDSVRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTKIQKQVIELLKQRKKSLLQKMFIPG
 9 GSHHHHHH
 10
 11 1 2 3 4 5 6



27 1- soluble cell extract, 2- Nickel column flow through, 3- Nickel column wash, 4- Nickel column
 28 5- eluate after conc. step and PD10 desalting, 6- Final concentrated protein
 29
 30
 31
 32
 33
 34

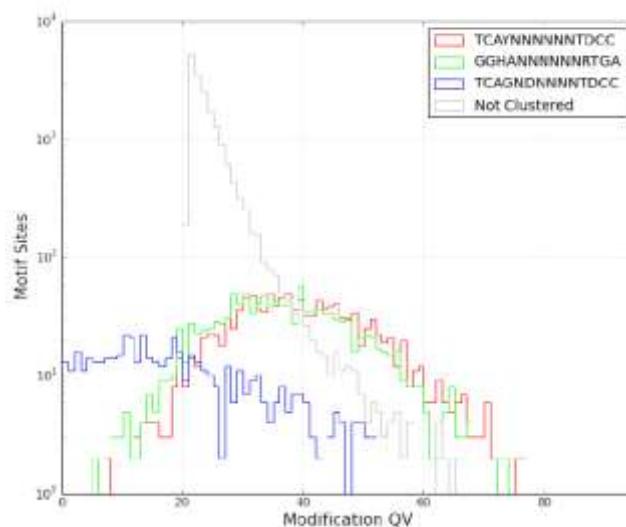
35 **Although this MTase was purified, it was only used in the SMRT**
 36 **sequencing assay.**

39 **Motifs**

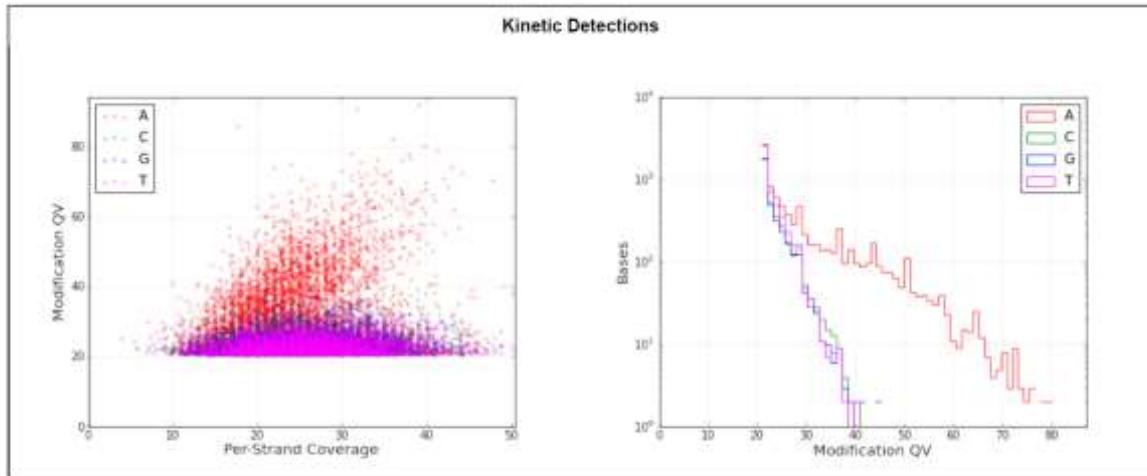
Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
TCAYNNNNNNNTDCC	3	m6A	77.39	989	1278	44.6	24.6	GGHANNNNNNRRTGA
GGHANNNNNNRRTGA	4	m6A	67.68	865	1278	43.2	24.7	TCAYNNNNNNNTDCC
TCAGNDNNNNNTDCC	3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

1
2 S . Saub*E GGHA-6-RTGA
3
4

5 Modification QV Histogram By Motif
6
7



28 Kinetic Detections
29
30



43 Motifs
44

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
TCAYNNNNNTDCC	3	m6A	77.39	989	1278	44.6	24.6	GGHANNNNNRTGA
GGHANNNNNRTGA	4	m6A	67.68	885	1278	43.2	24.7	TCAYNNNNNTDCC
TCAGNDNNNNNTDCC	3	m6A	21.57	110	510	41.0	25.2	
Not Clustered	0		0.01	928	9114200	36.7	27.3	

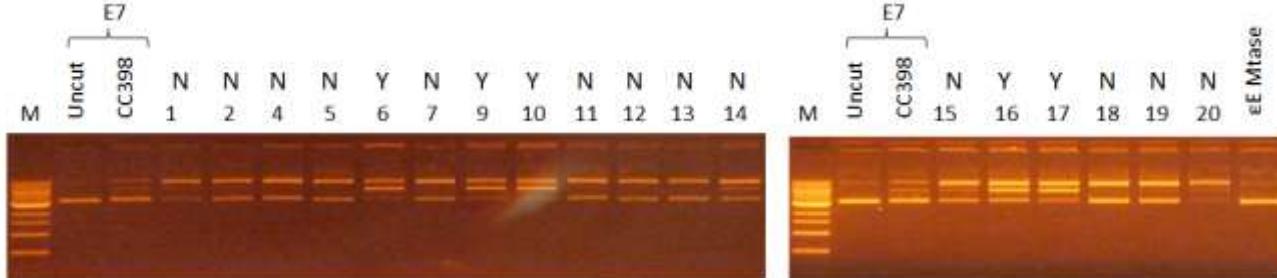
1
2 **S.Sau~~e~~^{*E} GAG-6-RTGA**

3 MSNTQKKNVPELRFPGEFEWEEKSISSFLKESKIKGSNGSHAKKLTVKLGKGVVPKKETFKGSD
 4 NTQYYKRKAGQLMYGKLDLNCAGIVPDSLNNYESTIDSPSFDFINGDSKFLLERIKLKSFYKKF
 5 GDIANGSRKAKRINQDTFLSLPVFAPKYDEQLRIGEFFSKLDRQIELEEQKLELLQQQKKGYMQKI
 6 FSQELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKDAITNGSYDFYVRSPIVYKINTFSYESEA
 7 ILTVGDGVGVGVFHYVNGKFDYHQRVYKISDFKNYYGLLFYYFSQNFKETKKYSAKTSVDSVR
 8 KDMIANMKVPRPIYIEQKKIGQFIKRVDNKTQKQVIELLKQRKKSLLQKMFIPGGSHHHHHH
 9
10 1 2 3 4 5 6



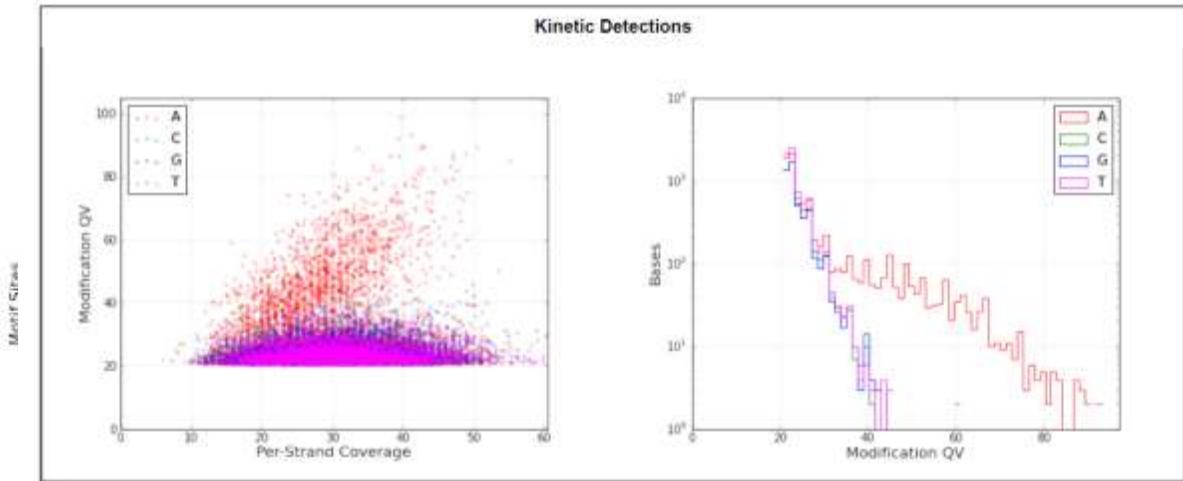
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28 1- soluble cell extract, 2- Nickel column flow through, 3- Nickel column eluate, 4- eluate after
 29 PD10 desalting, 5- Final concentrated protein, 6- RE purified protein as marker
 30
31

32 DNA cleavage assay.



1
2 S . Saue*E GAG-6-RTGA
34 Motifs
5

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	
DNNAGNDNNNNGAGA	5	m6A	18.56	36	194	39.8	32.7	
Not Clustered	0		0.01	914	9115229	34.6	36.3	

14 Kinetic Detections
1530 Motifs
31

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	
DNNAGNDNNNNGAGA	5	m6A	18.56	36	194	39.8	32.7	
Not Clustered	0		0.01	914	9115229	34.6	36.3	

1
2 **SUPPLEMENTARY INFORMATION FOR TABLE 4.**
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

MSNTQKKNVPELRFPGEFEWEEKKLGDLTTKIGSGKTPKGSENYTNKGIPFLRSQNIRNGKLNLDLVYISKDIDDEM
 KNSRTYYGDVLLNITGASIGRTAINSIVEIHANLNQHVCIIIRLKKEYYYNFFGQYLLSRKGKRKIFLAQSGGSREGLNFK
 EIANLKIPTPTIFEEQQKIGEFISKLDRQIELEEQKLELLQQQKKGYLQKIFSQELRFKDENGNDYPEWRFARFKDFMYK
 PINIRPAINISKSELLTVKLHCKGIEKANINRVLKLGATNYYKRFEQFIFIYGKQNFFNGAFDIVPKFDGLYSSSDVPAF
 EINTEKIEPNYFISYISRPSFYKSKEKYSTGTGSKRIHENTVLFNFSLHLPCCLNEQLKIASFVCFLNRKIELLERKIYLIK
 KQKQALLQQMFIPGGSSHHHHH

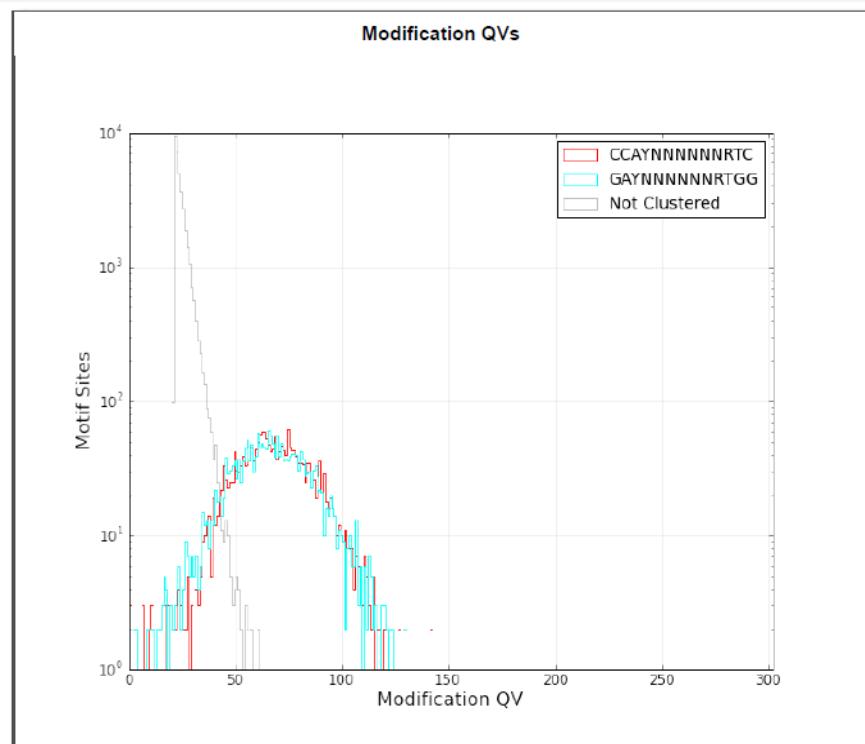
Wild Type S.SauAc*

MSNTQKKNVPELRFPGEFEWEEKQLGDLTTKIGSGKTPKGSENYTNKGIPFLRSQNIRNGKLNLDLVYISKDIDDEM
 KNSRTYYGDVLLNITGASIGRTAINSIVETHANLNQHVCIIIRLKKEYYYIFFGQYLLSRKGKRKIFLAQSGGSREGLNFK
 EIANLKIPTPTIFEEQQKIGKFSSKLDRQIELEEQKLELLQQQKKGYLQKIFSQELRFKDENGNDYPEWRFARFKDFMYK
 PINIRPAINISKSELLTVKLHCKGIEKANINRVLKLGATNYYKRFEQFIFIYGKQNFFNGAFDIVPKFDGLYSSSDVPAF
 EINTEKIEPNYFISYISRPSFYKSKEKYSTGTGSKRIHENTVLFNFSLHLPCCLNEQLKIASFVCFLNRKIELLERKIYLIK
 KQKQALLQQMFIPGGMF*

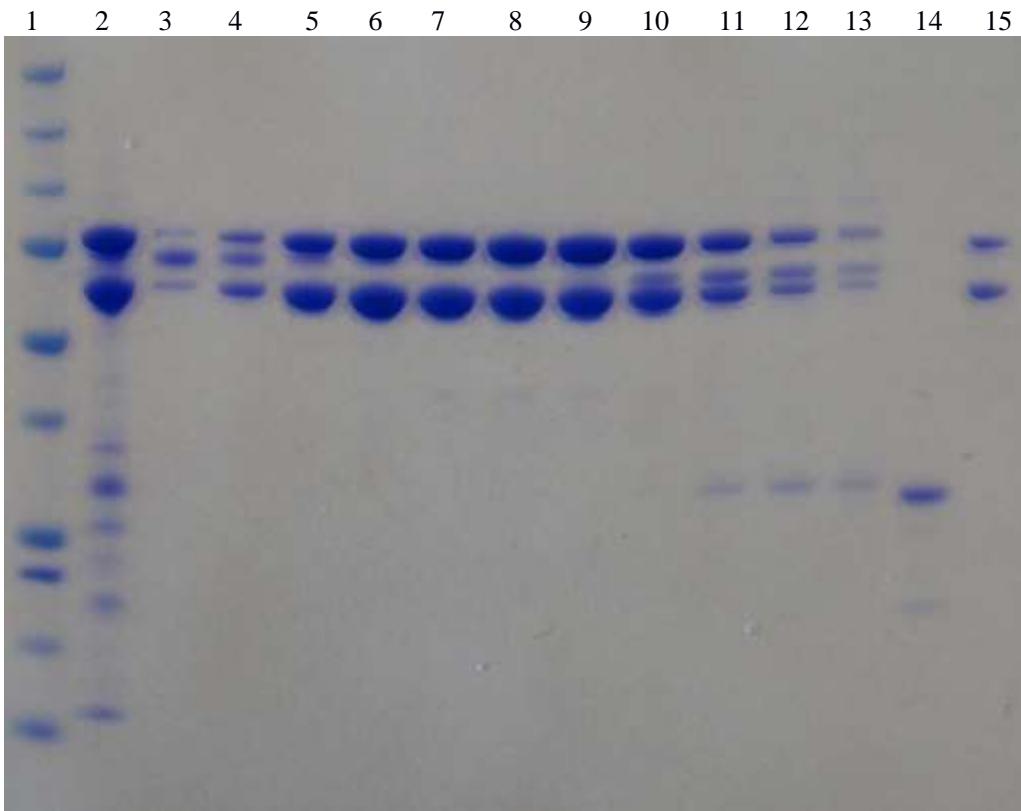
SMRT Cells: 1 Movies: 1

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

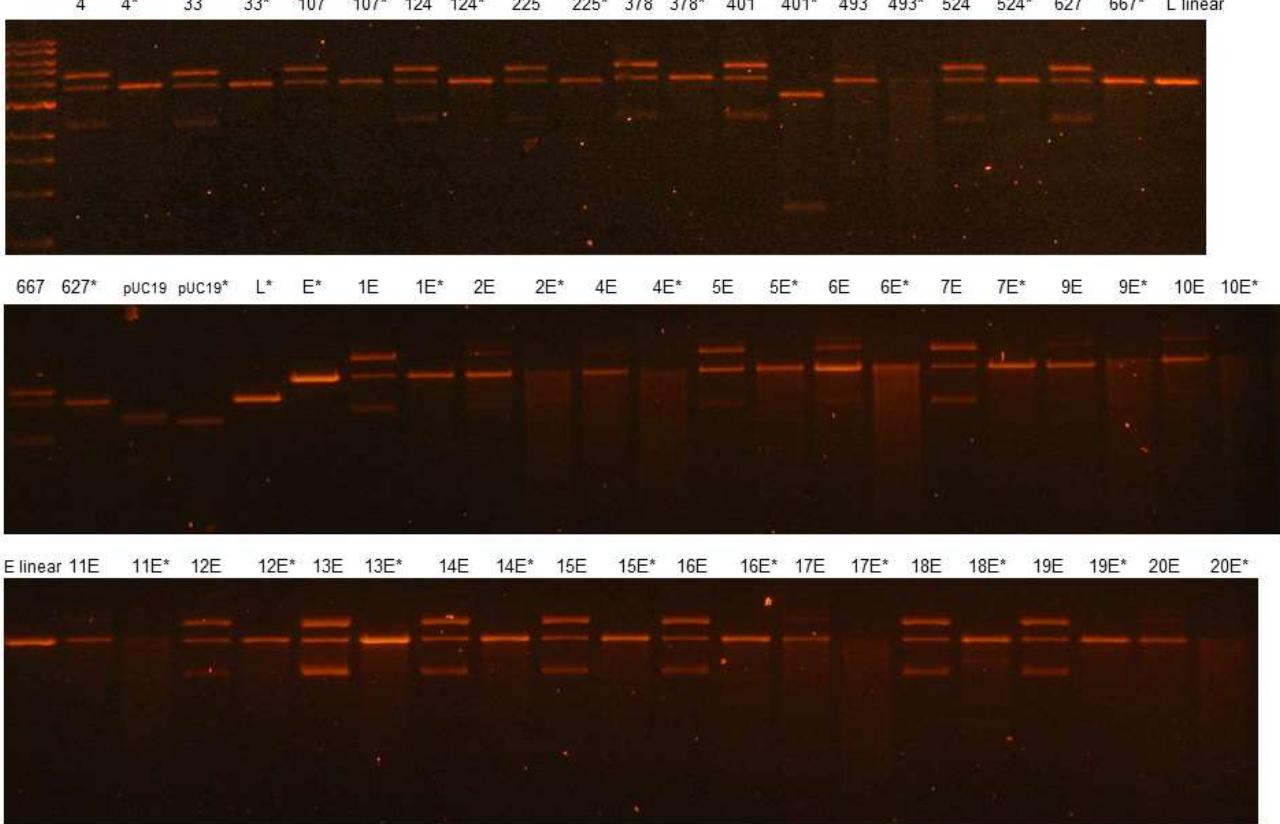


1
2 **S.SauBI-EGFP**
3 CC22-1 **AGG-6-TGAR**
4 This MTase was expressed and purified as a fusion with EGFP.
5 Nuclease assays and SMRT analysis gave the same target site.
6
7 MSNTQKKNVPELRFPGEWEEKKLGDLTDRVIRKNKNLESKKPLTISGQLGLIDQTEYFSKSVS
8 SKNLENYTLINKNGEFAYNKSYNSGYPLGAIKRLTRYDSGVLSLYICFSIKSEMSKDFMEAYFDST
9 HWYREVSGIAVEGARNHGLLNVSVNDFFTILIKYPSLEEQQKIGKFFSKLDRQIELEEQKLELLQQ
10 QKKGYMQKIFSQELRFKNENGNDYPDWERIKFFDVIDKVIDFRGRTPKKLNMEWSDEGYLALSAVN
11 VKKGYIDFNVEAKYGNLDLYTRWMRGNELYKGQVLFTTEAPMGNVAQVPDNKGYILSQRTIAFNSN
12 EKITDNFLASLLSENVYNDLLKLCGATAKGVSQKNLNRLYVTIPHISISEQEEIAEFFRKINQLV
13 ELQKYKIEHTKSQKVFLQKMFIGSMVSKGEELFTGVVPILVELGDVNGHKFSVSGEGERGATYG
14 KLTLKFICTTGKLPVPWPTLVTTLYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNY
15 KTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNNNSHNVYIMADKQKNGIKVNFKIRHNIE
16 DGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALKDPNEKRDHMVLLEFVTAAGITLGMDELYK
17 HHHHHH
18
19



47 1- marker 2- Nickel column eluate 3-14 Fractions from gel filtration column
48 15- CC5-1 Purified protein marker
49
50
51
52
53
54
55
56
57
58
59
60

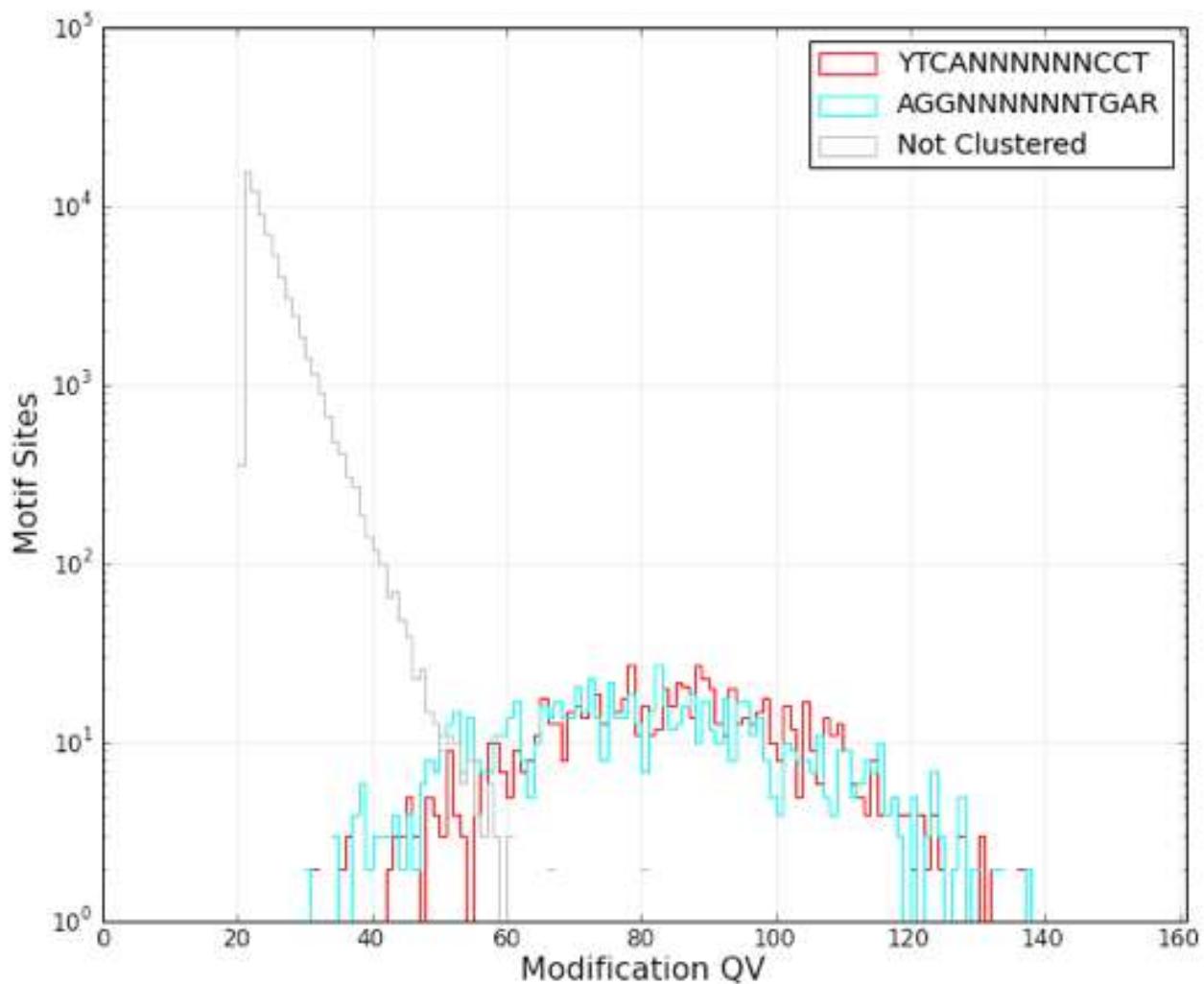
1
2 **S. SauBI-EGFP**
3 CC22-1 AGG-6-TGAR
4
5 DNA cleavage assay
6
7
8 4 4* 33 33* 107 107* 124 124* 225 225* 378 378* 401 401* 493 493* 524 524* 627 667* L linear
9
10
11
12
13
14
15
16
17 667 627* pUC19 pUC19* L* E* 1E 1E* 2E 2E* 4E 4E* 5E 5E* 6E 6E* 7E 7E* 9E 9E* 10E 10E*
18
19
20
21
22
23
24
25
26 E linear 11E 11E* 12E 12E* 13E 13E* 14E 14E* 15E 15E* 16E 16E* 17E 17E* 18E 18E* 19E 19E* 20E 20E*
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



1
2 S . SauBI - EGFP

3 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
AGGNNNNNNTGAR	1	m6A	99.24	919	926	83.9	55.7	YTCANNNNNCCT
Not Clustered	0		0.06	5230	9124356	34.8	61.5	

10
11
12
13
14
15 Modification QV Histogram By Motif16
17 Modification QV Histogram

1
2 **S.SauCE**
3 **ST425-1 GWAG-5-RTGA**

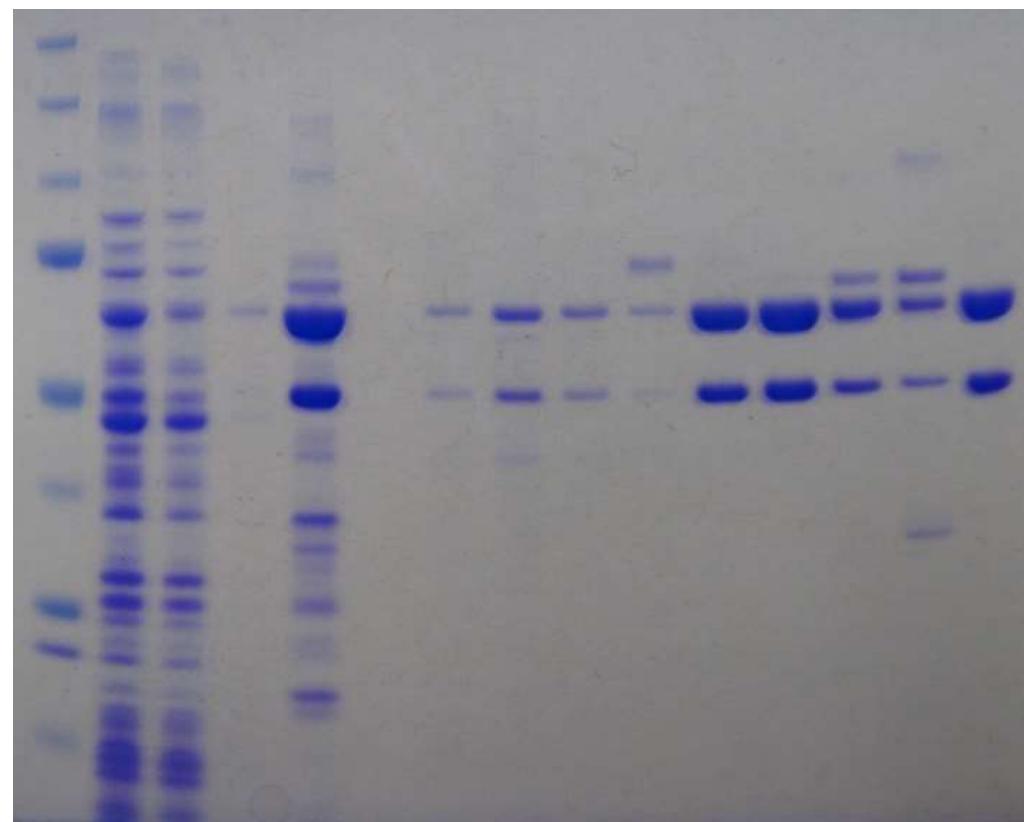
4 The recombinant enzyme with TRDs C and E was purified and used in
5 the nuclease assay. There are minor differences in amino acid
6 sequence between members of ST425-1.
7

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

9 MSNTQKKNVPELRFPGEFEWEEKKVGELLEFKNGLNKGKEYFGSGSSIVNFKDVFNNRSINTNNL
10 TGKVNVSKEKNYSVEKGDVFFTRTSEVIGEIGYPSVILNDPENTVFSFVLRGRPKSGIDLINN
11 NFKRYVFFTNSFRKEMITKSSMTTRALSGTAINRMKVIYPVSAKEQKKIGDFFSKLDRQIELEEQ
12 KLELLQQQKKGYMQKIFSQELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKDAITNGSYDFYVR
13 SPIVYKINTFSYEGERAILTVGDGVGVGVKFHYVNGKFDYHQRVYKISDFKNYYGLLFYYFSQNFL
14 KETKKYSAKTSVDSVRKDMIANMKVPRPIYIEQKKIGQFIKRVDNKTKIQQVIELLKQRKKSLLQ
15 KMFIPGGSHHHHHH

16
17 **Wild type S.SauCE**
18 MSNTQTKNVPELRFPGEFEWEEKQVGELLEFKNGLNKGKEYFGSGSSIVNFKDVFNNRSINTNNL
19 TGKVNVSKEKNYSVEKGDVFFTRTSEVIGEIGYPSVILNDPENTVFSFVLRGRPKSGIDLINN
20 NFKRYVFFTNSFRKEMITKSSMTTRALSGTAINRMKVIYPVSAKEQKKIGDFFSKLDRQIELEEQ
21 KLELLQQQKKGYMQKIFTQELRFKDENGNDYPEWEETTIKEIAQINTGKKDTKDAITNGSYDFYVR
22 SPIVYKINTFSYEGERAILTVGDGVGVGVKFHYVNGKFDYHQRVYKISDFKNYYGLLFYYFSQNFL
23 KETKKYSAKTSVDSVRKDMVANMKVPRPIYIEQEKIGQFIKKVDNKIKIQQQVIELLKQRKKALLQ
24 KMFIP*

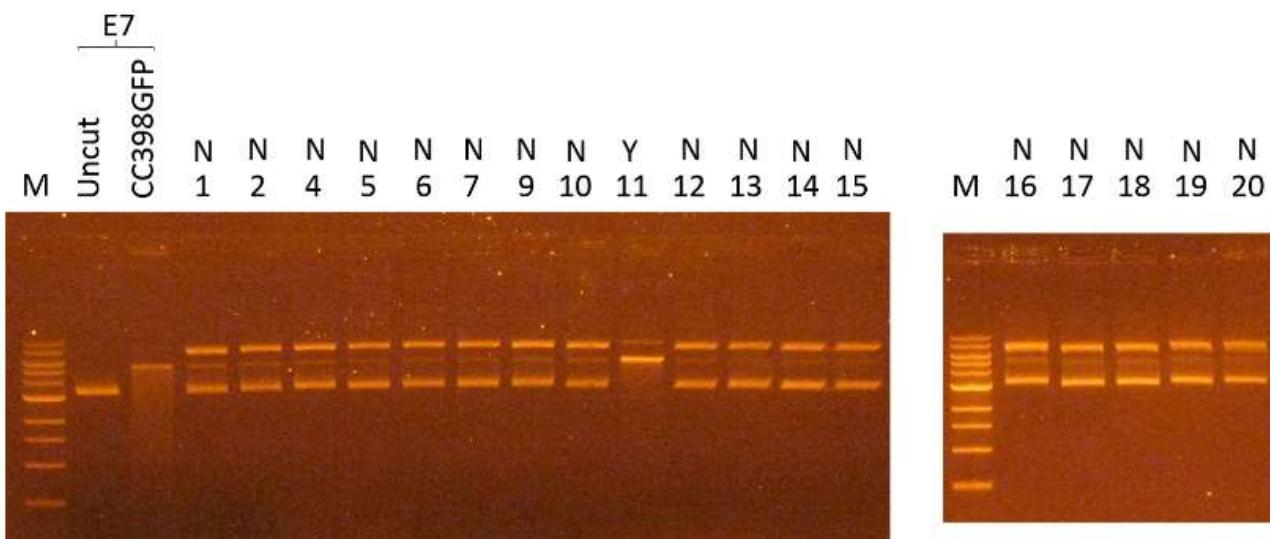
25
26 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15



55 1- marker 2- soluble cell extract 3- flow through from Nickel column 4- wash from Nickel
56 column
57 5- eluate from Nickel column 6-14 Fractions from gel filtration column 15- CC398-1 purified
58 protein marker
59
60

1
2 **S.SauCE**
3 **ST425-1 GWAG-5-RTGA**
4
5
6

DNA cleavage assay.



1
2 **S.SauJP**
3 **CC51 GGA-6-CCT**

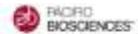
4 This MTase was used in the SMRT analysis of *E. coli* ER2796.
5 There are minor variations in the sequences of the S subunits in
6 CC51.

7 **Recombinant S.SauJP CC51-1**

8 MSNTQKKNVPELRFPGEFEWEEKKLDIIVNVNSGKDYKHLKGDI PVYGTGGYMTSVSEPLSEID
9 AVGIGRKGTINKPYLLEAPFWTVDTLFYCTPKKETDILFILSLFRKINWKVYDESTGVPSLSKQTI
10 NKINRFVPSNKEQQKIGEFFIKLDRQIELEEQQKLELFQQQKKGYMQKIFSQELRFKDESGNDYPDW
11 EEKELGEVADRVIRKNKNFESKKPLTISGQLGLIDQTEYFSKSVSSKNLENYTLINKGEFAYNKSY
12 SNGYPLGAIKRLTRYDSGVVLSSLYICFSIKSEMSKDFMEAYFDSTHWYREVSGIAVEGARNHGLLN
13 ISVNDDFTILIKYPSLEEQRKIGDFFIKLDRQIELEEQQKLELLQQRKKALLKSMLIPGGSHHHHHH
14 **Wild Type S.SauJP**

15 MSNTQTKNVPELRFPGEFEWEEKKLEDIIVNVNSGKDYKHLKGDI PVYGTGGYMTSVSEPLSEID
16 AVGIGRKGTINKPYLLEAPFWTVDTLFYCTPKKETDILFILSLFRKINWKVYDESTGVPSLSKQTI
17 NKINRFVPTNKEQQKIGKFFSKLDRQIELEEQQKLELFQQQKKGYMQKIFSQELRFKDESGNDYPDW
18 EEKELGEVADRVIRKNKNFESKKPLTISGQLGLIDQTEYFSKSVSSKNLENYTLINKGEFAYNKSY
19 SNGYPLGAIKRLTRYDSGVVLSSLYICFSIKSEMSKDFMEAYFDSTHWYREVSGIAVEGARNHGLLN
20 ISVNDDFTILIKYPSLEEQRKIGDFFIKLDRQIELEEQQKLELLQQRKKALLKSMLIPGGSHHHHHH
21

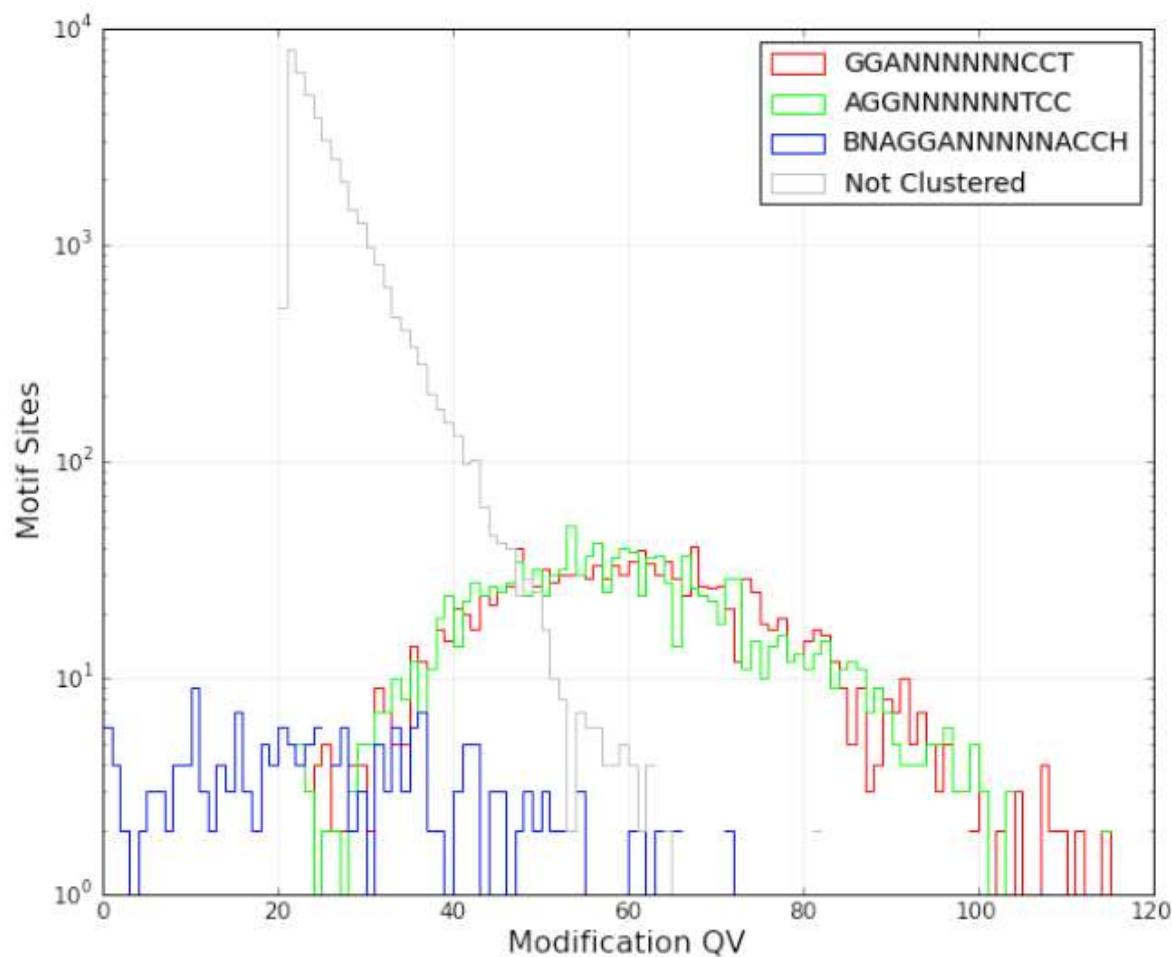
22 **Reports for Job Dryden_J_P_MODs**



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
AGGNNNNNNNTCC	1	m6A	97.58%	1333	1366	61.04	39.18	GGANNNNNNCCT
BNAGGAGNNNNACCH	3	m6A	46.26%	99	214	47.92	39.46	

1
2 S. SauJP
3 CC51
4
5

GGA-6-CCT

6 Modification QVs
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2 **S . SauCL-EGFP**3 **CC45-1 GWAG-6-TAAA**

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

7 **S . SauCL-EGFP "Expected" sequence**

8 MSNTQKKNVPELRFPGEFEWEEKKVGELLEFKNGLNKGKEYFGSGSSIVNFKDVFNNRSINTNNL
 9 TGKVNVSNSKELKNYSVEKGDVFFTRTSEVIGEIGYPSVILNDPENTVFSFVLGRPKSGIDLINN
 10 NFKRYVFFTNSFRKEMITKSSMTTRALSGTAINRMKV IYPVSAKEQKKIGDFFSKLDRQIELEEQ
 11 KLELLQQQKKGYMQKIFSQELRFKDENGNDYPNWRTIELKNILENIVDNRGKTPDNAPSEKYLLE
 12 VNALGYYRPAYIKVSKFVSENTYNNWFREHLKENDILFSTVGNTGIVSLMDNYKAVIAQNI VGLRV
 13 NNNNLPSFIYYMLSYKGNQKKIKRIQMGA VQPSVKVSQFKFIKYLVPIKDEQE KVAKLLIEIDKLV
 14 NKQLIKIELLQQRKKALLKSMFIGSMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYG
 15 KLTLKFICTTGKLPVPWPTLVTTLYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNY
 16 KTRAEVKFE GDTLVNRIELKGIDFKEDGNILGHKLEYNNNSHNVYIMADKQKNGIKVNFKIRHNIE
 17 DGSVQLADHYQQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMVLLEFVTAA GITLGMD ELYK
 18 HHHHHH

19 **S . SauCL-EGFP "Actual" sequence**

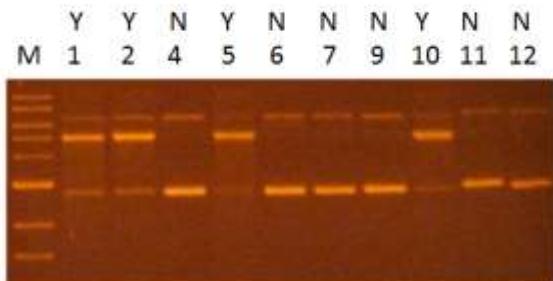
20 MSNTQKKNVPELRFPGEFEWEEKKVGELLEFKNGLNKGKEYFGSGSSIVNFKDVFNNRSINTNNL
 21 TGKVNVSNSKELKNYSVEKGDVFFTRTSEVIGEIGYPSVILNDPENTVFSFVLGRPKSGIDLINN
 22 NFKRYVFFTNSFRKEMITKSSMTTRALSGTAINKMKV IYPVSAKEQKKIGDFFSKLDRQIELEEQ
 23 KLELLQQQKKGYMQKIFSQELRFKDENGNDYPNWRTIELKNILENIVDNRGKTPDNAPSEKYLLE
 24 VNALGYYRPAYIKVSKFVSENTYNNWFREHLKENDILFSTVGNTGIVSLMDNYKAVIAQNI VGLRV
 25 NNNNLPSFIYYMLSYKGNQKKIKRIQMGA VQPSVKVSQFKFIKYLVPIKDEQE KVAKLLIEIDKLV
 26 NKQLIKIELLQQRKKALLKSMFIGSMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYG
 27 KLTLKFICTTGKLPVPWPTLVTTLYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNY
 28 KTRAEVKFE GDTLVNRIELKGIDFKEDGNILGHKLEYNNNSHNVYIMADKQKNGIKVNFKIRHNIE
 29 DGSVQLADHYQQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMVLLEFVTAA GITLGMD ELYK
 30 HHHHHH

31 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15



37 1- marker 2- Nickel column eluate 3-14 Fractions from gel filtration column
 38 15- CC5-1 purified protein marker

1
2 **S . SauCL-EGFP**
3 CC45-1 **GWAG-6-TAAA**
4 DNA cleavage assay.
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



1
2 **S. SauOE**
3 **CC15**
4 **Recombinant S. SauOE**

5 **CC15-1**

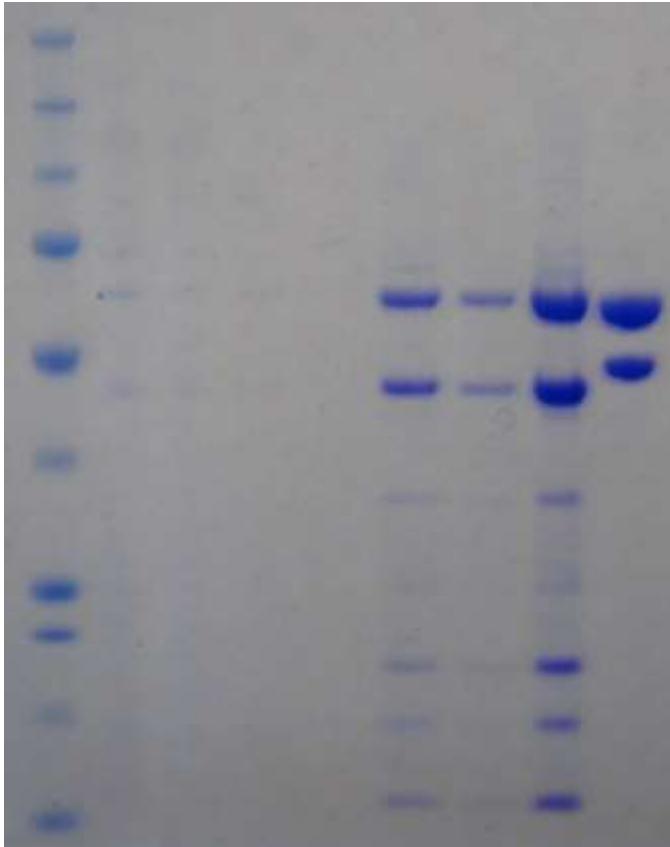
6 **CAAC-5-RTGA**

7 MSNTQKKNVPELRFPGEFEWEEKKLGEVGTFTSGGTPLSKSEYWNGDIPWITTGDIHNIKRENI
8 TNFITEKGLNESSAKLITNEAILIAMYGQGKTRGMSAILNFEATTNQACAIYQTQNQNINFVFQYFQ
9 KLYEFLRSLSNEGSQKNLSLSLLKEITLNYPNEQEQQKIGDFFSKLDRQIELEEQQKKG
10 YMQKIFSQELRFKDENGKDYPEWEETTIKEIAQINTGKKDTKDAITNGSYDFYVRSPIVYKINTFS
11 YEGEAILTVGDGVGVGVKFHYVNGKFDYHQRVYKISDFKNYYGLLLFSQNFLKETKKYSAKTS
12 VDSVRKDMIANMKVPRPIYIEQKKIGQFIKRVDNTKIQQVIELLKQRKSLLQKMFIPGGSHHH
13 HHH

14 **Wild Type S. SauOE**

15 MSNKQKKNVPELRFPGEFEWEEKKLGEVGTFTSGGTPLSKSEYWNGDIPWITTGDIHNIKRENI
16 TNFITEKGLNESSAKLITNEAILIAMYGQGKTRGMSAILNFEATTNQACAIYQTQNQNINFVFQYFQ
17 KLYEFLRSLSNEGSQKNLSLSLLKEITLNYPNEQEQQKIGDFFSKLDRQIELEEQQKKG
18 YMQKIFSQELRFKDENGNDYPEWEETTIKEIAQINXGKKDTKDAITNGSYDFYVRSPIVYKINTFS
19 YEGEAILTVGDGVGVGVKFHYVNGKFDYHQRVYKISDFKNYYGLLLFSQNFLKETKKYSAKTS
20 VDSVRKDMIANMKVPRPIYIEQKKIGQFIKRVDNTKIQQVIELLKQRKALLQKMF
21

22 1 2 3 4 5 6 7 8 9



50 1- marker 2- soluble cell extract 3- Nickel column flow through
51 4- Nickel column wash 1 5- Nickel column wash 2 6- Nickel column eluate
52 7- eluate after PD10 desalting 8- Final concentrated protein
53 9- CC398-1 purified protein marker

54
55 Although purified, this MTase was only used in SMRT sequencing.
56
57
58
59
60

1
2 S.SauOE
3
4 CC15
5

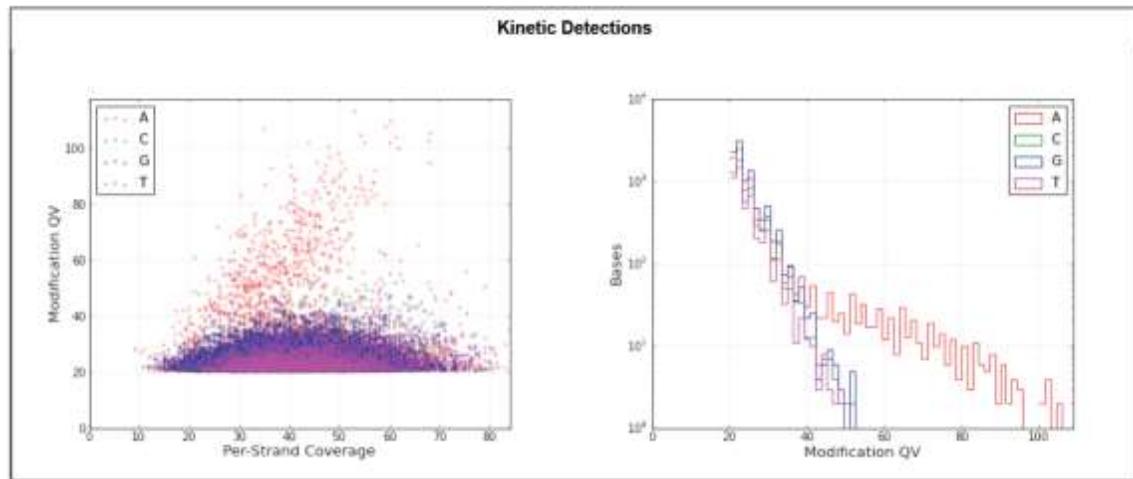
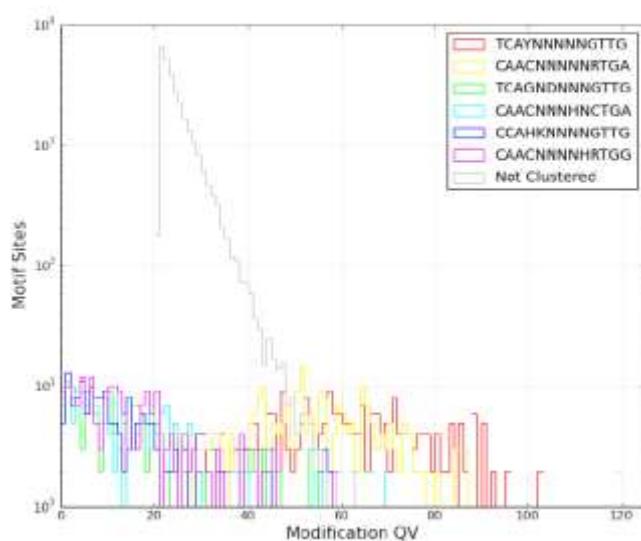
Recombinant S.SauOE

CC15-1

CAAC-5-RTGA

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	CAACNNNNRTGA
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
CCAHKNNNNNGTTG	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	9115968	34.9	47.5	

15 Modification QV Histogram By Motif

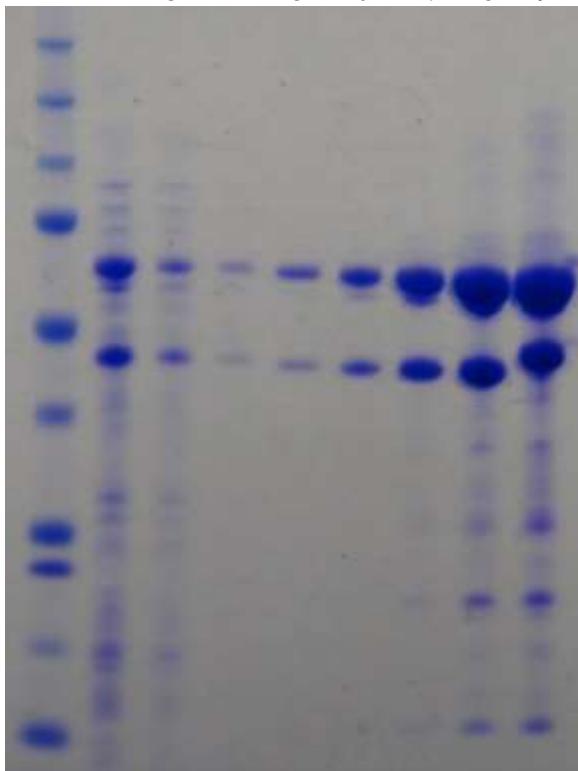


51 Motifs

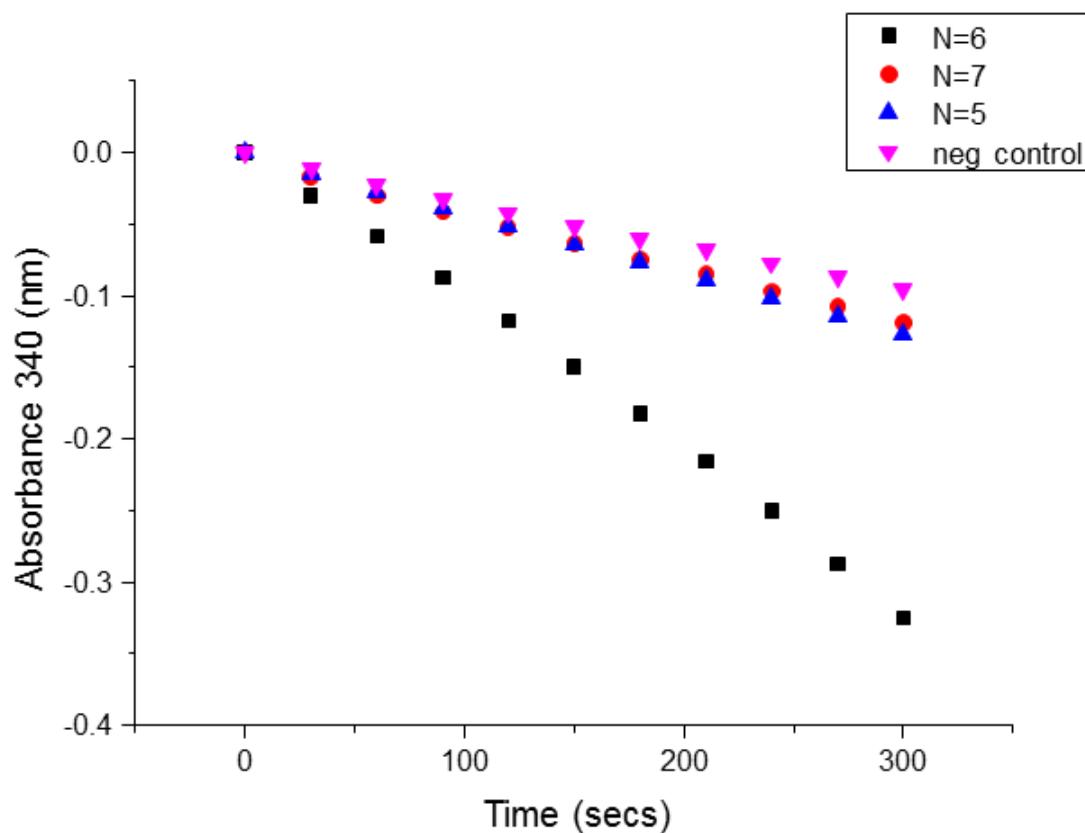
Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
TCAYNNNNNGTTG	3	m6A	96.67	261	270	63.4	38.3	CAACNNNNRTGA
CAACNNNNNRTGA	3	m6A	92.22	249	270	57.7	36.1	TCAYNNNNNGTTG
TCAGNDNNNGTTG	3	m6A	23.86	47	197	48.5	38.3	CAACNNNHNCTGA
CAACNNNHNCTGA	3	m6A	17.77	35	197	49.9	38.6	TCAGNDNNNGTTG
CCAHKNNNNNGTTG	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	9115968	34.9	47.5	

1
2 **S . SauJQ**3 **CC59**4 This enzyme was purified and analysed using the ATPase assay as
5 both TRD specificities were known and the DNA cleavage assay
6 showed cutting of all plasmids.7 **Recombinant S . SauJQ CC59-1 GGA-6-RTGT**8
9 MSNTQKKNVPELRFPGEFEWEERKLGDLIKVNNSGKDYKHLDKGDI
10 AVGIGRKGTINKPYLLEAPFWTVDTLFYCTPKKETDILFILSLFRKINWKVY
11 NKINRFVPSNKEQQKIGEFFIKLDRQIELEEQQQQKKGYMQKIFSQELRFK
12 EERRFADIFKFHNKLRLKPIKENLRVKGSYPYYGATGIIDYVDDFIFDGN
13 LVYLVNGKFWVNNHAHILSPLNGNIQYLQVAELVNYEKNTGTAQPKLN
14 EQQKIGSFLSKLDRQIDLEEQQKALLKSMFVPGGSHHHHH15 **Wild type S . SauJQ**16
17 MSNTQKKNVPELRFPFEFEWEERKLGDLIKVNNSGKDYKHLDKGDI
18 AVGIGRKGTINKPYLLEAPFWTVDTLFYCTPEKEADILFILSLFRKINWK
19 NKINRLVPTNKEQQKIGEFFSKLDRQIELEEQQQQKKGYMQKIFSQELRF
20 EERRFADIFKFHNKLRLKPIKENLRVKGSYPYYGATGIIDYVDDFIFDGN
21 LVYLVNGKFWVNNHAHILSPLNGNIQYLQVAELVNYEKNTGTAQPKLN
22 EQQKIGSFLSKLDRQIDLEEQQKALLKSMFV*

23 1 2 3 4 5 6 7 8 9

50 1- marker 2- soluble cell extract 3- Nickel column flow through
51 4- Nickel column wash 1 5- Nickel column wash 2 6- Nickel column eluate
52 7- eluate after conc. and PD10 desalting
53 8- final concentrated protein 9- XE purified protein marker

1
 2 S.SauJQ
 3 CC59
 4 Recombinant S.SauJQ CC59-1 GGA- 6-RTGT
 5 ATPase assay shows that N=6.
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34



Oligonucleotide name	DNA sequence (5' to 3')
JQ5for	AGATGATGTCATCAATGCGGATTACAGTGTGCCCTATACGATATAA
JQ5rev	TTATATCGTATAGGCACACTGTAATCCGCATTGATGACATCATCT
JQ6for	AGATGATGTCATCAATGCGGATTGACAGTGTGCCCTATACGATATAA
JQ6rev	TTATATCGTATAGGCACACTGTCAATCCGCATTGATGACATCATCT
JQ7for	AGATGATGTCATCAATGCGGATTAGACAGTGTGCCCTATACGATATAA
JQ7rev	TTATATCGTATAGGCACACTGTCTAACCGCATTGATGACATCATCT

1
2 **S . SauRQ**
3 **CC72**

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

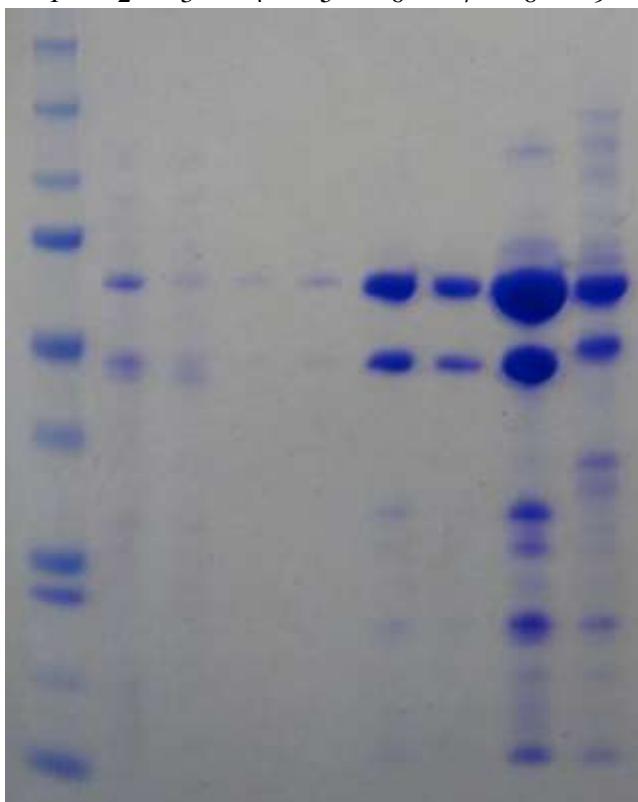
7 **Recombinant S . SauRQ CC72-1 GARA-6-RTGT**

8 MSNTQKKNVPELRFPGEFEWEEKKLGEVAKIYDGTHQTPKYTNEGIKFLSVENIKTLNSSKYISE
9 EAFEKEFKIRPEFGDILMTRIGDIGTPNIVSSNEKFAYYVSLALLTKNLNSYFLKNLILSSSIQN
10 ELWRKTLHVAFPKKINKNEIGKIKINYPKKQEQQKIGQFFSKLDRQIEEQKLELLQQQKKGYMQ
11 KIFSQELRFKDENGEDYSEWEERRFADIFKFHNKLRKPIKENLRVKGSYPYYGATGIIDYVDDFIF
12 DGNYLLIGEDGANIITRSAPLVYLVNGKFWVNNHAHILSPLNGNIQYLYQVAELVNYEKYNTGTAQ
13 PKLNIQNLKIINVVISTNLEEQQKIGSFLSKLDRQIDLEEQKLELLQQRKALLKSMFVPGGSHHH
14 HHH

15 **Wild type S . SauRQ**

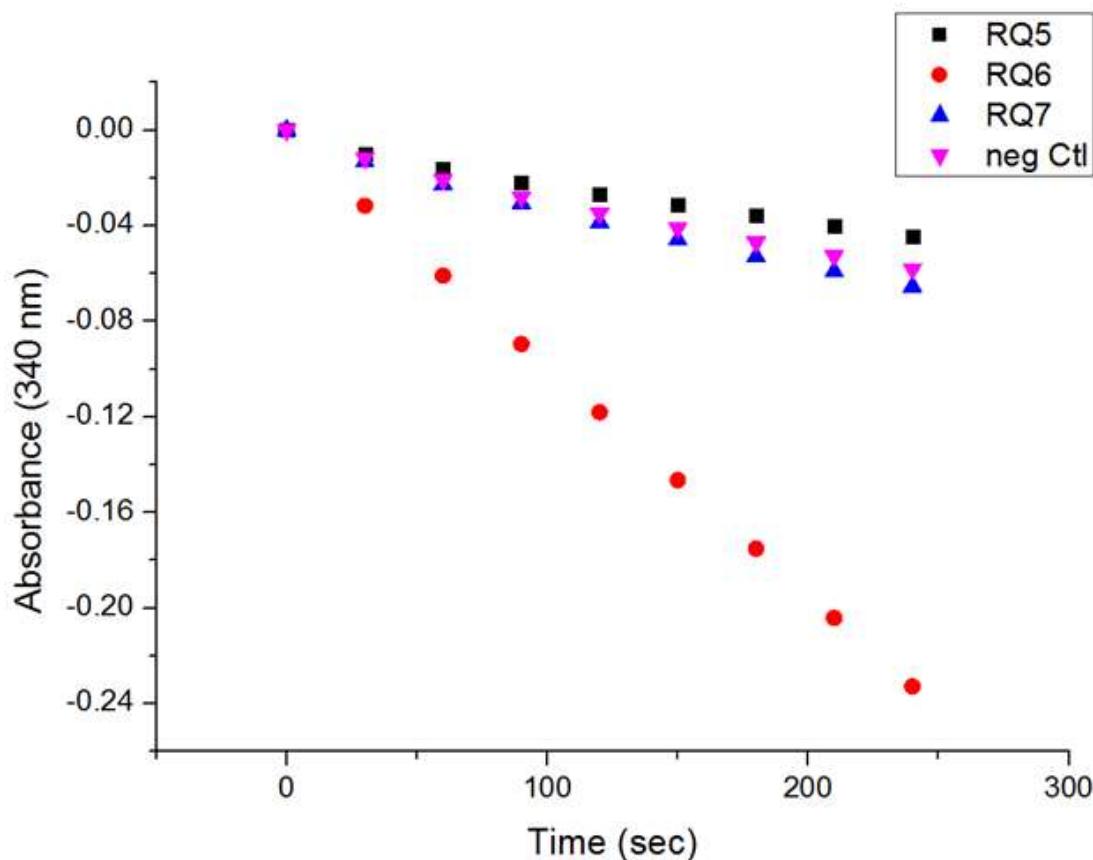
16 MSNTQKKNVPELRFPGEFEWEEKKLGEVAKIYDGTHQTPKYTNEGIKFLSVENIKTLNSSKYISE
17 EAFEKEFKIRPEFGDILMTRIGDIGTPNIVSSNEKFAYYVSLALLTKNLNSYFLKNLILSSSIQN
18 ELWRKTLHVAFPKKINKNEIGKIKINYPKKQEQQKIGQFFSKLDRQIEEQKLELLQQQKKGYMQ
19 KIFSQELRFKDENGNDYPEWEERRFADIFKFHNKLRKPIKENLRVKGSYPYYGATGIIDYVDDFIF
20 DGNYLLIGEDGANIITRSAPLVYLVNGKFWVNNHAHILSPLNGNIQYLYQVAELVNYEKYNTGTAQ
21 PKLNIQNLKIISVVISTNLEEQQKIGSFLSKLDRQIDLEEQKLELLQQRKALLKSMFV*

22 1 2 3 4 5 6 7 8 9



23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52 1- marker 2- soluble cell extract 3- Nickel column flow through
53 4- Nickel column wash 1 5- Nickel column wash 2 6- Nickel column eluate
54 7- eluate after PD10 desalting 8- Final protein after concentration
55 9- NP purified protein as marker

1
2 S . SauRQ
3 CC72
4 Recombinant S . SauRQ GARA-6-RTGT
5
6
7
8
9



35 N=6 shows activity.
36
37

Oligonucleotide name	DNA sequence (5' to 3')
RQ5for	AGATGATGGAATCAATGCGAGATTCCAGTGTGCCCTATA CGATATAA
RQ5rev	TTATATCGTATAGGCACACTGGAACTCTGCATTGATTCCATCATCT
RQ6for	AGATGATGGAATCAATGCGAGATGTCCAGTGTGCCCTATA CGATATAA
RQ6rev	TTATATCGTATAGGCACACTGGACATCTGCATTGATTCCATCATCT
RQ7for	AGATGATGGAATCAATGCGAGATGTACCAGTGTGCCCTATA CGATATAA
RQ7rev	TTATATCGTATAGGCACACTGGTACATCTGCATTGATTCCATCATCT

1
2 **S.SauJS**

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

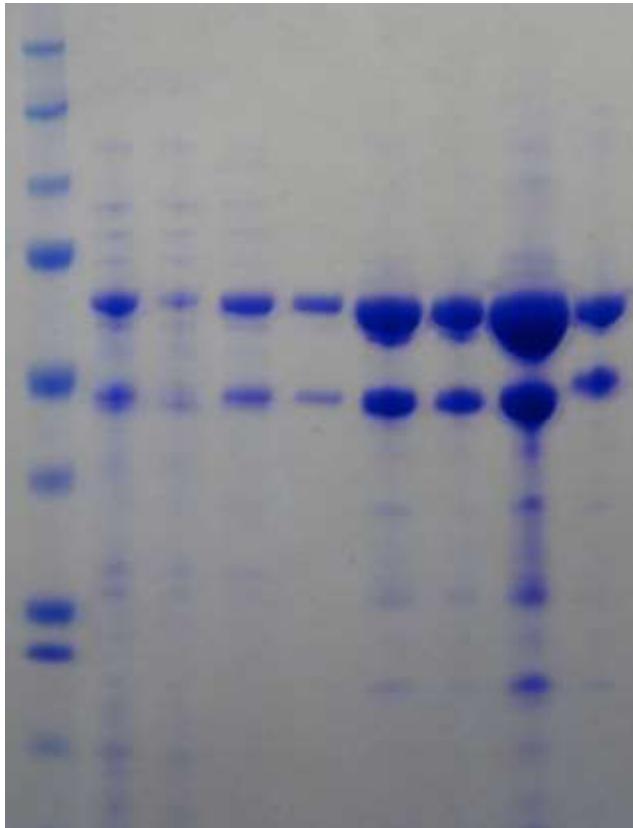
6 **CC72**7 **Recombinant S.SauJS**8 **CC72-2**9 **GGA-7-TGC**

10 MSNTQKKNVPELRFPGEFEWEEKKLGDLIKVNNSGKDYKHLEKGDI PVYGTGGYMTSVSEPLSEID
 11 AVGIGRKGTINKPYLLEAPFWTVDTLFYCTPKKETDILFILSLFRKINWKVYDESTGVPSLSKQTI
 12 NKINRFVPSNKEQQKIGEFFIKLDRQIELEEQQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPDW
 13 TNERLGEVTVTMGQSPKS VNYTDNSNDTVLIQGNADIENGLINPRIYTREVTKLIQKDEIILT
 14 APVGKLAMAQINACIGRGVCSIKGDKFLYYFLEWFATQNWKIRFSQGSTFESISGNDIRNIHIKIP
 15 VEDERTKIIKLLNSLDVLNSKTDLKIQNLKQRKQSLLQKIFVPGGSHHHHH

16 **Wild Type S.SauJS**

17 MSNTQKKNVPELRFPFEFEWEEKQLGNI IKVNNSGKDYKHLKGDI PVYGTGGYMTSVSEPLSEID
 18 AVGIGRKGTINKPYLLEAPFWTVDTLFYCTPKKETDILFILSLFRKINWKVYDESTGVPSLSKQTI
 19 NKINRFVPTNKEQQKIGKFFSKLDRQIELEEQQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPDW
 20 TNERLGEVTVTMGQSPKS VNYTDNSNDTVLIQGNADIENGLINPRIYTREVTKLIQKDEIILT
 21 APVGKLAMAQINACIGRGVCSIKGDKFLYYFLEWFATQNWKIRFSQGSTFESISGNDIRNIHIKIP
 22 VEDERTKIIKLLNSLDVLNSKTDLKIQNLKQRKQSLLQKIFV
 23

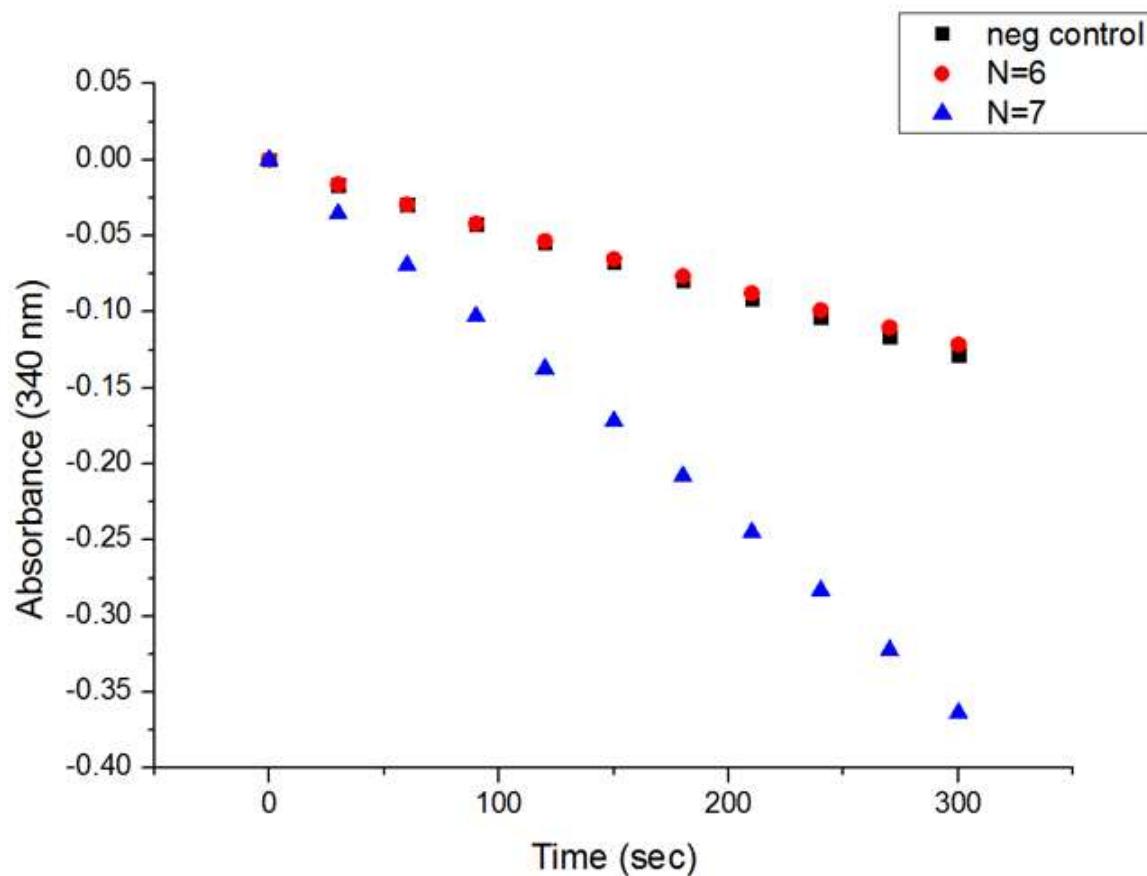
24 1 2 3 4 5 6 7 8 9



51
 52 1- marker 2- soluble cell extract 3- Nickel column flow through
 53 4- Nickel column wash 1 5- Nickel column wash 2 6- Nickel column eluate
 54 7- eluate after PD10 desalting 8- final protein after concentration
 55 9- NP purified protein as marker

1
 2 S.SauJS
 3 CC72
 4 Recombinant S.SauJS CC72-2 GGA-7-TGC
 5 N=7 shows activity.
 6

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



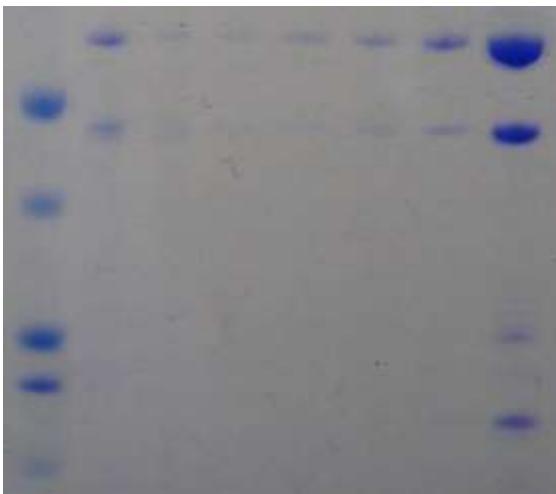
1
2 **S.SauTU**3 **CC75**4 **Recombinant S.SauTU**5 **CC75-1**6 **CAAG-5-RTC**

7 MSNTQKKNVPELRFPGFEGEWEEKELGEIFQIISGSTPLKSNEFYENGNIHWVKTTDLNNSKVTH
 8 SKEKITEYAMKSLKLKLPKNSVLIAMYGGFNQIGRTGLLKIDATINQAIASALLMNHETNPEFIQA
 9 FLNYQVKGWKRYAASSRKDPNITKKDIEQFKVPYVSINEQQKIGEFFSKIDHQIELEEQKLELLQQ
 10 QKKGYMQKIFSQELRFKDENGEDYPDWEVTTIQNITKYTSSKKSSNQYADKDNSKGYPVYDAVQEI
 11 GKDSNYDIEESYISILKDAGVGRLNLRPGKSSVIGTMGYIQSNNVDIEFLYYRMKVDFKKYIIG
 12 STIPHLYFKDYSKETLYIPSSIQEQAQIGMFISNLDKLIENKNLKLQQLQSMFIPGGS
 13 HHHHHH

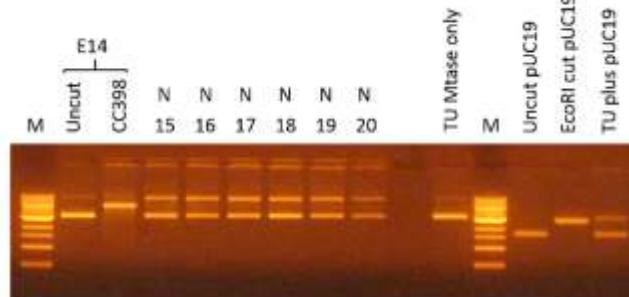
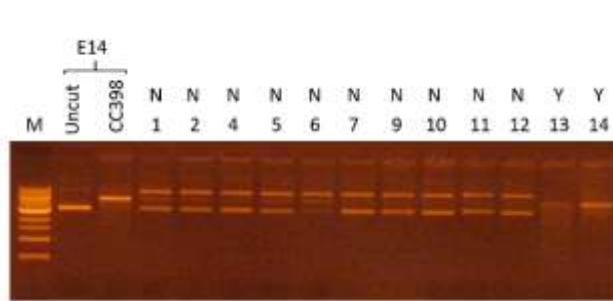
14 **Wild type S.SauTU**

15 MSNTQTKNVPELRFPGFEGEWEEKELGEIFQIISGSTPLKSNEFYENGNIHWVKTTDLNNSKVTH
 16 SKEKITEYAMKSLKLKLPKNSVLIAMYGGFNQIGRTGLLKIDATINQAIASALLMNHETNPEFIQA
 17 FLNYQVKGWKRYAASSRKDPNITKKDIEQFKVPYVSINEQQKIGEFFSKIDHQIELEEQKLELLQQ
 18 QKKGYMQKIFSQELRFKDENGEDYPDWEVTTIQNITKYTSSKKSSNQYADKDNSKGYPVYDAVQEI
 19 GKDSNYDIEESYISILKDAGVGRLNLRPGKSSVIGTMGYIQSNNVDIEFLYYRMKVDFKKYIIG
 20 STIPHLYFKDYSKETLYIPSSIQEQAQIGMFISNLDKLIENKNLKLQQLQSMFIPGGS

21 1 2 3 4 5 6 7 8



22
 23
 24 1- marker 2- soluble cell extract 3- Nickel column flow through 4- Nickel column wash 1 5-
 25 Nickel column wash 2 6- Nickel column eluate 7- eluate after conc. and PD10 desalting 8- final
 26 protein after concentration
 27
 28 DNA cleavage assay.



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

MSNTQKKNVPELRFPGEFEWEEKELRELNPDKYSYTGGPGSDLKKSODYTTDGIQIIQLQNIG
 DGYFYNNSNKVFTSNEKAEVLKSCNVFPGDIVIAKMADPIARAAIVPDNNIGKYLMASDGIRLSVDT
 VHFNTKFVLECINRKSFRKKVEDNSSGSTRMRIGLSTLGSLLKTTTKEQQKIGQFFSKLDRQIE
 LEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPD WEEKQLGELSQIVRGASPRPRIKDPKWFNK
 ESDIGWLRISDVTNQNGKIYHLEQKLSIEGQEKTRVLVTTHLLLSIAASIGKPVMNFVKTGVHDGF
 LIFLKPKFNLFMYYWLEYFKDKWSKYGQPGSQVNLNSEIVKSQTLNMPSNHEQEKGQFFNRNEK
 LIELQQEKIMYIKRCKQVLLQKMFIPGGSHHHHH

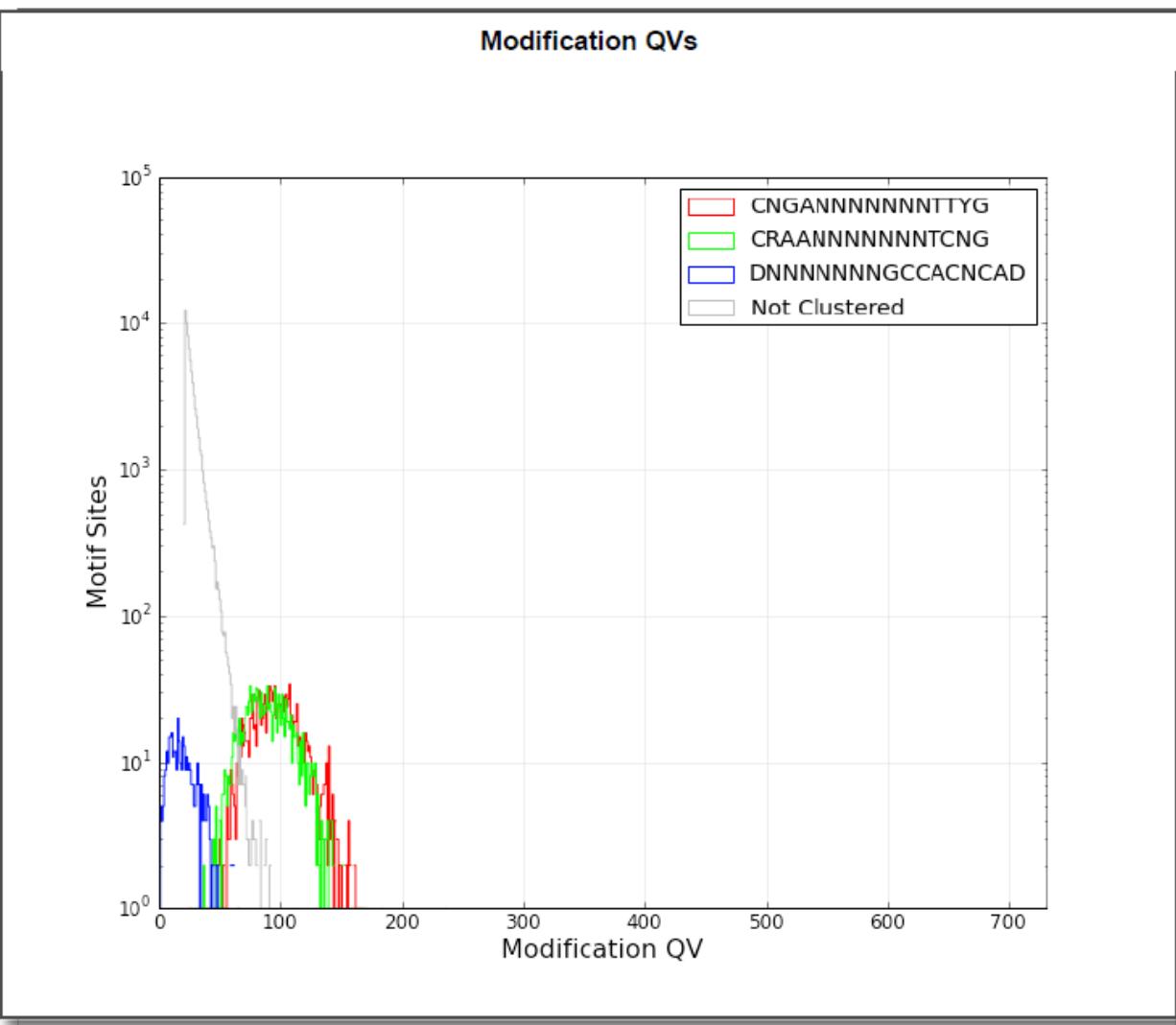
Wild Type S.SauVW

MSNTGKMNVPELRFPGEFEWEEKELRELNPDKYSYTGGPGSDLKKSODYTTDGIQIIQLQNIG
 DGYFYNNSNKVFTSNEKAEVLKSCNVFPGDIVIAKMADPIARAAIVPDNNIGKYLMASDGIRLSVDT
 VHFNTKFVLECINRKSFRKKVEDNSSGSTRMRIGLSTLGSLLKTTTKEQQKIGQFFSKLDRQIV
 LEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPD WEEKQLGELSQIVRGASPRPRIKDPKWFNK
 ESDIGWLRISDVTNQNGKIYHLEQKLSIEGQEKTRVLVTTHLLLSIAASIGKPVMNFVKTGVHDGF
 LIFLKPKFNLFMYYWLEYFKDKWSKYGQPGSQVNLNSEIVKSQTLNMPSNHEQEKGQFFNRNEK
 LIELQQEKIMYIKRCKQVLLQKMFIPGGSHHHHH*

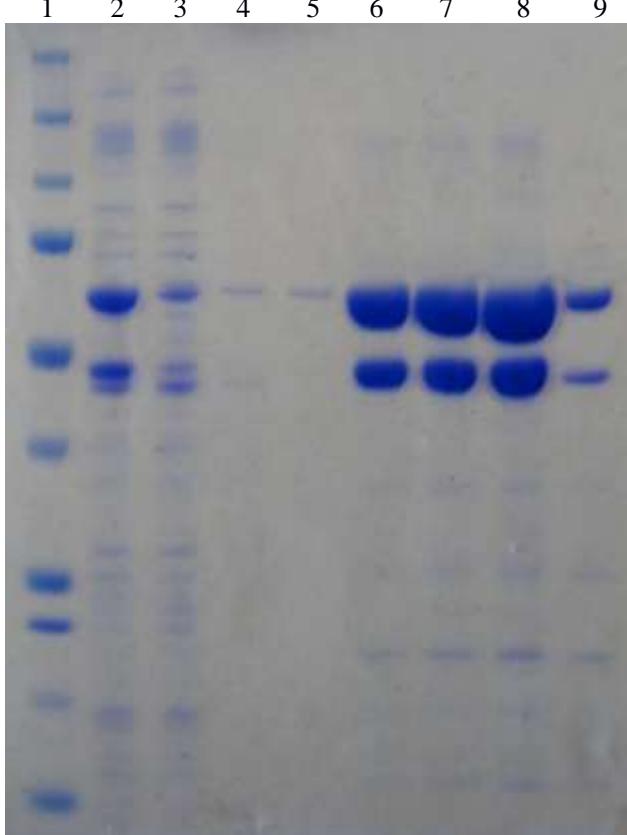
Reports for Job Dryden_V_W_MODs

Motifs	Modified Position	Type	% Motifs Detected	SMRT Cells 1		Mean Modification QV	Mean Motif Coverage	Partner Motif
				# Of Motifs Detected	# Of Motifs In Genome			
CNGANNNNNNNTYG	4	m6A	99.93%	1442	1443	97.87	66.11	CRAANNNNNNNTCNG
CRAANNNNNNNTCNG	4	m6A	99.86%	1441	1443	89.76	63.95	CNGANNNNNNNTYG
DNNNNNNNNGCACNCAD	9	unknown	19.1%	72	377	38.56	67.49	

1
2 S.SauVW
3 CC75
4 Recombinant S.SauVW CC75-2 CNGA7-TTYG
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



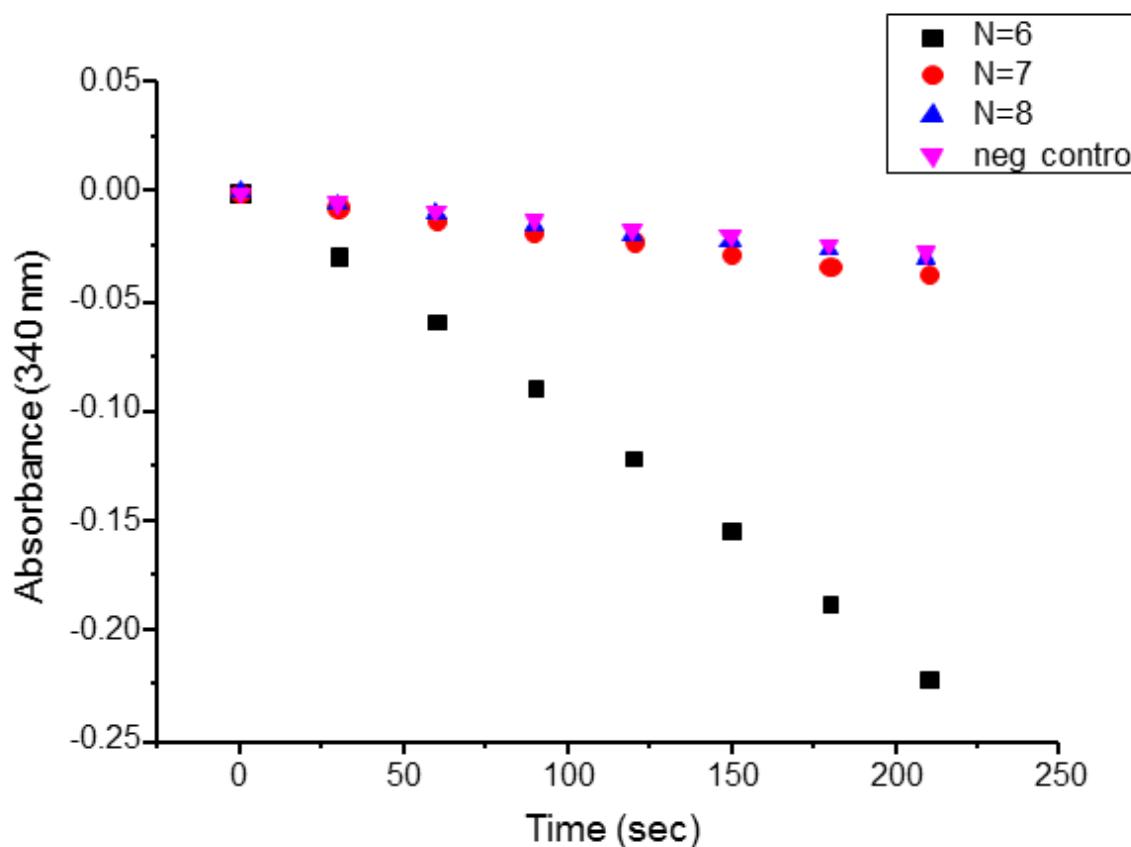
1
 2 **S.SauZW**
 3 CC80
 4 **Recombinant S.SauZW** **CC80-2** **GAC-6-TTYG**
 5
 6 MSNTQKKNVPELRFPGEFEYEYSLDIFGNLATNKSEKFNPQNENASIDIELDCIEQNTGRLIKIYNS
 7 KEFSSQKNKEFPQNVLYGKLRYLNKYYFTKKSGVCSSEIWVLKSTKEDKLLNLFLYYFIQTKRYS
 8 DVASKSAGSKMPRADWGLIENIRVYFPTELCEQQKIGQFFSKLDRQIELEEQKLELLQQQKKGYMQK
 9 IFSQELRFKDENGNDYPDWECKQLGELSQIVRGASPRPIKDPKWFNKESEDIGWLRISDVTNQNGKI
 10 YHLEQKLSIEGQEKRVLVTTHLLSIAASIGKPVMNFVKTGVHDGFLIFLKPKFNLFMYYWLEY
 11 FKDKWSKYGQPGSQVNLNSEIVKSQTLNMPSNHEQEKGQFFNRNEKLIELQQEKIMYIKRCKQVL
 12 LQKMFIPGGSHHHHHH
 13
 14 **Wild Type S.SauZW**
 15 MSNTQTKNVPELRFPGEFEYEYSLDIFGNLATNKSEKFNPQNENASIDIELDCIEQNTGRLIKIYNS
 16 KEFSSQKNKEFPQNVLYGKLRYLNKYYFTKKSGVCSSEIWVLKSTKEDKLLNLFLYYFIQTKRYS
 17 DVASKSAGSKMPRADWGLIENIRVYFPTELCEQQKIGQFFSKLDRQIELEEQKLELLQQQKKGYMQK
 18 IFSQELRFKDENGNDYPDWECKQLGELSQIVRGASPRPIKDPKWFNKESEDIGWLRISDVTNQNGKI
 19 YHLEQKLSIEGQEKRVLVTTHLLSIAASIGKPVMNFVKTGVHDGFLIFLNPKFNLFMYYWLEY
 20 FKDKWSKYGQPGSQVNLNTEIVKSQTLNMPSNHEQEKGQFFNRNEKLIELQQEKIMYLKRRKQVL
 21 LQKMFIPGGSHHHHHH
 22 LQKMFIPGGSHHHHHH
 23 1 2 3 4 5 6 7 8 9
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50



- 51 1- marker 2- soluble cell extract 3- Nickel column flow through
 52 4- Nickel column wash 1 5- Nickel column wash 2 6- Nickel column eluate
 53 7- eluate after conc. and PD10 desalting
 54 8- final protein after concentration 9- CC75-1 purified protein marker
 55
 56
 57
 58
 59
 60

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.

1
 2 S.SauZW
 3 CC80
 4 Recombinant S.SauZW CC80-2 GAC-6-TYG
 5 N=6 shows activity.
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32



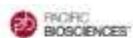
Oligonucleotide name	DNA sequence (5' to 3')
ZW6for	AGATGATGGAATCAATGCGACTTCCATT CGGCCCTACGATATAA
ZW6rev	TTATATCGTATAGGGCCGAAATGGAAGTCGCATTGATTCCATCATCT
ZW7for	AGATGATGGAATCAATGCGACTTCTCATT CGGCCCTACGATATAA
ZW7rev	TTATATCGTATAGGGCCGAAATGAGAAGTCGCATTGATTCCATCATCT
ZW8for	AGATGATGGAATCAATGCGACTTCTACATT CGGCCCTACGATATAA
ZW8rev	TTATATCGTATAGGGCCGAAATGTAGAAGTCGCATTGATTCCATCATCT

1
 2 **S.SauXf***
 3 **ST80**
 4 **Recombinant S.SauXf*** **CC80-3** **TCTA-6-RTTC**
 5 MSNTQKKNVPELRFPGEFEWEEKQFADFTKINQGLQIAINERKTEYSPELYFYITNEFLRPNSQT
 6 KYFIENPPQSVIANKE DILMTRTGNTGVVTVNGAFHNNFFKIKFDKNLYDRLFLVEVLNSSKIQ
 7 NKILSLAGSSTIPDLNHSDFYSISSSYPLLREQQKIGKFFSKLDRQIELEEQKLELLQQQKKGYMQ
 8 KIFSQELRFKDENGEDYPDWKEKKLGITEQSMYGIGASATRFDSKNIYIRITDIDEKSRKLNQYQN
 9 LTT PDELNNKYKLKRNDILFARTGASTGKSYIHKEEKDIYNYYFAGFLIKFKINEQNSPLFIYQFT
 10 LTSKFNKWKVMSVRSGQPGINSEYYAKLPLVLPNLEQQKIAKFLDRFDRQIELEKQKIEILQQQ
 11 KKGLLQSMFI PGGSHHHHHH
 12
 13
 14

15 **Wild Type S.SauXf***

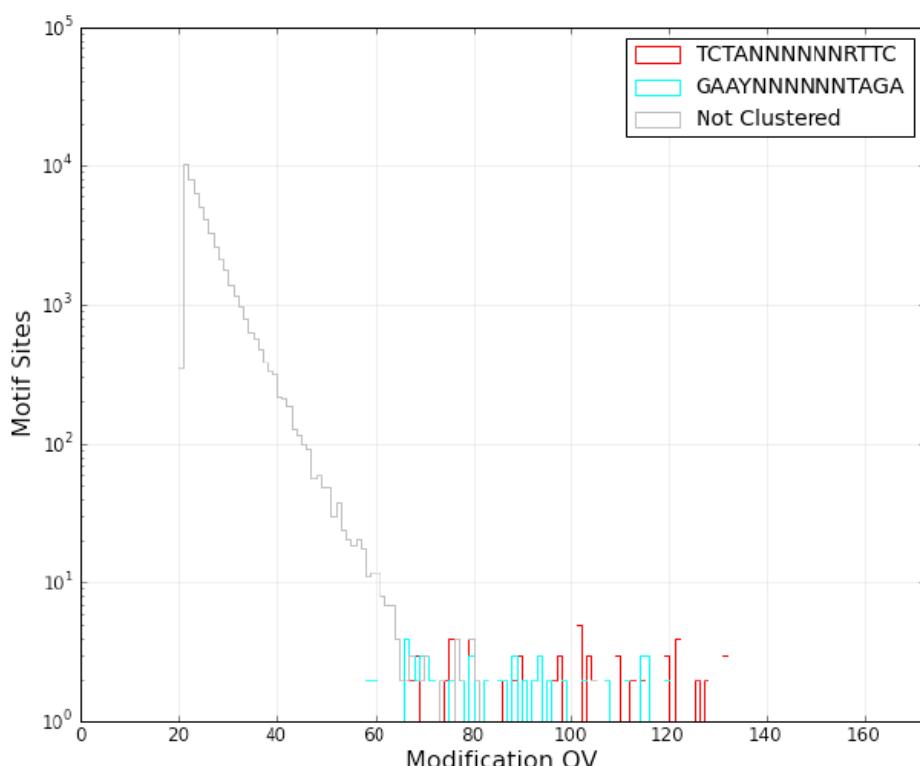
16 MSNTQKKNVPELRFPGEFEWEEKQFADFTKINQGLQIAINERKTEYSPELYFYITNEFLRPNSQT
 17 KYFIENPPQSVIANKE DILMTRTGNTGVVTVNGAFHNNFFKIKFDKNLYDRLFLVEVLNSSKIQ
 18 NKILSLAGSSTIPDLNHSDFYSISSSYPLLREQQKIGKFFSKLDRQIELEEQKLELLQQQKKGYMQ
 19 KIFSQELRFKDENGEDYPDWKEKKLGITEQSMYGIGASATRFDSKNIYIRITDIDEKSRKLNQYQN
 20 LTT PDELNNKYKLKRNDILFARTGASTGKSYIHKEEKDIYNYYFAGFLIKFKINEQNSPLFIYQFT
 21 LTSKFNKWKVMSVRSGQPGINSEYYAKLPLVLPNLEQQKIAKFLDRFDRQIELEKQKIEILQQQ
 22 KKGLLQSMFI
 23
 24

25 Reports for Job Dryden_X_zeta_MODs

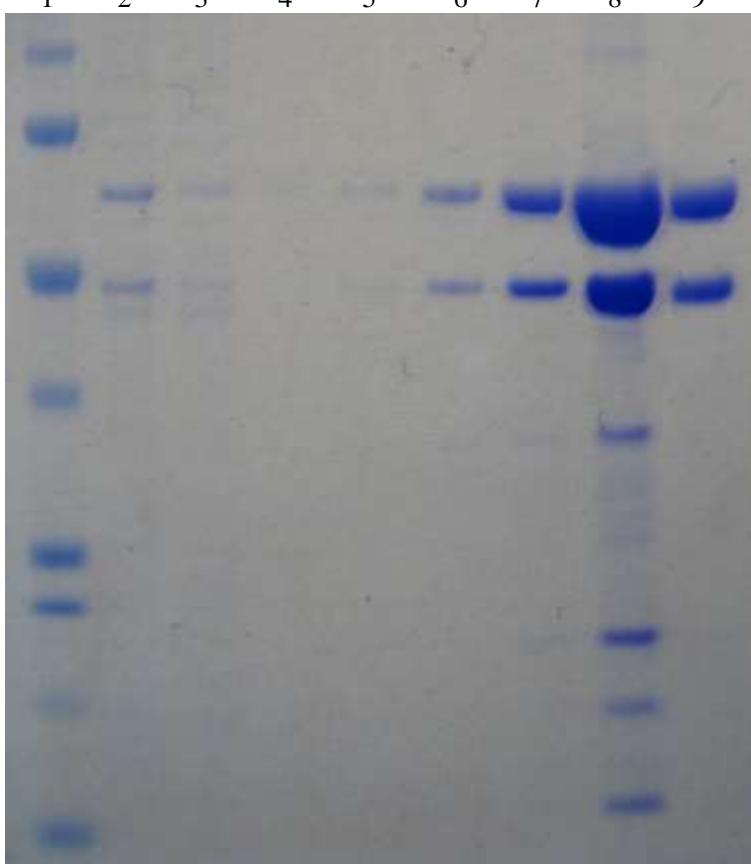


Motif Summary								
Motifs	Modified Position	Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
TCTANNNNNNRTTC	4	m6A	100.0%	92	92	96.27	61.85	GAAYNNNNNNNTAGA
GAAYNNNNNNNTAGA	3	m6A	100.0%	92	92	90.82	60.21	TCTANNNNNNRTTC

33 Modification QVs



1
 2 **S.Saue*D**
 3 CC873
 4 **Recombinant S.Saue*D** **CC873-1** **GAG-6-GAT**
 5
 6 MSNTQKKNVPELRFPGEFEWEEKSISSFLKESKIKGSNGSHAKKLTVKLGKGVVPKKETFKGSD
 7 NTQYYKRKAGQLMYGKLDLNCAGIVPDSLNNYESTIDSPSFDFINGDSKFLLERIKLKSFYKKF
 8 GDIANGSRKAKRINQDTFLSLPVFAPKYDEQLRIGEFFSKLDRQIELEEQKLELLQQQKKGYMQKI
 9 FSQELRFKDENSEDYPHWENSKIEKYLKERNERSDKGQMLSVTINSGIIFKSELDRKDNSSKDKSN
 10 YKVVRKNDIAYNSMRMWQGASGRSNYNGIVSPAYTVLYPTQNTSSLFIGYKFKTHRFMIHKFKINSQ
 11 GLTSDTWNLKYKQLKNINIDIPVLEEQEKGDFKKMDILISKQKIKIEILEKEKQSFQKMFLPG
 12 GSHHHHHH
 13
 14 **Wild Type S.Saue*D**
 15 MSNTQKKNVPELRFPGEFEWEEKSISSFLKESKIKGSNGSHAKKLTVKLGKGVVPKKETFKGSD
 16 NTQYYKRKAGQLMYGKLDLNCAGIVPDSLNNYESTIDSPSFDFINGDSKFLLERIKLKSFYKKF
 17 GDIANGSRKAKRINQDTFLSLPVFAPKYDEQLRIGEFFSKLDRQIELQKQKLELLQQQKKGYMQKI
 18 FSQELRFKDENGEDYPHWENSKIEKYLKERNERSDKGQMLSVTINSGIIFKSELDRKDNSSKDKSN
 19 YKVVRKNDIAYNSMRMWQGASGKSNYNGIVSPAYTVLYPTQNTSSLFIGYKFKTHRFMIHKFKINSQ
 20 GLTSDTWNLKYKQLKNINIDIPVLEEQEKGDFKKMDILISKQKIKIEILEKEKQSFQKMFL*
 21
 22 1 2 3 4 5 6 7 8 9

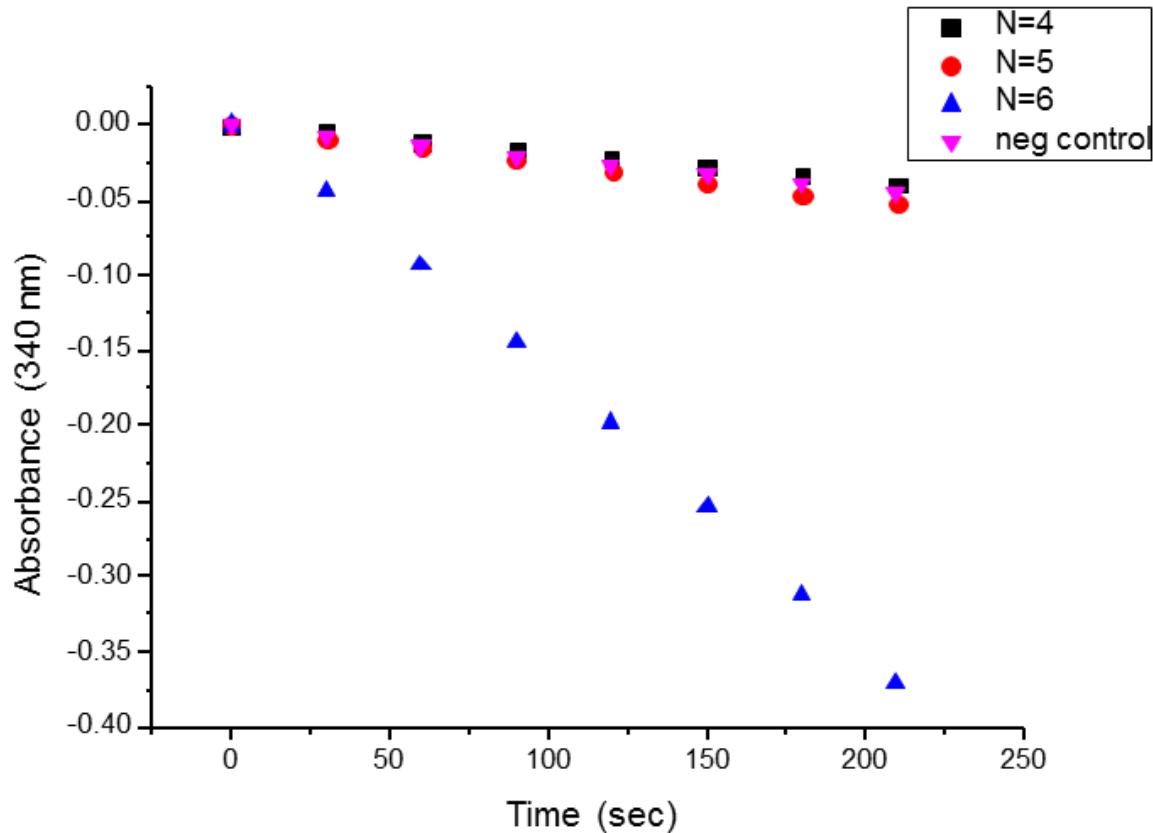


- 50 1- marker 2- soluble cell extract 3- Nickel column flow through
 51 4- Nickel column wash 1 5- Nickel column wash 2 6- Nickel column eluate
 52 7- eluate after PD10 desalting and concentration
 53 8- Final concentrated protein 9- CC398-1 purified protein marker

1
 2 S.Sau~~e~~*D
 3 CC873-1
 4 Recombinant S.Sau~~e~~*D GAG-6-GAT
 5
 6 e*D clearly digests pUC19 so the ATPase assay was used as we knew
 7 the specificities of both TRDs.
 8
 9 Likely site: GAG-N_x-GAT
 10
 11 GAG-4-GAT 2 sites in pUC19
 12
 13 GAG-5-GAT 0 sites in pUC19
 14
 15 GAG-6-GAT 2 sites in pUC19
 16
 17 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	AGATGATGGAATCAATGCGAGTTCCAGATGCCCTATACGATATAA
e*D4rev	TTATATCGTATAGGGCATCTGAACTCGCATTGATTCCATCATCT

N=6 shows activity.



1
2 SMRT results for S. aureus strains LGA251 and NCTC13435
3
4 LGA251

5

6 SMRT® Portal Print

7 Reports for Job Dryden_LGA_Mods PACIFIC BIOSCIENCES®

8

9

10 SMRT Cells: 2 Movies: 2

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

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

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

5

1
2 **SUPPLEMENTARY INFORMATION FOR TABLES 5 AND 6.**
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2 SUPPLEMENTARY INFORMATION FOR TABLES 5 AND 6.
34 By combining all TRD 1 with all TRD 2 amino acid sequences and
5 searching sequence databases, we found that some of our
6 "artificial hybrids" described in Table 3 were actually present in
7 real strains of *S. aureus*. We present several examples below.
8
910 S.SauAU
1112 A plasmid expressing S.SauAU with the M subunit was prepared but
13 not analysed further. The S.SauAU sequence matches that of the S
14 subunit of the Type I RM system in *S. schweitzeri* FSA084.
15
16 >S.SauAU
1718 MSNTQKKNVPELRFPGFEGEWEEKKLGDLTTKIGSGKTPKGGSNEYTNKGIPFLRSQNIRNGKLNL
19 NDLVYISKDIDDEMKNNSRTYYGDVLLNITGASIGRTAINSIVEIHANLNQHVCIIIRLKKEYYNFF
20 GQYLLSRKGKRKIFLAQSGGSREGLNFKEIANLKIFTPTIFEQQKIGEFISKLDRQIELEEQKLE
21 LLQQQKKGYMQKIFSQELRFKDENGEDYPDWEVTIIONITKYTSSKKSSNQYADKDNSKGYPVYDA
22 VQEIGKDSNYDIEESYISILDGAGVGRLNLRPGKSSVIGTMGYIQSNNDIEFLYRMRKVVDFFKK
23 YIIGSTIPHLYFKDYSKETLYIPSSIQEAKIGMFISNLDKLIENKNLKLNLCKQLKQGLLQSMFI
24 PGGSHHHHHH
25
2627 S. schweitzeri FSA084
2829 CLUSTAL O(1.2.1) multiple sequence alignment
30
3132 FSA084 msn-tqkkvpelrfpgfegewekklgevttkigsgktpkggsenytnkgipflrsqnir
33 S.SauAU MSNTQKKNVPELRFPGFEGEWEEKKLGDLTTKIGSGKTPKGGSNEYTNKGIPFLRSQNIR
34 *** :*:*****:*****:*****:*****:*****:*****:*****:*****:*****
35
36 FSA084 ngklnlndlvyiskdiddemknnsrtyygdvllnitgasigrtainsivethanlnqhvci
37 S.SauAU NGKLNLNLDLVYISKDIDDEMKNNSRTYYGDVLLNITGASIGRTAINSIVEIHANLNQHVCI
38 ***:*****:*****:*****:*****:*****:*****:*****:*****:*****
39
40 FSA084 irlkkeyyynffeqyllsrkgkrkiflaqsggsreglnfkeianlkiftstifeeqqkvg
41 S.SauAU IRLKKEYYYNFFGQYLLSRKGKRKIFLAQSGGSREGLNFKEIANLKIFTPTIFEQQKIG
42 ***:*****:*****:*****:*****:*****:*****:*****:*****:*****:
43
44 FSA084 kffskldrqielleqqkkgymqkifsqelrfkdengneypewkvtsiqdvtky
45 S.SauAU EFISKLDRQIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGEDYPDWEVTIIONITKY
46 :*:*****:*****:*****:*****:*****:*****:*****:*****:
47
48 FSA084 tsskkssnqyadkidskgypvydavreigkdsnydieesysisilkdgagvgrlnlrpeks
49 S.SauAU TSSKKSSNQYADKDNSKGYPVYDAVQEIGKDSNYDIEESYISILDGAGVGRLNLRPGKS
50 ***:*****:*****:*****:*****:*****:*****:*****:
51
52 FSA084 svigtmgylqannidleflyyrmkivdfkkyiigstiphlyfkdyketiyipssiqeqa
53 S.SauAU SVIGTMGYIQSNNDIEFLYRMRKVVDFFKKYIIGSTIPHLYFKDYSKETLYIPSSIQEQA
54 ***:*****:*****:*****:*****:*****:*****:
55
56 FSA084 kigkfisnlndkmienktrklnclkqlkqgllqgmfi-----
57 S.SauAU KIGMFISNLDKLIENKNLKLNLCKQLKQGLLQSMFIPGGSHHHHHH
58 ***:*****:*****:*****:*****:
59
60

1
2 **S.SauJE GGA-6-RTGA**
3 **Sub species 21262, a member of ST49**
4

5 CLUSTAL O(1.2.1) multiple sequence alignment
6 TRD R and TRD f* against EHO91218, the second HsdS in this strain.
7

8 CC80-3 -----
9 EHO91218 msntqkknvpelrfpgfegewekklevakiydgthqtpkytnegikflsveniktlns
10 CC72-1 MSNTQKKNVPPELRFPGFEGEWEKKLGEVAKIYDGTHQTPKYTNEGIKFLSVENIKTLNS
11
12 CC80-3 -----
13 EHO91218 skyiseeafekefkirpefgdilmtrigdigtpnivssnekfayyvslallktknlnsyf
14 CC72-1 SKYISEEEAFEKEFKIRPEFGDILMTRIGDIGTPNIVSSNEKFAYYVSLALLKTKNLNSYF
15
16 CC80-3 -----
17 EHO91218 lknlilsssiqnelwrktlhvafpkkinkneigkikinypkkqqkigqffskldrqie
18 CC72-1 LKNLILSSSIQNELWRKTLHVAFPKKINKNEIGKIKINYPKKQQKIGQFFSKLDRQIE
19
20 CC80-3 -----QELRFKDENGEDYPDWKEKKLGDIQEWSMYGIGASATR
21 EHO91218 leeqklellqqqkkgymqkifsqelrfkdengedypdwkekklgditeqsmgigasatr
22 CC72-1 LEEQKLELLQQQKKGYMQKIFS-----
23
24 CC80-3 FDSKNIYIRITDIDEKSRLNYQNLTPDELNNKYKLKRNDILFARTGASTGKSYIHKEE
25 EHO91218 fdskniyiritdideksrlnyqnltpdelnnkyklkrndilfartgastgksyihkee
26 CC72-1 -----
27
28 CC80-3 KDIYNYYFAGFLIKFKINEQNSPLFIYQFTLTTSKFNWKVMSVRSGQPGINSEEYAKLP
29 EHO91218 kdiynyyfagflikfeideqnplfiyqftlttskfnwkvmsvrsgqpginseeyaklp
30 CC72-1 -----
31
32 CC80-3 LVLPNKLEQQKIAKFLDRFDRQIELEKQKIEILQQQKKGLLQSMFI
33 EHO91218 lvpnkialeqqkiaeefldrfdqqielekqkieilqqqkkgllqsmfi
34 CC72-1 -----
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
 2 **S.SauJE GGA-6-RTGA**
 3 **S.SauJE against ST49 strain "Tager 104"**
 4 The ST49 Tager genome has the same TRD combinations as the ST49
 5 strain 21262.
 6
 7

8 **PATRIC db**

9 >fig|1381115.3.peg.1063|VBIStaAur301678_1063| Type I restriction-modification
 10 system, specificity subunit S (EC 3.1.21.3) [Staphylococcus aureus subsp.
 11 aureus Tager 104 | 1381115.3] This is TRD R+f*
 12 MSNTQKKNVPELRFPGEFEWEEKKLGEVAKIYDGTHQTPKYTNNEGIKFLSVENIKTLNS
 13 SKYISEEAFEKEFKIRPEFGDILMTRIGDIGTPNIVSSNEKFAYYVSLALLTKNLNSYF
 14 LKNLILSSSIQNELWRKTLHVAFPKKINKNEIGKIKINYPKKQEQQKIGQFFSKLDRQIE
 15 LEEQKLELLQQQKKGYMQKIFSQELRFKDENGEDYPDWKEKKLDITEQSMYGIGASATR
 16 FDSKNIYIIRITDIDEKSRSKLNQYQNLTTPDELNNKYKLKRNDILFARTGASTGKSYIHKEE
 17 KDIYNYYFAGFLIKFEIDEQNNPLFIYQFTLTSKFNKWVKVMSVRSGQPGINSEYYAKLP
 18 LVLPNKLEQQKIAEFLDRFDQQIELEKQKIEILQQQKKGLLQSMFI
 19 >fig|1381115.3.peg.2628|VBIStaAur301678_2628| Type I restriction-modification
 20 system, specificity subunit S (EC 3.1.21.3) [Staphylococcus aureus subsp.
 21 aureus Tager 104 | 1381115.3] This is TRD J+E
 22 MSNTQKKNVPELRFPGEFEWEEKKLEDIIKVNSGKDYKHLKDGDIPVYGTGGYMTSVSE
 23 PLSEIDAVGIGRKGTINKPYLLEAPFWTVDTLFYCTPKKETDILFILSLFRKINWKVYDE
 24 STGVPSLSKQTINKINRFVPTNKEQQKIGKFFSKLDRQIELEEQKIELLQQQKKGYIQKI
 25 FSQELRFKDENGDDYPEWEETTIQEIAQINTGKDKTDKDAITNGSYDFYVRSPIVYKINTF
 26 SYEGEAILTVGDGVGVGVFHYVNGKFDYHQRVYKISDFKNYYGLLFYYFSQNFLKETK
 27 KYSAKTSVDSVRKDMVANMKVPRPIYIEQEKIGQFIKKVDNKIKIQKQVIELLKQRKKAL
 28 LQKMF
 29
 30

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

32 CLUSTAL O(1.2.1) multiple sequence alignment

33
 34
 35 S.SauJE
 36 fig|1381115.3.peg.2628|VBIStaAur301678_2628|
 37
 38 S.SauJE
 39 fig|1381115.3.peg.2628|VBIStaAur301678_2628|
 40
 41 S.SauJE
 42 fig|1381115.3.peg.2628|VBIStaAur301678_2628|
 43
 44 S.SauJE
 45 fig|1381115.3.peg.2628|VBIStaAur301678_2628|
 46
 47 S.SauJE
 48 fig|1381115.3.peg.2628|VBIStaAur301678_2628|
 49
 50 S.SauJE
 51 fig|1381115.3.peg.2628|VBIStaAur301678_2628|
 52
 53 S.SauJE
 54 fig|1381115.3.peg.2628|VBIStaAur301678_2628|
 55
 56
 57
 58
 59
 60

MSNTQKKNVPELRFPGEFEWEEKKLGDLIKVNNSGKDYKHLKDGDIPVYGTGGYMTSVSE
 MSNTQKKNVPELRFPGEFEWEEKKLEDIIKVNSGKDYKHLKDGDIPVYGTGGYMTSVSE
 *****:*****:*****:
 PLSEIDAVGIGRKGTINKPYLLEAPFWTVDTLFYCTPKKETDILFILSLFRKINWKVYDE
 PLSEIDAVGIGRKGTINKPYLLEAPFWTVDTLFYCTPKKETDILFILSLFRKINWKVYDE
 *****:
 STGVPSLSKQTINKINRFVPSNKEQQKIGEFFIKLDRQIELEEQKIELLQQQKKGYMQKI
 STGVPSLSKQTINKINRFVPTNKEQQKIGKFFSKLDRQIELEEQKIELLQQQKKGYIQKI
 *****:*****:*****:
 FSQELRFKDENGDDYPEWEETTIQEIAQINTGKDKTDKDAITNGSYDFYVRSPIVYKINTF
 FSQELRFKDENGDDYPEWEETTIQEIAQINTGKDKTDKDAITNGSYDFYVRSPIVYKINTF
 *****:
 SYEGEAILTVGDGVGVGVFHYVNGKFDYHQRVYKISDFKNYYGLLFYYFSQNFLKETK
 SYEGEAILTVGDGVGVGVFHYVNGKFDYHQRVYKISDFKNYYGLLFYYFSQNFLKETK
 *****:
 KYSAKTSVDSVRKDMVANMKVPRPIYIEQEKIGQFIKKVDNKIKIQKQVIELLKQRKKSL
 KYSAKTSVDSVRKDMVANMKVPRPIYIEQEKIGQFIKKVDNKIKIQKQVIELLKQRKKAL
 *****:*****:*****:
 LQKMFIPGGSHHHHHH
 LQKMF-----

1 **S.SauNQ ACC-5-RTGT**
2
3 This TRD pair was found in strains KPL1845 (ST96) and 21343 (ST88).
4 Subspecies 21343 contains SauNQ and a novel TRD (NOVEL 1) paired
5 with TRD K.
6
78 **>EHQ67679 THIS IS TRD NOVEL 1 + TRD K**
9
10 MSNTQKKNVP~~ELRF~~PGFE~~GEEK~~WEEKKLGEVATFAKGKLGAKKDVSQNGVPVILYGELYTKY
11 GAIVSKIFS~~KTDI~~PENKLMAKKNDVLIPSSGETAIDIATASCIYLNGVAVGGDINILT
12 PQKQDGRFISLSINGINKNELSKYAQGKTVVHLYNNNDIKNLKIAFPSEEEQVRIGNFFS
13 KLDRQIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYPKWEKKIEDIASQVYGG
14 GTPNTKIKEFWNGDIPWIQSSDVKN~~VNDL~~LILQQCNKFISKNSIELSSAKLIPANSIAIVTR
15 VGVGKLC~~VE~~F~~DY~~ATSQDFSLSSLKYDKLYSLLYTMKKISANLQGTSIKGITKKEL
16 LDSIIKIPH~~N~~LEEQQKIGDLFYKIDKYISFNCKKIEILKSLQGLLKKMFI17 **>EHQ71248 THIS IS TRD N+Q ACC-5-RTGT**
18
19 MSNTQTKNVP~~ELKF~~PE~~FE~~GE~~EEK~~WEEKKLGEFAGKVTKKNVDKKYIETLTNSAELGIISQKDY
20 FDKEISNIDNIKKYYVVEENDFVYNPRISNYAPFGPVNRNKL~~G~~KKGVMSPLYTVFKIQNI
21 DLNFIEFYFKSSKWYRFMALNGDSGARADRF~~S~~IKNRTFMEMPLH~~I~~PCMDEQIKIGQFFSK
22 LDRQIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYP~~E~~WEERRFADIFKFHNKLR
23 KPIKENLRVKGSY~~P~~YYGATGIIDYVDDFIFDGNYLLIGEDGANIITRSAPLVYL~~V~~NGKFW
24 VNNHAHILSPLNGNIQYLYQVAELVNYEK~~T~~GAQPKLN~~I~~QNLKIISVVISTNLEEQQK
25 IGSFLSKLDRQIDLEEQKLELLQQRKALLKSMFV26 **SPECIES KPL1845 CONTAINS THREE *Sau1* S SUBUNITS.**27 **>ETD06224 THIS IS TRD N+Q ACC-5-RTGT**
28
29 MSNTQTKNVP~~ELKF~~PE~~FE~~GE~~EEK~~WEEKKLGEFAGKVTKKNVDKKYIETLTNSAELGIISQKDY
30 FDKEISNIDNIKKYYVVEENDFVYNPRISNYAPFGPVNRNKL~~G~~KKGVMSPLYTVFKIQNI
31 DLNFIEFYFKSSKWYRFMALNGDSGARADRF~~S~~IKNRTFMEMPLH~~I~~PCMDEQIKIGQFFSK
32 LDRQIELEEQKLELLQQQKKGYMQKIFSQELRFKDENGNDYP~~E~~WEERRFADIFKFHNKLR
33 KPIKENLRVKGSY~~P~~YYGATGIIDYVDDFIFDGNYLLIGEDGANIITRSAPLVYL~~V~~NGKFW
34 VNNHAHILSPLNGNIQYLYQVAELVNYEK~~T~~GAQPKLN~~I~~QNLKIISVVISTNLEEQQK
35 IGSFLSKLDRQIDLEEQKLELLQQRKALLKSMFV36 **>ETD11204 THIS HAS TWO NOVEL TRDS, NOVEL 2 + NOVEL 3.**37 MTEQINTPELR~~F~~PEFKNEWSYDLVSDVVTNKS~~K~~KFDPK~~KEEAKK~~DIELDSIEQNTGRLLD
38 TYISNDFTSQKNFKFNKG~~N~~VLYSKL~~R~~PYLN~~K~~YYATIDGVC~~S~~SEIWL~~N~~TLNKDV~~A~~NKFL
39 YYFIQTNR~~F~~SSVTNKSAGSKMPRAD~~W~~ELVKNIRLYKGSIEEQEKIGYFFSKLDRQIELEE
40 KKLELLEQQQKKGYMQKIFAQELRFKDENGNDYPDWVTKKL~~G~~DIGKVAMNKRIYKNETTEN
41 GEIPFYKIGNFGKNADTFITREKFDEYKEKYP~~P~~VGDI~~L~~ISASGSIGRTIEYTGEDAYY
42 QDSNIVWLHN~~H~~DEVINKYLKYFYKIVKWSGIEGTTIKRLYNKN~~I~~LN~~T~~KIELPTVEEQYKM
43 ANFLSKL~~D~~KI~~I~~DIQIEKIELLKQRKQGLLQKMFV44 **>ETD09130 THIS HAS A NOVEL TRD (NOVEL 4) PAIRED WITH TRD f***45 1MSNTQKKNVP~~ELRF~~PE~~FE~~GE~~EWK~~DVKFVS~~I~~FQEVSNK~~T~~SDL~~A~~KYPL~~F~~SLTVEKG~~I~~TPKTER
46 61YKRDFLVKKSDNF~~K~~IVEPRDIVNPMNVTLGAI~~D~~LSKY~~N~~YDIALSGYYHVMK~~I~~INSFNP~~D~~
47 121FISNFLKTEKMI~~I~~HYKKIATGSLMEKQRVHFSEFKNI~~I~~KKFPTNKEQQKIGDFFSKLDRQ
48 181IELQVQKLELLQQQKKGYMQKIFSQELRFKDENGEDY~~P~~DWKEKKLGDITEQSMY~~G~~IGASA
49 241TRFDSK~~N~~IYIRITDIDEKSRKLN~~Y~~QNL~~T~~TPDELNN~~K~~YKL~~R~~NDILFARTGASTG~~K~~SYIH~~K~~
50 301EEKDIYNYY~~F~~AGFLIKF~~E~~IDEQNNPL~~F~~IYQFTLTSKF~~N~~KWVKVMSVRSGQPG~~I~~NS~~E~~YAK
51 361LPLVLPN~~K~~LEQQKIAEFLDRFDQQIELEKQKIEILQQQKKGLLQSMF~~I~~

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:		5	
NOVEL_4_189	57 RLLDTY----ISNDFTSQKNKFNKGNVLYSKLRPY---LNKYYYATI---DGVCSSIEIWVLNTLNK-D	113	
Z_GAC_191	59 RLIKIY----NSKEFSSQKNKFNPQNVLGYKLRLPY---LNKYYFTKK---SGVCSSEIWVLKSTKE-D	115	
NOVEL2_194	59 ERYKRDFL---VVKSDNFK1VEPRDIVYNPMNVT---LGAILDSLKYN---YDIALSGYHVMIKIN---	115	
NOVEL1_199	62 AIVSK--IFS-KTDIPENKLKMAKKNDVLIPSSGETAIDIATASCIYLN---KGVAVGDDINILTPQ---	122	
R_GARA_192	61 SK----YIS--EEAFKEKFIRPEFGDILMTRIGDI---GTPNIVSSN---EKFAYYVSALLKTK---	114	
J_GGA_172	58 -----VSEPLSEIDAVGIRGKTI---NKPYLEA---PFWTVDLTYFCPTPEK---	99	
N_ACC_198	66 S-----NIDNIKKYYVVEENFVYNNPRMSNYAPFGVPNRNLKG---KGKVMSPLYTVFKIQ---	118	
O_CAAC_195	64 ENITN--FIT-EKGLENNESSAKLITNEALIAMIYGGK---TRGMSA1LNF---EATTNQACAIYQ----T	120	
T_CAAAG_199	65 THSKE---KIT-EYAMKSLKLKVPKNSVAMYGGFN---QI1GRTGLLK---IDATINQACASALMNH---	123	
C_GWAG_206	61 INTNNTLGKV--VNVSKEKLKNYSVEKGDFVFTTRTSVEIGEYGVSLNPD-ENTVFSGFVLGRPKSGID	128	
M_CAG_203	62 TVLDSD--GN1-PNIIEKAVFELIQKGDIVFADASEDYSDLGKAVMIDFEP-NSLISGLHTHLFRPLN---	125	
X_TCTA_192	65 QTKYF----IENP PQSVITANKEDILMTRTG---TGKVVTNV---FGAFHNNFFKIKBEDFKN---	115	
B_AGG_199	66 SSK-----NLENYTLIKNGEFAYNKSYNSGYPLGAIKRLTRY---DGSVLSSLYICFSIKS--	118	
A_CCAY_203	65 NLNDLV-YIS-KDIDDEMKNSRTYYGVDLNTITGAS---IGRTAINSIVE-THANLNQHVC1IRLKK--	125	
e*__GAG_190	62 F-----KGSDNTQYYRKAGQLMYGKLDL--NCAFGIVPDS--LNNYESTIDSPSPDFI--	112	
V_CNGA_210	71 YNSNKV-FTS-NEKAEVLKSCNVFPGDIVIAKMDP---IARA1AIPDPNNIGKYLMASD GIRLS VDT--V	133	
b*__GGHA_200	62 IISSDDRKISIDESDYKK1YKKNYKLEKGDDLLLTIVGTI---GRAAIVKNP--NNIAFQRSSVAILTKA--	122	
Consensus ss:	e ee eeeee eeeeeee eeee eeeee		

Conservation:		9	5	5	99	799	9999898998
NOVEL_4_189	VLANKFLYYFIQTNRFSS-VTNKSAG---	SKMPRADWELVKNIRLYKGS-IEEQEKIGYFFSKLDRQIE	177				
Z_GAC_191	KLNNLFLLYYFIQTKRYS-DVASKSAG---	SKMPRADWGLIENIRYVFPE-LCEQQKIGQFFSKLDRQIE	179				
NOVEL2_194	SFNPDFISNFLKTEKMIHYKKIATGS-	LMEQKQRVHFSEFKNI1KKFPT-NKEQQKIGDFFSKLDRQIE	182				
NOVEL1_199	KQDGRFISLSINGI-NKNELSKYAQG--	KTVVHLYNNNDIKNLKIAFPSEEEQVRIGNFFSKLDRQIE	187				
R_GARA_192	NLNLSYFLKLNLILSSSIQNELWRKTlhV--AFFPKINKNEGIKIKINYPK-KQEQQKIGQFFSKLDRQIE	180					
J_GGA_172	EADILFPLSFRLKRINWKL---YDES--	TGVPSLSKQTINKNIRLVP--NKEQQKIGEFFF SKLDRQIE	160				
N_ACC_198	IDLNLNFIEFFYFKSSWYRFLMALNGDSDA-RADFRSKLDRFTMEMPMLHIPC-MDEQKIQGFFSKLDRQIE	186					
O_CAAC_195	NQNINVFQYFQK--LYEFLRSLSNE--	GSQKNLNSLSSLKEITLNPN-EQEQQKIGDFFSKLDRQIE	183				
T_CAAAG_199	ETNPEIFIQAFLNQY-VKGWKRKYAASS--	RKDPNITKDDIEQFKPVPS-INEQOKIGEFFSKLDRQIE	187				
C_GWAG_206	LINNNFKRYYVFTNSRKEMITKSSM--	TTRALTSGTAINMKV1YIPVAKEQQKIGDFFSKLDRQIE	194				
M_CAG_203	NAISNLFIFYTKTLSYKKFIRQQCTG--	ISVLGISKSSKLNLNLNVLIPRSELEQQKIGQFFSKLDRQIE	191				
X_TCTA_192	LYDRFLVEVFLNNSKIQNKILSLAGS--	STIPDLNHSDFYISSYPL-LREQQKIGKFFSKLDRQIE	180				
B_AGG_199	EMSJKDFMEAYFDSTHWYREVSGIAVEGARNHGLLNVNSVNDKFLTLYKPS-LEEQQKIGKFFSKLDRQIE	187					
A_CCAY_203	EYYYIFFGQYLLSRGKRKIFLAQSG--	GSREGNLNFEIANLKITPTTIEEQQQKIGKFFSKLDRQIE	191				
e* _GAG_190	NGDSKFLLERIKLKSFYKKFGDIANGS--	RKAKRINQDTFLSLPWFAPK-YDEQLRIGEFFSKLDRQIE	178				
V_CNGA_210	HFTNTKFVLECINRKSFRKVEDNSSG--	STRMRIGLSTLGSLLTKTTT-LKEQQKIGQFFSKLDRQIV	198				
b* _GGHA_200	TYDVGIFIFQDFTQYFKNLLLRQKVVV--	SAQPGYLGDIRKIKISITNIIEEQRKIGEFFSKLDRQIE	188				
Consensus ss:	hhhhhhhhh bhhhhhhhhh	hhhhh ee bhhhhhhhhh bhhhhhhh					

Conservation:	977899999899
NOVEL_4_189_	178 LEEKKLELLEQQ 189
Z_GAC_191_	180 LEEQKLELLQQQ 191
NOVEL2_194	183 LQVQKLELLLQQQ 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_CAAAG_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

1 PROMALS alignment of all second TRDs.

2 Conservation: 9987999979999989696797 9 5

3 NOVEL3_205 1 KKGYMQK1FAQELRFKDENGNDYPDWTKLGLDIGKVAMNKRIYKNE-----TTENGEIPFYKIGNFG 63

4 S_GCA_200 1 KKGYMQK1FSQELRFKDENGNDYPDWTRNLGEVTTVTMQSPKSVN----YTDNSNDTVLIQGNADIE 65

5 d*_CYAA_220 1 KKGYMQK1FSQELRFKDENGNDYPDWTRNLGEVTTVTMQSPKSVN----YTDNSNDTVLIQGNADIE 65

6 a*_GAA_208 1 KKGYMQK1FSQELRFKDENGNDYPDWTRNLGEVTTVTMQSPKSVN----YTDNSNDTVLIQGNADIE 65

7 E_TCAY_194 1 KKGYMQK1FSQELRFKDENGNDYPDWTRNLGEVTTVTMQSPKSVN----YTDNSNDTVLIQGNADIE 65

8 W_CRAA_211 1 KKGYMQK1FSQELRFKDENGNDYPDWTRNLGEVTTVTMQSPKSVN----YTDNSNDTVLIQGNADIE 65

9 Q_ACAY_197 1 KKGYMQK1FSQELRFKDENGNDYPDWTRNLGEVTTVTMQSPKSVN----YTDNSNDTVLIQGNADIE 65

10 G ACA_196 1 KKGYMQK1FSQELRFKDENGNDYPDWTRNLGEVTTVTMQSPKSVN----YTDNSNDTVLIQGNADIE 65

11 f*_GAAY_224 1 KKGYMQK1FSQELRFKDENGNDYPDWTRNLGEVTTVTMQSPKSVN----YTDNSNDTVLIQGNADIE 65

12 L_TTA_213 1 KKGYMQK1FSQELRFKDENGNDYPDWTRNLGEVTTVTMQSPKSVN----YTDNSNDTVLIQGNADIE 65

13 Y_CTA_209 1 KKGYMQK1FSQELRFKDENGNDYPDWTRNLGEVTTVTMQSPKSVN----YTDNSNDTVLIQGNADIE 65

14 U_GAY_193 1 KKGYMQK1FSQELRFKDENGNDYPDWTRNLGEVTTVTMQSPKSVN----YTDNSNDTVLIQGNADIE 65

15 I_YTCA_220 1 KKGYMQK1FSQELRFKDENGNDYPDWTRNLGEVTTVTMQSPKSVN----YTDNSNDTVLIQGNADIE 65

16 K_CGA_212 1 KKGYMQK1FSQELRFKDENGNDYPDWTRNLGEVTTVTMQSPKSVN----YTDNSNDTVLIQGNADIE 65

17 D_ATC_204 1 KKGYMQK1FSQELRFKDENGNDYPDWTRNLGEVTTVTMQSPKSVN----YTDNSNDTVLIQGNADIE 65

18 c*_GAY_209 1 KKGYLQK1FSQELRFKDENGNDYPDWTRNLGEVTTVTMQSPKSVN----YTDNSNDTVLIQGNADIE 65

19 P_AGG_214 1 KKGYMQK1FSQELRFKDENGNDYPDWTRNLGEVTTVTMQSPKSVN----YTDNSNDTVLIQGNADIE 65

20 F_TAA_216 1 KKGYMQK1FSQELRFKDENGNDYPDWTRNLGEVTTVTMQSPKSVN----YTDNSNDTVLIQGNADIE 65

21 H_TAC_206 1 KKGYMQK1FSQELRFKDENGNDYPDWTRNLGEVTTVTMQSPKSVN----YTDNSNDTVLIQGNADIE 65

22 Consensus_ss: hhhhhh eeeeeeeeeeee eeeeeeeeeeee

23 Conservation: 5 7

24 NOVEL3_205 64 KNADTFITR--EKF---DEYKEKYPNVGD-TLISASG----SIGRTIETYTG--EDA-YYQDSNIV 117

25 S_GCA_200 66 NGLINPR-----YTREVTKLIQKDE-IIITVRA----PVGKlamaQin----ACIGRGVC 112

26 d*_CYAA_220 71 YDISNFRYY-INEN---KYKEMQSFVQPND-IIMSCSG----TIGRLALIPH-NYTK-GINQALI 126

27 a*_GAA_208 66 QKYL-KGN-KGIS---KDAAKNYMKVKNDT-LIMSFKL---TIGKLAIVKAP----LYTNEAIC 117

28 E_TCAY_194 63 YK-----INTFSYEGEAILTVGDGV----GVGFVPHVNNGK---FDYHQRVY 102

29 W_CRAA_211 68 NONGKIHYLEQKLS---IEGQEKTTRVLVTTH-LLSIAA----SIGKPVMNFVK---TGVHDGFL 121

30 Q_ACAY_197 65 DYV-----DDFIFDGNY-LLIGEDGANITRASPLVLYVNGK---FWNNNAH 108

31 G ACA_196 65 DYV-----KDYLNFNEERLLIGEDGAK-WGQFETSSFIANGQ---YWVNNNAH 108

32 f*_GAAY_224 63 EKSRKLNYQ-NLTT---PDELNNKYLKRND-ILFARTGA---STGKSYIHKEEKDIYNNYFAGFLI 121

33 L_TTA_213 64 YYRPAYIKV-SKFVS-ENTYNNWFREHLKEND-ILFSTVG---NTGIVSLMDNYK---AVIAQNIV 120

34 Y_CTA_209 66 NSKNIYESE-NKLT---QEGYKKARQLPENT-LLVTCIA----SIGKNAILRKQ---GSCNQQIN 118

35 U_GAY_193 66 GK-----DSNYDIEESY-ISILKDGA---GVGRNLNRPGK---SSVIGTMG 104

36 I_YTCA_220 68 KGYIDFNVE-AKGNLNDLYTRWMRGNELYKGQ-VLFTEA---PMGNVAQVPDNKG---YILSQRTI 126

37 K_CGA_212 66 VNDLILQQCNKFISK---NSIELSSAKLIPANS-IAIVTRV---GVGKLCVLEFD---YATSQDFL 121

38 D_ATC_204 57 IKFSELDR-KDNS---SKNKSNYKVVVRKND-IAYNSMRM---WQGASGKSNNY---GIVSPAYT 109

39 c*_GAY_209 60 KGEK-ANI-NRVL---KLGATNYKKRFEQG-FIYKGQNF---FNGAFDIVPKKFDG--LYSSSDVP 115

40 P_AGG_214 67 SKSVS---SKNLENLYLIKNGE-FAYNKSYSN---GYPGLAIKRLTRYDS---GVLSSLYI 117

41 F_TAA_216 63 INSTYNDQN-IRVN---KNKKTTEKYIILSKGD-LAMVNDKTKDGKIIIGRSIFIDKDNQ---YIYNQRTE 123

42 H_TAC_206 58 LILQSDYYKDRKTF---AESNIGYFILPKNH-ITYRSRS---DDGIFKFNLNLMDV-GIISKYYP 115

43 Consensus_ss: ee eeeee eeeeeee eeee ee

44 Conservation: 6 5 7 98 76 7

45 NOVEL3_205 118 WLHNND-EVINKYLKYFYKI---VKWSGIEG---TTIKRLYNKNILNTPKIEPT-VEEQYKMANFLS 176

46 S_GCA_200 113 SIKG-----DKDFLYFLWATQNKWIRFSQG---SEFTESISGNDIRNIH1KIPV-EDERTKIIKLLN 171

47 d*_CYAA_220 127 RFRTNH-KIRSEFFLIFMRSNQMRKILEANPG---SAITNLVPVKEKLIPFPLPV-KFEQDKISQFIH 191

48 a*_GAA_208 118 HF1WKVNKINTEPIYYLNSL---NISTFGVQA---VKGVTLNNDNSINSIIVKLPN-EEQNQIAKFL 179

49 E_TCAY_194 103 KISDFK-NYYGLLLFFYFSQN-FLKETKKYSAK---TSVDSVRKDMIANMKVPRPI-YIEQKKIGQFIK 165

50 W_CRAA_211 122 IFLKPK-FKNFLFMMWYLEYF---KDKWNSKYQGP---GSQVNLNSEIJKSQTLMNPS-NHEQERVKQFFF 182

51 Q_ACAY_197 109 ILSPL---NGNIQYLYQVAEL---VNVEKYNTG---TAQPKLNIQNLKIINVVISTNLEEQQKIGSFLS 168

52 G ACA_196 109 VVKSN---DHNLLFFMNYYLNF---KELRAFVTG---NAPAKLTHANCLNINLKIIPC-LTEQDKVSAALLK 167

53 f*_GAAY_224 122 KFKINE-QNSPLFLYQFTLTSKFNKWKVMSVR---SGQPGINSEYAKLPLVLPN-KLEQQKIAKFLD 185

54 L_TTA_213 121 GLRVVN-NNLPLPSIYIYMLSYKGQNQKKIARIQMG---AVQPSVKVSQFKFIKYLVPI-KDEQEKVAKLLI 184

55 Y_CTA_209 119 AVVPFE-NINIDLYLISDL---STFMKSIAGK---TATQIVNKNTFENLEIYIAP-FEEQNKIAIDLIS 180

56 U_GAY_193 105 YIQSN---NVDIEFLYRMMKVV---DFKYYIIG---STIPHLYFKDYSKETLYIPLLSEQAKIGMFIS 164

57 I_YTCA_220 127 AFNSNE-KITDNFLASSLSSENVYNDLLKLCG---ATAKGVSKQNLNRLYVTIIPHISSEQEEIAEFFR 191

58 K_CGA_212 122 SLSLLS---YDKLFLYKFTKMHFKKINSQGL-TSDTWNLKYKQLKNINIDIPV-LEEQEKGIDFFK 175

59 D_ATC_204 110 VLYPTQ-NTSSLFIGYKFKTHRMKIKFKKINSQGL-TSDTWNLKYKQLKNINIDIPV-LEEQEKGIDFFK 180

60 c*_GAY_209 116 AFEINTEKIPNFIYSLRSPFSYKSKKEKYSTG---TGSKRHENVTLNFSLHLP-C-LNEQQLKIASFVC 180

61 P_AGG_214 118 CFSIKS-EMSKDFMEAYFDSTHWYREVSGIAVEGARNHGLLNISVNDFFTILIKYPS-LEEQRKIGDFFI 185

62 F_TAA_216 124 RLIPFA-ENDNKFWLFLMTNDLIRNKKIGMMQG---ATQVYINYSSIKLISIQPL-LEEQQKIRGFLE 187

63 H_TAC_206 116 VFKG---DANQYYLTLHNLNYQ-LKKEYIYKATG---TSQLVLSQKDLQNIKTKLPS-YEEQQKIGDFFS 177

64 Consensus_ss: eee hhhhhhhh hhhhhhhh hhhh ee hhhhhhhhhh

65 Conservation: 5 75 56 97 697 786

66 NOVEL3_205 177 KLDKIIDQIEKIELLQKQGQLLQKMFV-----205

67 S_GCA_200 172 SLDVLNSKTDLKIQNLKQRKQSLQKIFV-----200

68 d*_CYAA_220 192 IINRRIESEKKEIRESLKNRQKQFLQKLFLV-----220

69 a*_GAA_208 180 EVDKTVNNQLVTKTLLKQRKQGQLLQRMFV-----208

70 E_TCAY_194 166 RVDNKTQKQVIELLQKQKALLQKMF-----194

71 W_CRAA_211 183 RNEKLIELQQECKIMYIKRKCKQVLLQKMF-----211

72 Q_ACAY_197 169 KLDRQIDLEEQKLELLQKQKALLKSMFV-----197

73 G ACA_196 168 SIDNMKNMNQMRRIELLKERKQKALLQKMF-----196

74 f*_GAAY_224 186 RFDRQIELEKQKIEILLQQKQKALLQKMF-----214

75 L_TTA_213 185 EIDKLVNQLKQKALLQKQKALLQKMF-----213

76 Y_CTA_209 181 SLEELIKEQASKLQKMKRSRQGMLQIMF-----209

77 U_GAY_193 165 NLDKLIENKNLKLNCQKQGQLLQSMFV-----193

78 I_YTCA_220 192 KINQVELQKYKIEHTKQSQKQVFLQKMF-----220

79 K_CGA_212 184 KIDKYISFNCKCIEMLKSLQKQGQLLKQMF-----212

80 D_ATC_204 176 KMIDLISKQKIKIELEKEKQSFQKQMF-----204

81 c*_GAY_209 181 FLNRKIELLERKQYIYLKKQKQALLQQMF-----209

82 P_AGG_214 186 KLDRQIELEQKLELLQKQKALLQKMF-----214

83 F_TAA_216 188 VLSGITTKQLHKIDQLKERKAFLQKMF-----216

84 H_TAC_206 178 EIDRLVEKQSSKVGRLKVRKKELLQKMF-----206

85 Consensus_ss: hhhhhhhhhhhhhhhhhhhhhhhhhhhhh

1
 2 PROMALS alignment of all TRDs.
 3 Conservation: 799997576 6 85 5
 4 CC80-3 f* 1 -----QELRFKDENGEDEDYPDWKEKKLGDIQEWSMYGIGASA-----TRFDSKNIYIRITDI 51
 5 CC45-1L 1 -----QELRFKDENGNNDYPNWRTIELKNILENIVDNRGKTP-----DNAPSEKYPLLEVNAL 52
 6 CC97 c* 1 -----QELRFKDENGNNDYPEWRFARFKDFMYKPINIRPAIN-----ISKSELLTVKLHCK-GI 52
 7 CC22-1I 1 -----QELRFKNEENGNDYPDWERIKFVDIKVIFRGRTPKK-----LNMEWSDEGYLALSAVN 56
 8 CC873D 1 -----QELRFKDENGEDEYPHWENSKEKYLKERNERSDKGQM-----LSVTIN--SGIJKFSEL 52
 9 CC5-1D 1 -----QELRFKDENGEDEYPHWENSKEKYLKERNERSDKGQM-----LSVTIN--SGIJKFSEL 52
 10 CC30-1D 1 -----QELRFKDENGEDEYPHWENSKEKYLKERNERSDKGQM-----LSVTIN--SGIJKFSEL 52
 11 CC5-2H 1 -----QELRFKDEEGNNYYKGWNKKQLKDVELEFSNKRTE-----NEYPVLTSSRQ 46
 12 CC133-2fromED133 d* 1 -----QELRFKDENGNNDYPWENVMQLQKVLDKTEGIKRGPFGG-ALKKIDIFVESGYAVYEQRNA 59
 13 CC72-2S 1 -----QELRFKDENGNNDYPDWNTNERLGEVITVTMQQSPKSVN-----YTDNSNDTQLQGNADI 54
 14 CC93-3 a* 1 -----QELRFKDENGNNDYPEWENKRIEDIANVNKGFTPSTNN-----NEYWDNNDKNWLISAGM 54
 15 CC93-2K 1 -----QELRFKDENGNNDYPKWEKKIKIEDIASQVYGGGTPTK-----IKEFWNGDIPWIQSSDV 54
 16 CC30-2K 1 -----QELRFKDENGNNDYPNWEKKIEEDIASQVYGGGTPTK-----IKEFWNGDIPWIQSSDV 54
 17 CC80-2W 1 -----QELRFKDENGNNDYPDWEKKQLGELSQIVRGASPPIKD-----PKWFNKESEDIGWLRISDV 56
 18 CC75-2W 1 -----QELRFKDENGNNDYPDWEKKQLGELSQIVRGASPPIKD-----PKWFNKESEDIGWLRISDV 56
 19 CC59Q 1 -----QELRFKDENGEDEYSEWERRFAIFKFHNKLRPIK-----ENLRVKGSYPYVGATGI 53
 20 CC72-1Q 1 -----QELRFKDENGNNDYPEWERRFAIFKFHNKLRPIK-----ENLRVKGSYPYVGATGI 53
 21 CC1-2G 1 -----QELRFKDENGEYPEWEWKFIKDFIFENRRPKIT-----SSLREKGLYPYYGATGI 53
 22 ST425-1E 1 -----QELRFKDENGNNDYPEWEETTIKEIAQINTGKDKTD-----AITNGSYDFYVRSP 51
 23 CC15TRD2E 1 -----QELRFKDENGNNDYPEWEETTIKEIAQINXGKDKTD-----AITNGSYDFYVRSP 51
 24 CC133_771E 1 -----QELRFKDENGNNDYPEWEETTIKEIAQINTGKDKTD-----AITNGSYDFYVRSP 51
 25 CC398-1E 1 -----QELRFKDENGDYPWEETTIKEIAQINTGKDKTD-----AITNGSYDFYVRSP 51
 26 CC80-1Y 1 -----QELRFKDENGNNDYPDWEKKKLKEIACVYTGNTPSKKE-----NIYWNKGEYVWVTPTDI 54
 27 CC75-1U 1 -----QELRFKDENGEDEDYPDWEVITIQNITKYTSSKKSSNQY-----ADKDNSKGYPVYDAVQE 54
 28 CC1-1F 1 KKGYMQKIFSQELRFKDEEGNDYPDWEVITIQNITKYTSSKKSSNQY-----ADKDNSKGYPVYDAVQE 61
 29 CC873 e* 1 MSNTQKKNVPELRFPGFE----GEWEKSISLLKESKIKGSNGS-----HAKKLTVKLWGKGVV 56
 30 CC80-2Z 1 MSNTQTKNVPELRFPGFE----GEYSLDIFGNLATNKSEKFNPQN-----ENASIDIELDCIEQNTG 58
 31 CC80-3XS.Sau11819ORF2227P 1 MSNTQKKNVPELRFPEFE----GEWEEEKQFADFTKINQGLQIAINE-----RKTEYSPELYFYITNEF 59
 32 CC80-1X 1 MSNTQKKNVPELRFPGFE----GEWEEKQFADFTKINQGLQIAINE-----RKTEYSPELYFYITNEF 59
 33 CC75-1T 1 MSNTQKKNVPELRFPGFE----GEWEEKELGEIFQIISGSTPLKSN-----KEFYENGNNIWVKTTL 59
 34 ST130-1T 1 MSNTQKKNVPELRFPGFE----GEWEEKKLGEIFQIISGSTPLKSN-----KEFYENGNNIWVKTTL 59
 35 CC93-3M 1 MSNTQTKNVPELRFPGFE----GEWEEKKLEDGLFQKSYSFSRA-----KEGNGKTKHIIHGDI 56
 36 CC133_771-1strain32320Hsd 1 MSNTQTKNVPELRFPGFE----GEWEEKKLGEDGLFQKSYSFSRA-----KEGNGKTKHIIHGDI 56
 37 CC133-2fromED133J 1 MSNTQKKNVPELRFPGFE----GEWEEKKLGEVAKIYDGTHTQTPK-----YTNNEGIKFLS 55
 38 CC72-2J 1 MSNTQKKNVPELRFPEFE----GEWEEKKLGEVAKIYDGTHTQTPK-----YTNNEGIKFLS 55
 39 CC51TRD1J 1 MSNTQKKNVPELRFPGFE----GEWEEKKLGEVAKIYDGTHTQTPK-----YTNNEGIKFLS 55
 40 CC30-2strainMRSA252HsdSJ 1 MSNTQTKNVPELRFPGFE----GEWEEKKLGEFAGKVTKQKNDKYY-----IETLTNSAELGIISQKD 60
 41 CC59-1J 1 MSNTQTKNVPELRFPGFE----GEWEEKQVGEELLEFKNGLNKGE-----YFGSSSIVNFKDV 55
 42 CC72-1R 1 MSNTQTKNVPELRFPGFE----GEWEEKVKGELLEFKNGLNKGE-----YFGSSSIVNFKDV 55
 43 CC15TRD1O 1 MSNTQKKNVPELRFPGFE----GEWEEKVKGELLEFKNGLNKGE-----YFGSSSIVNFKDV 55
 44 CC398-1strain398HsdSN 1 MSNTQKKNVPELRFPGFE----GEWEEKVKGELLEFKNGLNKGE-----YFGSSSIVNFKDV 55
 45 ST425-1C 1 MSNTQTKNVPELRFPGFE----GEWEEKVKGELLEFKNGLNKGE-----YFGSSSIVNFKDV 55
 46 CC30-1strainMRSA252HsdSC 1 MSNTQKKNVPELRFPGFE----GEWEEKVKGELLEFKNGLNKGE-----YFGSSSIVNFKDV 55
 47 CC45-1strain3067HsdSC 1 MSNTQKKNVPELRFPGFE----GEWEEKVKGELLEFKNGLNKGE-----YFGSSSIVNFKDV 55
 48 CC97A 1 MSNTQKKNVPELRFPGFE----GEWEEKVKGELLEFKNGLNKGE-----YFGSSSIVNFKDV 55
 49 CC1-2strainMW2HsdSA 1 MSNTQKKNVPELRFPGFE----GEWEEKVKGELLEFKNGLNKGE-----YFGSSSIVNFKDV 55
 50 CC1-1strainMW2HsdSA 1 MSNTQKKNVPELRFPGFE----GEWEEKVKGELLEFKNGLNKGE-----YFGSSSIVNFKDV 55
 51 CC5-2strainN315HsdSA 1 MSNTQTKNVPELRFPGFE----GEWEEKVKGELLEFKNGLNKGE-----YFGSSSIVNFKDV 55
 52 CC75-2V 1 MSNTGKMNVPRLFPGE----GEWEEKELRELNPDKYSYTGPFGSDLKKSDDTTDGQIQLQNI 65
 53 CC22-1strain5096HsdSB 1 MSNTQKKNVPELRFPGFE----GEWEEKKLGEDLTDTRVIRKKNLES-----KKPLTISGQLGLIDQTEY 60
 54 CC51TRD2P 1 -----QELRFKDESGNDYPDWEKKELGEVADRVIRKKNFES-----KKPLTISGQLGLIDQTEY 55
 55 CC5-1strainN315HsdSB 1 MSNTQKKNVPELRFPGFE----GEWEEKVKGELLEFKNGLNKGE-----YFGSSSIVNFKDV 55
 56 CC93-2 b* 1 MSNTQKKNAPELRFPEFE----GEWKEKLEDLEFIKDGTHGTH-----ENVNNNGPWLSSAKNI 56
 57 Consensus_ss: 1 eeeeeeeeeeee eeeeeeeeeeee e
 58
 59
 60

1
2 Conservation:
3 CC80-3 f*
4 CC45-1L
5 CC97 c*
6 CC22-1I
7 CC873D
8 CC5-1D
9 CC30-1D
10 CC5-2H
11 CC133-2fromED133 d*
12 CC72-2S
13 CC93-3 a*
14 CC93-2K
15 CC80-2W
16 CC75-2W
17 CC59Q
18 CC72-1Q
19 CC1-2G
20 ST425-1E
21 CC15TRD2E
22 CC133_771E
23 CC398-1E
24 CC80-1Y
25 CC75-1U
26 CC1-1F
27 CC873 e*
28 CC80-2Z
29 CC80-3XS.Sau11819ORF2227P
30 CC80-1X
31 CC75-1T
32 ST130-1T
33 CC93-3M
34 CC133_771-1strain32320Hsd
35 CC133-2fromED133J
36 CC72-2J
37 CC51TRD1J
38 CC30-2strainMRSA252HsdSJ
39 CC59-1J
40 CC72-1R
41 CC15TRD1O
42 CC398-1strain398HsdSN
43 ST425-1C
44 CC30-1strainMRSA252HsdSC
45 CC45-1strain3067HsdSC
46 CC97A
47 CC1-2strainMW2HsdSA
48 CC1-1strainMW2HsdSA
49 CC5-2strainN315HsdSA
50 CC75-2V
51 CC22-1strain5096HsdSB
52 CC51TRD2P
53 CC5-1strainN315HsdSB
54 CC93-2 b*
55 Consensus_ss:
56
57 5 5
58 52 DEKSRKLN-YQNLTPP---DELNNKYKLKRNDILFARTGAST----GKS-YIHKEEKDIYNYYFAGFL 110
59 53 GYYRPAYI-KVSFKVSE-NTYNNWFRHLKENDILFSTVGNT----GIV-SLMDN---YKAVIAQNI 109
60 53 EKANINRV-----LKLGATNYYKRFEGQFYIGKQNFFN----GAF-DIVPKK--FDGLYSSSDV 104
61 57 KKGYIDFNVEAKYGNLD-LYTRWMRGNELYKGQVLFTTEAPM----GNV-AQVPD---NKGYILSQR 115
62 53 DRKDND-----SSKNKSNYKVVRKNDIAYNSMRMWQ----GAS-GKSNY---NGIVSPAY 98
63 53 DRKDND-----SSKDKSNYKVVRKNDIAYNSMRMWQ----GAS-GKSNY---NGIVSPAY 98
64 53 DRKDND-----SSKDKSNYKVVRKNDIAYNSMRMWQ----GAS-GRSNY---NGIVSPAY 98
65 47 GLLILQSD---YYKDRKT-FAESNIGYFILPKNHITYRSRSDD----GIFKFNLNLIM-IDVGIISKY 104
66 60 IYDISNF---RYYINE-NKYKEMQSFSVQPNDIIMSCSGTI----GRL-ALIPHNE-YTKGIINQAL 115
67 55 ENGL-----INP-RIYTREVTKLIQKDEIILTTRAPV----GKL-AMAQI---NACIGRGRV 101
68 55 NQKYLYK---GNKGIS---KDAAKNYMKVKNNDLIMSFKLT---GKL-AIVKA---PLYTNEAI 106
69 55 KVNDLILQ-QCNKFISK-NSIELSSAKLIPANSIAIVTRGV-----GKL-CLVEF---DYATSQDF 110
70 55 KVNDLILR-QCNKFISK-NSIELSSAKLIPANSIAIVTRGV-----GKL-CLVEF---DYATSQDF 110
71 57 TNQNQKIX-HLEQKLS---IEGQEKTTRVLVTTHLLSIAASI----GKP-VMNFV---KTGVHDGF 110
72 57 TNQNQKIX-HLEQKLS---IEGQEKTTRVLVTTHLLSIAASI----GKP-VMNFV---KTGVHDGF 110
73 54 IDXV-----DDFIFDGNYLIGEDGA-NIITRSAPLVYLVNG---KFWVNNHA 97
74 54 IDYV-----DDFIFDGNYLIGEDGA-NIITRSAPLVYLVNG---KFWVNNHA 97
75 54 IDYVK-----DYLFNNEERLLIGEDGA-KWQQFETSS-FTANG---QYWVNNHA 97
76 52 VYKI-----NTFSYEGEAILTVGDGVGV---GKV-FHYVN---GKFDYHQRV 91
77 52 VYKI-----NTFSYEGEAILTVGDGVGV---GKV-FHYVN---GKFDYHQRV 91
78 52 VYKI-----NTFSYEGEAILTVGDGVGV---GKV-FHYVN---GKFDYHQRV 91
79 52 VYKI-----NTFSYEGEAILTVGDGVGV---GKV-FHYVN---GKFDYHQRV 91
80 55 NNSKNIY-ESENKLT---QEGYKKARQLPENTLVTCAIASI----GKN-AILRK---QGSCNQQI 107
81 55 IGK-----DSNYDIEEYSISILKDGAGGV---GRI-NLRPG---KSSVIGTM 93
82 62 SINSTYN--DQNIRVN--KNKKTEKYIILSKGDLAMVLNDKTDGKIIGRS-IFIDK---DNQYIYNQRT 122
83 57 PPKETF-----KGSDNTQYYKKRAGQQLMYGKLDFLN---CAF-GIVPD---SLNNYESTID 105
84 59 RLIKIYN---SKEFSSQKNKFNPQNVLYGKLRPYL---NKY-YFTKK---SGVCSEI 106
85 60 LRPNS-----QTKY-FIENPPQSVIANKEIDLMTRTGNT----GKV-VTNVF---GAFHNNFF 108
86 60 LRPNS-----QTKY-FIENPPQSVIANKEIDLMTRTGNT----GKV-VTNVF---GAFHNNFF 108
87 60 NNSKVTH--SKEKITE-YAMKSLKLKLVPKNSVLIAMYGGFNQI---GRT-GLLKI---DATINQAI 116
88 60 NNSKVTH--SKEKITE-YAMKSLKLKLVPKNSVLIAMYGGFNQI---GRT-GLLKI---DATINQAI 116
89 57 HSFKTV--LSDSGNIP-NIIEKAVFELIQKGDIVFADASEDYSDSL---GKA-VMIDFE--PNSLISGLHT 118
90 57 HSFKTV--LSDSGNIP-NIIEKAVFELIQKGDIVFADASEDYSDSL---GKA-VMIDFK--PNSLISGLHT 118
91 55 MTS-----VSEPLSEIDIAVGIGRKGTI----NKP-YLEA---PFWTVDTL 92
92 55 MTS-----VSEPLSEIDIAVGIGRKGTI----NKP-YLEA---PFWTVDTL 92
93 55 MTS-----VSEPLSEIDIAVGIGRKGTI----NKP-YLEA---PFWTVDTL 92
94 55 MTS-----VSEPLSEIDIAVGIGRKGTI----NKP-YLEA---PFWTVDTL 92
95 56 KTLNSS---KYISE-EAEFEKEFKirPEFGDILMTRIGDI---GTP-NIVSS---NEKFAYYVSL 108
96 59 HNIKREN---ITNFIKE-KGLNESSAKLTNEAILIAMYQGKTR---GMS-AILNF---EATTNQAC 115
97 61 FDKEIS-----NIDNIKYYVVEENDFVYNPRMSNYAPF--GPV-NRNLK---GKKGVMSPLY 112
98 56 FNNRSINT--NNLTGKVN-VNSKELKNYSVEKGDVFFTRTSEVIGEI--GYP-SVILND--PENTVFSGFV 118
99 56 FNNRSINT--NNLTGKVN-VNSKELKNYSVEKGDVFFTRTSEVIGEI--GYP-SVILND--PENTVFSGFV 118
100 56 FNNRSINT--NNLTGKVN-VNSKELKNYSVEKGDVFFTRTSEVIGEI--GYP-SVILND--PENTVFSGFV 118
101 60 RNGKLNL--NDLVYISK-DIDDEMKNSTRYYGDVLLNITGASI----GRT-AINSIV--ETHANLNQHV 118
102 60 RNGKLNL--NDLVYISK-DIDDEMKNSTRYYGDVLLNITGASI----GRT-AINSIV--ETHANLNQHV 118
103 60 RNGKLNL--NDLVYISK-DIDDEMKNSTRYYGDVLLNITGASI----GRT-AINSIV--ETHANLNQHV 118
104 66 GDGYFYN--SNKVFTSN-EKAEVLKSCNCVFPGDIVIAKMDPI----ARA-AIVPDN-NIGKYLMASDG 125
105 61 FSKSVS-----SKNLENYTLIKNGEFAYNKSYSNGYPL--GAI-KRLTR---YDSGVLSSLY 111
106 56 FSKSVS-----SKNLENYTLIKNGEFAYNKSYSNGYPL--GAI-KRLTR---YDSGVLSSLY 106
107 61 FSKSVS-----SKNLENYTLIKNGEFAYNKSYSNGYPL--GAI-KRLTR---YDSGVLSSLY 111
108 57 KNNKIIIS-SDDRKISESDYKKIYKNEYKLGDLLLTGTI----GRA-AIVKN---PNNIAFQRSV 115

e ee eeeeeeeeee eeee e

1
2 Conservation:
3 CC80-3 f*
4 CC45-1L
5 CC97 c*
6 CC22-1I
7 CC873D
8 CC5-1D
9 CC30-1D
10 CC5-2H
11 CC133-2fromED133 d*
12 CC72-2S
13 CC93-3 a*
14 CC93-2K
15 CC30-2K
16 CC80-2W
17 CC75-2W
18 CC59Q
19 CC72-1Q
20 CC1-2G
21 ST425-1E
22 CC15TRD2E
23 CC133_771E
24 CC398-1E
25 CC80-1Y
26 CC75-1U
27 CC1-1F
28 CC873 e*
29 CC80-2Z
30 CC80-3XS.Sau11819ORF2227P
31 CC80-1X
32 CC75-1T
33 ST130-1T
34 CC93-3M
35 CC133_771-1strain32320Hsd
36 CC133-2fromED133J
37 CC72-2J
38 CC51TRD1J
39 CC30-2strainMRSA252HsdSJ
40 CC59-1J
41 CC72-1R
42 CC15TRD1O
43 CC398-1strain398HsdSN
44 ST425-1C
45 CC30-1strainMRSA252HsdSC
46 CC45-1strain3067HsdSC
47 CC97A
48 CC1-2strainMW2HsdSA
49 CC1-1strainMW2HsdSA
50 CC5-2strainN315HsdSA
51 CC75-2V
52 CC22-1strain5096HsdSB
53 CC51TRD2P
54 CC5-1strainN315HsdSB
55 CC93-2 b*
56 Consensus_ss:
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
559
560
561
562
563
564
565
566
567
568
569
569
570
571
572
573
574
575
576
577
578
579
579
580
581
582
583
584
585
586
587
588
589
589
590
591
592
593
594
595
596
597
598
599
599
600
601
602
603
604
605
606
607
608
609
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
659
660
661
662
663
664
665
666
667
668
669
669
670
671
672
673
674
675
676
677
678
679
679
680
681
682
683
684
685
686
687
688
689
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
709
710
711
712
713
714
715
716
717
718
719
719
720
721
722
723
724
725
726
727
728
729
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
759
760
761
762
763
764
765
766
767
768
769
769
770
771
772
773
774
775
776
777
778
779
779
780
781
782
783
784
785
786
787
788
789
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
809
810
811
812
813
814
815
816
817
818
819
819
820
821
822
823
824
825
826
827
828
829
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
859
860
861
862
863
864
865
866
867
868
869
869
870
871
872
873
874
875
876
877
878
879
879
880
881
882
883
884
885
886
887
888
889
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
959
960
961
962
963
964
965
966
967
968
969
969
970
971
972
973
974
975
976
977
978
979
979
980
981
982
983
984
985
986
987
988
989
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1839
1840
1

1
 2 Conservation: 75 6 5 76 66 597 568568
 3 CC80-3 f* 173 FLDRFDRQIELEKQKIEIILQQKKGLLQSMFI 204
 4 CC45-1L 172 LLIEIDKLVNKQLIKIELLQQRKALLKSMFI 203
 5 CC97 c* 168 FVCFLNRKIELLERKIVYLIKQKQALLQQMF 199
 6 CC22-1I 179 FFRRKINQLVELQKYKIEHTKSQKVFLQKMF 210
 7 CC873D 163 FFKKMDILISKQKIKIEILEKEKQSFLQKMF 194
 8 CC5-1D 163 FFKKMDILISKQKMKIEILEKEKQSFLQKMF 194
 9 CC30-1D 163 FFKKMDILISKQKIKIEILEKEKQSFLQKMF 194
 10 CC5-2H 165 FFSEIDRLIVEKQSSKVGRLKVRKELLQKMFV 196
 11 CC133-2fromED133 d* 179 FIHIINRRIEQSEKKIESLKNRKQGFLQKLFV 210
 12 CC72-2S 159 LNSLDVLNSKTDLKIQNLKQRKQSLLQKIFV 190
 13 CC93-3 a* 167 FLLEVDKTVNNQQLVKTKLLKQRKKGLLQRMFV 198
 14 CC93-2K 171 LFYKIDKVIISFNKCKIEIMLKLSQLQGLLKKMF 202
 15 CC30-2K 171 LFYKIDKVIISFNKCKIEIMLKLSQLQGLLKKMF 202
 16 CC80-2W 170 FFNRNEKLIELQQEKIMYLKRRKQVLLQKMF 201
 17 CC75-2W 170 FFNRNEKLIELQQEKIMYIKRCKQVLLQKMF 201
 18 CC59Q 156 FLSKLDRQIDLEEQKLELLQQRKALLKSMFV 187
 19 -----
 20 CC72-1Q 155 LLKSIDNKMNQMNRIEELLKERKELLQKMF 186
 21 CC1-2G 153 FIKKVDNKKIKIQQKQVIELLKQRKKALLQKMF 184
 22 ST425-1E 153 FIKKVDNKTQKQVIELLKQRKKALLQKMF 184
 23 CC15TRD2E 153 FIKKVDNKTQKQVIELLKQRKKALLQKMF 184
 24 CC133_771E 153 FIKKVDNKTQKQVIELLKQRKKALLQKMF 184
 25 CC398-1E 153 FIKKVDNKTQKQVIELLKQRKKSLLQKMF 184
 26 CC80-1Y 168 LISSLEELIEKQASKLKLQGMLQIMFI 199
 27 CC75-1U 152 FISNLDKLJENKNLKLNLKQLQGLLQSMFI 183
 28 CC1-1F 185 FLEVLSGITTKQLHKIDQLKERKKAFLQKMF 216
 29 CC873 e* 169 FFSKLDRQIELQKQKLELLQQQKKGYMQKIFS 200
 30 CC80-2Z 170 FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS 201
 31 CC80-3XS.Sau11819ORF2227P 171 FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS 202
 32 CC80-1X 171 FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS 202
 33 CC75-1T 178 FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS 209
 34 ST130-1T 178 FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS 201
 35 CC93-3M 182 FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS 213
 36 CC133_771-1strain32320Hsd 182 FFSKLDRQIELEEQKLELLQQQKKGYIQQKIFS 213
 37 CC133-2fromED133J 151 FFSKLDRQIELEQKLELLQQQKKGYMQKIFS 182
 38 CC72-2J 151 FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS 182
 39 CC51TRD1J 151 FFSKLDRQIELEEQKLELFQQQKKGYMQKIFS 182
 40 CC30-2strainMRSA252HsdSJ 151 FFIKLLDRQIELEEQKLELLQQQKKGYMQKIFS 182
 41 CC59-1J 151 FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS 182
 42 CC72-1R 171 FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS 202
 43 CC15TRD1O 174 FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS 205
 44 CC398-1strain398HsdSN 177 FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS 208
 45 ST425-1C 185 FFSKLDRQIELEEQKLELLQQQKKGYMQKIFT 216
 46 CC30-1strainMRSA252HsdSC 185 FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS 216
 47 CC45-1strain3067HsdSC 185 FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS 216
 48 CC97A 182 FFSKLDRQIELEEQKLELLQQQKKGYLQKIFS 213
 49 CC1-2strainMW2HsdSA 182 FFSKLDRQIELEEQKLELLQQQKKGYMQKIFT 213
 50 CC1-1strainMW2HsdSA 182 FISKLDRQIELEEQKLELLQQQKKGYMQKIFS 213
 51 CC5-2strainN315HsdSA 182 FFSKLDDQIELEEQKLELLQQQKKCYIQKIFS 213
 52 CC75-2V 189 FFSKLDRQIVLEEEQKLELLQQQKKGYMQKIFS 220
 53 CC22-1strain5096HsdSB 178 FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS 209
 54 CC51TRD2P 173 FFIKLLDRQIELEEQKLELLQQRKALLKSMFI 204
 55 CC5-1strainN315HsdSB 178 FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS 209
 56 CC93-2 b* 179 FFSKLDRQIELEEQKLELLQQQKKGYMQKIFS 210
 57 Consensus_ss: hhhhhhhhhhhhhhhhhhhhhhhhhhhhh
 58
 59
 60