

**Anton Slyvka**

Rozprawa doktorska

## **Specyficzne rozpoznawanie i przetwarzanie zmodyfikowanych zasad cytozyny w DNA**

### **STRESZCZENIE**

5-metylocytozyna (5mC) jest modyfikacją występującą w DNA organizmów należących do wszystkich królestw życia, a także w DNA niektórych wirusów. U eukariontów, zwłaszcza u ssaków, 5mC, a także jej utleniona forma, 5-hydroksymetylocytozyna (5hmC), odgrywają istotną rolę w regulacji ekspresji genów. Aby przekazać informację epigenetyczną obie modyfikacje muszą być specyficznym wiązane przez odpowiednie białka. Molekularne podstawy takiego wiązania pozostają słabo poznane. W świecie prokariotycznym bakteriofagi mogą włączać 5mC i 5hmC do swojego genomowego DNA w celu zablokowania aktywności enzymów restrykcyjnych gospodarza. Natomiast bakterie i archeony wytwarzają enzymy restrykcyjne zależne od modyfikacji (MDRE), które specyficznie degradowują zmodyfikowane DNA.

EcoKMcrA z *Escherichia coli* jest enzymem z grupy MDRE specyficznym wobec 5mC i 5hmC, który został odkryty ponad czterdzieści lat temu. Badania *in vivo* oraz *in silico* sugerowały, że EcoKMcrA zawiera domenę N-końcową oraz domenę nukleazy typu HNH. Na podstawie eksperymentów genetycznych i modelowania molekularnego zasugerowano, że domena N-końcowa EcoKMcrA rozpoznaje modyfikowaną cytozynę, natomiast domena HNH bierze udział w cięciu DNA. Jednak do niedawna brakowało dowodów biochemicznych i strukturalnych potwierdzających zaproponowany model. Niniejsza praca obejmuje kompleksowe badania aktywności EcoKMcrA *in vitro* oraz analizę strukturalną jego specyficzności wobec 5mC i 5hmC. Struktura krystaliczna enzymu o pełnej długości potwierdziła występowanie C-końcowej domeny typu HNH. Wyniki zamieszczone w pracy potwierdzają również, że domena N-końcowa odpowiada za specyficzne rozpoznawanie modyfikowanego DNA. Domena N-końcowa ma unikalną strukturę i rozpoznaje 5mC i 5hmC w obrębie dwuniciowego DNA, nie powodując wywinięcia nukleotydów (ang. nucleotide flipping). Struktura krystaliczna kompleksu domeny N-końcowej EcoKMcrA z metylovanym i hydroksymetylovanym DNA pokazuje, że ma ona dwie kieszenie do rozpoznawania zmodyfikowanej zasady odpowiednio w każdej nici DNA. Mutacje punktowe, które zaburzają wiązanie domeny N-końcowej *in vitro*, silnie obniżają aktywność pełnego enzymu *in vivo*.

U eukariontów 5hmC występuje jako produkt enzymatycznego utleniania 5mC przez dioksygenazy TET. 5hmC może ulegać dalszemu utlenianiu, które prowadzi do powstania 5-formylocytozyny (5fC) i 5-karboksycytozyny (5caC). 5fC i 5caC mogą być wycięte z DNA i zastąpione niezmodyfikowaną cytozyną na drodze naprawy przez wycinanie zasady (ang. base excision repair, BER), zainicjowanej przez glikozylazę TDG. Proces utleniania 5mC i usuwania jej utlenionych pochodnych nazywany jest aktywną demetylacją DNA i jest szczególnie ważny w

procesie rozwoju ssaków. TDG jest glikozylazą jednofunkcyjną, która może wycinać zasadę, ale do dalszego przecięcia łańcucha głównego DNA potrzebuje endonukleazy APE1. Wyniki badań genetycznych sugerowały, że glikozylazy NEIL mogą niezależnie od TDG wycinać utlenione formy 5mC. Glikozylazy NEIL to enzymy dwufunkcyjne, zdolne zarówno do wycinania zasad jak i przecięcia łańcucha głównego DNA dzięki aktywności  $\beta/\delta$ -liazowej. W innej publikacji nie potwierdzono jednak zdolności białek NEIL do samodzielnego wycinania utlenionych form 5mC. Wykazano natomiast, że enzymy NEIL1 i 2 stymulują aktywność TDG, usuwając miejsca AP (apurynowe/apirymidynowe) generowane przez TDG. Wyniki badań będące przedmiotem niniejszej rozprawy sugerują, że obie koncepcje są częściowo poprawne. Z jednej strony wykazano, że NEIL1 może bezpośrednio wycinać 5caC, ale nie 5fC, a następnie przecinać łańcuch główny DNA dzięki aktywności  $\beta/\delta$ -liazowej. Z drugiej strony potwierdzono, że NEIL1 oddziałuje z TDG stymulując jej aktywność glikozylazową wobec 5fC i 5caC. Ponadto wykazano, że TDG stymuluje aktywność  $\beta/\delta$ -liazową, ale nie glikozylazową białka NEIL1. Co więcej, otrzymane dane sugerują, że oddziaływanie NEIL1-TDG jest specyficzne dla aktywnej demetylacji DNA, ponieważ NEIL1 nie stymuluje, ale raczej blokuje TDG-zależne wycinanie tyminy błędnie sparowanej z guaniną.

Utlnienie 5mC, katalizowane przez białka TET, odgrywa kluczową rolę w aktywnej demetylacji DNA. Białka TET mogą utlenić węgiel grupy metylowej trzykrotnie, co odpowiada liczbie wodorów, znajdujących się na  $C\alpha$ . Utlnienie przebiega w trzech rundach. W pierwszej powstaje 5hmC, w drugiej - 5fC, a w trzeciej - 5caC. Niedawno zaobserwowano aktywność ludzkiego TET2 wobec wydłużonych alkilowych analogów 5mC (5-etylocytozyna, 5-winylocytozyna, 5-etylocytozyna i 5-propylocytozyna), co sugerowało, że białka TET mogą być znacznie bardziej elastyczne katalitycznie niż oczekiwano. Niniejsza praca miała na celu poszerzenie wiedzy na temat aktywności katalitycznej białek TET na nietypowych substratach. Aktywność mysiego TET1 i paralogu TET z pierwotniaka *Naegleria gruberi* była przetestowana na zestawie alkilowych analogów 5mC o różnej długości i typach hybrydyzacji (5-(prop-2-enylo)cytozyna, 5-(but-2-ynylo)cytozyna i 5-(6-azydoheks-2-ynylo)cytozyna). Wykazano, że białka TET mogą utleniać wszystkie analogi 5mC zarówno w jedno-, jak i dwuniciowym DNA i potwierdzono utlenianie 5-etylocytozyny. Wykazano także odwrotną korelację pomiędzy aktywnością TET a rozmiarem łańcucha alkilowego. Co ważne, pokazano, że kiedy enzymy TET wyczerpią wszystkie atomy wodoru na  $C\alpha$ , mogą utleniać  $C\beta$ , a także dalsze atomy węgla. Ponadto zbadano wpływ wydłużonych łańcuchów alkilowych na aktywność glikozylaz NEIL1 i TDG. Podczas gdy obie glikozylazy są nieaktywne w przypadku nieutlenionych analogów, produkty utleniania przez TET cytozyn z dłuższymi łańcuchami są wycinane przez NEIL1, natomiast TDG wykazuje słabą aktywność wobec produktów utleniania 5-etylocytozyny.

21.12.2020 Anton Szyvka

**Anton Slyvka**

Doctoral thesis

**Specific recognition and processing of modified cytosine bases in DNA**

**ABSTRACT**

5-methylcytosine (5mC) is found in DNA of the organisms from all kingdoms of life, as well as in DNA of some viruses. In eukaryotes, especially in mammals, 5mC and its oxidized form 5-hydroxymethylcytosine (5hmC) play an important role in gene expression regulation. Both 5mC and 5hmC have to be specifically recognized by proteins to convey their epigenetic message. The molecular details of such specific recognition remain poorly understood. In the prokaryotic world, bacteriophages can incorporate 5mC and 5hmC in their genomic DNA in order to block the restriction enzymes of the host. However, bacteria and archaea developed modification dependent restriction enzymes (MDREs) that can specifically degrade modified DNA.

*Escherichia coli* EcoKMcrA is 5mC and 5hmC specific MDRE that was discovered more than forty years ago. *In vivo* and *in silico* studies suggested that it contains an N-terminal domain and an HNH nuclease domain. Based on the genetic experiments and molecular modeling it was suggested that EcoKMcrA senses the modified cytosine with its N-terminal domain, and cleaves DNA with the HNH domain. However, the *in vitro* and structural proofs of these theories were missing until recently. This work presents the comprehensive study of EcoKMcrA activity *in vitro* and the structural basis of its 5mC and 5hmC specificity. Crystal structure of the full-length enzyme confirms the HNH fold of the C-terminal domain. The work also proves the modification sensing role of the N-terminal domain, which has a unique structure and recognizes 5mC and 5hmC in the context of double-stranded DNA without nucleotide flipping. Crystal structure of the N-terminal domain with methylated and hydroxymethylated DNA reveals that it has two pockets for the recognition of the modified base in each DNA strand. Point mutations that abrogate the N-terminal domain binding *in vitro* severely compromise the activity of the full-length enzyme *in vivo*.

In eukaryotes, 5hmC is a product of enzymatic oxidation of 5mC by TET dioxygenases. It can be further oxidized to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) that can be excised from DNA and replaced with the unmodified cytosine by TDG glycosylase mediated base excision repair pathway. The process of 5mC oxidation and removal of the oxidized derivatives is called active DNA demethylation and it is particularly important in mammalian development. TDG is a monofunctional glycosylase that can only excise the base and requires APE1 endonuclease for the subsequent DNA backbone cleavage. NEIL glycosylases were proposed to act redundantly to TDG in the excision of oxidized 5mC forms. NEIL glycosylases are bifunctional, and after the base is excised, they can incise the DNA backbone using their  $\beta/\delta$ -lyase activity. However, a follow-up work has shown that NEIL enzymes are unable of independent excision of oxidized 5mC species.

Instead, it was demonstrated that NEIL1 and 2 stimulate TDG activity by removing the abasic sites generated by TDG. The biochemical data presented in this dissertation show that both reports are partially correct. On the one hand, NEIL1 can directly excise 5caC, but not 5fC, and subsequently cleave the DNA backbone by its  $\beta/\delta$ -lyase activity. On the other hand, NEIL1 interacts with TDG and stimulates its glycosylase activity on 5fC and 5caC. In addition, TDG stimulates the  $\beta/\delta$ -lyase activity of NEIL1 but not its glycosylase activity. Moreover, NEIL1 does not stimulate, but rather inhibits TDG activity on T/G mismatches, suggesting that NEIL1-TDG cooperation is specific for the active DNA demethylation.

TET catalyzed 5mC oxidation has a central role in active DNA demethylation. TETs can oxidize methyl group carbon ( $C\alpha$ ) three times, which corresponds to the number of hydrogens that can be abstracted from it. The oxidation occurs in three rounds, first resulting in 5hmC, second – in 5fC, and third – in 5caC. Recently, human TET2 activity on the extended 5mC alkyl analogues (5-ethynylcytosine, 5-vinylcytosine, 5-ethylcytosine and 5-propylcytosine) was reported, suggesting that TETs may be much more catalytically flexible than anticipated. In this thesis, the activity of mouse TET1 and TET paralogue from protist *Naegleria gruberi* was tested on the set of extended alkyl analogues that varied in length and electronic environment (5-(prop-2-enyl)cytosine, 5-(but-2-ynyl)cytosine, and 5-(6-azidohex-2-ynyl)cytosine) in order to further explore the substrate promiscuity of TET enzymes. The results indicate that TETs can oxidize all tested substrates in both single- and double-stranded DNA and confirm 5-ethylcytosine oxidation. TET activity is inversely correlated with alkyl chain size. Remarkably, when TETs exhaust all hydrogens on  $C\alpha$ , they can oxidize  $C\beta$  and further carbons. The effects of extended alkyl chains on the excision by NEIL1 and TDG glycosylases were also studied. Whereas both glycosylases are inactive on the untreated set, NEIL1 can excise products of TET oxidation of bulky substrates, and TDG has a weak activity on products of 5-ethylcytosine oxidation.

21.12.2020 Anton Slyzka