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

Citation for published version:
Cooper, LP, Roberts, GA, White, JH, Luyten, YA, Bower, EKM, Morgan, RD, Roberts, RJ, Lindsay, JA & Dryden, DTF 2017, 'DNA target recognition domains in the Type I restriction and modification systems of Staphylococcus aureus' Nucleic Acids Research. DOI: 10.1093/nar/gkx067

Digital Object Identifier (DOI):
10.1093/nar/gkx067

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Nucleic Acids Research

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
DNA target recognition domains in the Type I restriction and modification systems of Staphylococcus aureus

Laurie P. Cooper1,†, Gareth A. Roberts1,†, John H. White1,†, Yvette A. Luyten2,†, Edward K.M. Bower1, Richard D. Morgan2, Richard J. Roberts2, Jodi A. Lindsay3 and David T.F. Dryden4,*

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

Received December 21, 2016; Revised January 19, 2017; Editorial Decision January 20, 2017; Accepted February 03, 2017

ABSTRACT

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

INTRODUCTION

Type I DNA restriction–modification (RM) systems are found in about half of the sequenced prokaryotic genomes (1–4). They present a formidable barrier to the invasion of the host cell by foreign DNA whether by transduction, transformation or conjugation and thus exercise control over horizontal gene transfer (HGT) (1,4–8). As an example of their effectiveness, less than 1 in 10^4 or 10^5 phage infections can successfully avoid the classical EcoKI Type I RM system of Escherichia coli K12. In some circumstances, such as when anti-restriction 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) recognizes the target sequence, HsdM (or M) recognizes 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 adenosine triphosphate (ATP) hydrolysis and DNA translocation (12). An M2S1 complex can act solely as a methyltransferase (MTase) (13). Type I RM enzymes almost always recognize 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.
These genomes contain two hsdM and two hsdS and share a single hsdR, although some S. aureus strains have different numbers of hsdM and hsdS (Figure 1A). The presence of only a single hsdR is not a problem as it can interact with each hsdMhsdS pair. It has long been known that S. aureus displays a clonal population structure (20) in which HGT between different clonal complexes is exceedingly rare. Multilocus 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 realized 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 recognized 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 E. 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 hybridization, antibody cross reactivity and subunit complementation (32,33), although now it is more usual to use the high levels of sequence identity (over 90%) in HsdM and HsdR to define a family in silico. Although the name Sau1 for this family of Type I RM systems in S. aureus is an informal one not following the usual conventions (34), we retain it as it is established in the literature. However, it is important to note that some strains of S. aureus show additional Type I RM systems, which show limited amino acid sequence identity to the HsdR, HsdM and HsdS of Sau1 (Figure 1A). For instance, Monk et al. (35) identified an active Type I RM system, SauJKDIII, in S. aureus JKD6159 which showed low sequence identity to members of the Sau1 family. This is clearly a member of a new and different Type I RM family whose subunits will be unable to interact with the Sau1 HsdM and HsdR (D. T. F. Dryden, J. A. Lindsay and M. T. G. Holden, in preparation).

The Sau1 Type I RM systems are so effective because they show great variability in the target sequences recognized thus preventing HGT between CC but allowing HGT between strains within a CC (31,35,36). This variability in target sequences is due to the modular construction of the Type I RM systems (Figure 1B). The S subunit contains two target recognition domains (TRDs) each of which recognizes one half of a bipartite target, for example the first Type I RM system in CC1, given the generic name CC1-1, recognizes CCAYNNNNNTTTA (adenine methylation sites

Figure 1. Staphylococcus aureus genomes showing the genes and the typical organization of target recognition domains (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 organization of the HsdS specificity subunit. The conserved regions (cr) are common to all S subunits within a family. The two TRDs (TRD1 and TRD2) define the target sequences recognized by the RM enzyme and can be swapped between S subunits of the same family to generate new specificities.
are underlined) (35,36). Swapping TRDs between S subunits generates new targets, for example the second Type I RM enzyme in CC1, termed CC1-2, couples the first TRD of CC1-1 with a different second TRD to recognize CCAYNNNNNNNTGT. This swapping is easy because the DNA for S subunits contain conserved sequences bounding each TRD. Most S. aureus strains have two copies of hsdS, two of hsdM and one of hsdR. Thus, there are often four TRDs in each CC, which define the restriction barrier against HGT. Some Type I RM enzymes have half-size HsdS incorporating only a single TRD. It has been shown that these products are often able to dimerize and recognize symmetric target sequences (37–39). We have been able to recapitulate these results on ‘half-HsdS’ enzymes by manipulating the CC398-1 S. aureus system (E. K. M. Bower and D. T. F. Dryden, unpublished results).

Previously we have identified the target sequences recognized by several common community-associated, hospital-associated and livestock-associated MRSA clonal complexes (31,36) and recently several more have been identified (3,35,40). Monk et al. (35) and Jones et al. (40) have used this information to prepare DNA methylated by the MTase M2S1 component enzymes to aid the transformation of S. aureus strains that are usually resistant to transformation. The identification of further targets recognized by the S subunits of S. aureus Type I RM systems would in principle allow more CC to be transformed for genetic analysis. In addition, further understanding of the structural requirements for TRDs to recognize different specific DNA sequences is of intense interest as the Type I RM systems are very widespread in bacteria and archaea (1,4) and exert considerable pressure on HGT and the evolution of prokaryotes. For instance, the use of multiple TRDs being exchanged between strains has been observed in Helicobacter (41), Mycoplasma (42,43), Streptococci (44,45) and Bacteroides (46).

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

### MATERIALS AND METHODS

#### Nomenclature for expression plasmids encoding new MTases

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

### Preparation of M.SauBI (CC22-1), M.SauCD (CC30-1), M.SauJK (CC30-2) and M.SauCL (CC45-1)

These four MTases were prepared as EFGP-His tag fusions as described in Roberts et al. (31). pSauBI-EFGP (CC22-1, genomic DNA from MRSA5906), pSauCD-EFGP (CC30-1, genomic DNA from MRSA252), pSauJK-EFGP (CC30-

| Table 1. TRD targets shown from 5′ to 3′ |
|----------------------------------------|
| TRD1 code letter and target | TRD2 code letter and target |
| A | CCAY | D | ATC |
| B | AGG | E | TCAY |
| C | GWAG | F | TTAA |
| J | GGA | G | ACA |
| M | CAG | H | TAC |
| N | ACC | I | YTC |
| O | CAAC | K | CGA |
| R | GAR | L | TT |
| T | CAAG | P | AGG |
| V | CNGA | Q | ACAY |
| X | TCTA | S | GCA |
| Z | GAC | U | GAY |
| b* | GGHA | W | CRA |
| e* | GAG | Y | CT |

The underlined A indicates the site of adenine methylation. TRD pair B and P and TRD pair U and e* recognize 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.

2, genomic DNA from MRSA252) and pSauCL-EFGP (CC45-1, genomic DNA from strain 70642) were all constructed by the polymerase chain reaction (PCR) with their hsdS fused to DNA encoding EFGP and a His-tag, with the following locus-specific oligonucleotides priming from the 3′ end of the genes encoding the S subunits:

CC22-1 BI BS 5’GATCGAATTCCGGATCCAATAAAACCTTTTGTAAAAAACAC3’
CC30-1 CD BS 5’GATCGAATTCCGGATCTAAAGAATTCAGCTTGGCTGTTT
CC30-2 JK BS 5’GATCGAATTCCGGATCTAAAGAATTCAGCTTGGCTGTTT
CC30-2 JK BS 5’GATCGAATTCCGGATCTAAAGAATTCAGCTTGGCTGTTT

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

#### A new vector for MTase expression: pJF118his

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

pJFShisTS 5’AGCCTTCGAAGGTCCCGATCCATCATCATAATTAGAATTACGCTGTTT
TGGCGG3' and pJFMSEGFPHisBS 5'GAGTGAATCC CCGGGGATCCGTCGACC3'.

The resulting PCR product was cut with BamHI and unimolecular religation gave pJF118 his into which the hsd dMS operon could be ligated as BamHI fragments and from which all subsequent MTase clones were descended.

Construction of an MTase plasmid to allow TRD swaps: pSauDeltaXmA1

A PCR-based strategy was devised to allow free pairwise assortment of desired TRDs in HsdS. Many, but not all of the HsdS subunits, including that encoded by the Type I system in CC398 (36), have a predicted proline-glycine sequence near the N-terminus. This dipetide can be encoded by CCCGAGG, which would be a target site for SmaI or XmaI. Oligonucleotides were designed which would introduce this motif in the N-terminus (a replacement with no amino acid changes) and at the C terminus (an insertion of two amino acids) of the S subunit of the CC398 system (36), by a two stage PCR fusion. Thus, primary PCR products were generated by reactions primed by: PromoterJF 5'GGTCTTCCGCTCAGGCAACCC3' with 398SmaIOligoBS 5'CCCATTCGCCTTCAAACCCG GGAATCTCAACTGCAC3' and 398SmaIOligoTS 5'GTGCCAGAGTTGAGATTTAAGTAAGGC3' and 398SmaIBamHI 5'GATCGATCGGATCCCCGGGAATAAACAC3'

The purified PCR products were then fused in a secondary PCR reaction primed by PromoterJF with 398SmaIBamHI. The product was then cut with BamHI, and ligated into the BamHI site of pJF118 his as pSauNE- XmA1. This mutated form of the CC398-1 MTase, could assemble the complete restriction enzyme that proved to be active in endonucleolytic cleavage (36). This indicated that insertion of a proline and glycine toward the C-terminus did not affect the function of the enzyme. Subsequently, on re-analyzing the DNA sequence, a single PCR mutation was discovered within the XmaI fragment. This caused a mutation A50S but this clearly did not affect the specificity or function of the S subunit in our assays. Digestion of pSauNE-XmA1 with XmaI followed by intramolecular religation of the vector fragment generates pSauDeltaXmA1, into which any pairwise combination of TRDs with XmaI cohesive ends may be inserted.

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

The DNA for each TRD of these S subunits was fused to the DNA for the reciprocal TRD of S.SauNE (CC398-1). This was achieved by creating primary PCRs with a short area of homology, which then allowed base pairing of single strands of each PCR, in a secondary PCR. For example, S.SauBE TRD B was generated from an appropriate plasmid template by PCR with oligonucleotides, TRD1FOR398SmaIOligoTS 5'GTGCCAGAGTTGAGATTTCCCGGGGTTTGGAA GCGAATGGG3' paired with TRD1inearuniversal 5'GTTCCTTCTAATTTCAATTGTG3'. TRD E was similarly generated by PCR from plasmid template with oligonucleotides TRD2inearuniversal 5'ACAAATTGAGAAGAAC3' and 398SmaIBamHI 5'GATCGATCGGATCCCCGGGAATAAATCC TTTGAAATGATGAC3'. The final insert was then generated by PCR with the two gel-purified primary oligonucleotides and TRD1FOR398SmaIOligoTS 5'GTGCCAGAGTTGAGATTTCCCGGGGTTTGGAA GCGAATGGG3' and 398SmaIBamHI 5'GATCGATCGGATCCCCGGGAATAAACAC3'

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

Additional hsdS sequences were obtained as synthetic genes from GeneArt (ThermoFisher Scientific) with sequences optimized for expression in E. coli (Supplementary Data). All the first TRDs begin with 5'CCCGGGTTTGAAGG CGAATGGAG3', except that for CC80-2 which begins with 5'CCCGGGTTTGAAGGCGAATATCT3'. Each of the first TRDs end with 5'CAAAATTGAGAAGAACAAG3'. All the second TRDs begin with 3'CAAAATTGAGAAGAACAAG3' and have a universal reverse oligonucleotide, Trd2unirev 5'GATCGATCGGATCCCCGGG TATAAAATTTTTTTAAGTATCCTGT3' and TRD2unirev 5'GATCGATCGGATCCCCGGG TATAAAATTTTTTTAAGTATCCTGT3'. All the primers used were cut with XmaI and ligated into the XmaI site of pSauDeltaXmA1.

Expression and purification of MTases

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

Nucleic Acids Research, 2017
Nucleate and ATPase assays

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

Preparation of genomic DNA for SMRT sequencing

The expression plasmids harboring the various MTases were used to transform a non-methylating (dam− dcm−) strain of *E. coli* ER2796 (30). Single colonies from the transformation plate of Lysogeny Broth (LB) agar medium supplemented with 10 μg/ml kanamycin, 10 μg/ml tetracycline as well as 100 μg/ml carbenicillin, which acted as a selection marker for the expression construct, were picked and used to inoculate 5 ml of LB containing the same cocktail of antibiotics. The cultures were incubated overnight with shaking at 37°C and 1 ml aliquots of the overnight culture were then pelleted by centrifugation (6000 × g/15 min, 4°C). The culture medium was carefully removed and the cell pellets stored at −20°C until required. Genomic DNA was prepared from each cell pellet using the Wizard Genomic DNA purification kit (Promega, Madison, WI, USA) according to the manufacturer’s instructions. The quality of the genomic DNA preparations was initially assessed by agarose gel electrophoresis and from the shape of the absorbance profile from 240 to 340 nm. Genomic DNA from *S. aureus* strains LGA251 (a kind gift from Mark Holmes) and NCTC13435 (a kind gift from Angela Kearns) was prepared by using the PurElut Bacterial Genomic Kit (EdgeBio, Gaithersburg, MD 20877, USA). The DNA library for SMRT sequencing was prepared and subsequently analyzed as described in Anton *et al.* (30).

Methylation of plasmids using *M.EcoGII*

*M.EcoGII* was kindly supplied by Dr Iain Murray (New England Biolabs) and used to modify plasmids E2, E5, E10, E11 and E12 previously described (31) and plasmid pCN36 (47). A total of 0.45 μg DNA was methylated using 2.0 U of *M.EcoGII* for 100 min at 37°C in a 50 μl volume. The reaction was in 1 x NEB4 buffer (50 mM potassium acetate, 20 mM Tris acetate, 10 mM Mg acetate, 1 mM dithiothreitol, pH 7.9, 25°C) supplemented with 320 μM S-adenosyl-L-methionine (SAM). As a negative control, DNA was incubated in the same buffer without *M.EcoGII*. The DNA samples were then supplemented with ATP (20 μM) and additional SAM (160 μM) and then digested with a Type I enzyme (CC5-1, CC5-2, CC30-1, CC45-1 or the NY TRD hybrid) for 14 min at 37°C. As a control, methylated and unmethylated DNA was digested with EcoRI.

RESULTS AND DISCUSSION

Assigning TRDs to target sequences

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

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

In contrast, TRDs U and c* are both examples of the second TRD in the HsdS subunit recognizing 5′-GAY-3′ but the level of identity between them is much lower (~36%) (Figure 2B). This level of identity between TRDs recognizing the same target is expected if the TRDs are from different Type I RM families so the low level of identity observed here is unusual. Despite this low level of sequence identity, the predicted secondary structure elements are the same as expected from the early work of Sturrock and Dryden (49). In fact, all of the TRDs in the *Saut* family of RM systems align well when secondary structure elements are taken into consideration (50) and they will have the same protein fold (Supplementary Data: PROMALS alignments). Therefore, it should in future be possible to predict the precise amino acid to nucleotide contacts involved in sequence recognition as was done for the Type IIG TRDs (51,52).

Determination of complete target sequences recognized by pairs of TRDs

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

Table 2 contains systems from a range of CC investigated previously as well as several examined in this study. It is important to note that in our work those systems containing *M.SauMRSII* plus S. SauMR-
Figure 2. Amino acid sequence and secondary structure alignment of two pairs of TRDs recognizing the same DNA target. The TRD sequences are highlighted in yellow. Consensus secondary structure shows 'h' for a helix and 'e' for a sheet. (A) TRDs B and P are examples of a first and a second TRD respectively recognizing 5′-AGG-3′. (B) TRDs U and c* are both examples of second TRDs with the same specificity, 5′-GAY-3′. The long predicted α 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.

SII, M.Sau133ORF1794P plus S.Sau133ORF1794P and M.SauMRSI plus S.SauMRSI are paired with the HsdR (SauN315ORF189P) from the N315 strain of CC5 in DNA cleavage and ATPase assays. Those shown in Tables 3 and 4 are studied as HsdS paired with the HsdM (M.SauSTORF499P) from strain S0385 of CC398 and the HsdR (SauN315ORF189P) from the N315 strain of CC5 (if used in DNA cleavage or ATPase assays). Therefore, these HsdS are not examined in the context of their natural genome, but since they are all from the Saul family of Type I RM systems and the HsdM and HsdR of these RM systems are essentially identical in all of the strains, it is reasonable to assume that the target specificities identified are those that would be recognized in their natural host.

Identifying the complete target recognized by a member of the Saul Type I RM family when both TRDs have un-
The DNA sequences for further pairs of TRDs found in a wide range of CC and ST groups were then inserted after the hsdM of CC398-1 in our expression vector and examined to ascertain the spacer sequence in the natural system (Table 4).

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

### Table 2. The *S. aureus* RM systems with published recognition sequences

| 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                | CCAAY-5-TTAA          | AF            | CC1-1                  | g, s, a             | g (31)                                     |
|                                 | CC2                            | S.SauMW2I                | CCAAY-6-TTG           | AG            | CC1-2 (CC8-2)          | g, s, a             | a (36)                                     |
| N315 (54)                       | CC5                            | S.SauN315I               | CCAAY-6-GTA           | AH            | CC5-2                  | s                   | (CC5-1 and CC5-2 in strain NRS384 are from ref. 35) |
|                                 |                                 | S.SauN315I               | AGG-5-GATA            | BD            | CC5-1 (CC8-1)          | g, s, a             | s (35)                                     |
| MRSA252 (55)                    | CC30                           | S.SauMR5II               | GWAG-5-GAT-2-TGC      | CD            | CC30-1                 | g, s, a             | g (35), s (this work)                       |
|                                 |                                 | S.SauMR5II               | GGA-7-TGC             | JK            | CC30-2                 | s                   | s (35) (this work)                         |
| JKD6159 (56)                    | CC93                           | S.SauJKDII               | GAAAG-5-TAC or complement GGHA-7-TGC | Not a *S. aureus* system | CC93-3 | s                   | s (35)                                     |

### Table 3. The ‘artificial’ *S. aureus* systems containing novel pairings of TRDs

| Recognition sequence | TRDs assigned | Experimental method | Recognition sequence | TRDs assigned | Experimental method |
|----------------------|---------------|---------------------|----------------------|---------------|---------------------|
| AGG-5-RTGA           | BE            | a                   | ACC-6-TTC            | Na*           | s                   |
| GGA-6-RTGA           | JE            | g, s                | ACC-6-RTC            | Ne*           | s                   |
| ACC-6-TGAR           | N1            | g                   | ACC-6-TRG            | Nd*           | g, s                |
| ACC-6-TGG            | NK            | g                   | GARA-6-RTGA          | RE            | s                   |
| ACC-6-TII            | NL            | g                   | CAAG-6-RTGA          | TE            | s                   |
| ACC-6-TCT            | NP            | s                   | CNGA-6-RTGA          | VE            | s                   |
| ACC-5-RGT            | NQ            | g, s                | TCTA-6-RTGA          | XI            | g, s                |
| ACC-6-TGC            | NS            | s                   | GAC-6-RTGA           | ZE            | a                   |
| ACC-6-TGG            | NU            | g, s                | GXC-6-TGC            | ZS            | a                   |
| ACC-6-TYG            | NW            | g, s                | GGHA-6-RTGA          | b*E           | s                   |
| ACC-6-TAG            | NY            | g, s                | GAG-6-RTGA           | e*E           | g, s                |

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 Data. *S. aureus* strain S.Sau133ORF1794P is characterized in this work but is included here as it is part of the RM system found in strain ED133. SauMR5II and SauMR5I characterized by Monk *et al.* and SauSTORF499P characterized by Chen *et al.* are also further characterized in this work.

**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 Data.**

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

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

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

It is apparent that the number of base pairs separating the adenines targeted for methylation and the number of base pairs in the non-specific spacer between the sequences recognized by the TRDs is not constant, with the former varying between 7 and 9 bp and the latter varying between 5 and 7 bp. This variation makes it very difficult to predict a Type I RM recognition sequence if one knows only the targets recognized by the two TRDs as the length of the non-specific spacer separating them is unknown. As mentioned above, although the sequences recognized 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.

Linking TRD 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 Data) (69). As sequence databases expand, more and more of the possible TRD combinations based on the TRDs in Table 1 will be found. For example, 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 3. Strain KPL1845 contains ‘NOVEL 2’ paired with ‘NOVEL 3’ and ‘NOVEL 4’ paired with TRD F*. Undoubtedly further TRDs will be found as sequencing continues.

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 3. 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 recognized by the *SauI* 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...
Table 5. Further TRD pairs found in sequenced strains of *Staphylococcus aureus*

| TRD pair       | Example strain                  | Clonal complex or sequence type of example strain | REBASE name               |
|----------------|---------------------------------|--------------------------------------------------|---------------------------|
| AD             | FDAARGOS,159                    | ST5                                              | S.Sau159ORF12345P         |
| AL             | K1280375                        | ST692                                            | S.Sau375ORF9DP            |
| AU             | *Staphylococcus schweitzeri* FSA084 |                                                 |                           |
| AW             | FDA239P                         | ST464                                            | S.Sau280A4ORF355P         |
| BG             | MRSN611                         | ST8                                              | S.Sau8611ORF11430P        |
| BH             | PLAC6019                        | ST5                                              | S.Sau6019ORF851P          |
| BU             | SA-083                          | ST101                                            | S.Sau083SORF9680P         |
| BY             | *Staphylococcus argenteus* M260-MSHR |                                                 |                           |
| BC*            | SA-083                          | ST101                                            | S.Sau083SORF1720P         |
| JE             | Tager 104                       | ST49                                             | S.Sau104ORF1112P          |
| JL             | W56227                          | ST45                                             | S.Sau56227ORF970P         |
| JW             | CIG290                          | ST45                                             | S.SauCIG290ORF24085P      |
| JW             | APS211                          | ST45                                             | S.SauAPS211ORF9230P       |
| MW             | FSA037                          | ST1872                                           | S.SauFSA037ORF2487P       |
| NQ             | KPL1845                         | ST96                                             | S.SauKPL1845ORF2596P      |
| O*             | USA300-TCH59                    | ST1159                                           | S.SauUSA300-TCH59ORF2596P |
| R*             | Tager 104                       | ST49                                             | S.SauTager104ORF2433P     |
| TY             | M21126                          | ST2250                                           | S.SauM21126ORF1063P       |
| XF             | 21334                           | ST109, CC9                                      | S.Sau21334ORF1353P        |
| Xf             | RK14                            | ST27                                             | S.SauRK14ORF1905P         |
| XW             | ST10-356                        | ST45                                             | S.SauST10-356ORF9230P     |
| ZY             | DI39                            | ST145                                            | S.SauDI39ORF2470P         |
| b*W            | ST20130941                      | CC15                                             | S.SauST20130941ORF4130P   |
| e*TI           | SA-120                          | ST425                                            | S.SauST20130941ORF4130P   |

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.

Table 6. New TRD pairs associated with pairs shown in Tables 2, 3 and 4

| Subspecies 21343 Bioproject accession: PRJNA35699 | Clonal complex or sequence type of example strain | REBASE name               |
|---------------------------------------------------|--------------------------------------------------|---------------------------|
| Species KPL1845. Bioproject accession: PRJNA169473 |                     |                           |

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).

and Hia5 (70) and EcoGII recognizing 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 target the sequences of the *S. aureus* Type I enzymes and have been previously described (31). These plasmids were then mixed with various purified *S. aureus* Type I restriction enzymes or, as a control, the EcoRI restriction enzyme. After one hour of methylation by M.EcoGII, the plasmids were completely resistant to digestion by EcoRI and by the *S. aureus* restriction enzymes (Figure 3). Furthermore, the shuttle vector pCN36 (47) was also protected from digestion by these same enzymes (data not shown). Subsequent experiments using the methylated pCN36 to transform *S. aureus* were unfortunately entirely unsuccessful (J. A. Lindsay, unpublished results using strains HO5096 (CC22), JE2 (CC8) and RN4220 (CC8, *hsdR*). The reason for the failure of transformation with the highly methylated pCN36 when it should be resistant to all *Sau* RM systems is not clear. This result may imply a further unrecognized barrier to transformation of *S. aureus* or some aspect of the physical properties of highly methylated DNA. Nevertheless
CONCLUSIONS

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

The data obtained here will allow the construction of new *E. coli* strains for preparing methylated shuttle vectors (35) and MTase reagents for *in vitro* methylation of DNA (40) to assist transformation of further *S. aureus* strains. However, these approaches are time consuming and it is worth noting that the common shuttle vector used for transformation of *S. aureus*, pCN36 (47), contains a target site for almost every TRD pair investigated in this paper. This means that pCN36 is inevitably a poor vector for transformation of *S. aureus*. The construction of new shuttle vectors completely lacking *Sau* targets via DNA synthesis, coupled with careful analysis of the fragments to be ligated into the vector so that they also lack targets, may be an effective way forward to improve transformation of *S. aureus* now that so many target specificities have been determined. Obviously, the avoidance of the sequence AN6-9T, although difficult to achieve without altering protein coding sequences in a vector, would be a general method to negate the effect of the Type I RM systems in *S. aureus* and other prokaryotes.

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

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

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

FUNDING

Biotechnology and Biological Sciences Research Council (BBSRC) [BB/K005804/1 to D.T.F.D.]; Wellcome Trust [GR080463MA, 090288/Z/09/Z to D.T.F.D., J.A.L.]; Institute of Advanced Study, Durham University Fellowship (to D.T.F.D.). Funding for open access charge: Wellcome Trust; BBSRC.

Conflict of interest statement. None declared.

REFERENCES

1. Oliveira, P.H., Touchon, M. and Rocha, E.P. (2014) The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.*, 42, 10618–10631.
2. Roberts, R.J., Vincze, T., Posfai, J. and Macelis, D. (2015) REBASE - a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.*, 43, D298–D299.
3. Blow, M.J., Clark, T.A., Daum, C.G., Deutschbauer, A.M., Fomenkov, A., Fries, R., Froula, J., Kang, D.D., Malmstrom, R.R., Morgan, R.D. et al. (2016) The epigenomic landscape of prokaryotes. *PLoS Genet.*, 12, e1005854.
4. Oliveira, P.H., Touchon, M. and Rocha, E.P. (2016) Regulation of genetic flux between bacteria by restriction-modification systems. *Proc. Natl. Acad. Sci. U.S.A.*, 113, 5658–5663.
20. Feil, E.J., Cooper, J.E., Grundmann, H., Robinson, D.A., McCarthy, A.J. and Lindsay, J.A. (2012) The distribution of plasmids
25. McCarthy, A.J., Witney, A.A. and Lindsay, J.A. (2012)
26. McCarthy, A.J., Loeffler, A., Witney, A.A., Gould, K.A., Lloyd, D.H. and Lindsay, J.A. (2014) Extensive horizontal gene transfer during Staphylococcus aureus co-colonization in vivo. Genome Biol. Evol., 6, 2497–2708.
27. Corvalía, A.R., Franciosí, P., Hernandez, D., Petrok, K., Linder, P. and Schrenzel, J. (2010) A Type III-like restriction endonuclease functions as a major barrier to horizontal gene transfer in clinical Staphylococcus aureus strains. Proc. Natl. Acad. Sci. U.S.A., 107, 11954–11958.
28. Xu, S.Y., Corvaglia, A.R., Chan, S.H., Zheng, Y. and Linder, P. (2011) A new type IV modification-dependent restriction enzyme SauUSI from Staphylococcus aureus subsp. aureus USA300. Nucleic Acids Res., 39, 5597–5610.
29. Anton, B.P., Mongodin, E.F., Agrawal, S., Fomenkov, A., Byrd, D.R., Roberts, J.R. and Raleigh, E.A. (2015) Complete genome sequence of ER2796, a DNA methyltransferase-deficient strain of Escherichia coli K-12. PLoS One, 10, e0127446.
30. Roberts, G.A., Houston, P.J., White, J.H., Chen, K., Stephanoiu, A.S., Cooper, L.P., Dryden, D.T.F. and Lindsay, J.A. (2013) Impact of target site distribution for Type I restriction enzymes on the evolution of methicillin-resistant Staphylococcus aureus (MRSA) populations. Nucleic Acids Res., 41, 7472–7484.
31. Barcus, V.A., Titheradge, A.B. and Murray, N.E. (1995) The diversity of restriction enzymes: an analysis prompted by molecular and genetic data for type ID restriction and modification systems. Nucleic Acids Res., 23, 4195–4205.
32. Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., Blumenthal, R.M., Degtaryev, S.K., Dryden, D.T.F. et al. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res., 31, 1805–1812.
33. Monk, I.R., Tree, J.J., Howden, B.P., Stinear, T.P. and Foster, T.J. (2015) Complete bypass of restriction systems for major Staphylococcus aureus lineages. Mbio, 6, e00308–e00315.
34. Chen, K., Stephanoiu, A.S., Roberts, G.A., White, J.H., Cooper, L.P., Houston, P.J., Lindsay, J.A. and Dryden, D.T.F. (2016) The Type I restriction enzymes as barriers to horizontal gene transfer: determination of the DNA target sequences recognised by livestock-associated methicillin-resistant Staphylococcus aureus clonal complexes 155/SST771 and 398. Adv. Exp. Med. Biol., 915, 81–97.
35. Abadji, E.A., Patel, J., Webb, M., Zinkevich, V. and Firman, K. (1993) A deletion mutant of the type IC restriction endonuclease EcoR241 expressing a novel DNA specificity. Nucleic Acids Res., 21, 4435–4443.
36. Meister, J., MacWilliams, M., Hübner, P., Jütte, H., Skrzypczek, E., Pekarowicz, A. and Bickle, T.A. (1993) Macroevolution by transposition: drastic modification of DNA recognition by a type I restriction enzyme following Tn5 transposition. EMBO J., 12, 4585–4591.
37. MacWilliams, M.P. and Bickle, T.A. (1996) Generation of new DNA binding specificity by truncation of the type IC EcoDXXI hsds gene. EMBO J., 15, 4775–4783.
38. Jones, M.J., Donegan, N.P., Mkhipyeva, I.V. and Cheung, A.L. (2015) Improving transformation of Staphylococcus aureus belonging to the CC1, CCS and CC8 clonal complexes. PLoS One, 10, e0149478.
39. Furuta, Y., Kawai, M., Uchiuma, J. and Kobayashi, J. (2011) Domain movement within a gene: a novel evolutionary mechanism for protein diversification. PLoS One, 6, e18189.
40. Dybvik, K., Sitarman, R. and French, C.T. (1998) A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements. Proc. Natl. Acad. Sci. U.S.A., 95, 13923–13928.
41. Xiao, L., Prior, E., Osborne, J.D., Crab, D.M., Simmons, W.L., Lefkowitz, E.J., Waite, K.B., Atkinson, T.P. and Dybvik, K. (2015) Comparative genome analysis of Mycoplasma pneumoniae. BMC Genomics, 16, doi:10.1186/s12866-015-1801-0.
42. Li, J., Li, J.W., Feng, Z., Wang, J., An, H., Liu, Y., Wang, Y., Wang, K., Zhang, X., Miao, Z. et al. (2016) Epigenetic switch driven by DNA
