Construction and study of altered specificity DNA methyltransferases

Ph. D. thesis summary

Edit Tímár

Supervisors: Pál Venetianer, Ph. D.
Professor Emeritus

Antal Kiss, Ph. D. scientific adviser

Biological Research Center of the Hungarian Academy of Sciences

Ph. D. School of Biology

University of Szeged, Faculty of Science and Informatics

2016

Szeged

Type II modification enzymes are ideally suited to studying the molecular mechanisms of sequence-specific protein–DNA interactions as these enzymes are able to recognize, and act upon, 2–8 bp long DNA sequence elements with an astonishing degree of accuracy (Cheng, 1995). The members of one subgroup of these enzymes, the cytosine-C5 methyltransferases (C5-MTases), display a basically similar architecture. Comparison of the amino acid sequences of C5-MTases showed that they share 6 strongly conserved and 4 weakly conserved sequence motifs arranged in every case in the same order. Motifs VIII and IX are separated by a long variable stretch of amino acids with little or no homology between different enzymes. This structure suggested the intuitively attractive assumption that the common motifs might be responsible for the common enzymatic reaction, whereas the variable region would determine the recognition specificity that distinguishes these enzymes from each other. Although changing sequence specificity by transferring only the variable region has not been reported for any of the monospecific C5-MTases, in two cases chimeric C5-MTases could be constructed whose methylation specificity strongly supported the role of the variable region in the DNA substrate recognition (Klimasauskas et al, 1991; Mi and Roberts, 1992). The three-dimensional structures of the specific enzyme–DNA complexes of M.HhaI (Cheng et al, 1993) and M.HaeIII (Reinisch et al, 1995) also confirmed the notion that the variable region determines specificity. The structures show that the enzymes consist of a larger and a smaller domain forming a cleft where the substrate DNA fits. In both X-ray models, the large

domain containing most of the conserved motifs faces the minor groove of the DNA, whereas the small domain contains the variable region and motif IX are faces the major groove. Contacts mediating sequence-specific DNA recognition were shown to occur between the small domain of the MTase and the major groove surface of the DNA. However, experiments from this laboratory have shown that at least in the case of two C5-MTases (M.SinI and M.EcoRII, recognition sequences are $GG^{A}/_{T}CC$ and $CC^{A}/_{T}GG$, respectively) the ability of the enzyme to distinguish between $^{A}/_{T}$ and $^{G}/_{C}$ base pairs in the middle of the recognition sequence depends on interaction with the minor groove of the DNA (Kiss et al, 2001). This observation suggested that, if these enzymes have a similar three-dimensional structure to M.HhaI and M.HaeIII, the determining elements must be in the large domain, but not in the variable region. This model was consistent with the results from random mutagenesis of a large segment (including the variable region) of the M.SinI gene. Two mutants with relaxed specificity for the central base pair of the recognition sequence were isolated. None of these mutations was found in the variable region.

Aims

In the previous study, only the C-terminal two-third of the M.SinI gene was mutagenized, we decided to try a new approach to isolate relaxed specificity mutants of M.SinI. Our experimental strategy used random mutagenesis of the entire M.SinI gene combined with 'DNA-shuffling' (*in vitro* random recombination), and a strong *in vitro* selection for the required phenotype.

In the second part of my thesis we tested the level of protection provided by the M.SinI mutants

(isolated from this work and earlier work, Kiss et al 2001) against a restriction enzyme cleaving both GGWCC and GGSCC sites. We describe experiments using an *in vivo* system in which the mutant MTases were coexpressed with the GGNCC-specific Sau96I endonuclease, whose expression could be tightly controlled, and viability of the cells was quantitatively estimated.

Methods

- PCR techniques (mutagenesis, DNA-shuffling)
- Recombinant-DNA techniques for the synthesis of plasmids
- Transformation of bacterial cells
- Protein purification
- Determination of the steady-state kinetic parameters
- DNA sequence analysis

Results

Selection and characterization of the mutant SinI MTase

By a combination of *in vitro* mutagenesis, DNA shuffling and strong selection for the required phenotype, we have isolated a mutant SinI MTase, which in addition to its canonical substrate sequence GGWCC, efficiently methylates also GGSCC sites. DNA sequencing revealed that the mutant gene contained nine point mutations, seven transitions and two transversions. Five mutations (C101T, A130C, T197C, T641C and T685C) resulted in amino acid replacements (A34V, K44Q, M66T, L214S and Y229H). Three of these replacements

were located in the N-terminal part, preceding the first conserved structural motif;and two replacements were in the middle of the molecule between conserved blocks VI and VII. No substitution was found in the variable region. We found that altered sequence specificity of the mutant enzyme was mainly caused by one or both of the two replacements between motifs VI and VII, and contribution to the phenotype of the three N-terminal mutations were minimal.The enzyme carrying only the two internal mutations was soluble and could be purified to near homogeneity. Steady-state kinetic parameters of the double-mutant SinI MTase were determined. M.SinI(L214S + Y229H) had a 4.5-fold lower k_{cat}/K_m value for the canonical substrate GGWCC than the WT enzyme. This change is mainly due to the decrease in k_{cat}. The mutant has 2-fold higher K_m than the WT enzyme for the GGSCC substrate, however, this loss is more than compensated by the large (40-fold) increase in k_{cat}.

In vivo DNA protection by relaxed-specificity SinI DNA methyltransferase variants

The role of the modification MTase in a restriction-modification system is to protect the host DNA from the cognate restriction endonuclease. In this work we tested the level of protection provided by the M.SinI mutants against a restriction enzyme cleaving both GGWCC and GGSCC sites. We used an *in vivo* system in which the mutant MTases were coexpressed with the GGNCC-specific Sau96I endonuclease, whose expression could be tightly controlled, and viability of the cells was quantitatively estimated. A plasmid (pOB-RSau96I) expressing R.Sau96I under the control of the *araBAD* promoter was constructed. Transcription of the Sau96I REase gene in pOB-RSau96I can be repressed

by adding glucose or induced by adding arabinose to the medium. pOB-RSau96I is stable in cells containing pSTC-MSau96I, which expresses the Sau96I MTase, but is lethal in m^- cells even in the presence of glucose.

To assess the viability of cells coexpressing the mutant MTases and Sau96I endonuclease, *E. coli* DH10B cells were transformed with three plasmids: pJAT13*araE*, pSin5 or its mutant derivatives, and pOB-RSau96I. Plasmid pJAT13*araE*, which constitutively expresses an arabinose transporter protein and is compatible with both ColE1 and p15A replicons, was used to ensure uniform expression of R.Sau96I in all cells of the culture.

Viability was tested by plating aliquots of cultures grown in the presence of glucose onto agar plates containing glucose or different concentrations of arabinose. Viability was quantitatively assessed by determining plating efficiency at 0.01% arabinose. At this arabinose concentration, the plating efficiency of cells producing the N172S or V173L mutant MTases was approximately fivefold higher than that of cells producing the WT or the 5mut enzyme, indicating that the N172S and V173L mutants of M.SinI can provide significant protection against Sau96I cleavage *in vivo*. Interestingly, although the methylation level of GGSCC sites in plasmid containing 5 amino acid change mutation was not dramatically lower than that in the two single mutants, the clone producing the variant with five substitutions displayed no increase in plating efficiency relative to the WT.

The experiments described above demonstrated that the N172S and V173L mutant SinI MTases provide significant protection *in vivo* against a restriction

endonuclease with GGNCC specificity. However, when the colonies obtained on the arabinose-containing plates were transferred onto fresh arabinose-containing plates, they usually did not grow, indicating that the level of methylation at GGSCC sites was insufficient to support long-term survival.

Effect of elevated level of DNA ligase

At first glance, even the limited protection provided by the N172S and V173L mutants seems surprising, as methylation of the GGSCC sites in the plasmid DNA, and presumably in the genomic DNA, was not complete, probably leaving many of the 4218 GGSCC sites in the *E. coli* genome (J. Pósfai, personal communication) unprotected. We assumed that viability under these conditions was due to DNA ligase-mediated repair of Sau96I endonuclease-inflicted DNA scissions. This model was suggested by two previous observations (Heitman et al,1989; Smith et al, 1992). To test the hypothesis that DNA ligase can repair DNA scissions occurring at incompletely modified GGSCC sites, two plasmids encoding SinI-Sau96I hybrid R-M systems [pMSin-RSau and pMSin(V173L)-RSau] and a compatible plasmid (pOK-ligA) encoding *E. coli* DNA ligase were constructed. pMSin-RSau carries the WT SinI MTase and the Sau96I REase genes, whereas pMSin(V173L)-RSau carries the V173L mutant allele of the SinI MTase gene and the Sau96I REase gene. Because of unprotected GGSCC sites, these plasmids can be maintained only in cells containing the Sau96I MTase gene on a compatible plasmid. When plasmid preparations containing pMSin-RSau and pSTC-MSau96I were used to transform ER1398 (pOK-ligA) cells at 42°C, a temperature nonpermissive for pSTC-MSau96I, Ap^RKn^R colonies of various sizes were

obtained. Of 1,000 $Ap^RKn^RCm^S$ transformants, only two grew when transferred onto fresh Ap-Kn plates or inoculated into LB-Ap-Kn liquid medium. Analysis of these two clones revealed that they did not contain Sau96I endonuclease. In contrast, when ER1398(pOK-ligA) cells were transformed with the same plasmid preparation [pMSin(V173L)-RSau/pSTC-MSau96I], approximately 10% of the 200 $Ap^RKn^RCm^S$ clones tested were stable enough to grow up to high density in a liquid culture. Digestion of plasmid DNA purified from these clones showed the characteristic partial fragmentation pattern with Cfr13I, and cell extracts prepared from the clones contained Sau96I endonuclease, proving that the increased viability was not due to the presence of the Sau96I MTase or the absence of the Sau96I endonuclease.

These results and previous observations suggest the generalization that the same mechanism (ligase-mediated repair) saved *E. coli* cells from self-destruction in all cases where the puzzling phenomenon of viable r^+ m^- clones was observed. In this context, it is worth to call attention to the so-far-perhaps-unrecognized fact that, to our knowledge, in all R-M systems where viable r^+ m^- clones have been reported, the restriction enzyme produces cohesive ends, which (unlike blunt ends) are substrates for *E. coli* DNA ligase.

List of publications:

E Tímár, G Groma, A Kiss, P Venetianer (2004) Changing the recognition specificity of a DNA-methyltransferase by in vitro evolution *Nucleic acids research* **32** (13), 3898-3903

E Tímár, P Venetianer, A Kiss (2008) In vivo DNA protection by relaxed-specificity SinI DNA methyltransferase variants *Journal of bacteriology* **190** (24), 8003-8008

K Ślaska-Kiss, E Tímár, A Kiss (2012) Complementation between inactive fragments of SssI DNA methyltransferase *BMC molecular biology* **13** (1), 1

B Csörgő, T Fehér, E Tímár, FR Blattner, G Pósfai (2012) Low-mutation-rate, reduced-genome Escherichia coli: an improved host for faithful maintenance of engineered genetic constructs *Microbial cell factories* **11** (1), 1