Research Article

Systemic Mutational Rescue in Escherichia Coli Elicited by a Valency Dependent, High Affinity Protein DNA Interaction.

Hurd, P.J.[1] Al-Swailem, A.M.[2] Bin Dukhyil, A.A.A.[3] Sheikh, Q.I.[4] Al-Ghanim, A.A.[5] Alfageih, L.[6] Matin, M.[7] Yueh Ting, L.u.[8] Abdalgelel, A.[9] Florence, J.[10] Mohammed Al-Shemirti,[11] Al Harbi, S.[12] Brown, P.E.[13] Hornby, D.P.*[14]

Abstract

The controlled formation of sequence specific DNA protein complexes is a fundamental feature of most genetic transactions. Studies of the impact of point mutations on the function of individual components, such as repressors, remains a key aspect of many systems and synthetic biology research programmes. One of the most dramatic systemic consequences of a point mutation is exhibited by the monomeric DNA methyltransferases M.HhaI and M.EcoRII, where substitution of a single, catalytic cysteine by either glycine or alanine, creates a lethal gain of function phenotype. In vivo expression of these point mutants promotes the deposition of high affinity nucleoprotein complexes that arrest replication in vivo, causing cell death. Interestingly, it appears that a systemic response to expression of these mutant enzymes is dramatically enhanced when they are expressed as synthetic dimers. A previously unreported form of "mutational rescue" appears to be triggered as a result of networked crosslinking of host cell DNA resulting from an increased valency of nucleoprotein complex formation. This finding may have significance for developing molecular interventions for the controlled regulation of genome function and for the development of novel antimicrobial and anticancer strategies.

Affiliation:

[1]Current address School of Biological and Chemical Sciences, Queen Mary University of London, London, E1 4NS, UK.

[2]Current address Chief executive officer Saudi Authority for Intellectual Property6531 Al Okaya Rd, Al Sahafah P. O. Box 13321 Riyadh 3059 Saudi Arabia.

[3]Current address Medical Laboratories Sciences Department, College of Applied Medical Sciences, Majmaah University, P.O. Box 25529, Riyadh 11476, Saudi Arabia

[4]The Krebs Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Western Bank, Sheffield, S10 2TN, UK.

[5]Current address Medical Laboratories Sciences Department, College of Applied Medical Sciences, Qassim University. P.O. Box 6699, Qassim 51452, Saudi Arabia.

[6]The Krebs Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Western Bank, Sheffield, S10 2TN.

[7]Current address Department of Biology, Faculty of Sciences, Ferdowsi University of Mashhad, Iran.

[8]The Krebs Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Western Bank, Sheffield

[9]The Krebs Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Western Bank,

[10]The Krebs Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Western Bank, Sheffield, S10 2TN.

[11]Current address, Faculty of Science, The University of Kufa, Najaf, Iraq

[12]Department of Medical Laboratory Science, College of Applied Medical Sciences Al- Quwayiyah, Shaqra University, Kingdom of Saudi Arabia.

[13]The Krebs Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Western Bank, Sheffield, S10 2TN.

[14]The Krebs Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Western Bank, Sheffield, S10 2TN. d.hornby@shef.ac.uk

*Corresponding author:
D.P. Hornby, The Krebs Institute, School of Biosciences, University of Sheffield, Western Bank, Sheffield, S10 2TN.

Email: d.hornby@shef.ac.uk.

Keywords: Mutation; rescue; DNA methyltransferase; deletion; insertion; (indel); transposon mutagenesis; genome engineering; antimicrobial resistance.

Introduction

The origins of contemporary molecular medicine date back to the pioneering work of Linus Pauling [1] and the subsequent demonstration by Vernon Ingram [2] that sickle cell anaemia can result from a single missense mutation in the β globin gene. However, while the outcome of mutational events is often disadvantageous to an organism, genome stability and mutation are generally considered as two sides of the same coin. On the one hand, the stable and faithful inheritance of genetic information is essential for the viability of a species, on the other, a measured level of change drives evolution at the genetic level. In the simplest examples of genetic mutation, the replication machinery provides the primary "engine" that balances the conflicting demands of stability and change. Our understanding of DNA damage, repair and general genome maintenance has revealed that changes to the genome are brought about by an assortment of mechanisms that range from individual nucleotide events to those involving a form of genetic recombination [3]. For example, transposition and deletion events can lead to genome remodelling, promoting both positive and negative outcomes.

Furthermore, all organisms encode DNA modifying enzymes, whose activity can impart an additional level of influence on both genome integrity and phenotype. Ongoing research into these phenomena at a systems level and the therapeutic opportunities that may emerge, remain a high priority.

Just as the lac operon has proven to be an enduring workhorse for both molecular biology [4] and more recently systems biology [5], the cytosine- specific C-5 DNA methyltransferases (C5-DNMTs) represent a well-studied class of DNA modifying enzymes, most commonly associated with restriction and modification in prokaryotes, but also as facilitators of a range of epigenetic phenomena [6]. In vivo, these prokaryotic monomeric enzymes recognise a short DNA sequence (typically between 2-8 base pairs) and methylate one strand of a hemi methylated DNA duplex, by flipping out the unmethylated, target base [7]. Following S-adenosyl-L-methionine dependent methylation, the target base is returned to the DNA duplex and the enzyme subsequently dissociates. A small number of heterodimeric C5-DNMTs have been described; M.AquI [8] and M.EcoHK31I [9], but importantly all possess a single active site. Most bacterial C5-DNMTs form part of a Type II restriction and modification operon. In eukaryotes however [10-12], and some bacteriophage [13] these enzymes are expressed independently of a cognate restriction endonuclease, and their biological roles are diverse, and not always understood [6].

Many mutants of this class of enzymes have been described, but one unusual mutant stands out: the substitution of the catalytic Cys residue (generally preceded by Pro), for Gly, referred to here as ProGly. Wyszynski et al [14] were the first to report the "toxic" phenotype of the ProGly mutation in M.EcoRII, and Mi and Roberts [15] subsequently demonstrated that the same mutation in M.HhaI dramatically reduced the dissociation constant for its cognate DNA sequence. Henderson and Kreuzer [16] went on to confirm, that a key element of the "toxic" phenotype of this mutation results from compromised DNA replication, induced by the expression of abnormally high affinity protein complexes, which in turn inhibit DNA replication. The effects of the ProGly mutant have been compared to the molecular events that follow the administration of anti-cancer drugs including azacytidine (vidaza) [17] and zebularine [18] or the fluoroquinolone family of antibiotics [19]. These compounds promote the irreversible accumulation of nucleoprotein complexes at the sites of DNA recognition, which forms the basis of their application in epigenetic medicine (vidaza) and in the treatment of microbial infections (ciprofloxacin). While the specific modes of action of vidaza and zebularine involves incorporation into genomic DNA, followed by the covalent trapping of C5- DNMTs; the fluoroquinolones prevent the completion of the catalytic cycle required for the segregation of bacterial chromosomes [19]. In this context, Henderson and Kreuzer [16] have shown that there are clear differences in the cellular responses arising as consequence of the formation of these two persistent nucleoprotein complexes.

During a comparative investigation of the molecular characteristics of drug induced complexes and those formed by the ProGly mutant enzymes, we constructed a range of recombinant C5-DNMTs exhibiting different sequence specificities. In most cases, the expressed enzymes were fused to the C-terminus of a naturally homodimeric Glutathione-S-transferase [20,21]. One of the most surprising consequences of expressing the C5- DNMTs in this way, was the apparent neutralisation of the "toxic" effect of the ProGly mutation. We report here that the oligomeric status of a point mutation in a DNA modifying enzyme has a profound impact on the systemic response to its expression in vivo. These results may provide new insights into our understanding of the role of genome topology in genome surveillance and maintenance, and point to important considerations in understanding the interplay between the sequence and the operational valency of a protein in a wider evolutionary context.

Results

Synthetic oligomerisation directs the host response to high affinity DNA recognition

A range of plasmids encoding the prokaryotic C5-DNMT enzymes, M.HhaI, M.MSPRI and M.MspI were generated and provided the starting point for all genetic constructs in the experiments described here. Their general properties are summarised in Table 1. Comparative transformation efficiency was used to assess the DNA modification proficiency of all plasmid encoded enzymes, using the protocol described by Raleigh and Wilson [22]. As expected, plasmid transformation of *E.coli* strains AB1157 (*mcr+*) and DH5αMCR (*mcr-*) gave the anticipated results with respect to DNA methylation proficiency of each plasmid encoding a monospecific C5-DNMT (see Table 1). The expression of missense mutants where catalytic Cys to Gly (ProGly mutation) is incorporated, significantly reduces host viability, as demonstrated by relative transformation efficiencies shown in Figs 1 a and b.

These results are in full agreement with the observations of others [14] and [15]. Furthermore, plasmids encoding ProGly mutations yielded negligible numbers of colonies. Superficially, the impact of the ProGly mutation appears similar to the restriction phenomenon observed when appropriately methylated plasmids are eliminated from mcr+ strains [22]. As part of a series of experiments investigating the active site of several bacterial C5-DNMTs, plasmid expression vectors encoding Glutathione-S-transferase (GST) were constructed as described elsewhere for M.MspI [23]. The general features of this group of plasmids and their expression products are provided in Table 1. As with the WT monomeric forms of the encoded enzymes, the

Table 1: Plasmid constructs used in this work. The relative transformation efficiencies are expressed using pSPR-stop and pLITMUS28 (New England Biolabs) as reference plasmids. *These plasmids were introduced following Kan insertion to inactivate the C5- DNMT during ProGly mutant construction and subsequent removal (by restriction and subsequent ligation, as shown in Fig 2b).

Plasmid	DNA Methyltransferase	AB1157 mcrABC+ CFU	DH5αMCR mcrABC- CFU	DNA recognition Valency
pGMspI	M.MspI	<10	10^7	Dimer
pGMspIC	M.MspI	<10	10^7	Dimer
pGMspICGly	Mutant M.MspI	10^3	10^3	Dimer
pHhaI-1	M.HhaI	<10	10^7	Monovalent
pMspI-1	M.MspI	<10	10^7	Monovalent
pSPRI-1	M.SPRI	<10	10^7	Monovalent
pLITMUS28	None	10^7	10^7	N/A
pSPR-stop	None	10^7	10^7	N/A
pHhaI-2	M.HhaI	<10	10^7	Dimer
pMspI-2	M.MspI	<10	10^7	Dimer
pSPRI-2	M.SPRI	<10	10^7	Dimer
pNSPRI-ZIP	None	10^5	10^5	N/A
pCSPRI-ZIP	None	10^7	10^7	N/A
pCSPRI-ZIP + pNSPRI-ZIP	M.SPRI	<10	10^5	Multimer
*pSPRIGly-1	Mutant M.SPRI	<10	<10	Monomer
*pMspIGly-1	Mutant M.MspI	<10	<10	Monomer
*pSPRIGly-2	Mutant MSPRI	10^4	10^4	Dimer
*pMspIGly-2	Mutant M.MspI	10^5	10^5	Dimer
pNSPRIGly-ZIP + pCSPRI-ZIP	Mutant M.SPRI	10^6	10^6	Multimer
pSPRIGly-2-Stop	None	10^7	10^7	N/A

polypeptides retain DNA methylation proficiency as artificial dimers (a consequence of the obligate dimerization of GST). All plasmids expressing GST fusions of M.MspI, M.SPRI and M.HhaI are fully methylated and are consequently restricted by mcrABC+ strains (eg AB1157), but can be stably maintained in E.coli DH5αMCR [23]. Surprisingly, following substitution of the ProCys motif at the active site of M.MspI with ProGly (expressed from pGMspICGly, illustrated in Fig. 2a), and in contrast to the toxicity observed following transformation with plasmids encoding monomeric ProGly mutant C5-DNMTs, transformation efficiency was only moderately reduced (see Table 1 and Fig.1).

Subsequent extraction and purification of the GST fusion as described elsewhere [23], from colonies harbouring the plasmid pGMspICGly revealed an enzymatically inactive GST fusion protein of the expected Mr, as determined by SDS polyacrylamide gel electrophoresis (data not shown). However, nucleotide sequencing of pGMspICGly revealed a 30bp in-frame deletion (eliminating 10 amino acids from the catalytic domain) in all plasmids recovered from transformed colonies. Similarly, two other ProGly mutant C5-DNMTs also showed only marginal reductions in transformation

efficiencies in both AB1157 and DH5αMCR hosts (see Table 1 and Fig.1). In these experiments, the ProGly mutation was introduced into the gene after a first step in which a kanamycin cassette was inserted to disrupt the C5-DNMT gene (as shown in Fig. 2b). The protocol used to remove the kanamycin cassette in this way led to a reduction in transformation efficiencies compared with supercoiled plasmid controls, as would be expected, but proved a more efficient method for reproducible evaluation of transformation efficiencies. A profile of transformation efficiencies of all plasmids used in this work is given in Table 1 and Figure 1. It is clear from these data that earlier observations [14] and [15], only apply to plasmids encoding monomeric ProGly mutant C5-DNMTs. Therefore, synthetic, GST dimers overcome the barrier to transformation observed when plasmids encode monomeric C5-DNMT ProGly mutants. It should be noted that, in view of the incorporation of plasmid mediated ampicillin selection in the experimental design, only cells retaining replication competence and antibiotic resistance are expected to be recovered in these experiments. Given the sequence specific nature of C5-DNMTs, we reasoned that output plasmids might show patterns of sequence similarity in the vicinity of the mutational rescue process.

1a

1b
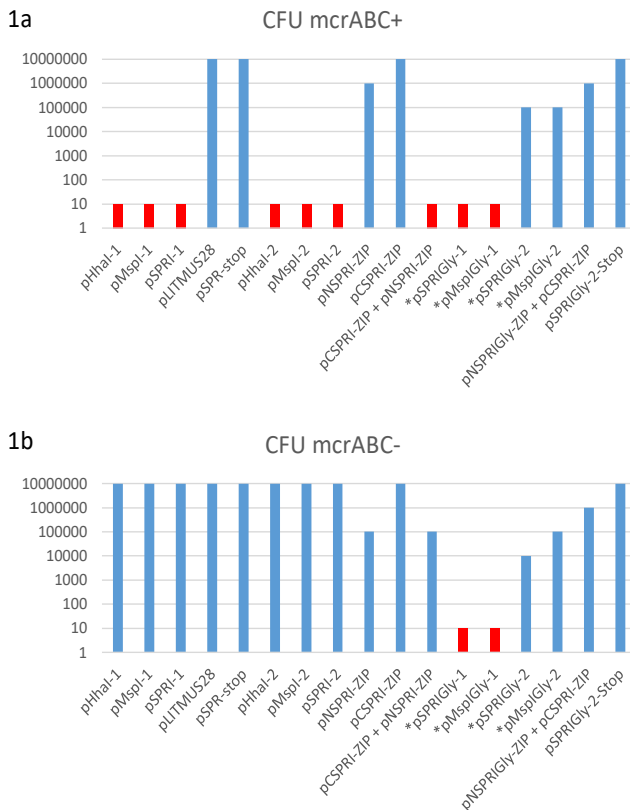
Figure 1: Log plot of the transformation efficiencies of plasmids used in this work. (a) E.coli AB1157 (mcrABC+) was the host for the transformation in each case of 10ng of supercoiled plasmid DNA. Plasmids indicated by and asterisk contain a ProGly mutation and were digested with a single restriction endonuclease to excise a kanamycin gene cassette, followed by re-ligation. In these cases, all CFUs reported are ampicillin resistant and kanamycin sensitive. Transformation experiments involving the monomeric ProGly mutant encoding plasmids are coloured red. The same experiments in (a) are shown in (b) with *E.coli* DH5αMRC acting as recipient.

However, this is not the case and is exemplified by the analysis of a series of mutants recovered from a pSPRIGly-2 transformation experiment. As can be seen in Fig.3, the nucleotide sequences of 5 classes of deletion mutants showed no similarities in sequence at the boundaries of the deletion, and the deletion events are unrelated to the recognition sequences of this multi-specific C5-DNMT (GGCC, CCGG and CCA/TGG). To date, we have not detected any patterns following nucleotide sequencing of over 100 plasmids encoding bivalent ProGly mutant C5-DNMTs.

The lethality of ProGly mutants is dependent on multivalent expression of C5-DNMTs

In order to further investigate the role of increased valency of high affinity nucleoprotein complex formation, a pair of compatible plasmids were constructed in which the catalytic domain of M.SPRI was expressed independently from one plasmid (pNSPRIGly-ZIP) and the C-terminal DNA

target recognition domain (TRD) expressed from a second (pCSPRIZIP). Both segments were expressed as GST fusions, and functional assembly of the two separated domains was achieved by tagging the C-terminus of the catalytic domain with the b-ZIP sequence from rat C/EBPα. The same sequence was appended to the TRD region. One additional advantage of this pair of constructs lay in their ability to reconstitute a functional C5-DNMT or a ProGly mutant via sequential transformation of supercoiled plasmids, thereby eliminating the kanamycin excision step. Interestingly, both transposable and IS element insertions were the hallmark of the loss of C5-DNMT ProGly "toxicity" in the b-ZIP constructs (see Fig. 4). However, while insertion mutants dominated transformation plates, deletions were also recovered. This experiment independently confirms the requirement for multivalent DNA complex formation, since the combinatorial nature of b-ZIP transcription factors is well established [24]. Moreover, because of the promiscuity of b-ZIP motifs, it is likely that a spectrum of possible permutations of oligomers are expressed in vivo, including not only monomeric and dimeric forms, but also polyvalent DNA binding proteins. The results of all transformation experiments are summarised schematically in Fig. 5.
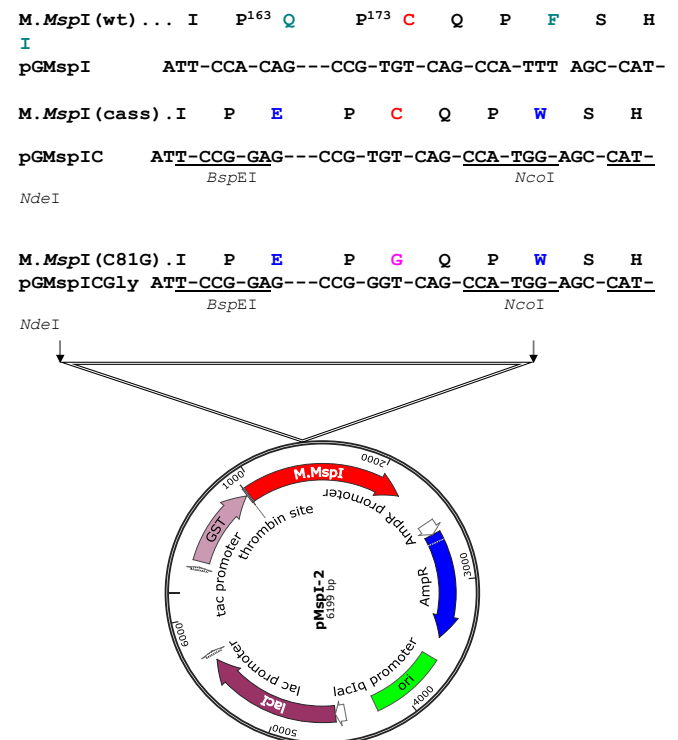

Figure 2a: Construction of the ProGly mutant in pMspI-2. The sequence encoding the catalytic loop of M.MspI was modified by oligonucleotide exchange methodology. The resultant plasmid contained unique BspEI and NcoI sites, facilitating the codon change as shown in a second restriction and re-ligation step.
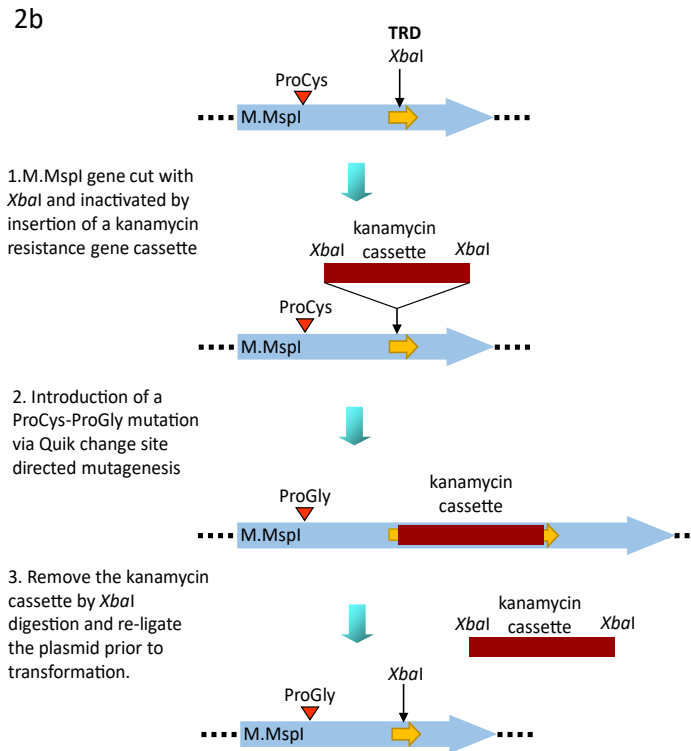
Figure 2b: Alternative methodology for the generation of the ProGly mutants. In this case the plasmid encodes a GST fusion of M.MspI. An XbaI gene cassette encoding the kanamycin resistance gene from pUC4K (GE Healthcare) was inserted into the unique XbaI restriction site in the region encoding the Target Recognition Domain (TRD) of the C5-DNMT. Following site directed mutation of the catalytic Cys codon Gly, the plasmid was re-cut with XbaI, re-ligated and transformants selected for ampicillin resistance and kanamycin sensitivity.
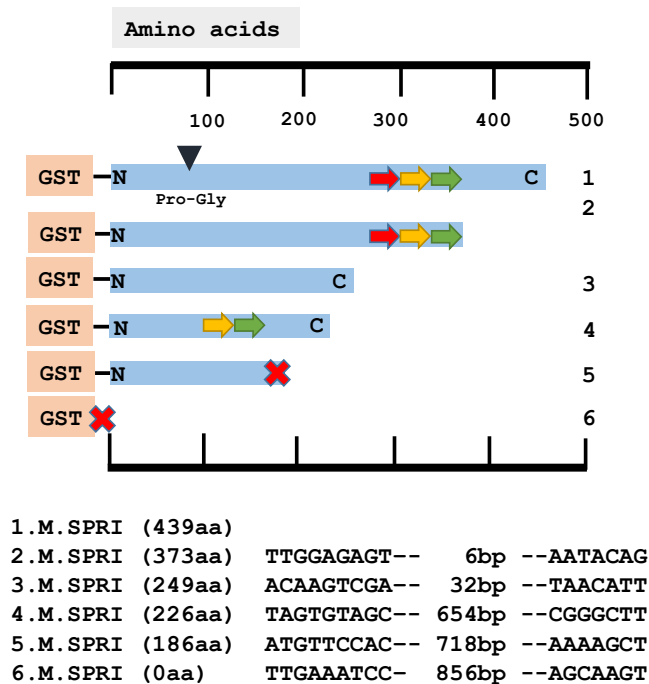
Figure 3: Comparative nucleotide sequence analysis of plasmids recovered following transformation of pSPRIGly-2. The coding sequence is indicated by the blue bar, with the N and C termini labelled and the mutation of the catalytic Cys identified by the black triangle as Pro-Gly. The three TRDs are shown as coloured arrows. The sequences at the boundaries of the deletion events are shown below for each plasmid. The large red cross indicates the presence of a premature stop codon.
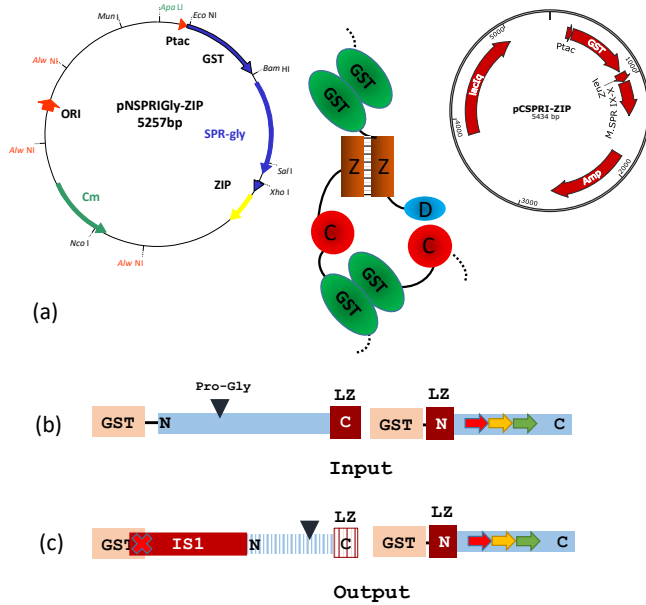
Figure 4: A typical result obtained following transformation of AB1157 with a pair of compatible plasmids designed to generate a heterodimeric M.SPRI enzyme, carrying the ProGly mutation.

(a) The assembled enzyme is shown schematically with the b-ZIP motifs enhancing assembly.

(b) The schematic diagram shows the input open reading frames expressed from the complementary plasmids. The output ORFs indicate that ISI insertion leads to premature termination of the N-terminal GST catalytic domain (blue stripes), with retention of the C terminal domain.

(c) The nucleotide sequence of an IS1 element, inserted within the GST orf was identified by nucleotide sequencing. This disrupts dimer formation.

Mutational rescue is restricted to plasmids and not the genome

In order to establish whether the enhanced levels of mutation induced by the dimeric ProGly mutants, is restricted to plasmid DNA, several colonies were subjected to whole genome sequencing. In this experiment, genomic DNA from E.coli DH5αMCR carrying a deletion of the mcrBC locus was analysed alongside that from AB1157 strains transformed with pSPRIGly-2, with and without the kanamycin gene cassette. Genome sequencing was carried out by MicrobesNG at the University of Birmingham UK and sequences compiled and analysed using Integrative Genomics Viewer Software (version 2.6.3). Alignment of the genome sequences showed no evidence for genome wide indel events (see Fig. 6). The presence of the mcrBC deletion in strain DH5αMCR acted as a reference point for genome comparisons. BLAST was used to identify the pSPRIGly-2 plasmid in each genome assembly: a search was carried out on the partially assembled genome files, using the pSPRIGly-2 plasmid file as the query. The resultant matches identified in which nodes the plasmid could be located, in each of the genome sequences. From

this, the plasmids could be identified as encoding either the ProGly mutant, or the kanamycin insertion construct which had not been primed by removal of the insertion sequence. All plasmid sequences revealed inactivating mutations as seen in the earlier plasmid isolation and sequencing experiments. In addition, the kanamycin sequence is clearly present in transformants that were not processed in advance of transformation. A further comparative BLAST analysis of 50 E.coli genes showed no sequence variation across all experimental strains. This result is not unexpected, given that the cells recovered retain viability under ampicillin selection as judged by standard overnight growth protocols. An additional evaluation of the location of transposable elements was carried out, and in this respect there appears to be no difference between the genomes sequenced. Through the comparison of these genomes, it is clear that there are no large rearrangements, deletions or insertions present in strains that have recovered from expression of a bivalent ProGly mutant C5-DNMT.
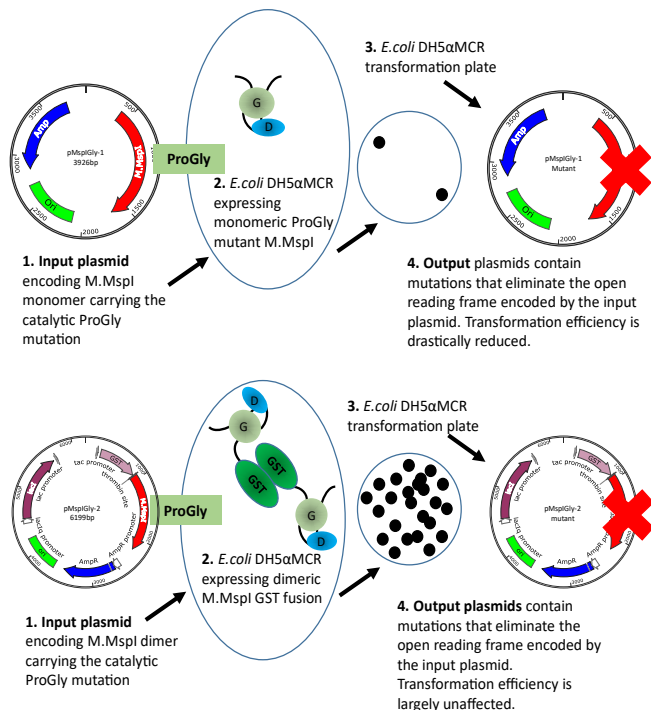
Mutants that abrogate mutational rescue

The catalytic pathway of all C5-DNMTs is known to involve flipping of the target cytosine from the DNA helix into the catalytic site, followed by the formation of a transient covalent complex. The difficulties in isolating
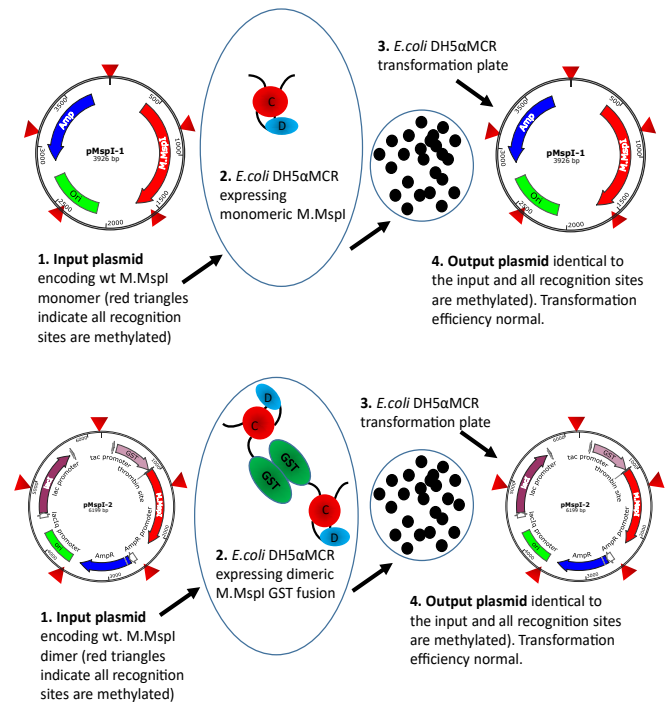
Figure 5a: Schematic representation of the transformation experiments exemplified with M.MspI. The input plasmids both lead to the expression of methylation proficient enzymes. In the case of pMspI-1 the enzyme is monomeric whereas the enzyme expressed from pMspI-2 is a GST fusion and therefore dimeric. Transformation efficiencies are comparable and the output plasmids are unchanged.

Citation: Hurd, P.J. Al-Swailem, A.M. Bin Dukhyil, A.A.A. Sheikh, Q.I. Al-Ghanim, A.A. Alfageih, L. Matin, M. Yueh Ting, L.u. Abdalgelel, A. Florence, J. Mohammed Al-Shemirti, Al Harbi, S. Brown, P.E. Hornby, D.P. Systemic Mutational Rescue in Escherichia Coli Elicited by a Valency Dependent, High Affinity Protein DNA Interaction. Journal of Bioinformatics and Systems Biology. 6 (2023): 97-109.

sufficient quantities of ProGly mutants, have made structure determination an insurmountable challenge to date. However, it is not unreasonable to speculate that the unusually high affinity for DNA exhibited by ProGly mutants, results from the formation of an irreversible base flipped complex. The determinants of sequence specific base flipping reside in the C-terminal domain of C5-DNMTs and comprise the TRD and conserved motifs IX and X [25-27]. In order to investigate the possibility that the ProGly mutants show high affinity recognition through an interplay between the catalytic domain and the TRD domain, a series of M.HhaI mutants were generated using error-prone PCR [28] and the library recombined with a plasmid (pHhaI-2) carrying the ProGly mutation. Colonies recovered from this "library" were initially screened for plasmid integrity and were subsequently sequenced. Two classes of plasmids were obtained: those retaining the characteristics of the input plasmid while neutralizing the ProGly effect, and those modified by the mutational rescue phenomenon described above. Inspection of the former class of plasmids revealed that the secondary mutations found in the sequence encoding the Cterminal

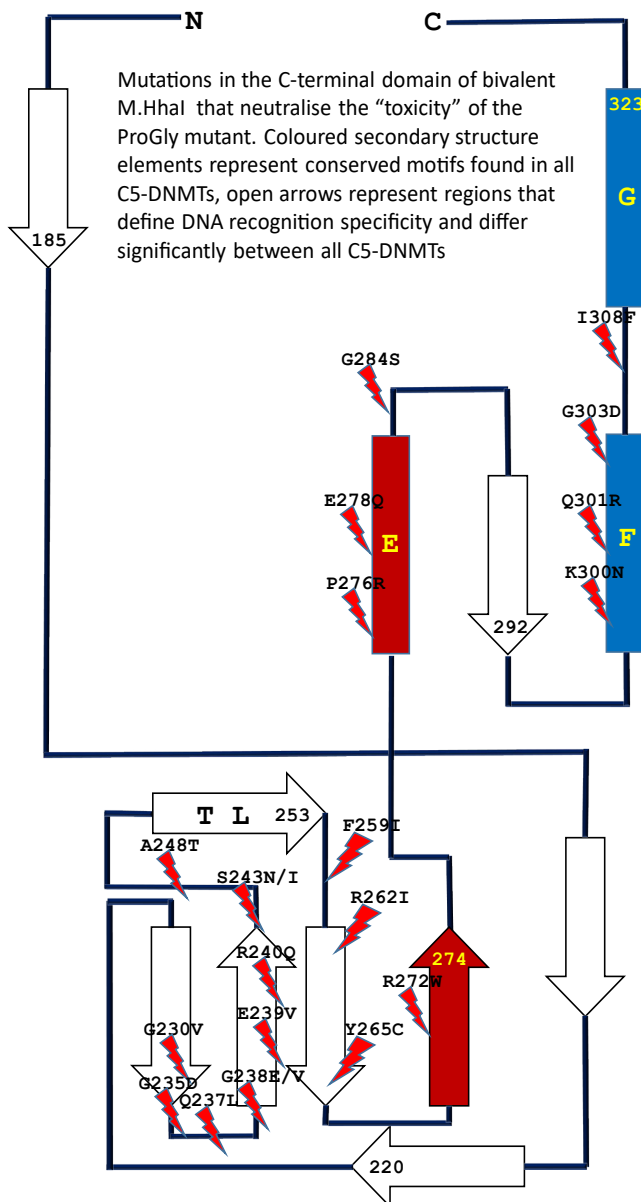| Plasmid	Genome	Sequence code
pSPRI-2	DH5aMCR	5678
pSPR-stop	AB1157	5679
pSPR1Gly-2 (1)	AB1157	5680
pSPR1Gly-2 (Kan)	AB1157	5681
pSPR1Gly-2 (Kan)	AB1157	5682
pSPR1Gly-2 (2)	AB1157	5683
pSPR1Gly-2 (3)	AB1157	5684
pSPR1Gly-2 (Kan)	AB1157	5685

Figure 6: Comparative analysis of genomic DNA following transformation with ProGly mutants expressed from the plasmid pSPRI-2, with or without excision of the kanamycin cassette as indicated. Genome sequences were constructed using the E. coli DH1 strain (NC_017625) genome as a template. The number of reads at each point of the genome were plotted using IGV software and compared. Because some regions of the genome were well covered, while others were not, a logarithmic scale was used.

domain were associated with amino acids known to play a role in the stabilisation of the flipped cytosine ring or in stabilising the structure of the entire domain. Several of these mutants are shown in the context of the secondary structure of M.HhaI in Fig. 7: there is a clustering of mutations around the structural elements associated with stabilisation of the flipped base and also in the C-terminal, conserved structural elements that in mono-specific DNMTs interact with the TRD and confer sequence specificity on the enzyme [29].

Discussion

The substitution of the catalytic Cys in monomeric C5-DNMTs by Gly, results in the formation of high affinity nucleoprotein complexes, which under certain conditions have lethal consequences for host cells [14,15]. In contrast, when a plasmid encoding the same mutation is expressed as a synthetic dimer, there appears to be only a minor impact on cell viability. However, comparisons of cell growth and plasmid transformation efficiencies, conceal a profound difference in the host response to the expression of these two forms of the same point mutation. Previous studies of this unusual mutant have shown that dissociation from its recognition sequence is approximately 50-fold lower than wild type enzyme and other point mutations [15]. A similar effect has also been observed when the catalytic Cys is replaced by Ala [14]. However, Gly and Ala are the only two proteogenic amino acids that share a lack of nucleophilic character and possess a side chain volume smaller than that of Cys. The

Figure 5b: The diagram illustrates the different consequences arising from the transformation of plasmids encoding the monomeric ProGly mutant in the enzyme M.MspI (pMspIGly-1) and the dimeric form (pMspIGly-2). In both cases output plasmids differ from the input plasmids, but importantly, the transformation efficiencies observed when the ProGly mutant is presented in vivo in a bivalent form are comparable with those observed using wt plasmids. (The bold red cross indicates that plasmids carry either, deletion,insertion, missense or nonsense mutations. Revertants in which the ProGly sequence is altered to ProCys would be viable, but have not been recovered to date.)

base pairs in the genome trap C5- DNMTs, the mode of action of the fluoroquinolones involves replication stalling arising from the irreversible formation of DNA gyrase related chromosomal complexes prior to postreplication segregation [19]. This suggests that the cell marshals subgroups of genes in a combinatorial manner to mount a response to the various types of nucleoprotein complexes that threaten normal genetic transactions.

The valency of macromolecular ligand:protein interactions has been widely discussed in the context of immunoglobulins [30] and considered from a thermodynamic perspective by Mammen et al [31] and Errington et al [32]. In considering the different outcomes of the binding of monovalent and bivalent polypeptides to specific sequences of DNA, the net fraction of a given concentration of protein bound to a site is simply a function of the number of specific base pairs in the recognition site (a 4 bp sequence occurs on average once every 256 base pairs (44) etc) and the distance between the two recognition sites in three-dimensional space [33,34]. In one example, the vertebrate CTCF DNA binding protein creates an extensive three-dimensional network of interactions through multivalent binding via its zinc finger domains [35,36]. In prokaryotes, the topological status of the chromosome is also known to be a major factor in genome maintenance, repair and several genetic transactions reviewed in [37]. Against this background, it is perhaps unsurprising that a multivalent, high-affinity DNA binding protein, such as a GST-C5- DNMT ProGly mutant, elicits a different "intensity" of DNA repair, compared with a monovalent form of the same protein. Most missense mutations have little or no impact on the function of a protein coding gene, unless the base change introduces a stop codon or replaces a critical active site or structural amino acid. The resultant reduction in activity associated with most mutant proteins is only partial, and rarely presents a life or death outcome for the cell. Moreover, where a gene encodes an oligomeric protein, such as a homo-tetramer (α4), the same is usually the case, unless subunit interactions play a key role in the regulatory function of that protein. However, as demonstrated here, the substitution of the catalytic Cys by an amino acid of shorter chain length, and lacking nucleophilic character, has a major impact on whole cell physiology. The results presented here point to a profound difference in the host cell response to a lethal gain of function point mutant resulting from differences in the oligomeric state of the mutant enzyme. In this context, all prokaryotic C5-DNMTs are monomeric apart from a small number of enzymes where their active site is constructed from two polypeptide chains (which has been mimicked in this work with the split domain, b-ZIP constructs). In vivo, bacterial C5-DNMTs catalyse sequential methylation reactions, mostly on hemi-methylated, newly replicated genomic DNA, unlike the Type II restriction endonucleases which cleave two palindromic half sites and are almost always closely coupled homodimers. A series of
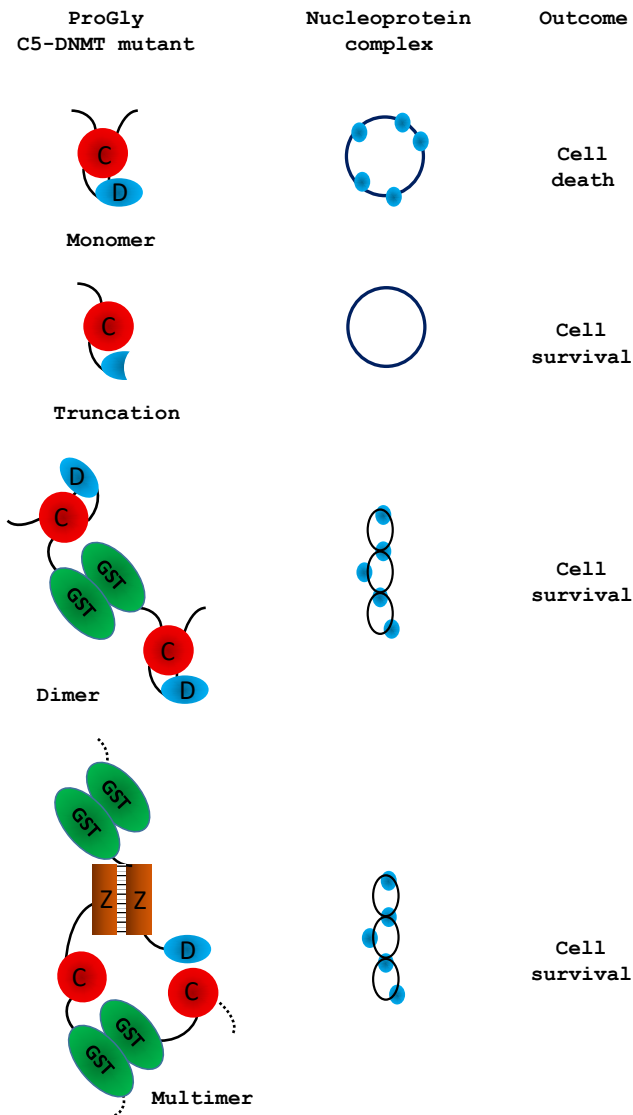

Figure 7: Schematic diagram showing the locations of point mutations recovered from cells in which the sequence encoding the catalytic N-terminal domain of an M.HhaI ProGly mutant, expressed as a GST fusion, is combined with a library of random mutants obtained by error-prone amplification of the sequence encoding the C-terminal domain of M.HhaI.

ProGly mutant meets all of the general criteria for a lethal, gain of function mutant. In a detailed study of the regulated, low level expression of a Pro-Ala mutant of M.EcoRII, [16] the authors used a wide range of repair deficient mutants to compare responses to these types of nucleoprotein lesion. These results point to the involvement of genetically distinct pathways in the resolution of M.EcoRII lesions compared with those involved in the elimination of similar covalent C5-DNMT chromosomal complexes induced by administration of 5- azacytidine or fluoroquinolone. While G:5 azacytidine

Figure 8: Schematic representation of the observed effect of expression of various forms of the ProGly mutants following transformation of cells. Only multivalent complexes elicit the mutational rescue required for cell survival: these are illustrated by a twisted circle: the ProGly mutants are shown as simple blue circles when bound to DNA.

events, consistent with the results presented here is proposed in Fig. 8.

Following transformation of a host cell with a plasmid encoding a monovalent ProGly mutant, transcription and translation generate a low level of the ProGly monomer. As the concentration of ProGly mutant rises, high affinity nucleoprotein lesions will accumulate on plasmid and genomic DNA, where there are several thousand potential recognition sites. It is well established that both plasmid and genomic DNA is a substrate for restriction and modification enzymes. Failure to recover from this state, under selective pressure from an applied antibiotic, is manifest by extremely poor transformation efficiencies and by growth inhibition [14,15].Henderson and Kreuzer [16] have shown that this observed "toxicity" arises through a major block to replication. In contrast, the negative impact of multivalent ProGly mutants (eg GST-HhaIC81G) has only a minor impact on transformation efficiency. One explanation for the results observed with multivalent C5- DNMTProGly mutants is shown in Fig. 8. Expression of a multivalent C5-DNMT mutant leads to the formation of both cis and trans nucleoprotein complexes. Multivalent ProGly mutant proteins will make monovalent-like interactions with both plasmid and chromosomal DNA; they will also make a second interaction which, if the site is on the same plasmid molecule or within the same chromosome, can be considered as intramolecular, and will contribute to a net reduction in the dissociation constant associated with multivalent compared with monovalent nucleoprotein complex formation [18]. One of the most striking observations in these experiments is the substantial levels of structural rearrangement observed in those plasmids encoding multivalent ProGly DNMT mutants. Moreover, colonies recovered from a single petri dish typically contain several plasmid variants: multiple experiments revealed similar phenomena, and in each experiment different "structural" classes of mutants emerged following each transformation (inferred from restriction mapping). Finally, the nucleotide sequences of over 100 recovered plasmids were obtained, providing evidence for a general mutational response, removing the need to invoke a specialised mechanism to explain the results presented here. Evidence that the gain of function associated with the ProGly mutation is associated with irreversibly base flipping is strongly supported by the results of a secondary mutagenesis experiment. Mutational rescue from a bivalent ProGly M. HhaI mutant can be mitigated by secondary point mutations that map onto residues associated with DNA recognition and stabilisation of the flipped target cytosine as shown schematically in Fig. 7. In the original structure determination of M.HhaI complexed with a cognate DNA duplex, residues 233-240 and 250-260 were shown to form the interface with the major groove of DNA. The contribution of one specific residue, Gln237 has been the subsequent focus of a detailed structural analysis and is one of the mutants recovered in this work. Daujotyt et al [38] have presented evidence for a role for this residue in M.HhaI in lowering the energy barrier to eversion of the target cytosine. It seems reasonable to propose that the mutants isolated in this screen neutralize the ProGly phenotype, acting as intragenic suppressors of inappropriate high affinity recognition of DNA which in turn eliminates the need for a mutational rescue event. It is clear from these data that there are many mutagenic strategies available to the host that can overcome the lethality of the ProGly mutant phenotype, but that it is only when the gain of function mutant is expressed in a multivalent context that these solutions are exploited effectively. In 2001, Loeb described the concept of

a mutator phenotype [39] in which such a cell is characterised by higher than normal levels of mutagenesis, associated with aberrant DNA replication and repair functions. It seems clear that the presence of the ProGly mutant C5- DNMT, when expressed as a bivalent, high affinity DNA binding protein, induces such a phenotype in the host. We suggest that there is a much more rapid (or at least more efficient) induction of error prone repair pathways in response to multivalent compared with monovalent high affinity nucleoprotein complex formation and we refer to this phenomenon as "mutational rescue". As a result of mutational rescue, transformation efficiencies approach WT levels, and while all surviving cells contain plasmids conferring antibiotic resistance, the GST-C5-DNMT open reading frame carries a variety of inactivating mutations. Interestingly, we observed no evidence for any widespread mutagenesis across the chromosome. Since the formation of ProGly complexes is not expected to be restricted to plasmid sites, the possibility exists that mutagenesis related events that result in the elimination of sequences encoding the toxic multivalent ProGly protein, occur rapidly following plasmid transformation, thereby minimising chromosomal damage, or that replication and subsequent cell division are delayed A more extensive set of genome sequencing would be necessary to test this more thoroughly and may have some relevance to the emerging observations of catastrophic genome remodeling in the cancer related phenomenon of chromothripsis [40]. In the majority of experiments described here, expression of ProGly mutants was achieved following plasmid transformation and analysis was carried out using a number of strategies to optimise transformation efficiency. However, while plasmid analysis provides an insight into the frequency and nature of the mutational rescue phenomenon, it overlooks events on the chromosome. Indeed, following expression of the high affinity C5-DNMT mutant in host cells, nucleoprotein complexes are just as likely to form at genomic sites, since restriction and modification is a well-established genome wide phenomenon and is not restricted to bacteriophage or extra-chromosomal elements. Analysis of the genome sequences of colonies harbouring plasmid indels, revealed no evidence for significant genome wide perturbation. Clearly, non-viable mutants would not have been recovered, but there was no indication of any significant indel events. Fig. 6 shows an alignment of the genomes analysed, with the strain E.coli DH5alphaMCR acting as a reference point. The mcr-phenotype arises from a deletion of this locus as confirmed by the whole genome sequence analysis. In addition, there was no evidence for anomalous levels of transposition as judged by the alignment of know transposable element sequences. In this work we have shown that in contrast to the dominant negative phenotype associated with in vivo expression of the ProGly mutation in monomeric DNMTs, when the mutants are expressed as synthetic dimers (or higher order oligomers), the lethal effect is abrogated by a raised level of general mutagenesis. While transformants are readily obtained following introduction of the ProGly mutant in artificially oligomerised C5-DNMTs, the input plasmids are often significantly altered and contain a complete spectrum of mutations ranging from missense mutations, to nonsense mutations and a variety of indels. Using several different C5-DNMTs, recognising different DNA sequences and by engineering in multivalency through different oligomerisation mechanisms; the efficient recovery of deactivated ProGly encoding plasmids was shown to be a robust phenomenon. Application of the fundamental principles underpinning Molecular Biology is currently being "stress tested" by Systems Biologists which in turn will improve outcomes in the field of Synthetic Biology. The successful delivery of programmable synthetic organisms and novel synthetic macromolecules draws heavily on the original model of bacterial gene regulation proposed by [4], and the subsequent validation work on the lac operon and the gene regulatory networks in bacteriophage Lambda. In a more recent study, [5] present a systems-based mathematical model based on the relationship between the specificity and interaction affinities of a series of Lac repressor mutants. These kinds of experiments will undoubtedly provide the basis of a more robust framework for designing controllable microbial systems. As the authors [5] point out, while the impact of point mutations on individual genes has provides us with considerable insight into their function, obtaining a wider systems level perspective is now technologically possible and will undoubtedly strengthen future strategies for incorporating programmable functions into cells via genome engineering. The conversion of wild type C5-DNMTs into ultra-high affinity DNA binding proteins, provides a novel platform for investigating the cellular responses arising from interference with genomic transactions. In contrast, from an evolutionary perspective, the benefits of the intrinsic bivalency of immunoglobulins provides the most compelling example of how valency can impart a positive impact on biological function. In the series of experiments, the naturally homo-dimeric GST moiety provides a simple means of tethering together two otherwise independent DNA recognition elements. Assuming equilibrium binding to DNA remains below a certain threshold, both catalytically active and inactive DNMT proteins, monovalent of multivalent, do not threaten cell survival. However, beyond this threshold, systems are in place in E.coli that allow cells to respond to the threat from multivalent high affinity DNA recognition far more efficiently than the monovalent threat. This finding could also provide a novel approach to investigating genes involved in the coordination of DNA repair and chromosome packaging in maintenance of genome integrity.

Experimental Section

Construction of Mutants

The genes encoding M.MspI, M.HhaI (a kind gift from

Citation: Hurd, P.J. Al-Swailem, A.M. Bin Dukhyil, A.A.A. Sheikh, Q.I. Al-Ghanim, A.A. Alfageih, L. Matin, M. Yueh Ting, L.u. Abdalgelel, A. Florence, J. Mohammed Al-Shemirti, Al Harbi, S. Brown, P.E. Hornby, D.P. Systemic Mutational Rescue in Escherichia Coli Elicited by a Valency Dependent, High Affinity Protein DNA Interaction. Journal of Bioinformatics and Systems Biology. 6 (2023): 97-109.

Dr. R. J. Roberts, New England Biolabs, USA) and M.SPRI (a generous gift from Prof. T.A. Trautner, Berlin, Germany) were amplified via the polymerase chain reaction (PCR) using primers to facilitate subsequent ligation of the PCR products into the glutathione-S-transferase encoding vectors (Life Technologies). The M.MspI PCR product was initially inserted into the BamHI and EcoRI sites in pGEX 2T, the M.HhaI PCR product was ligated into the EcoRI and XhoI sites in pGEX 4T whilst the M.SPRI gene was inserted into the BamHI and XhoI sites in pGEX KG. The genes encoding the above of C5-cytosine-specific DNA methyltransferases (DNMTs) (see Table 1) were constructed using standard protocols (as described earlier (Taylor et al, 1993). All C5-DNMTgenes expressed from pGEX vectors are dimeric, whilst all other constructs yield monomeric (the typical form) enzymes. The details of all constructs are available to the corresponding author.In order to introduce the apparently "toxic" ProGly mutation, a number of strategies were employed. In one approach, following the introduction of a pair of unique restriction sites flanking sequence encoding the catalytic Cys residue in M.MspI (Fig.2a), insertion of an oligonucleotide duplex facilitated the substation to create a ProGly mutant. The details are as follows: an engineered BspEI site was added 30bp upstream of the Pro Cys codons and an engineered NdeI site, 5 bp downstream of the Pro Cys codons as shown in Fig. 2a. Alternatively, the C5-MTase was first inactivated by insertion of a kanamycin cassette into the XbaI site of the M.MspI gene (or similarly into the SalI site of the M.SPRI gene or an engineered XbaI site in the M.HhaI gene). Finally, the catalytic and DNA recognition domains were expressed from separate compatible plasmids shown schematically for M.SPRI in Fig 4a. The N-terminal sequence up to the boundary of the Target Recognition element was appended via PCR with a sequence encoding the bZIP sequence from C/EBPα, [LELTSDNDRLRKRVEQLSRELDTLRGIFRQL] and inserted into pACYC184. The sequence encoding the C-terminal domain of M.SPRI was first appended with an N-terminal bZIP sequence and then inserted downstream of the GST open reading frame in the vector pGEXKG, creating an in-frame fusion as shown schematically in Fig. 4. pNSPRIGly-ZIP was maintained in E.coli AB1157 cell, which were then made competent prior to a second transformation by pCSPRI-ZIP. Plasmids were used to transform the various strains of E.coli following ligation of the appropriate mutant constructs (Fig.2b). Transformation efficiencies of all strains were monitored and normalised using a stock solution of pLITMUS28 (New England Biolabs, USA). In addition, pGEX 2T encoding a frameshifted M.SPRI mutant (pSPRIGly-2-Stop) was used to determine the efficiency of transformation of each strain. The properties of each plasmid employed in this investigation are listed in Table I and all novel plasmids were characterized by nucleotide sequencing either using the Krebs Institute Biomolecular Synthesis laboratory or the University of Sheffield Medical School DNA sequencing facility (using an Applied Biosystems Model 373A sequencer).

Transformation, plasmid and genome analysis

All transformations were carried out using strains prepared using the rubidium chloride method described in (23) and transformation efficiencies were routinely evaluated using a standard batch of the commercial supercoiled plasmid pLITMUS38 (NEB). All E.coli strains used were obtained from New England Biolabs, except for E.coli AB1157 which was a kind gift from Barry Hall (currently Bellingham Research Institute). Plasmids were recovered using mini-prep kits from Qiagen. Restriction analysis, plasmid nucleotide sequencing and DNA methylation analysis was carried out as described elsewhere [23] and genome sequencing was carried out either by the genome sequencing services at Queen Mary University or the University of Birmingham.

Acknowledgments

References

1. Pauling L, Itano H, Singer S, et al. Sickle Cell Anemia, a Molecular Disease". Science 110(1949): 543-548.

2. Ingram VM. A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. Nature 178(1956): 792-794.

3. Friedberg EC, Walker GC, Siede W, et al. DNA repair and mutagenesis. American Society for Microbiology Press 19 (2005).

4. Jacob F and Monod J. Genetic regulatory mechanisms in the synthesis of proteins. Journal of molecular biology 3(1961): 318-356.

5. Chure G, Razo-Mejia M, Belliveau NM, et al. Predictive shifts in free energy couple mutations to their phenotypic consequences. Proceedings of the National Academy of Sciences 116(2019): 18275-18284.

6. Edwards JR, Yarychkivska O, Boulard M, et al. DNA methylation and DNA methyltransferases. Epigenetics & chromatin 10 (2017): 23.

7. Hong S and Cheng X. DNA methyltransferases-role and function. Springer 16 (2016): 321-341.

8. Karreman C, de Marsac NT and De Waard A. Isolation of a deoxycytidylate methyl transferase capable of protecting DNA uniquely against cleavage by endonuclease R. Aqu I (isoschizomer of Ava I). Nucleic acids research 14(1986): 5199-5205.

9. Lee K-F, Kam K-M and Shaw P-C. A bacterial methyltransferase M. Eco HK31I requires two proteins for in vitro methylation. Nucleic acids research 23(1995): 103-108.

10. Bester TH. Cloning of a mammalian DNA methyltransferase. Gene 74(1988): 9-12.

11. Malagnac F, Wendel B, Goyon C, et al. A gene essential for de novo methylation and development in Ascobolus reveals a novel type of eukaryotic DNA methyltransferase structure. Cell 91(1997): 281-290.

12. Kouzminova E and Selker EU. dim-2 encodes a DNA methyltransferase responsible for all known cytosine methylation in Neurospora. The EMBO journal 20 (2001): 4309-4323.

13. Trautner T, Pawlek B, Behrens B et al. Exact size and organization of DNA target-recognizing domains of multispecific DNA-(cytosine-C5)-methyltransferases. The EMBO journal, 15(1996): 1434-1442.

14. Wyszynski MW, Gabbara S, and Bhagwat AS. Substitutions of a cysteine conserved among DNA cytosine methylases result in a variety of phenotypes. Nucleic acids research 20(1992): 319-326.

15. Mi S and Roberts RJ. The DNA binding affinity of HhaI methylase is increased by a single amino acid substitution in the catalytic center. Nucleic acids research 21(1993): 2459-2464.

16. Henderson ML and Kreuzer KN. Functions that protect Escherichia coli from tightly bound DNA-Protein complexes created by mutant EcoRII methyltransferase. PloS one 10(2015).

17. Ganesan A, Arimondo PB, Rots MG, et al. The timeline of epigenetic drug discovery: from reality to dreams. Clinical Epigenetics 11(2019): 174.

18. Zhou L, Cheng X, Connolly B, et al. Zebularine: a novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. Journal of molecular biology 321(2002): 591-599.

19. Bax BD, Murshudov G, Maxwell A, et al. DNA topoisomerase inhibitors: trapping a DNA-cleaving machine in motion. Journal of molecular biology 431(18) (2019): 3427-3449.

20. Henderson RM, Schneider S, Li Q, et al. Imaging ROMK1 inwardly rectifying ATP-sensitive K+ channel protein using atomic force microscopy. Proceedings of the National Academy of Sciences 93(1996): 8756-8760.

21. McTigue MA, Williams DR and Tainer JA. GST-Induced Dimerization of DNA-Binding Domains Alters Characteristics of Their Interaction with DNA. Elsevier 14(1995): 208-220.

22. Raleigh EA and Wilson G. Escherichia coli K-12 restricts DNA containing 5-methylcytosine. Proceedings of the National Academy of Sciences 83(1986): 9070-9074.

23. Taylor C, Ford K, Connolly, BA et al. Determination of the order of substrate addition to Msp I DNA methyltransferase using a novel mechanism-based inhibitor. Biochemical Journal 291(1993): 493-504.

24. Rodriguez-Martinez JA, Reinke AW, Bhimsaria D, et al. Combinatorial bZIP dimers display complex DNA-binding specificity landscapes. Elife 6 (2017): e19272.

25. Lauster R, Trautner TA and Noyer-Weidner M. Cytosine-specific type II DNA methyltransferases: (1989): 305-312.

26. Pósfai J, Bhagwat AS, Pósfai G, et al. Predictive motifs derived from cytosine methyltransferases. Nucleic acids research 17(1989): 2421-2435.

27. Klimašuaskas S, Nelson JL and Roberts RJ. The sequence specificity domain of cytosine- C5 methylases. Nucleic acids research 19(1991): 6183-6190.

28. Biles BD and Connolly BA. Low-fidelity Pyrococcus furiosus DNA polymerase mutants useful in error-prone PCR. Nucleic acids research 32(2004): e176-e176.

29. Klimasauskas S, Kumar S, Roberts RJ et al. HhaI methyltransferase flips its target base out of the DNA helix. Cell 76(1994): 357-369.

30. Karush F. Contemporary topics in molecular immunology. Springer 16 (1976): 217-228.

31. Mammen M, Choi SK and Whitesides GM. Polyvalent interactions in biological systems: implications for design and use of multivalent ligands and inhibitors. Angewandte Chemie International Edition 37(1998): 2754-2794.

32. Errington WJ, Bruncsics B and Sarkar CA. Mechanisms of noncanonical binding dynamics in multivalent protein–protein interactions. Proceedings of the National Academy of Sciences 116(2019): 25659-25667.

33. Saiz L and Vilar JM. DNA looping: the consequences and its control. Current opinion in structural biology 16 (2006): 344-350.

34. Laurens N, Rusling DA, Pernstich C, et al. DNA looping by FokI: the impact of twisting and bending rigidity on protein-induced looping dynamics. Nucleic acids research 40 (2012): 4988-4997.

35. Heath H, De Almeida CR, Sleutels F, et al. CTCF regulates cell cycle progression of αβ T cells in the thymus. The EMBO journal, 27 (2008): 2839-2850.

36. Wang DC, Wang W, Zhang L, et al. Seminars in cell & developmental biology. Elsevier 90 (2019): 4-11.

37. Uphoff S and Sherratt DJ. Single-molecule analysis of bacterial DNA repair and mutagenesis. Annual review of biophysics 46 (2017): 411-432.

38. Daujotyt D, Serva S, Vilkaitis G, et al. HhaI DNA methyltransferase uses the protruding Gln237 for active flipping of its target cytosine. Structure 12(2004): 1047-1055.

39. Loeb LA. A mutator phenotype in cancer. Cancer research 61(2001): 3230-3239.

40. Ly P and Cleveland DW. Rebuilding chromosomes after catastrophe: emerging mechanisms of chromothripsis. Trends in cell biology 27(2017): 917-930.