



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## The evolutionary pathway from a biologically inactive polypeptide sequence to a folded, active structural mimic of DNA

**Citation for published version:**

Kanwar, N, Roberts, GA, Cooper, LP, Stephanou, AS & Dryden, DTF 2016, 'The evolutionary pathway from a biologically inactive polypeptide sequence to a folded, active structural mimic of DNA' *Nucleic Acids Research*, vol. 44, no. 9, pp. 4289-303. DOI: 10.1093/nar/gkw234

**Digital Object Identifier (DOI):**

[10.1093/nar/gkw234](https://doi.org/10.1093/nar/gkw234)

**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](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# The evolutionary pathway from a biologically inactive polypeptide sequence to a folded, active structural mimic of DNA

Nisha Kanwar, Gareth A. Roberts, Laurie P. Cooper, Augoustinos S. Stephanou and David T.F. Dryden\*

EaStCHEM School of Chemistry, University of Edinburgh, The King's Buildings, Edinburgh EH9 3FJ, UK

Received November 06, 2015; Accepted March 24, 2016

## ABSTRACT

The protein Ocr (overcome classical restriction) from bacteriophage T7 acts as a mimic of DNA and inhibits all Type I restriction/modification (RM) enzymes. Ocr is a homodimer of 116 amino acids and adopts an elongated structure that resembles the shape of a bent 24 bp DNA molecule. Each monomer includes 34 acidic residues and only six basic residues. We have delineated the mimicry of Ocr by focusing on the electrostatic contribution of its negatively charged amino acids using directed evolution of a synthetic form of Ocr, termed *pocr*, in which all of the 34 acidic residues were substituted for a neutral amino acid. *In vivo* analyses confirmed that *pocr* did not display any antirestriction activity. Here, we have subjected the gene encoding *pocr* to several rounds of directed evolution in which codons for the corresponding acidic residues found in Ocr were specifically re-introduced. An *in vivo* selection assay was used to detect antirestriction activity after each round of mutation. Our results demonstrate the variation in importance of the acidic residues in regions of Ocr corresponding to different parts of the DNA target which it is mimicking and for the avoidance of deleterious effects on the growth of the host.

## INTRODUCTION

Nature is replete with examples of mimicry, both at the macro and molecular level. Phage- or plasmid-encoded antirestriction proteins that act as structural mimics of DNA alleviate the effects of host restriction during invasion of the host by phage or plasmid DNA (1–6). To date, the bacteriophage T7 protein Ocr is probably the best characterised structural mimic of double stranded DNA (7–18). Ocr is

able to competitively inhibit Type I restriction modification (RM) systems found in eubacteria.

Type I RM systems are complex oligomeric multifunctional enzymes that can mediate both restriction and modification of a DNA substrate (19–22). Recent analyses have confirmed that Type I RM systems are widely found in diverse eubacteria and provide a formidable barrier to the exchange of DNA by horizontal gene transfer or phage infection (23). The Type I RM system operates by modifying the host DNA *via* methylation of specific bases at an asymmetric bipartite recognition sequence containing a central non-specific spacer region which marks it as being ‘self’ (24). For example, the EcoKI Type I RM enzyme recognises 5'-AACNNNNNNGTGC-3' within a larger footprint and determines the methylation status of the adenine at the underlined bases (25,26). DNA entering the cell (e.g. during phage infection) that lacks the appropriate pattern of modification will be recognised as ‘foreign’ and rendered a target for degradation by the endonuclease component of the Type I RM system. The Type I RM enzyme comprises three different subunits, S, M and R, encoded by three discrete genes, *hsdS*, *hsdM* and *hsdR*, respectively. Depending on the methylation status of the DNA substrate, the RM complex functions as either a restriction endonuclease or a methyltransferase. Specifically, the M<sub>2</sub>S<sub>1</sub> complex acts as a sequence-specific methyltransferase on hemimethylated DNA, whereas the complete enzyme, R<sub>2</sub>M<sub>2</sub>S<sub>1</sub>, displays endonuclease activity on unmethylated DNA (27–29). The S subunit identifies the DNA recognition sequence, the two M subunits contain the methyltransferase activity and the two R subunits house the ATP-dependent translocase and endonuclease activity. *S*-adenosylmethionine (SAM) is the methyl donor for the methyltransferase activity, whilst the endonuclease activity requires ATP, SAM and Mg<sup>2+</sup>.

Bacteriophage T7 gene 0.3, which encodes Ocr, is the first phage gene to be expressed during the infection of *Escherichia coli* (1,2). The newly produced Ocr then binds, almost irreversibly, to the Type I RM system and protects the unmodified phage DNA as it enters the host cell by in-

\*To whom correspondence should be addressed. Tel: +44 131 650 4735; Fax: +44 131 650 6453; Email: david.dryden@ed.ac.uk  
Present address: Nisha Kanwar, University of Minnesota, BioTechnology Institute, 1479 Gortner Avenue, 140 Gortner Laboratory, St. Paul, MN 55108, USA.

hibiting the endonuclease activity (7). The structure of Ocr reveals the protein to be a remarkable mimic of bent B-form DNA approximately 24-bp in length (8,9,11). Ocr is a homodimer of a polypeptide of 116 amino acids which includes an overabundance of acidic residues (30,31). The structure features several alpha helices and loops (Figure 1a). The overall shape of the protein resembles that of a bent DNA duplex ( $\sim 34^\circ$  in the helical axis) and the numerous surface-exposed carboxylic acid groups are thought to mimic the phosphate backbone of the DNA molecule (10,11). Furthermore, the acidic sidechains can be divided into regions corresponding to the regions of DNA target sequence recognised by the Type I RM enzyme (15), namely a central region corresponding to the non-specific central part of the target sequence (coloured orange in Figure 1b) flanked by two regions corresponding to the sequence-specific parts of the target sequence (coloured green in Figure 1b) and outermost are two regions corresponding to non-specific DNA making up the remainder of the large footprint of the RM enzyme (coloured purple in Figure 1b). The bent shape of Ocr is believed to simulate the bend induced in the DNA substrate upon binding to the Type I RM enzyme (32). In this way, the 'pre-bent' Ocr molecule is energetically more favoured as a binding partner to the Type I RM enzyme than an unbent DNA molecule. Hence, the overall binding affinity of the Type I RM system EcoKI for Ocr is approximately 50-fold greater than for DNA (7,8,10,16–18). Thus, a combination of the shape component together with the electrostatic mimicry facilitates the efficient inhibition of the Type I RM system. Although Ocr is a highly effective mimic of DNA, the mimicry is not sequence specific (3,7,11). Hence, Ocr is able to inhibit all Type I RM systems, rather than one particular enzyme.

We previously investigated the electrostatic component of the mimicry displayed by Ocr. Initially, we attempted to analyse the electrostatic contribution of the mimicry by chemically modifying Ocr to neutralise the surface exposed carboxylates whilst simultaneously retaining the overall structure of the protein (13). We found that Ocr requires a minimum of  $\sim 16$  acidic residues per monomer to act as an effective antirestriction protein. However, the chemical modification procedure is essentially a random process in which surface exposed carboxylates are subject to neutralisation. As such, the positional context of acidic residues within Ocr that are particularly important for its interaction with M.EcoKI (the  $M_2S_1$  core of the RM enzyme) could not be established using this approach.

Following the chemical modification experiments, we substituted acidic residues for neutral amino acids in specific structural regions of Ocr using site-directed mutagenesis (14,15). Mutagenesis of up to 17 amino acids per monomer resulted in no significant alteration in the protein fold as based on CD analysis and structural stability studies. Examination of these mutant proteins, both *in vitro* and *in vivo*, indicated that acidic residues located in the central loop region (loop 2) of Ocr are particularly important in terms of its interaction with the archetypal Type I RM system EcoKI (Figure 1). However, these mutagenesis studies were obviously limited to analysing a relatively small number of mutated forms of Ocr. We therefore sought an inde-

pendent method to identify acidic residues of Ocr that are principally involved in its DNA mimicry.

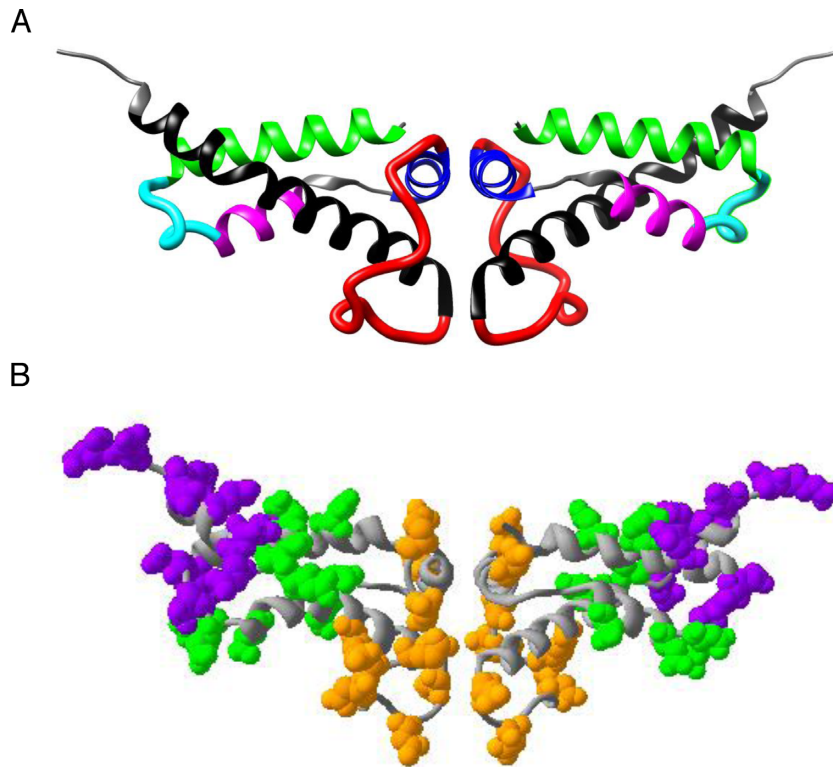
During the course of the mutagenesis studies (15) we generated a mutant form of Ocr, termed pocr, in which all 34 acidic residues had been substituted for the corresponding isogenic neutral amino acid (i.e. Asp to Asn and Glu to Gln). The pocr gene was then engineered for heterologous expression in *Escherichia coli*. However, despite numerous attempts, we were unable to detect the accumulation of recombinant pocr, suggesting the protein is unable to properly fold and/or is inherently unstable. Moreover, phage restriction assays showed pocr to display no detectable antirestriction activity. These findings were unsurprising given that pocr comprises a mutation frequency of almost 30% of the residues found in Ocr. Nonetheless, we reasoned that the gene encoding pocr could act as an appropriate starting point for a directed evolution study aimed at generating an Ocr-like protein with antirestriction activity.

The directed evolution study described in this paper was designed to introduce codons for acidic amino acids into appropriate sites in the pocr gene in order to specifically substitute neutral residues found in pocr for acidic residues found in Ocr (i.e. Asn to Asp; Gln to Glu as required). Conditions were chosen to mutate four or five codons per round of mutation in a random unbiased manner. This process can be viewed as a progressive substitution of neutral residues in pocr, which displays no antirestriction activity, towards the generation of Ocr, which obviously possesses antirestriction activity. By subjecting libraries of such pocr mutants to a powerful plate-based selection procedure after each round of mutagenesis we hoped to identify an Ocr-like protein with antirestriction activity containing a much reduced number of acidic residues.

## MATERIALS AND METHODS

### Materials

Unless otherwise stated all the chemicals used in this study were purchased from Sigma-Aldrich (St Louis, MO, USA). pBRsk1 (33) is an engineered version of pBR322 (4361 bp) in which one of the two EcoKI sites (4024–4036) has been removed by site-directed mutagenesis. Plasmid pET24a was purchased from Novagen (Madison, WI, USA) and pTrc99A was a gift from Dr Angela Dawson (School of Physics and Astronomy, University of Edinburgh). A synthetic version of the gene encoding Ocr in which all the codons for acidic acid residues had been mutated to codons for the corresponding isogenic neutral residues (Asp and Glu to Asn and Gln, respectively), termed pocr, was created by GeneArt (Regensburg, Germany). Codon usage was optimised for *E. coli* (Supplementary Figure S1). *E. coli* XL1 Blue supercompetent cells were obtained from Stratagene (La Jolla, CA, USA) and *E. coli* BL21(DE3) was from Invitrogen Life Technologies (Carlsbad, CA, USA). *E. coli* NM1261 ( $r^-m^-$ ; no RM system) and *E. coli* NM1049 ( $r^+m^+$ ; EcoKI RM system) were a kind gift of Professor Noreen Murray (School of Biology, University of Edinburgh, UK). These two NM strains were converted to DE3 lysogens as described in (6). *E. coli* NM1041 (*E. coli* MG1655  $r^+m^+$ ; EcoKI RM system  $clpX::kan$ ) and *E. coli* NM1057 (*E. coli* MG1655  $rm^+$ ; no RM system,



**Figure 1.** (A) The structure of the dimeric Ocr protein (PDB 1S7Z) showing the alpha helices (A in green, B in magenta, C in blue and D in black), Loop 1 (cyan) and Loop 2 (red). The N termini of the monomers are at the dimer interface and the C termini at the outer edges of the dimer. (B) The structure of the Ocr dimer highlighting the acidic amino acids overlapping with different regions of the DNA target recognised by Type I RM enzymes. The orange region overlaps with the central non-specific region, green overlaps with the two specific parts of the target sequence and purple overlaps with DNA flanking the target sequence.

$\Delta$ hsdR<sub>EcoKI</sub>, clpX::kan) were kind gifts from Dr Angela Dawson. Virulent unmodified bacteriophage lambda  $\lambda_{v,o}$ , or EcoKI-modified lambda  $\lambda_{v,k}$  were kindly provided by Professor Noreen Murray. Isoelectric point, pI, values were calculated using the ExPaSy compute pI/Mw tool in the bioinformatics research portal.

### Library construction

Standard DNA manipulation procedures were used throughout (34). Directed evolution of pocr was performed using a modified version of the ISOR protocol described by Herman and Tawfik (35,36). The libraries were constructed with desalted oligonucleotides obtained from Invitrogen (Paisley, UK). Initially, the pocr gene ligated into pET24b was amplified with oligonucleotides that primed from within the vector to give a PCR product of 500 bp. The resulting flanking regions of the PCR product covering part of the vector sequence facilitated manipulation of the relatively short pocr gene and assisted in the PCR reassembly step. DNase I digestion of the PCR product was conducted and the mixture was resolved on a 2% agarose gel before extracting fragments of <100 bp in length. Mutagenic oligonucleotides were designed to mutate codons for neutral amino acids in pocr to the corresponding codons for acidic residues found in Ocr (i.e. Asn to Asp; Gln to Glu). To circumvent problems associated with overlapping mutagenic oligonucleotides, the ISOR procedure was

conducted in four separate reactions targeting four discrete sets of non-contiguous codons. In order to achieve an average mutation frequency of four or five mutated codons per round of ISOR, the total concentration of mutagenic oligonucleotides was fixed at 200 nM. Equimolar amounts of each oligonucleotide were used throughout. The sequences of the four pools of mutagenic oligonucleotide primers used in this study are shown in Supplementary Table S1. The four pools of DNA obtained by ISOR were then shuffled by DNase I digestion and PCR reassembly to generate the library for screening. Primers Nfor, 5'-GTATCACCATGGCTATGTCTAACATG-3' and Nrev 5'-GATCCCCCGTCAAGCTTA-3' were designed to amplify the assembled pocr library and introduce NcoI and HindIII restriction sites (underlined) in a nested PCR reaction to facilitate insertion of the DNA into pTrc99A.

In all, three rounds of directed evolution were performed. For the final round of ISOR, degenerate oligonucleotides incorporating more than one mutation were also used. This step was necessary because the shuffling procedure was inefficient and led to under-representation in the number of consecutively mutated codons. Consequently, a bias emerged in the library after the second round that favoured mutations which were well separated from each another. Unfortunately, selecting shorter DNA fragments after DNase I digestion in the shuffling procedure made the PCR reassembly inefficient. To circumvent such problems, the third round of mutation included degenerate oligonu-

cleotides covering successive mutation sites (Supplementary Table S2). Using this strategy, a degree of randomness in the construction of the library was maintained. There was, however, a marked reduction in the fidelity of library construction in terms of the number of unwanted alterations during this round of directed evolution. Specifically, we noticed that ~10% of the library after the third round of directed evolution was made up of pocr constructs containing either deletions or frameshifts. It is unclear why the introduction of oligonucleotides containing more than one mutation should result in this reduced fidelity, although it may be related to mispriming events. Nonetheless, we judged that the library was sufficiently diverse to proceed with the selection assay.

### Selection assay procedure

Each plasmid library was initially used to transform XL-Gold Supercompetent cells at high efficiency by following the manufacturer's instructions (Stratagene). The complete transformation mix was plated to ensure the diversity of the respective library was fully represented. The resulting colonies from all of the plates were pooled using 3 ml of LB per plate and the plasmid DNA library was prepared using a Qiagen Maxiprep kit (Qiagen, Hilden, Germany). Each library was quantified using a ND1000 nanodrop spectrophotometer (Thermo Scientific, Waltham, MA, USA). The plasmid library (1 or 4 ng) was then used to transform *E. coli* NM1041 cells. Quenching experiments were carried out by plating known amounts of *E. coli* NM1041 cells with increasing concentrations of 2-aminopurine (2AP). A 2AP concentration of 80 µg/ml was found to be effective for selecting 2000 colonies of *E. coli* NM1041 cells per plate (90 mm diameter Petri dish). Therefore 100 ng of plasmid DNA was used to transform 200 µl of *E. coli* NM1041 chemically competent cells and 100 µl aliquot (~2000 colonies) of the transformation mix was spread on each plate. To analyse  $1.2 \times 10^6$  clones for each library, the selection process was split into three batches. For each batch, 20 separate transformations of *E. coli* NM1041 with the plasmid DNA library were plated onto LB-agar plates supplemented with 2AP (80 µg/ml) and incubated overnight at 37°C. The plates were then left at room temperature for three days. Positive clones from the initial selection screen were re-streaked onto 2AP-containing plates (80 µg/ml) for a second screen. After overnight incubation at 37°C, any positive clones were picked and plasmid DNA prepared. The plasmid was used to transform *E. coli* NM1041 which were selected for 2AP resistance as before. Mutated versions of pocr isolated from the 2AP-resistant cells were subsequently sequenced (Genepool, Edinburgh University, Edinburgh, UK).

### *In vivo* phage restriction and modification assay

The activity of the selected pocr mutants was assessed using an *in vivo* assay with virulent bacteriophage  $\lambda_{v.o}$  or  $\lambda_{v.k}$  as described previously (5).

### Protein purification

Recombinant protein was obtained using *E. coli* BL21(DE3) as host and purified as described previously (13,14). M.EcoKI and EcoKI were prepared as previously described (27,37). All pocr and Ocr protein concentrations refer to the dimeric form of the protein.

### Glutaraldehyde crosslinking

Chemical crosslinking was used to assess whether each of the pocr mutant proteins exist as a homodimer in solution. A 25 µg aliquot of each protein was processed using the method described in (15).

### Circular dichroism for assessment of pocr folding

Far UV (190–260 nm) circular dichroism (CD) analysis was performed on a Jasco model J-180 spectropolarimeter (Jasco Corporation, Tokyo, Japan) at 25°C in 10 mM sodium phosphate buffer pH 8.0, containing 50 mM NaF and 7 mM mercaptoethanol at a protein concentration of 30 µM using a 1 mm pathlength cell. Spectra were corrected for buffer contribution. Each spectrum was an accumulation of four individual scans.

### Thermal denaturation

The transition melting temperature  $T_m$  of the protein was measured by a thermal denaturation assay. The assay measures the fluorescence emitted by the dye SYPRO orange (Invitrogen, Burlington, Canada), which is strongly quenched in an aqueous environment. The SYPRO orange (5000-fold dilution of stock provided by Invitrogen) was added to the protein samples in 10 mM HEPES buffer, pH 8.0 and the temperature was controlled using a qPCR machine.

### Isothermal titration calorimetry

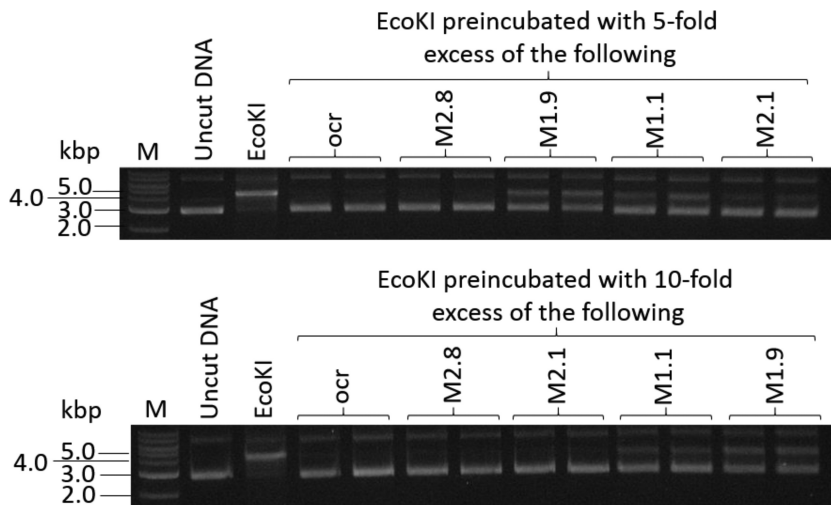
The binding of selected pocr mutants to M.EcoKI was assessed by isothermal titration calorimetry (ITC) as described in (15) but with the substitution of 20 mM HEPES buffer, pH 7.0.

### *In vitro* DNA cleavage assay

The ability of each pocr mutant to inhibit the endonuclease activity of EcoKI was assessed *in vitro* as described in (15) with the substitution of 10 mM HEPES buffer, pH 7.0 and the use of 5:1 and 10:1 ratios of pocr mutant to EcoKI and a reaction time of 10 minutes.

### Microscopy and bacterial growth curve experiments

These experiments were performed as described in (15). Growth curves were determined using *hsdR*<sup>-</sup> and *hsdR*<sup>+</sup> strains *E. coli* NM1261 and *E. coli* NM1041.



**Figure 2.** Endonuclease assay to analyse the inhibition of EcoKI nuclease activity by the pocr mutants. Unmethylated pBRsk1 was digested with EcoKI to give a linear species of  $\sim 4.4$  kb. The experiment was also repeated using EcoKI that had been preincubated with either Ocr or a pocr mutant prior to the reaction. In each case the reactions were performed in duplicate. The ratio of EcoKI to Ocr or pocr was either 1:5 (top gel) or 1:10 (bottom gel). The reaction products were resolved on a 0.9% agarose gel. Size marker; 1 kbp ladder (NEB).

## RESULTS

### Directed evolution of pocr

We aimed to progressively mutate all 34 targeted codons of the pocr gene in an unbiased manner. To achieve this objective, we employed a modified version of the ISOR methodology (35,36) in which codons for acidic residues replaced codons for neutral amino acids in a stepwise fashion to generate a series of libraries. ISOR is a DNA shuffling technique, which generates mutations by assembly of the gene of interest using oligonucleotides containing the desired mutation at the appropriate position. In general, conditions were chosen that favoured the introduction of an average of four or five mutated codons for each round of directed evolution.

To the best of our knowledge, ISOR has not previously been used to mutate small genes, such as pocr, where the targeted sites for mutation are often in close proximity to one another. As far as possible, mutagenic oligonucleotides were designed with a similar melting temperature ( $T_m$ ) and GC content in order to reduce bias during the assembly process. For each round of mutation, separate mutagenesis reactions were carried out in which four different sets of codons were independently targeted (Supplementary Figure S2 and Supplementary Table S1). Using this strategy, codons in close proximity to one another could be mutated without the mutagenic oligonucleotides overlapping. Herman and Tawfik (35) reported that the mutation frequency in this procedure could be enhanced by increasing the concentration of mutagenic oligonucleotides. In our experiments, a desired mutation rate of approx. four to six mutations per round of mutagenesis was achieved using an oligonucleotide concentration of 200 nM (oligonucleotide concentrations  $\geq 360$  nM inhibited the reaction). To reduce the possibility of introducing bias, an equimolar amount of each oligonucleotide was used in all cases.

The mutations introduced in each of the four separate reactions produced four sub-libraries, which were subsequently mixed and shuffled by DNase I digestion followed

by PCR reassembly. Analysis of the resulting library determined whether the shuffling process was efficient. A total of 30 clones were randomly picked from the library and plasmid DNA was prepared from each and sequenced in order to monitor the average number of introduced mutations. In this way, we were able to assess the overall fidelity of the procedure. Our analysis suggested an average of five mutations were generated during the first round of mutation with no clones containing undesired changes (e.g. non-targeted mutations, frameshifts, truncations, duplications, etc.). A small proportion ( $\sim 1$ –3%) of the library was made up of plasmid with no insert (empty vector). This was not deemed to be a problem because empty vector was lost from the process during the subsequent amplification of the pocr insert prior to the next round of mutagenesis. The shuffled library was then subjected to the first round of the selection procedure. A plate-based assay was used to screen approximately  $1.2 \times 10^6$  clones from the first library. No clones displaying antirestriction activity were detected using this initial selection process. Therefore, total DNA from the first library was pooled and subjected to further rounds of ISOR mutagenesis followed by selection.

In all, three rounds of mutation using the ISOR procedure were required before the appearance of clones with detectable antirestriction activity. Sequence analysis of the second and third round libraries showed the average number of mutated codons was 6–13 and 8–17, respectively. Approximately  $1.2 \times 10^6$  variants from each library were screened after each successive round of directed evolution.

### Selection process

For screening, the library was ligated into the NcoI and HindIII sites of the pTrc99A vector and the ligation mix was used to transform *E. coli* NM1041. This strain has a functioning Type I RM system (EcoKI) but lacks the ClpXP protease, which destroys the R subunit of EcoKI when DNA damaging agents are present. As such, the *E.*

*coli* NM1041 strain is exquisitely sensitive to the presence of 2AP, which is both an adenine analogue and a powerful mutagen. When 2AP is present, unmodified sites occur on the host chromosome and EcoKI cleaves the chromosome. As such, *E. coli* NM1041 plated on solid medium containing 2AP at a concentration of  $>20 \mu\text{g/ml}$  will not survive. However, the cells are able to grow if the strain harbours an inhibitor of the restriction endonuclease, such as Ocr.

Thus, the selection process involved plating out the *E. coli* NM1041 transformed with the library of pocr mutants onto solid LB medium supplemented with  $50 \mu\text{g/ml}$  carbenicillin and  $80 \mu\text{g/ml}$  2AP. The presence of a pocr mutant that can act as an effective DNA mimic will inhibit the restriction endonuclease and enable survival. Any clones that grew were picked and then subjected to a second round of screening using 2AP.

No positive clones were isolated from the first or second rounds of directed evolution. However, the third round of screening generated seven positive clones (M1.1, M1.4, M1.7, M1.9, M2.1, M2.8 and M2.11). The clones were verified by retransformation of *E. coli* NM1041 with the isolated plasmids and then selection using plates supplemented with 2AP. Plasmid DNA was isolated from each of the seven positive clones and the insert was sequenced. Clones M1.7 and M1.9 were found to be identical to each other as were clones M1.4, M2.1 and M2.11. Thus, our analysis revealed four of the seven clones contained unique sequences. In all four cases, mutations of pocr were confined to the targeted codons with no other unanticipated changes. Moreover, each of the introduced point mutations resulted in a Glu for Gln or Asp for Asn substitution and the total number of mutations in the four isolated clones were 16, 16, 14 and 17 for M2.1, M1.1, M1.9 and M2.8, respectively.

### ***In vivo* phage restriction assay**

The activity of the selected pocr mutants was assessed *in vivo* using a restriction assay (Table 1). Infection was carried out with  $\lambda_{v.o}$  of two different *E. coli* strains, transformed with pTrc99A or the plasmids expressing the pocr mutants or Ocr as a control. *E. coli* NM1049(DE3) contains the EcoKI RM system ( $r^+m^+$ ) while *E. coli* NM1261(DE3) lacks the RM system ( $r^-m^-$ ). The efficiency of plating (eop) of  $\lambda_{v.o}$  on *E. coli* NM1049(DE3) pTrc99A compared to *E. coli* NM1261(DE3) pTrc99A was  $5 \times 10^{-5}$  indicating that EcoKI was active. Comparing the two strains when transformed with the plasmid expressing the wild-type Ocr gave a similar number of plaques and an eop of 4, demonstrating that Ocr had inhibited the Type I RM system as anticipated. Assaying the four selected pocr mutants in the same way showed that all were fully active with an eop ranging from 1 to 3. Control experiments using  $\lambda_{v.k}$  showed there was no cut back in the number of plaques on the restrictive versus the non-restrictive strain.

### ***In vivo* phage modification assay**

To assess the methylation activity of each of the mutants, a phage modification assay was carried out.  $\lambda_{v.o}$  was used to infect the modifying host *E. coli* NM1057 ( $r^-m^+$ ) transformed with either pTrc99A or with pTrc99A expressing

Ocr or the selected pocr mutants. *E. coli* NM1057 expresses M.EcoKI so phages recovered from a single plaque after infection should have been methylated by the action of M.EcoKI unless the pocr mutants successfully inhibited the methyltransferase activity. Activity was assessed by determining the eop of the phages recovered from a single plaque on *E. coli* NM1057. The eop was determined on the restrictive host *E. coli* NM1041 ( $r^+m^+$ ,  $clpX^-$ ) relative to the modifying host *E. coli* NM1057 ( $r^-m^+$ ,  $clpX^-$ ). If the pocr mutant displays antimodification activity similar to Ocr then the titre of recovered phages per ml after plating  $\lambda_{v.o}$  on *E. coli* NM1041 would be lower than the titre on *E. coli* NM1057. The results showed that wild-type Ocr and all of the selected pocr mutants were active in preventing DNA modification by M.EcoKI as their eop was  $1 \times 10^{-4}$  to  $3 \times 10^{-4}$  (Table 1).

### **Protein purification**

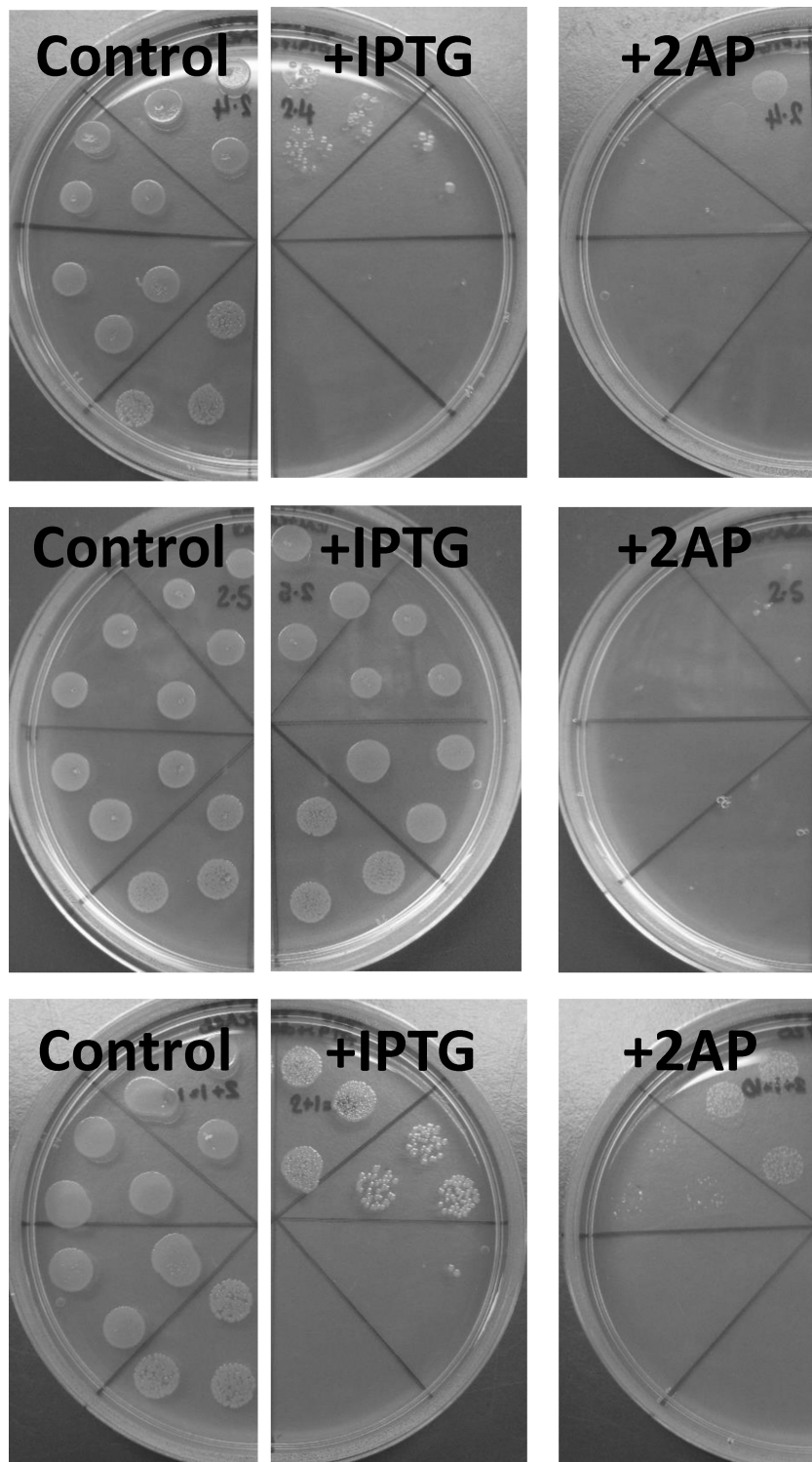
All four of the unique pocr clones selected from the third round of mutation gave similar levels of expression, equivalent to that observed for wild-type Ocr and were purified to apparent homogeneity using the same protocol employed for the purification of wild-type Ocr (data not shown, 13,14).

### **Evaluation of the protein fold of the pocr mutants**

Far UV CD analysis of each of the four pocr mutants was performed along with Ocr for comparison (Supplementary Figure S3). The spectra were analysed using the Dichroweb online secondary structure deconvolution program (38). The CDSSTR method gave a  $\alpha$ -helical content of  $\sim 60\%$  for Ocr, which is in good agreement with the crystal structure. Similar values ( $\pm 2\%$ ) were obtained for each of the pocr mutants, indicating that the absence of approximately half of the acidic residues present on Ocr has no significant effect on its secondary structure. Thus, it appears that many of the acidic residues on the wild-type Ocr protein have no substantial bearing on its overall structural integrity. Wild-type Ocr is known to exist as a homodimer in solution. Glutaraldehyde crosslinking confirmed that the selected pocr mutants also exist as dimers (Supplementary Figure S4). The thermal stability of the pocr mutants was assessed using a fluorescence-based denaturation assay. Wild-type Ocr, M1.1, M2.1 and M2.8 all gave a similar  $T_m$  value of  $54\text{--}57^\circ\text{C}$  under the conditions used in our assay. Intriguingly, however, mutant M1.9 displayed a  $T_m$  of  $67^\circ\text{C}$ , indicating an enhanced thermal stability (Supplementary Figure S5). The curious enhancement in thermal stability of M1.9 may be related to an increase in the distance between discrete carboxylates on the protein surface, which reduces electrostatic repulsion. Indeed, the closeness of the carboxylates in Ocr is approximately equal to the Bjerrum length, leading to electrostatic repulsion energies of the order  $k_B T$  (39).

### **Isothermal titration calorimetry**

ITC was used to analyse the interaction between M.EcoKI and each of the four unique pocr proteins. Wild-type Ocr was used as a positive control. Each injection peak was integrated and the heat of dilution subtracted. A plot of the



**Figure 3.** Cell growth of *E. coli* NM1041 ( $r^+$ ) after transformation with different mutants from the ISOR library. Top: a partially active mutant (mutant 2.4). Middle: an inactive mutant (mutant 2.5). Bottom: a toxic mutant (mutant 21+10). From left to right each panel shows growth on LB-agar supplemented with ampicillin (50  $\mu\text{g}/\text{ml}$ ), with 50  $\mu\text{g}/\text{ml}$  IPTG and ampicillin, and lastly with 50  $\mu\text{g}/\text{ml}$  IPTG, 20  $\mu\text{g}/\text{ml}$  2AP and ampicillin. Serial dilutions of cultures were spotted in triplicate in an anticlockwise or clockwise fashion.



**Table 1.** *In vivo* antirestriction and antimodification activity of Ocr and selected pocr mutant proteins

Plasmid	Efficiency of plating on strain transformed with plasmid	Efficiency of plating after recovery from modifying strain transformed with plasmid
Vector pTrc99A	$5 \times 10^{-5}$	1
Ocr	4	$1.4 \times 10^{-4}$
M2.8	2	$2.7 \times 10^{-4}$
M2.1	2	$1.7 \times 10^{-4}$
M1.1	3	$2.1 \times 10^{-4}$
M1.9	1	$1.7 \times 10^{-4}$

The titre of phage  $\lambda_{v.o}$  per millilitre was determined in *E. coli* NM1261(DE3) ( $r^-m^-$ ) and *E. coli* NM1049(DE3) ( $r^+m^+$ ) transformed with either the vector alone (pTrc99A) or plasmid expressing Ocr or pocr mutant (M2.8, M2.1, M1.1 or M1.9). The ratio of phage per ml in *E. coli* NM1049(DE3) to phage per ml in *E. coli* NM1261(DE3) was calculated to obtain the efficiency of plating (eop). For the antimodification assays, unmodified phage  $\lambda_{v.o}$  grown on *E. coli* NM1057 ( $r^-m^+$ ,  $clpX^-$ ) in the presence or absence of Ocr or pocr mutant were recovered and a phage stock generated. The degree of antimodification was assessed by determining the eop of each phage stock on a restriction-proficient strain (*E. coli* NM1041 ( $r^+m^+$ ,  $clpX^-$ )), relative to a non-restricting strain (*E. coli* NM1057 ( $r^-m^+$ ,  $clpX^-$ )). Errors in the number of plaques were  $\pm 20\%$  from full plate assays performed in triplicate.

heat of interaction *versus* the molar ratio of Ocr dimer to M.EcoKI was then generated (Supplementary Figure S6). As found previously (10,13), the interaction between wild-type Ocr and M.EcoKI was highly exothermic and exhibited essentially stoichiometric binding (one Ocr dimer per M.EcoKI) within the concentration range used in these experiments (Supplementary Figure S6). However, the binding of Ocr to M.EcoKI was too strong to allow determination of the dissociation constant, which has been estimated as  $\sim 50$  pM using other methods (10). For the pocr mutants the enthalpy of interaction with M.EcoKI varied with each mutant. The pocr mutant proteins M2.1, M2.8 and M1.1 displayed a highly exothermic interaction with M.EcoKI similar to that observed for wild-type Ocr under the same experimental conditions. By contrast, the interaction of M1.9 with M.EcoKI showed a markedly reduced enthalpy of interaction (Supplementary Figure S6).

### Nuclease assay

*In vitro* nuclease assays were performed to analyse the level of inhibition exerted by the mutated forms of pocr (Figure 2). Unmethylated pBRsk1, containing a single EcoKI recognition sequence, was used as substrate. A 10-fold excess of Ocr/pocr was pre-incubated with EcoKI prior to the addition of plasmid substrate and cofactors. Linearisation of the plasmid DNA was assessed by subjecting the reaction mixture to agarose gel electrophoresis. Using a molar ratio of 1:10 of nuclease to antirestriction protein, complete inhibition of EcoKI by Ocr, M2.8 and M2.1 was observed. By contrast, the pocr mutants M1.1 and M1.9 displayed only partial inhibition of the endonuclease. The experiment was then repeated using a reduced molar ratio of 1:5 nuclease to antirestriction protein. Under these conditions, partial inhibitory activity was observed for pocr mutant M2.1 as well as M1.9 and M1.1. However, the mutant M2.8 together with wild-type Ocr retained full inhibitory activity.

### Analysis of further mutants from the ISOR libraries

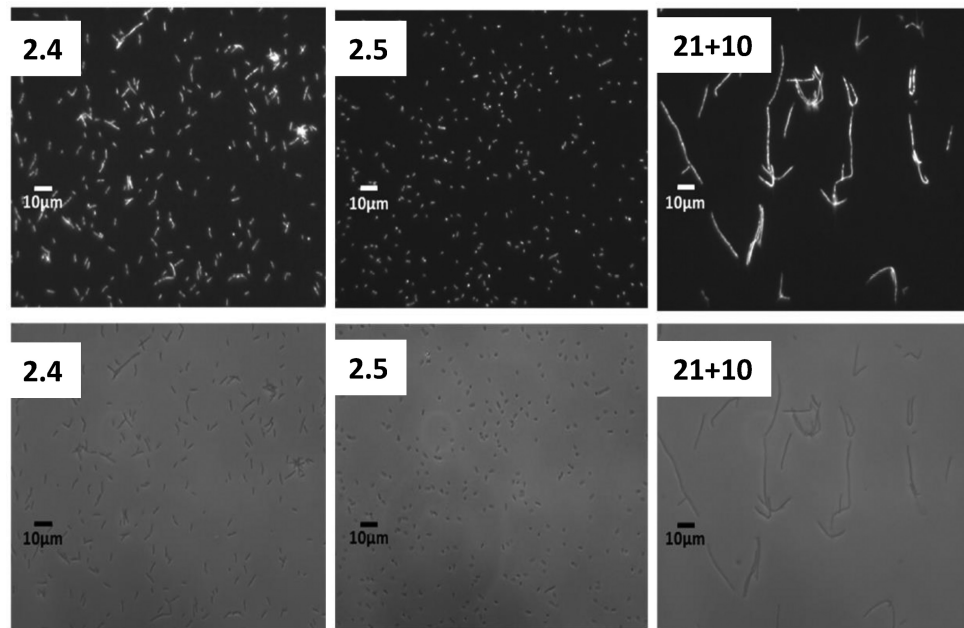
The recovery of only four active protein sequences from the millions of sequences generated by the ISOR procedure was disappointing and suggests that the evolutionary path from pocr to Ocr is difficult. Thus we decided to return to the first three libraries generated and to select at random a

group of 65 sequences for further analysis using our range of *in vivo* techniques (Supplementary Table S3). Eleven of the group proved to give irreproducible results when growing on media containing 20  $\mu\text{g/ml}$  2AP, sometimes growing and sometimes not, so these were not analysed further. This group contained two, seven and two members from libraries 1, 2 and 3 respectively.

The remaining 54, made up of 21, 15 and 18 representatives from libraries 1, 2 and 3 respectively, gave reproducible results in the 2AP assay showing several different behaviors in a strain expressing HsdR, namely, limited growth in the presence of IPTG and even more limited growth in the presence of 20  $\mu\text{g/ml}$  2AP (without filamentation—see later), growth in the presence of IPTG but not when 20  $\mu\text{g/ml}$  2AP was also present and limited growth in the presence of IPTG and even more limited growth in the presence of 20  $\mu\text{g/ml}$  2AP (with filamentation – see later) (Figure 3). We term these three situations as Partially Active (PA), Inactive (IN) and Toxic, respectively. Previously we have found fully active mutants and three partially active mutants, Mut12, Mut16 and Ocr/pocr, but the three partially active mutants were not able to grow at all in the presence of 20  $\mu\text{g/ml}$  2AP (15). These partially active mutants are therefore distinct from the current PA mutants, which do grow a little in the presence of 2AP and are therefore listed as PA\* in Supplementary Table S3.

The ISOR mutants in groups PA, IN and Toxic were tested using the antirestriction and antimodification assays and this revealed that, apart from one instance, these three groups possessed no antirestriction or antimodification activity (Supplementary Table S3). The mutant 3.11 showed full antirestriction and antimodification activity (antirestriction value of 0.37 and antimodification value of  $1.9 \times 10^{-4}$  for comparison to values in Table 1) despite being classified as Toxic in the presence of IPTG. Leaky expression in the absence of IPTG has been found previously to be sufficient for antirestriction and antimodification activity (11) so this apparently contradictory behaviour of mutant 3.11 can be reconciled. The antirestriction and antimodification activity of this mutant indicates that the protein is folded.

Microscopy revealed that the IN and Toxic mutants induced filamentation indicative of a deleterious effect on cell division and suggestive that they are interacting with other



**Figure 4.** Microscopy images of *E. coli* NM1041 ( $r^+$ ) transformed with different mutants from the ISOR library. The top row shows DAPI fluorescence from stained chromosomes while the bottom row shows phase contrast images of the whole cells. From the left to right the cells are transformed with a partially active mutant (mutant 2.4), an inactive mutant (mutant 2.5) and a toxic mutant (mutant 21+10).

proteins (Figure 4 and Supplementary Table S3). This occurred in both *hsdR*<sup>-</sup> and *hsdR*<sup>+</sup> strains.

Analysis of growth curves for the three main groups of mutants in *hsdR*<sup>-</sup> and *hsdR*<sup>+</sup> strains showed that the PA mutants grew normally in the absence of HsdR but grew slowly in its presence, thus indicating that the mutants are capable of interacting with HsdR even though not sufficiently strongly to inhibit restriction in the antirestriction assay (Figure 5, Supplementary Table S3). These findings suggest this group of mutant proteins are folded *in vivo*. The IN group of mutants had no effect on the growth curves irrespective of whether HsdR was present or not (Figure 5, Supplementary Table S3). Thus, it is not clear from this experiment whether these mutant proteins are capable of folding. However, the observation of filamentation (Figure 4) suggests that they are capable of interacting with other proteins in the cell so they may be folded. In addition, we have previously found that inactive mutant forms of Ocr can be purified and are folded (15). Thus, we suggest that at least some members of the IN group are folded. The mutants in the Toxic group showing no antiRM activity demonstrated slowed cell growth in both *hsdR*<sup>-</sup> and *hsdR*<sup>+</sup> strains (Figure 5, Supplementary Table S3). This observation indicates that these mutant proteins are folded and capable of interacting with HsdR as well as other protein species within the cell.

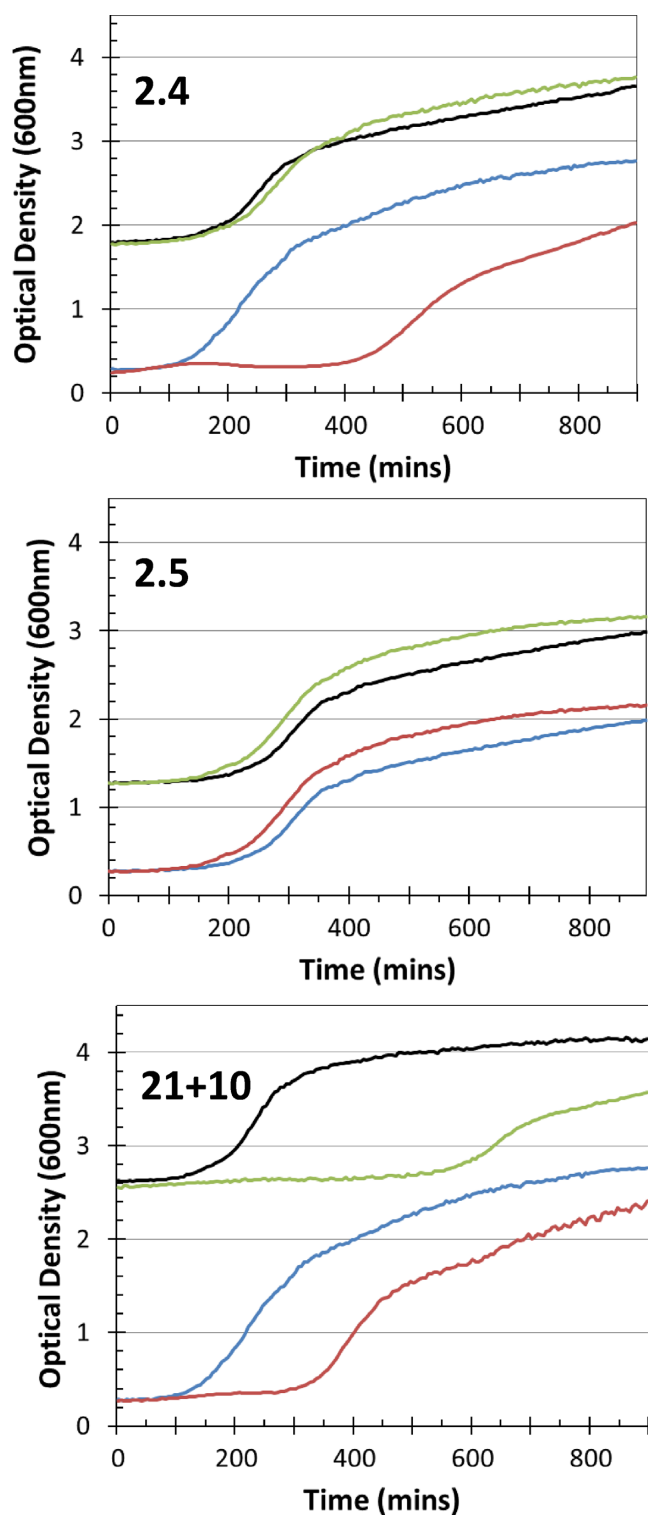
Taken together, we conclude that all of the PA and Toxic mutants selected for analysis are folded *in vivo*. Moreover, some of these mutant forms of pocr are capable of interacting with the EcoKI Type I enzyme and others can also interact with alternative proteins to interfere with cell division. We also conclude that a non-zero proportion of the IN mutants are folded.

#### Amino sequence analyses

Sequence alignments were performed on each group of ocr/pocr mutant forms, namely Active, PA\*, PA, Inactive and Toxic. The complete alignments for each group of variants are given in Supplementary Figure S7. Although it is apparent that sequences associated with non-active variants of the wild-type Ocr protein tend to have fewer acidic residues, no obvious straightforward pattern associated with the groupings derived by our experimental results is immediately discernible. A further alignment derived from a BLASTP search of non-redundant protein sequences using the wild type Ocr sequence was performed, Supplementary Figure S8. This alignment revealed that homologues to Ocr contained very similar patterns and numbers of acidic residues to Ocr.

To pull out any trend in the arrangement or number of negative residues in a given Ocr or pocr variant, a number of further analyses were performed. Firstly, the number of negative residues in the complete sequence, in the three sections mimicking the different parts of the DNA target shown in Figure 1b, and in the two loops (loops 1 and 2) were determined for each Ocr and pocr variant, Supplementary Table S3. Secondly, the number of sequences in each group containing a given number of negative residues in a particular region was plotted on a bar chart, Figure 6. Lastly, the number of negative charges for each Ocr or pocr variant in each region was plotted against the total number of negative charges in the Ocr or pocr variant, Figure 7.

The bar charts, Figure 6, split the various groups of activities with it being particularly clear that an active ocr variant requires at least 14 negative residues in total and that these should be located in the ‘orange’ region shown in Figure 1b. The other groups of activities tend to overlap with each



**Figure 5.** Growth curves of *E. coli* transformed with different mutants from the ISOR library. Top: a partially active mutant (mutant 2.4). Middle: an inactive mutant (mutant 2.5). Bottom: a toxic mutant (mutant 21+10). *E. coli* NM1261 (*r*<sup>-</sup>) in green is supplemented with 50 µg/ml IPTG and in black is without IPTG. *E. coli* NM1041 (*r*<sup>+</sup>) in red is supplemented with IPTG and in blue is without IPTG. The *E. coli* NM1261 curves are offset from the zero.

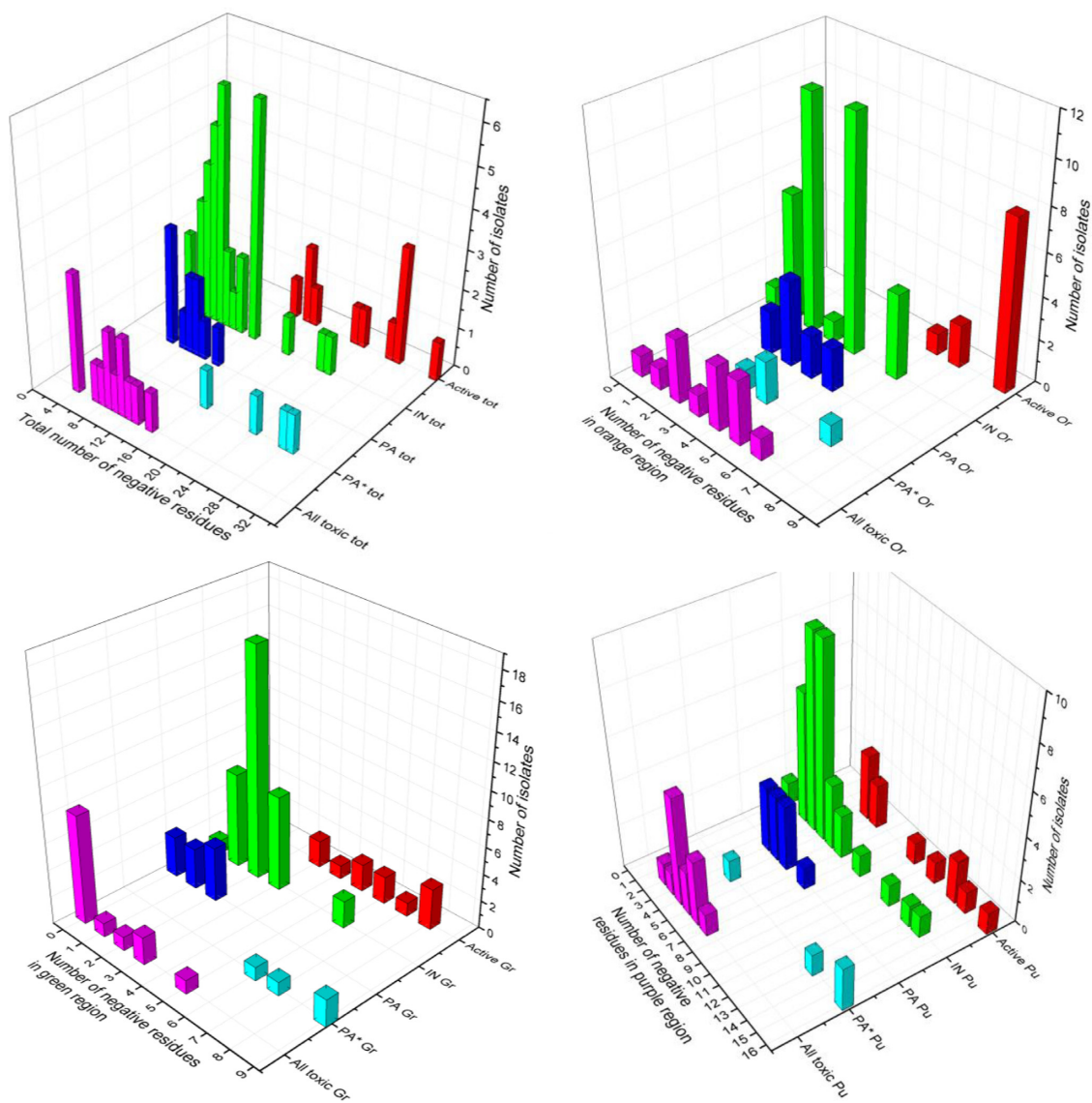
other so no further trend is very obvious in these bar charts. Figure 7 pulls out some further features of the number of negative residues to be found in each group of activities. To be active, in addition to at least 14 negative residues in total, Ocr requires an excess of charges in the orange region (at least 6 of a possible 9) and no more than average in the purple region (no >47% of the total). The number of negative charges in the green region is not too important as long as it is at least three. The active variants found in the ISOR library are all close to the boundary between inactivity and activity while those engineered previously (15) are closer to the wild type Ocr charge composition as expected from their biased construction method. This biased construction method also places some inactive variants (Mut10, Mut11 and Mut13) and the PA\* variants (Mut3, Mut12, Mut16 and ocr/pocr) within the space occupied by the active Ocr variants.

The unusual pocr variant, 3.11, which shows antiRM activity but also toxicity lies within the group of active variants isolated from the ISOR library. A range of Inactive, PA and Toxic variants from the ISOR library possess 13 negative charges and lie just below the boundary for Ocr activity (MM6, C2, 21+8, 2.4, DG3 and DG12). These particular mutant forms of Ocr show fewer than expected charges in the green region, more than expected in the orange region and no correlation in the purple region.

## DISCUSSION

The DNA mimicry of Ocr, which is crucial for its ability to act as an antirestriction protein, comprises two main features; namely, shape and charge. Here, we have investigated the charge mimicry of Ocr. The Ocr protein is replete with acidic residues, which generate a highly negatively charged surface that mimics the phosphate backbone of the DNA duplex. ITC data for the interaction between Ocr and M.EcoKI under different salt conditions is consistent with the idea that the binding is driven, at least to a large extent, by electrostatics (13). Previously, we have investigated the charge mimicry by removing charge from the wild type Ocr protein focussing either on one or two residues at a time (14) or on contiguous blocks of charge (15) and our results indicated that loops 1 and 2 included within the regions mimicking the DNA target of the RM enzyme were important for activity (Figure 1). However, this methodology is strongly biased by our knowledge of the protein structure. Here, we have approached the evolution of Ocr from the opposite direction by starting from a gene encoding no acidic residues, but with the remainder matching Ocr, and then subjecting the gene to directed evolution until activity was detected. In this way the two approaches will overlap to reveal at what point activity appears.

In order to gradually evolve the pocr gene in the desired fashion, we chose to utilise a modified version of the ISOR methodology (35,36). We were able to gradually mutate the pocr gene and select for antirestriction activity after each round of ISOR. In all, three rounds of mutagenesis were required before clones showing antirestriction activity could be detected. Seven individual clones (M1.1, M1.4, M1.7, M1.9, M2.1, M2.8 and M2.11) were selected but only four, M1.1, M1.9, M2.1 and M2.8 were unique. These protein



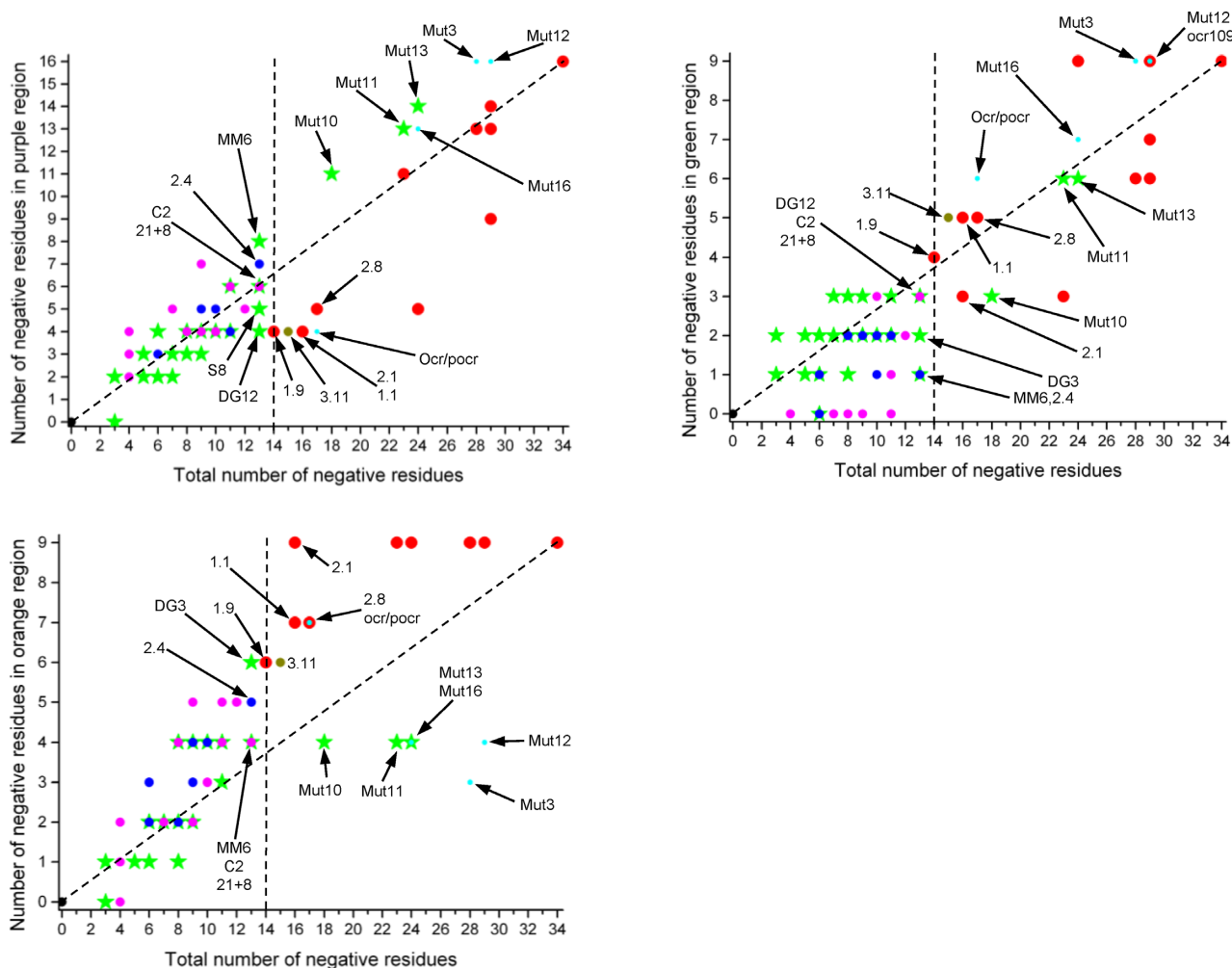
**Figure 6.** Bar charts showing the number of sequences analysed in each group of activities having a particular number of negative charges in various structural regions of Ocr. The bars are coloured as follows: Active (red), Inactive (IN, green), partially active (PA, blue), partially active from reference 15 (PA\*, cyan) and toxic (magenta).

sequences were folded and active in a range of *in vivo* and *in vitro* experiments. Thus, the evolutionary process to introduce negative charges into pocr had generated mutants of Ocr that overlap with those made by removing negative charges from the wild type Ocr. The effectiveness of the pocr proteins as antirestriction proteins follows the order: wild type Ocr = M2.8 > M2.1 > M1.1 > M1.9.

A sequence comparison of these pocr clones along with that of Ocr is shown in Figure 8. The four pocr mutant sequences display some striking similarities to one another. Firstly, the number of acidic residues in each pocr mutant is similar i.e., 14, 16, 16 and 17 for M1.9, M1.1, M2.1 and M2.8, respectively. At least 50% of the acidic residues introduced into the four pocr clones are clustered in the two loop regions of Ocr (Figures 1a and Figure 8). Sequence alignments reveal M2.1 contains all seven acidic residues within loop 2, and mutant M2.8 includes all five acidic residues in

loop 1 (Figure 8). These findings support the proposal that acidic residues in the two loop regions (particularly loop 2) are important for antirestriction activity as previously suggested on the basis of site directed mutagenesis studies (15). If from these results and the structural features highlighted in Figure 1, we propose that the acidic residues in the two loops or alternatively, the green and orange regions identified as mimicking the sequence-specific and the central non-specific parts of the Type I RM enzyme target sequence, are particularly significant for the interaction of Ocr with Type I RM enzymes, then one might anticipate that these amino acids would be well conserved. A sequence alignment of Ocr homologues reveals that there is indeed a very high degree of conservation of acidic residues in these loops and DNA-mimicking regions (Supplementary Figure S8).

However, it must be borne in mind that pocr itself does not appear to fold into a stable protein. Hence, some of

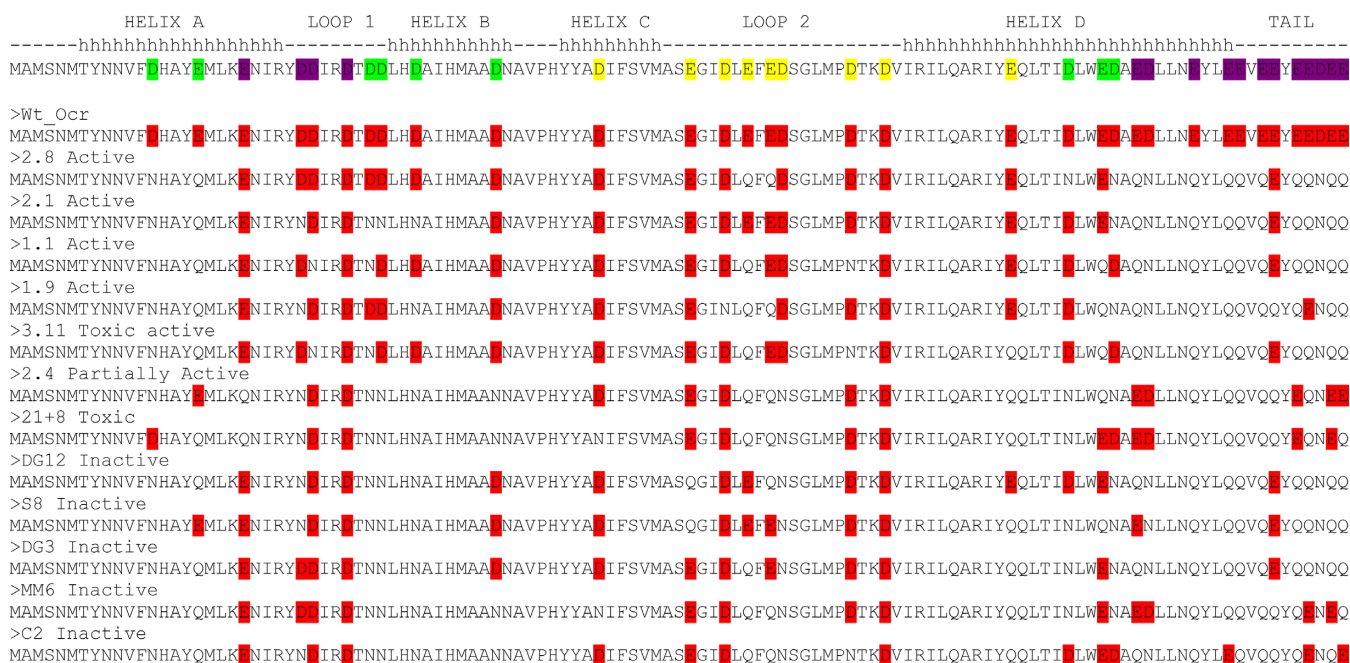


**Figure 7.** Plots of the number of negative charges in the three regions defined in Figure 1b against the total number of negative charges in the whole sequence for each Ocr and pocr variant. The points are coloured as follows: Active (red), Inactive (green), Partially Active (blue), Partially Active from reference 15 (PA\*, cyan), Toxic (magenta), Toxic mutant 3.11 (olive) and pocr (black). Various mutants are highlighted including the MutX mutants from reference 15. The vertical dashed line indicates a total negative charge of  $-14$  while the other dashed line joins pocr to wild type Ocr.

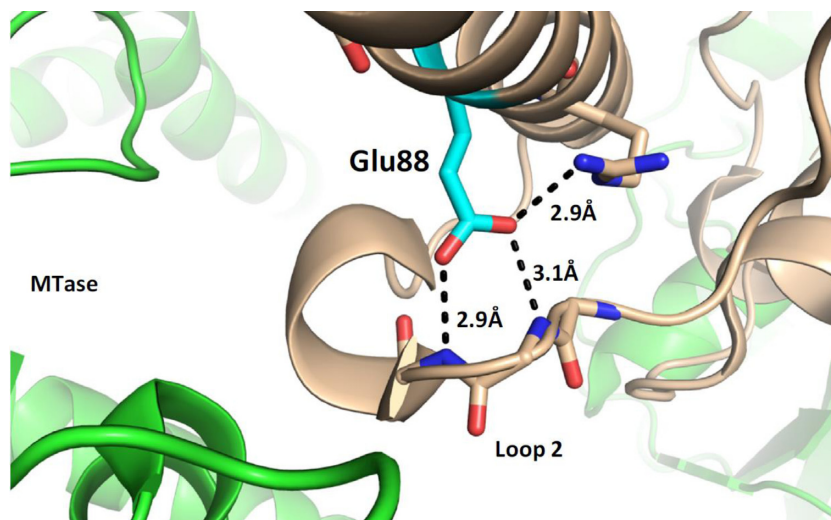
the amino acid substitutions introduced into selected pocr mutants during the process of directed evolution will be structurally important. Initially, we focussed on evolving antirestriction activity, but only a small number of positive mutants were isolated. To fill in the gap from inactive pocr to active mutants M1.1, M1.9, M2.1 and M2.8, we re-examined the three libraries produced in the first rounds of ISOR. A survey of a range of pocr mutants from these libraries revealed several behaviours falling short of full antirestriction activity, which were grouped as Inactive, Toxic or partially active (PA) (Figures 3-5, Supplementary Table S3). Given that most of these mutant forms of pocr elicited a significant effect on the growth characteristics of their hosts, it is clear the proteins are being expressed and are able to fold. The analysis of fully active pocr and Ocr mutants and homologous proteins indicates that the amount of charge in the regions mimicking the DNA target sequence and the non-specific central part of the target (the green and orange regions of Figure 1b, respectively) and the overall total charge are important for activity, Figures 6 and 7.

Figure 7 shows plots of all of our mutants derived from Ocr (15) and from pocr using axes for total negative charge and total negative charge in various defined regions of the structure. These plots reveal a continuum of mutants stretching from the inactive pocr to the fully active Ocr. The three groups of mutants, inactive, PA and toxic, lacking antirestriction and antimodification activity all overlap with each other indicating that the fine details of the location of the negative charges can have a large effect on the behaviour of the polypeptide. These three groups are separate from the active mutants.

Of particular interest is the region at which the four active pocr mutants found in this work appear. In the middle of this group lies a very unusual pocr mutant, 3.11, which showed full antirestriction and antimodification activity but whose overexpression resulted in toxicity and cell death. Just next to this active group and mutant 3.11 lie several Inactive, PA and Toxic variants. A sequence alignment of these variants with 3.11 and the active Ocr sequences is given in Figure 8. Crucially all these inactive, PA or toxic



**Figure 8.** Sequence alignment of the active pocr clones selected from the third round of directed evolution and those found to be inactive, toxic or partially active but being very close in charge distribution to active mutants. Acidic residues are highlighted by red shading. The secondary structure elements of Ocr (helix = h, loop = dash) and the regions corresponding to different parts of the DNA target (highlighted in green, orange and purple) as shown in Figure 1b are also given.



**Figure 9.** The structural model of the complex of M.EcoKI (green) bound to Ocr (light brown, 26) focussing on the possible hydrogen bonding interactions in the Ocr structure between E88 (cyan) in Helix D with the amide groups of the main peptide chain of Loop 2 (PDB:2Y7C). The residues and main chains are displayed using sticks and the hydrogen bonds using black dotted lines. The bond lengths are all  $\sim 3$  Å.

mutants are lacking negative charges on Helix B in the green region of the structure and lack aspartate at position 88 (except DG12). This E88 location is conserved in active forms of the mutants and in naturally occurring relatives of Ocr (Figure 8, Supplementary Figure S8 and Supplementary Table S3). Thus, based on sequence analysis; (i) activity requires sufficient negative charge in the green and orange regions, and (ii) toxicity is avoided by maintaining aspartate at position 88. The avoidance of toxicity would also provide the evolutionary pressure to move away from this small re-

gion of the plots in Figure 7 by acquiring further negative charge.

Examination of the crystallographic structure of Ocr shows that E88, which is part of Helix D, can potentially make several hydrogen bonds to residues in loop 2 (Figure 9). If the ability to make these hydrogen bonds is removed, as in the toxic mutant 3.11, then loop 2 will be much more flexible. This may allow loop 2 to form unwanted interactions with proteins involved in cell division and hence cause growth problems for the host cell. It is worth recalling that

Ocr has been shown to bind to the RNA polymerase in *E. coli* so the ability to form extraneous interactions is not unprecedented (40).

The number of potential variants of ocr containing 34 possible negative or neutral amino acids is  $2^{34}$  or  $1.7 \times 10^{10}$  and we screened  $\sim 1.2 \times 10^6$  of this number. In our screen we found only four unique active sequences or 0.0003% of the  $1.2 \times 10^6$  sequences screened. If this 0.0003% of active sequences is true for the  $2^{34}$  variants then we expect the  $2^{34}$  variants to include 0.0003% of  $2^{34}$  active sequences or  $\sim 2^{16}$  active sequences indicating that an active Ocr sequence should contain about 16 negative charges. This is approximately the same as found experimentally. Obviously Ocr has not really evolved from pocr but will have started from some other location in the plots shown in Figure 7 and most probably with an 'average' number of negative and positive side chains (12 of each would be expected rather than the 34 and 6 respectively actually observed per monomer). However, whatever the evolutionary origins of Ocr, the surfeit of acidic residues ensures that both modification and restriction are efficiently inhibited and toxicity is avoided.

In summary, a detailed examination of our data shows that to obtain an active antirestriction and antimodification Ocr protein capable of being expressed without deleterious effects on the host cell one has to achieve several conditions as follows:

The pI must be 4.5 or less.

The number of negative charges in the orange region must be at least 6.

In the purple region, there should be no more than 47% of the total number of negative charges.

The green region should contain at least three negative charges.

The total number of negative charges on the whole sequence must be at least 14.

Position 88 must be aspartate to form hydrogen bonds to loop 2 thereby constraining its conformation and removing its ability to interact with other proteins in the cell involved in cell division.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

N.K. was supported by funding from the EPSRC/BBSRC RASOR Doctoral Training Centre in Cell & Proteomic Technologies. We would like to acknowledge the late Prof. Noreen Murray (Institute of Cell and Molecular Biology, University of Edinburgh) for kindly providing the NM strains and phage. ITC experiments were conducted at the Edinburgh Protein Production Facility (EPPF, University of Edinburgh). We also thank Dr Liat Rockah-Shmuel (Department of Biological Chemistry, Weizmann Institute of Science, Rehovot, Israel), Prof. Tom McLeish (Biophysical Sciences Institute, Durham University) and Dr Angela Dawson (School of Physics and Astronomy, University of Edinburgh) for insightful discussions and the Institute of Advanced Study, Durham University for providing an environment conducive to writing this paper.

## FUNDING

Wellcome Trust [GR080463MA to D.T.F.D.]; Biotechnology and Biological Sciences Research Council (BBSRC) [BB/D001870/1 and BB/C511599/1 to D.T.F.D.]. Funding for open access charge: Wellcome Trust; BBSRC.

*Conflict of interest statement.* None declared.

## REFERENCES

- Studier, F.W. (1975) Gene 0.3 of bacteriophage T7 acts to overcome the DNA restriction system of the host. *J. Mol. Biol.*, **94**, 283–295.
- Kruger, D.H., Schroeder, C., Hansen, S. and Rosenthal, H.A. (1977) Active protection by bacteriophages T3 and T7 against *E. coli* B- and K-specific restriction of their DNA. *Mol. Gen. Genet.*, **153**, 99–106.
- Tock, M.R. and Dryden, D.T.F. (2005) The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.*, **8**, 466–472.
- Dryden, D.T.F. (2006) DNA mimicry by proteins and the control of enzymatic activity on DNA. *Trends Biotechnol.*, **24**, 378–382.
- Serfiotis-Mitsa, D., Roberts, G.A., Cooper, L.P., White, J.H., Nutley, M., Cooper, A., Blakely, G.W. and Dryden, D.T.F. (2008) The Orf18 gene product from conjugative transposon Tn916 is an ArdA antirestriction protein that inhibits type I DNA restriction–modification systems. *J. Mol. Biol.*, **383**, 970–981.
- McMahon, S.A., Roberts, G.A., Johnson, K.A., Cooper, L.P., Liu, H., White, J.H., Carter, L.G., Sanghvi, B., Oke, M., Walkinshaw, M.D. *et al.* (2009) Extensive DNA mimicry by the ArdA anti-restriction protein and its role in the spread of antibiotic resistance. *Nucleic Acids Res.*, **37**, 4887–4897.
- Bandyopadhyay, P.K., Studier, F.W., Hamilton, D.L. and Yuan, R. (1985) Inhibition of the type I restriction–modification enzymes EcoB and EcoK by the gene 0.3 protein of bacteriophage T7. *J. Mol. Biol.*, **182**, 567–578.
- Atanasiu, C., Byron, O., McMiken, H., Sturrock, S.S. and Dryden, D.T.F. (2001) Characterisation of the structure of Ocr, the gene 0.3 protein of bacteriophage T7. *Nucleic Acids Res.*, **29**, 3059–3068.
- Blackstock, J.J., Egelhaaf, S.U., Atanasiu, C., Dryden, D.T.F. and Poon, W.C.K. (2001) Shape of Ocr, the gene 0.3 protein of bacteriophage T7: modelling base on light scattering experiments. *Biochemistry*, **40**, 9944–9949.
- Atanasiu, C., Su, T.-J., Sturrock, S.S. and Dryden, D.T.F. (2002) Interaction of the Ocr gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme. *Nucleic Acids Res.*, **30**, 3936–3944.
- Walkinshaw, M.D., Taylor, P., Sturrock, S.S., Atanasiu, C., Berge, T., Henderson, R.M. *et al.* (2002) Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. *Mol. Cell*, **9**, 187–194.
- DelProposto, J., Majmudar, C.Y., Smith, J.L. and Brown, W.C. (2009) Mocr: a novel fusion tag for enhancing solubility that is compatible with structural biology applications. *Protein Expr. Purif.*, **63**, 40–49.
- Stephanou, A.S., Roberts, G.A., Cooper, L.P., Clarke, D.J., Thomson, A.R., MacKay, C.L., Nutley, M., Cooper, A. and Dryden, D.T.F. (2009a) Dissection of the DNA mimicry of the bacteriophage T7 Ocr protein using chemical modification. *J. Mol. Biol.*, **391**, 565–576.
- Stephanou, A.S., Roberts, G.A., Tock, M.R., Pritchard, E.H., Turkington, R., Nutley, M., Cooper, A. and Dryden, D.T.F. (2009b) A mutational analysis of DNA mimicry by Ocr, the gene 0.3 antirestriction protein of bacteriophage T7. *Biochem. Biophys. Res. Commun.*, **378**, 129–132.
- Roberts, G.A., Stephanou, A.S., Kanwar, N., Dawson, A., Cooper, L.P., Chen, K., Nutley, M., Cooper, A., Blakely, G.W. and Dryden, D.T.F. (2012) Exploring the DNA mimicry of the Ocr protein of phage T7. *Nucleic Acids Res.*, **40**, 8129–8143.
- Zavilgelsky, G.B., Kotova, V.Y. and Rastorguev, S.M. (2008) Comparative analysis of anti-restriction activities of ArdA (ColIb-P9) and Ocr (T7) proteins. *Biochemistry (Mosc.)*, **73**, 906–911.
- Zavilgelsky, G.B., Kotova, V.Y. and Rastorguev, S.M. (2009) Antirestriction and antimodification activities of the T7 Ocr protein: effect of mutations in interface. *Mol. Biol. (Mosc.)*, **43**, 93–100.

18. Zavilgelsky, G.B., Kotova, V.Y. and Rastorguev, S.M. (2011) Antimodification activity of the ArdA and Ocr proteins. *Russ. J. Genet.*, **47**, 139–146.
19. Murray, N.E. (2000) Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol. Mol. Biol. Rev.*, **64**, 412–434.
20. Loenen, W.A.M., Dryden, D.T.F., Raleigh, E.A., Wilson, G.G. and Murray, N.E. (2014) Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res.*, **42**, 3–19.
21. Loenen, W.A.M., Dryden, D.T.F., Raleigh, E.A. and Wilson, G.G. (2014) Type I restriction enzymes and their relatives. *Nucleic Acids Res.*, **42**, 20–44.
22. 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.
23. 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.
24. Murray, N.E. (2002) 2001 Fred Griffith review lecture. Immigration control of DNA in bacteria: self versus non-self. *Microbiology*, **148**, 3–20.
25. Powell, L.M., Connolly, B.A. and Dryden, D.T.F. (1998) The DNA binding characteristics of the trimeric EcoKI methyltransferase and its partially assembled dimeric form determined by fluorescence polarisation and DNA footprinting. *J. Mol. Biol.*, **283**, 947–961.
26. Powell, L.M., Dryden, D.T.F. and Murray, N.E. (1998) Sequence-specific DNA binding by EcoKI, a type IA DNA restriction enzyme. *J. Mol. Biol.*, **283**, 963–976.
27. Dryden, D.T.F., Cooper, L.P., Thorpe, P.H. and Byron, O. (1997) The *in vitro* assembly of the EcoKI type I DNA restriction/modification enzyme and its *in vivo* implications. *Biochemistry*, **36**, 1065–1076.
28. Kennaway, C.K., Obarska-Kosinska, A., White, J.H., Tuszynska, I., Cooper, L.P., Bujnicki, J.M., Trinick, J. and Dryden, D.T.F. (2009) The structure of M.EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein. *Nucleic Acids Res.*, **37**, 762–770.
29. Kennaway, C.K., Taylor, J.E., Song, C.F., Potrzebowski, W., Nicholson, W., White, J.H., Swiderska, A., Obarska-Kosinska, A., Callow, P., Cooper, L.P. *et al.* (2012) Structure and operation of the DNA-translocating Type I DNA restriction enzymes. *Genes Dev.* **26**, 92–104.
30. Dunn, J.J., Elzinga, M., Mark, K.K. and Studier, F.W. (1981) Amino acid sequence of the gene 0.3 protein of bacteriophage T7 and nucleotide sequence of its mRNA. *J. Biol. Chem.*, **256**, 2579–2585.
31. Mark, K.K. and Studier, F.W. (1981) Purification of the gene 0.3 protein of bacteriophage T7, an inhibitor of the DNA restriction system of *Escherichia coli*. *J. Biol. Chem.*, **256**, 2573–2578.
32. Su, T.-J., Tock, M.R., Egelhaaf, S.U., Poon, W.C.K. and Dryden, D.T.F. (2005) DNA bending by M.EcoKI methyltransferase is coupled to nucleotide flipping. *Nucleic Acids Res.*, **33**, 3235–3244.
33. Davies, G.P., Kemp, P., Molineux, I.J. and Murray, N.E. (1999) The DNA translocation and ATPase activities of restriction-deficient mutants of EcoKI. *J. Mol. Biol.*, **292**, 787–796.
34. Sambrook, J. and Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edn. Cold Spring Harbor Laboratory Press, NY.
35. Herman, A. and Tawfik, D.S. (2007) Incorporating synthetic oligonucleotides via gene reassembly (ISOR): a versatile tool for generating targeted libraries. *Protein Eng. Des. Sel.*, **20**, 219–226.
36. Rockah-Shmuel, L., Tawfik, D.S. and Goldsmith, M. (2014) Generating targeted libraries by the combinatorial incorporation of synthetic oligonucleotides during gene shuffling (ISOR). *Methods Mol. Biol.*, **1179**, 129–137.
37. Dryden, D.T.F., Cooper, L.P. and Murray, N.E. (1993) Purification and characterization of the methyltransferase from the type I restriction and modification system of *Escherichia coli* K12. *J. Biol. Chem.*, **268**, 13228–13236.
38. Whitmore, L. and Wallace, B.A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.*, **32**, W668–W673.
39. Bloomfield, V.A., Crothers, D.M. and Tinoco, I. (2000) In: Connolly, B. (ed). *Nucleic Acids: Structures, Properties and Functions*. University Science Books, Sausalito, pp. 335–443.
40. Ratner, D. (1974) The interaction of bacterial and phage proteins with immobilized *Escherichia coli* RNA polymerase. *J. Mol. Biol.*, **88**, 373–383.