International Journal of Biochemistry Research & Review

29(6): 1-25, 2020; Article no.IJBCRR.57357
ISSN: 2231-086X, NLM ID: 101654445

Analyses of Homing Endonucleases and Mechanism of Action of CRISPR-Cas9 HNH Endonucleases

Peramachi Palanivelu[1*]

[1]Department of Molecular Microbiology, School of Biotechnology, Madurai Kamaraj University, Madurai – 625 021, India (Retd.).

Author's contribution

The sole author designed, analysed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/IJBCRR/2020/v29i630192
Editor(s):
(1) Dr. Chunying Li, Georgia State University, USA.
Reviewers:
(1) Yasir Mehmood, GCUF, Pakistan.
(2) Sana Zahoor, Virtual University of Pakistan, Pakistan.
Complete Peer review History: http://www.sdiarticle4.com/review-history/57357

Original Research Article

Received 02 April 2020
Accepted 08 June 2020
Published 22 June 2020

ABSTRACT

Aim: To analyze different HNH endonucleases from various sources including the HNH endonuclease regions of CRISPR-Cas9 proteins for their conserved motifs, metal-binding sites and catalytic amino acids and propose a plausible mechanism of action for HNH endonucleases, using CRISPR-Cas9 as the model enzyme.
Study Design: Multiple sequence analysis (MSA) of homing endonucleases including the CRISPR-Cas9 using Clustal Omega was studied. Other biochemical, Site-directed mutagenesis (SDM) and X-ray crystallographic data were also analyzed.
Place and Duration of Study: School of Biotechnology, Madurai Kamaraj University, Madurai, India, between 2007 and 2013.
Methodology: Bioinformatics, Biochemical, SDM and X-ray crystallographic data of the HNH endonucleases from different organisms including CRISPR-Cas9 enzymes were analyzed. The advanced version of Clustal Omega was used for protein sequence analysis of different HNH endonucleases from various sources. The conserved motifs identified by the bioinformatics analysis were analyzed further with the data already available from biochemical and SDM and X-ray crystallographic analyses of this group of enzymes and to confirm the possible amino acids involved in the active sites and catalysis.
Results: Different types of homing endonucleases from various sources including the HNH endonuclease regions of CRISPR-Cas9 enzymes exhibit different catalytic regions and metal-

**Corresponding author: E-mail: ppmkupp@gmail.com;*

binding sites. However, the catalytic amino acid, i.e., the proton acceptor histidine (His), is completely conserved in all homing endonucleases analyzed. From these data, a plausible mechanism of action for HNH endonucleases, using CRISPR-Cas9 from *Streptococcus pyogenes*, as the model enzyme is proposed. Furthermore, multiple sequence alignment (MSA) of various homing endonucleases from different organisms showed many highly conserved motifs also among them. However, some of the HNH endonucleases showed consensus only around the active site regions. Possible catalytic amino acids identified among them belong to either -DH---N or -HH--N types. There are at least two types of metal-binding sites and bind Mg^{2+} or Zn^{2+} or both. The CRISPR-Cas9 enzyme from *S. pyogenes* belongs to the -DH- based HNH endonucleases and possesses –DxD- type metal-binding site where it possibly binds to a Mg^{2+} ion. The other HNH enzymes possess one or two invariant Zn binding CxxC/ CxxxC motifs.

Conclusions: The CRISPR-Cas9 enzymes are found to be -DH- type where the first D is likely to involve in metal-binding and the second invariant H acts as the proton acceptor and the N in –HNH- Cas9 confers specificity by interacting with the nucleotide near the catalytic region. In this communication, a metal-bound water molecule is shown as the nucleophile initiating catalysis. Homing endonucleases may be used as novel DNA binding and cleaving reagents for a variety of genome editing applications and Zinc finger nucleases have already found applications in genome editing.

Keywords: Homing endonucleases; HNH endonucleases; CRISPR-Cas9; Colicins; Pyocins; group II intron reverse transcriptases; CRISPR-Cas9-HNH endonucleases: Conserved motifs; active sites; mechanism of action.

1. INTRODUCTION

The CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) is the most well-characterized prokaryotic adaptive immunity mechanism that provides RNA-mediated protection from viruses and other mobile genetic elements. CRISPR–Cas9 system (Cas9, CRISPR associated enzyme 9) based genome editing technology is revolutionizing all areas of modern biology. This is because the CRISPR-Cas9 system allows researchers to perform precise and easy-to-use genome-editing in almost all organisms from prokaryotes to eukaryotes. It is simple, rapid, cost-effective, efficient and precise in modifying both genetic and epigenetic regions of target cells enabling researchers to generate genetically modified cells or organisms by both *in vivo* and *ex vivo* editing of human somatic cells for therapeutic applications. The Cas9 uses a guide ribonucleic acid (gRNA) to guide the enzyme to target any specific region on the genome. By engineering the gRNA sequence, it is possible to edit any DNA target in any type of cell. Therefore, the CRISPR-Cas9 based genome editing has become a workhorse in modern molecular biology, genetic engineering and biomedical research. As mentioned earlier, the CRISPR-Cas9 system is nothing but a naturally occurring adaptive immune system, originally discovered by Ishino et al. [1] in 1987 in *E. coli.* Now it is reported from a large number of eubacteria and almost in all the archaebacteria as well [2-5]. CRISPR-Cas9 system also holds great promise in stem cell and tissue engineering, translational medicine, next-generation gene therapy and engineering trait enhancement and disease-resistant plants [6,7].

The HNH endonuclease is a small nucleic acid binding motif (~30 amino acids in length) with an associated cleavage module. Such modules are commonly widespread in α-α-β-metal finger endonucleases. The main function of these enzymes is to promote the lateral transfer of their own coding and flanking DNA regions between genomes, by a recombination-dependent process known as 'homing'. They are reported in all kingdoms of life and are particularly very common in the genomes of bacteria and their phages and organellar genomes. These homing endonucleases have a long recognition sequence of ~20 amino acids to prevent random cleaving of the host genome. The homing occurs when the endonuclease makes a double-stranded break (DSB) in the genome as the first step, which is followed by the host repair mechanism which uses the intron-containing allele as the template and inserts the intron-containing the endonuclease region into the intronless allele. The free-standing ones are inserted in intergenic regions. It is interesting to note that these homing endonuclease genes are usually placed in phenotypically neutral locations on the host chromosomes. The term 'homing' is

used to describe the movement of these genes and the homing usually occurs when two genomes are juxtaposed but only one possesses the mobile element.

1.1 The CRISPR-Cas9 System

The CRISPR-Cas9 system essentially consists of an array of *Cas* genes with spacers (procured from various phages and genetic elements) arranged in-between direct repeats in the bacterial genome. Fig. 1 shows the schematic arrangement of the CRISPR-Cas9 system in bacteria. The spacer sequences, procured from various phages and genetic elements, ranges from 20-58 bp in length and placed between direct repeat sequences (21-40 bp) on a CRISPR locus. Each spacer occurs only once in a given CRISPR locus. The number of repeats and spacers per locus varies from 2 to 120 [8]). Most prokaryotes have only one CRISPR locus [8] but some contain up to eight CRISPR loci [9].

Unlike restriction enzymes, the CRISPR-cas9 recognition sequences are much longer (~20 nt) and make a double-stranded break at a specific region on the genome. Therefore, it has found much wider applications in genome-editing techniques.

1.2 Major Types of CRISPR-Cas Systems in Bacteria

Three major types of CRISPR-Cas systems (I–III) have been functionally identified across a wide range of microbial species [10-12], and each contains a cluster of CRISPR-associated (*Cas*) genes and its corresponding CRISPR array as shown in Fig. 1. The type I and III CRISPR-Cas systems employ a multi-protein complex to make a DSB on the target DNA [13-15], whereas, type II systems use a single protein, RNA-guided nuclease, viz. the Cas9 enzyme, to target DNA recognition as well as a double-stranded cleavage [16]. Though there are many CRISPR associated proteins involved in the CRISPR-Cas system, the Cas9 protein is the most abundant across the bacterial kingdom and widely used in genome editing techniques. However, the Cas9 enzyme itself vary widely both in sequence and size. All known Cas9 enzymes contain two endonuclease domains, viz. a HNH endonuclease domain which cleaves the DNA strand complementary to the gRNA sequence (otherwise known as the target DNA strand), and a RuvC endonuclease domain, which cleaves the non-complementary strand (otherwise known as the non-target DNA strand), resulting in a DSB on the genomic DNA [16]. The Cas9 enzyme from *S. pyogenes* (SpyCas9) is one of the first studied enzymes and consists of a well-conserved HNH and RuvC domains and widely used in genome editing techniques.

The 3D structure of the Cas9 enzyme is available now [17] and found it is made up of a bilobed structure, i.e., composed of a nuclease lobe (NUC) and recognition lobe (REC). The NUC is placed juxtaposed to RuvC and HNH nuclease domains. In addition to, a variable alpha-helical lobe is also identified which is likely to be involved in nucleic acid binding. The RuvC domain is made up of three discontinuous segments (RuvC-I, RuvC-II and RuvC-III), whereas the HNH domain is inserted in between the RuvC-II and RuvC-III segments (Fig. 2). The RuvC domain is much larger and forms the structural core of the nuclease lobe, with six-stranded β sheets surrounded by four α helices. All three subdomains, viz. RuvC I, II and III are highly conserved and form the active site and harbour the catalytic residues.

1.3 CRISPR-Cas9 Enzyme and Its Substrate DNAs

S. pyogenes Cas9 is one of the most well studied and was the first enzyme used for targeted mutagenesis and is still the most widely used genome editing tool [18,19]. The CRISPR-Cas9 enzyme system requires two conditions to be met to cut a specific DNA sequence: 1) a 20-nucleotide (nt) target sequence, also known as the protospacer sequence and 2) a Protospacer Adjacent Motif (PAM), (a triad, viz. 5'-NGG-3'). The PAM lies immediately at the 3´ region from the targeting crRNA (CRISPR-RNA)/proto-spacer sequence. CrRNAs are also known as gRNAS as it guides the Cas9 endonuclease to the specific site on the DNA to be cleaved. (The gRNAs are small molecular weight RNAs that can be easily pre-designed to bind a specific gene of interest within a cell). Once these two conditions are met, Cas9 will bind the DNA sequence complementary to the target sequence and make a DSB 3–4 nt, 5´ of the PAM sequence. By identifying and binding the gene of interest, gRNAs direct Cas9 to the precise region of DNA that needs to be deleted or edited. Endogenous DNA DSB repair mechanisms will then repair the DSB by using any one of the two methods, i.e., either by repairing the DSB by means of Non-Homologous End Joining (NHEJ) method or Homology Directed Repair (HDR) which uses

similar DNA sequences to repair the DSB via the incorporation of exogenous DNA to function as the repair template (In fact, supplying a pre-designed DNA template to the cell can alter a gene as desired or correct a mutation resulting in repairing the disease-causing genes).

1.4 Location of the Two Endonuclease Domains in CRISPR–Cas9 Enzymes

The CRISPR–Cas9 system which belongs to Type II system is the most widely used for genome-editing techniques. (Type V (CRISPR-Cas12a) and Type VI (CRISPR-Cas13) are also finding applications for specialized genome-editing techniques [15]. In the CRISPR–Cas9 system, there are two different types of endonuclease domains, viz. i) HNH and ii) RuvC. The RuvC and HNH domains are well-conserved in all Cas9 enzymes (Fig. 9). However, the RuvC endonuclease gene sequence is not contiguous and split into 3 domains, approximately residues are 1-59 (RuvC-I), 718-769 (RuvC-II) and 909-1098 (RuvC-III) which recognizes and cleaves the target DNA non-complementary to gRNA. In contrast, the HNH endonuclease domain is present as a single unit (approximately residues are 775-908) and cleaves the target DNA complementary to gRNA (Fig 2). As discussed elsewhere, the crystal structure of the SpyCRISPR-Cas9 enzyme showed two distinct lobes, viz. a recognition lobe (REC) and a nuclease lobe (NUC) which are connected by an R-rich bridge helix. The two nucleases, viz. the RuvC I-III and HNH make the NUC lobe. The PAM interacting domain interacts with the 3' tail of the sgRNA [17].

1.5 Homing Endonucleases and Their Types

Homing endonucleases are encoded by open reading frames that are found embedded predominantly in plastid and phage group I introns; mitochondrial, plastid, and eubacterial group II introns, plasmids archaebacterial introns. They are also reported from a large group of yeast intron 1 proteins, MutS, bacterial colicins, pyocins, maturases (highly specific cofactors for the RNA splicing reactions), T4 resolvase, etc. Some of them are also found in inteins (intervening sequences that are spliced and excised post-translationally). Most of them not only display extremely high DNA-binding specificities but also very long DNA target sites (14-40 bp), and also tolerant to a variety of sequence variations in these sites. They initiate a transfer of the embedded elements and/or themselves along with the additional ectopic sites (recognition sites for future mobility) by generating DSBs in cognate alleles that lack the intervening sequence. (Some are also found to be non-specific like colicins, pyocins, etc.)

There are four major families of homing endonuclease genes (HEGs) identified based on the presence of highly conserved amino acid sequence motifs in them. They are known as - LAGLIDADG, HNH, GIY-YIG and His-Cys box types [20].

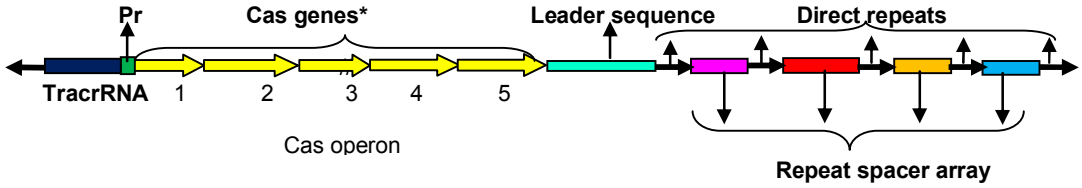


Fig. 1. A schematic diagram of a typical CRISPR-Cas array system
*NB: *Cas genes number, order and direction may be different in different organisms; Pr, promoter region; TracrRNA, Trans-activating CRISPR-RNA*

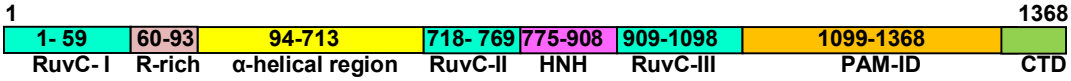


Fig. 2. A schematic diagram of the CRISPR-Cas9 enzyme and its various domains
R-rich, Arg rich region; RuvC and HNH, endonuclease domains; CTD, Carboxy Terminal Domain; PAM-ID, Protospacer Adjacent Motif Interacting Domain

In this communication, only the HNH domain of various homing endonucleases including the CRISPR-Cas9 is analyzed. The HNH endonuclease signature is found in viral, prokaryotic and eukaryotic proteins. The HNH motif is a small nucleic acid binding and cleavage module and adopts "β-β-α-Metal finger and widespread in metal finger endonucleases. This motif mainly creates zinc finger domains with completely conserved Cs. In HNH the first His acts as a proton donor and the last His involve in metal-binding, which is replaced in some cases with a second Asn creating a HNN motif that also acts in the same way as HNH.

2. MATERIALS AND METHODS

Protein sequence analyses have become a powerful tool to decode the structure-function relationships in proteins and enzymes from the highly conserved motifs among the related proteins/enzymes. A complete protein and nucleic acid sequences for a large number of HNH type endonucleases from viruses, eubacteria and eukaryotes are available in various databases. Complete protein sequence data for the CRISPR-Cas9 enzymes from various sources are also available now. These data were retrieved from SWISS-PROT and PUBMED sites and analyzed using Clustal Omega, an accurate, fast and widely accepted algorithm, available on their website.

The HNH endonuclease of CRISPR-Cas9 from *S. pyogenes* is used as the model enzyme as this is one of the most well-studied enzymes and therefore, a large amount of data on biochemical, SDM and X-ray crystallographic analyses of this enzyme are available. These data along with the MSA data were used for delineating the DNA cleavage mechanism of HNH endonucleases of CRISPR-Cas9 enzymes. For MSA analysis of other homing endonucleases such as colicins, pyocins, mcrA endonucleases and group II intron endonucleases, the protein sequences were also retrieved from SWISS-PROT and PUBMED databases and were analyzed using Clustal Omega programme.

3. RESULTS AND DISCUSSION

3.1 -HH- type Homing Endonucleases

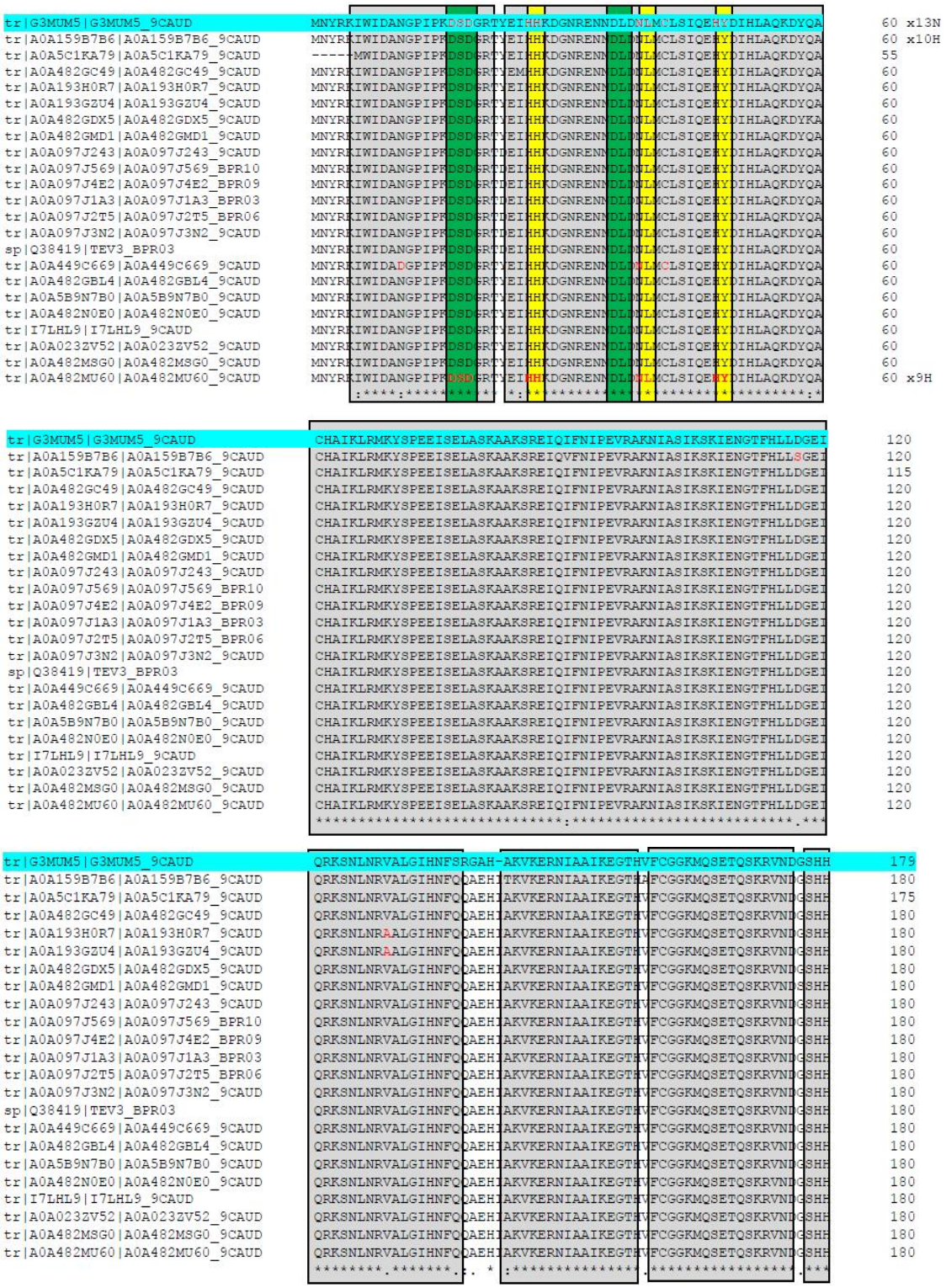
Based on the sequence analysis, the HNH endonuclease families may be broadly classified into two groups, mainly based on the immediate amino acid adjacent to the proton acceptor (His) i.e., either as **DH**-based or **HH**-based. (Some exceptions are also observed).

Fig. 3 shows the results of the MSA of –HH-based HNH endonucleases from various organisms. This group of enzymes is highly conserved from N-terminal to C-terminal (exhibit close to 99% homology) with a few minor changes. Such an exceptional homology among them suggests they did not undergo much diversity during evolution because of their important homing function. The active site amino acids, viz. –HH- followed by -N- and –H- are located in the N-terminal region and are highlighted in yellow. The active site amino acids are placed in the N-terminal region of the enzymes (highlighted). Two DxD types of metal-binding motifs are found near the active site region (highlighted in green) and a completely conserved H in all of them. CxxC or CxxxC type of Zn binding motifs are not seen around the catalytic region but two CxxC type motifs are found in the N-terminal region (highlighted). The additional metal ions are implicated to play a role in the structure and stability of these enzymes. Interestingly, there is a complete distance conservation between the proton acceptor -H^{27} and NTP binding -N^{39} amino acids and the distance is maintained at about 13 amino acids. The third H in the HNH endonuclease is completely conserved and placed at about 10 amino acids downstream from the nucleotide binding N and is followed by an invariant Y (Numbering from I-*Tev*III homing endonuclease of Enterobacteria phage Bp7 and highlighted in light blue).

3.2 -HH- Type in Modified Cytosine Restriction-A (mcrA) Endonucleases

McrA, a small molecular weight protein of ~31 kDa, belongs to type IV site-specific endonucleases, which is one of the four restriction systems evolved in bacteria to defend the bacterial cells against bacteriophage DNAs. Unlike the other restriction endonucleases, viz. I-III, it specifically recognizes 5-methylcytosine (5mC) and 5′-hydroxymethylcytosine (5hmC) residues in DNA and degrades T-even phages containing non-glucosylated DNA (RglA = "restricts glucoseless DNA") containing these residues. The mcrA is encoded by the *E. coli* chromosome and is localized in the outer membrane [21].

CLUSTAL O (1.2.4) MSA of homing endonucleases of -HH- type



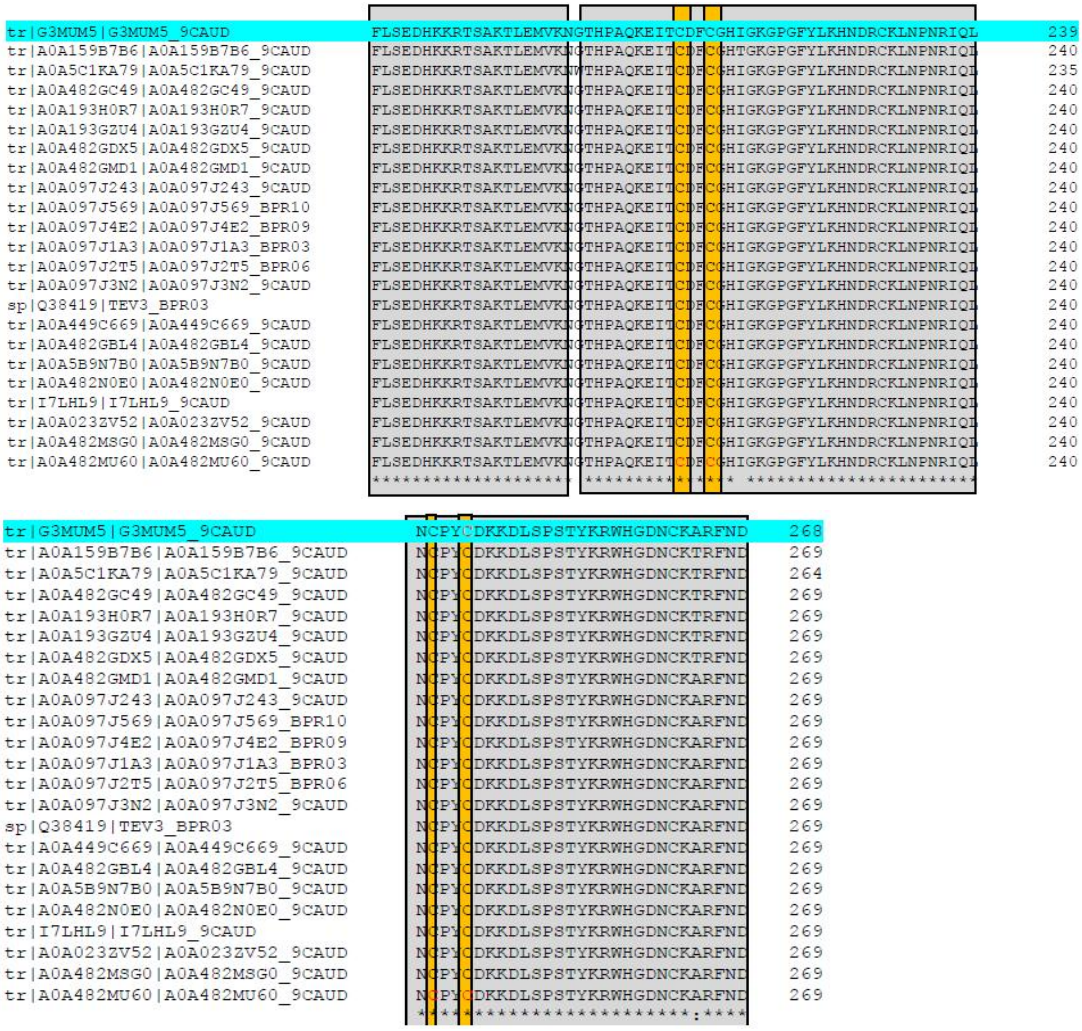


Fig. 3. MSA of HNH endonucleases –HH- type of enzymes

G3MUM5_9CAUD I-*Tev*III Homing endonuclease, Enterobacteria phage Bp7
A0A159B7B6_9CAUD Putative intron-associated endonuclease 3, *Escherichia* phage HY03
A0A482GC49_9CAUD Putative HNH endonuclease, *Escherichia* phage vB_EcoM_G2285
A0A193H0R7_9CAUD Homing endonuclease, *Shigella* phage SHFML-26
A0A193GZU4_9CAUD Homing endonuclease, *Shigella* phage SHFML-11
A0A482GDX5_9CAUD Uncharacterized protein, *Escherichia* phage vB
A0A482GMD1_9CAUD Uncharacterized protein, *Escherichia* phage vB_EcoM_G2540-3
A0A097J243_9CAUD Homing endonuclease, Enterobacteria phage RB5
A0A097J569_BPR10 Homing endonuclease, Enterobacteria phage RB10
A0A097J4E2_BPR09 Homing endonuclease, Enterobacteria phage RB9
A0A097J1A3_BPR03 Homing endonuclease, Enterobacteria phage RB3
A0A097J2T5_BPR06 Homing endonuclease, Enterobacteria phage RB6
A0A097J3N2_9CAUD Homing endonuclease, Enterobacteria phage RB7
Q38419|TEV3_BPR03 Intron-associated endonuclease 3, Enterobacteria phage RB3
A0A449C669_9CAUD Phage-associated homing endonuclease, *Yersinia* phage fPS-2
A0A482GBL4_9CAUD Uncharacterized protein, *Escherichia* phage vB_EcoM_G2469
A0A5B9N7B0_9CAUD Intron-associated endonuclease 3, *Shigella* phage JK45
A0A482N0E0_9CAUD Uncharacterized protein, *Escherichia* phage vB_EcoM_KAW3E185
I7LHL9_9CAUD Uncharacterized protein, *Yersinia* phage phiD1
A0A023ZV52_9CAUD Intron-associated endonuclease 3, *Escherichia* phage vB_EcoM_112
A0A482MSG0_9CAUD Uncharacterized protein, *Escherichia* phage vB_EcoM_WFL6982
A0A482MU60_9CAUD Uncharacterized protein, *Escherichia* phage vB_EcoM_WFK

Generally, the bacteriophage DNAs, specifically lack the methyl-tag in its genomes, which when enter the bacterial cells, serve as a signal that trigger the bacterial restriction endonucleases belonging to types I to III. However, to protect their DNA from cleavage, the bacteriophages also acquired the ability to incorporate modified bases such as 5mC and 5hmC in their genomic DNAs mimicking the bacterial DNAs and avoiding restriction endonucleases belonging to types I to III. However, to overcome this problem, the bacteria developed the type IV restriction system, viz. the mcr system to degrade even the modified bases incorporated in bacteriophage DNAs.

Structurally, mcrA enzymes are also similar to the ββα-Me superfamily of HNH endonucleases and belong to the -HH- type of HNH endonucleases by this classification; but the active site is placed towards the C-terminal end unlike other HNH enzymes (Fig. 4) where the HNH endonuclease domain is placed towards the N-terminal region (Fig. 3). The mcrA type enzymes are almost completely conserved. It possesses the typical CxxC/CxxxC type, possible Zn binding motifs around the catalytic region but there were no DxD type of motifs. The third H in the mcrA HNH endonucleases is also completely conserved but placed far away, i.e., about 28 amino acids downstream from the nucleotide binding N, which is immediately followed by an invariant Y as in others (Fig. 3).

3.3 -DH- Type Homing Endonucleases

Fig. 5 shows HNH endonucleases of –**DH**- type and these enzymes follow HNN type with completely conserved Ns towards the downstream region from the proton acceptor His. Furthermore, they have two completely conserved–CxxC- motifs, one in front of the proton acceptor and the other after the proton acceptor and very close to the second invariant N suggesting a Zn binding site. Other possible metal-binding regions are marked in green D/ExD/E. The SpyCas9 HNH sequence region is highlighted in red and AnsCas9 HNH sequence

CLUSTAL O (1.2.4) multiple sequence alignment mcrA type HNH endonucleases

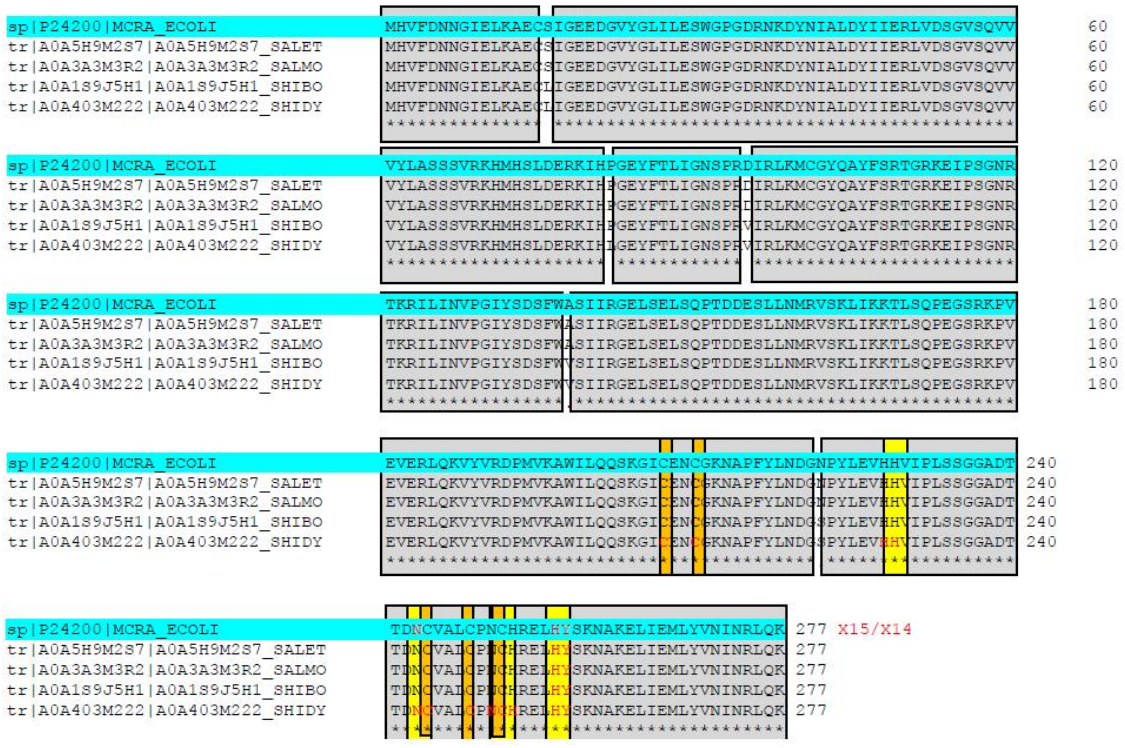


Fig. 4. MSA of HNH endonucleases –HH- type of enzymes belonging to mcrA types
P24200|MCRA_ECOLI, *Escherichia coli* (K12); A0A5H6R221_SALON, *Salmonella oranienberg* A0A3V1QJZ2_SALET, *Salmonella enterica* I**;** A0A1S9J5H1_SHIBO, *Shigella boydii* A0A403M222_SHIDY, *Shigella dysenteriae*

region is in magenta. It is interesting to note that the SpyCas9 HNH sequence region did not show any CxxC motif in contrast to all the other enzymes analyzed and but surprisingly with no C in the entire sequence region. However, it showed at least two metal-binding sites, viz. –DxD- motifs one near DH region (marked in green) and another one at the end as -ExD-. The HNH endonuclease from *Escherichia* phage did not show any DxD type motif near the catalytic diad –DH- but possessed only one CxxC motif close to the second invariant N similar like other enzymes. The Intron-associated endonuclease 3 from Enterobacteria phage showed a -KH- diad instead of the regular –DH- diad in the catalytic region and replaced R instead of first invariant N and D instead of the second invariant N.

CLUSTAL O (1.2.4) MSA of all DH based enzymes+ HNH Ana & Spy

Fig. 5. MSA of HNH endonucleases –DH- type with HNH SpyCas9 and AnaCas9 sequences

SpyCas9 HNH endonuclease sequence
QBB00100.1 HNH endonuclease (plasmid), *Klebsiella pneumoniae*
AVX35624.1 HNH endonuclease (plasmid), *Escherichia coli*]
ARX61598.1 putative HNH endonuclease (plasmid), *Escherichia coli*
ARX61672.1 HNH endonuclease family protein (plasmid) [*Escherichia coli*
Q6QGL2|TFLIV_BPT5 HNH endonuclease F-*Tfl*IV, *Escherichia* phage T5
Q38419|TEV3_BPR03 Intron-associated endonuclease 3, Enterobacteria phage
AXY85329.1 HNH endonuclease, *Escherichia* phage LL2
AKL98006.1 HNH endonuclease, *Endomicrobium proavitum*
ACL33437.1 HNH endonuclease, *Glaesserella parasuis* SH0165
STO81700.1 HNH endonuclease, *Glaesserella parasuis*
5H0M|A Chain A, HNH Endonuclease
Q08724_9CYAN Group II intron-contained ORF from *Calothrix* sp.
EFX42782.1 HNH endonuclease domain protein, *Helicobacter suis* HS1
AFZ55703.1 HNH endonuclease, *Anabaena cylindrica* PCC 7122
Q38112_BPR1T ORF26 (Fragment), *Lactococcus* phage r1t
AnaCas9 HNH endonuclease sequence

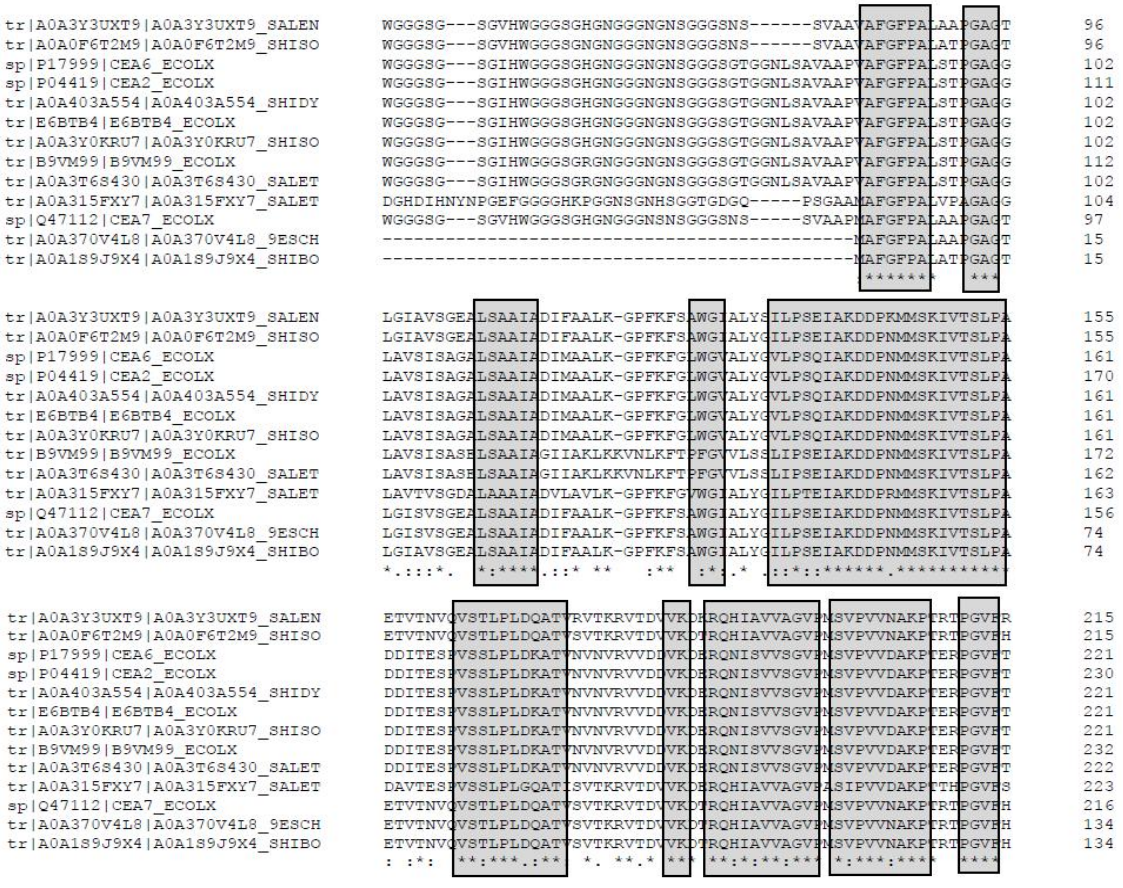
3.4 MSA Analysis of Colicins

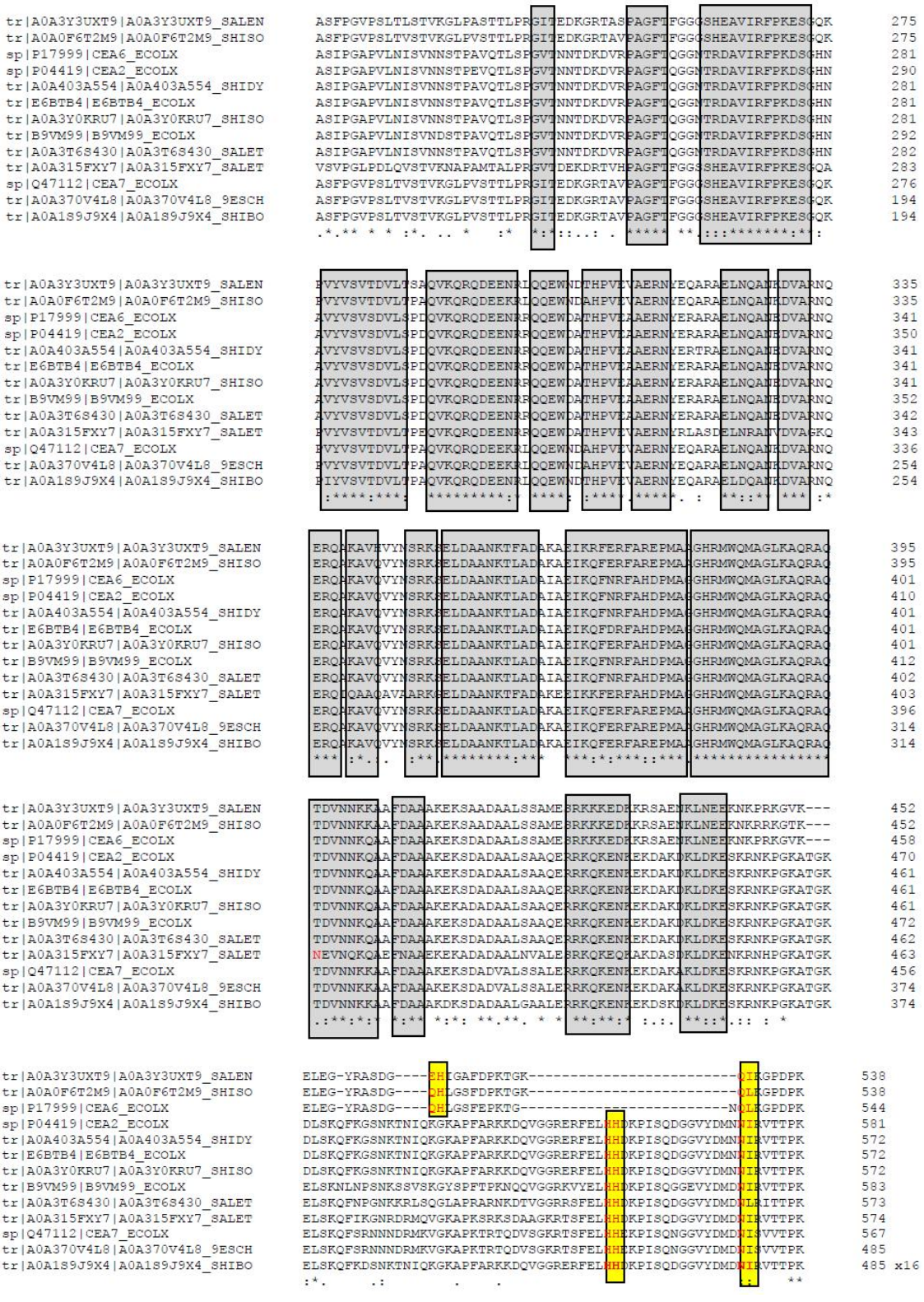
Colicins, a type of bacteriocin, are antimicrobial proteins produced by *Escherichia coli*. Upon secretion from the host, colicin binds to its receptors on the outer membrane of susceptible bacterial cells and kills them by forming pores in the inner membrane and also degrades cellular DNA and RNA *nonspecifically* which is in quite contrast to the other site-specific HNH endonucleases. Thus, colicins provide a competitive advantage to the colicin producing *E. coli* over other species for complete nutrient utilization and growth. It is interesting to note that the colicin-producing host cells are protected by an immunity protein that binds and blocks the activity of its cognate colicin. Bacteriocins are named after their species of origin, e.g., colicins

are produced by *E. coli*. In fact, due to their unique bacterial cell-killing activities, the bacteriocins are considered as viable alternatives to conventional antibiotics, as they are not toxic to humans and their cytotoxicity is effective only on bacteria that produce specific receptor proteins on their membranes which are not present in human cells. In fact, several bacterial toxins, including colicin E7 (ColE7), also contain the 30 amino acid HNH motif in their nuclease domains and uses a single metal ion active site with a Zn atom.

Fig. 6 shows the MSA of colicins from various sources. The active sites are found in the C-terminal region of the enzyme unlike in other HNH homing endonucleases (Fig. 3). This is found to be a –HH- based endonucleases as -HH –NL/I- and –H- are completely conserved in all the colicins analyzed here. (In some cases the second N from the proton acceptor is replaced by an equivalent amino acid, viz. Q (Fig. 6). Furthermore, the second H in HNH shows some sequence tolerance, i.e., it is replaced by an N in some organisms as HNN. The colicin-E3 from *Shigella sonnei* shows an S but is preceded by an invariant R as in all cases. Huang and Yuan [22] have studied the nuclease domain of ColE7 (N-ColE7) by SDM to find the role of the conserved N and H residues in the HNH motif of colicins. Interestingly, DNA cleavage activity of H^{545}→N-ColE7 mutant was completely abolished while activities of N^{560} and H^{573} mutants varied from 6.9% to 83.2% of the wild-type activity. These results suggest that the highly conserved N in the HNH motif, in general, plays a structural role in constraining the loop in the metal finger structure and keeping the general base H and scissile phosphate in the correct position for DNA hydrolysis. In most of the cases, the second H is followed by an Ile. The first three sequences did not harbour the typical –HH- diad but the H is conserved among them at the expected distance from the base binding N. Wy et al. [23] have shown that the zinc ion in the HNH motif of the endonuclease domain of colicin E7 is not required for DNA binding but is essential for DNA hydrolysis.

CLUSTAL O (1.2.4) MSA of Colicins





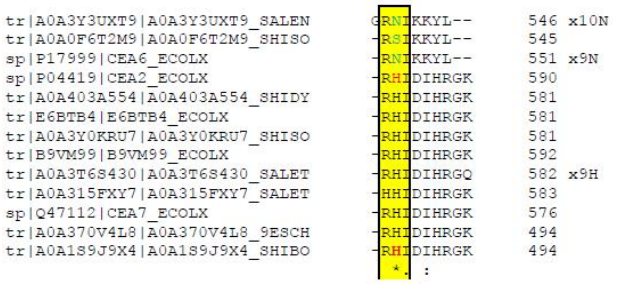


Fig. 6. MSA of Colicins from different organisms

A0A3Y3UXT9_SALEN Colicin, *Salmonella enteritidis*
A0A0F6T2M9_SHISO Colicin-E3, *Shigella sonnei*
P17999|CEA6_ECOLX Colicin-E6, *Escherichia coli*
P04419|CEA2_ECOLX Colicin-E2, *Escherichia coli*
A0A403A554_SHIDY Colicin, *Shigella dysenteriae*
E6BTB4_ECOLX Colicin-E3, *Escherichia coli MS 85-1*
A0A3Y0KRU7_SHISO Colicin, *Shigella sonnei*
B9VM99_ECOLX Colicin E9, *Escherichia coli*
A0A3T6S430_SALET Colicin, *Salmonella enteric*
A0A315FXY7_SALET Colicin, *Salmonella enteric*
Q47112|CEA7_ECOLX Colicin-E7, *Escherichia coli*
A0A370V4L8_9ESCH Colicin-E7, *Escherichia marmotae*
A0A1S9J9X4_SHIBO Colicin, *Shigella boydii*

3.5 MSA Analysis of Pyocins

The bacteriocins produced by *Pseudomonas* species are called pyocins, and, in contrast to colicins whose genes are plasmid-borne, the pyocin genes are found on the chromosome itself [24,25]. Pyocins are produced by more than 90% of *Pseudomonas aeruginosa* strains and each strain may synthesize several pyocins. Structurally the pyocins contain three domains, viz. N-terminal receptor-binding domain, translocation domain and C-terminal DNase domain. It is interesting to note that in colicins also the DNA cleavage domain is located in the C-terminal domain only, as in pyocins (Fig. 6). Pyocins not only cause breakdown of chromosomal DNA but also involve in the complete inhibition of lipid biosynthesis in sensitive cells. Three types of pyocins are reported and are known as R, F and S types.

1) R-type pyocins resemble non-flexible and contractile tails of bacteriophages. They depolarize the cytoplasmic membrane in relation with pore formation.
2) F-type pyocins also resemble phage tails, but with a flexible and non-contractile rod-like structure.
3) S-type pyocins are colicin-like, protease-sensitive proteins. They are constituted of two components. The large component carries the killing activity (DNase activity in pyocins S1, S2, S3, AP41; tRNase activity in pyocin S4 and channel-forming activity in pyocin S5). The killing domains of S1, S2, AP41 pyocins show a close evolutionary relationship with E2 group colicins, S4 pyocin with colicin E5, and S5 pyocin with colicins Ia, and Ib [25].

Fig. 7 shows the MSA of pyocins from different organisms. These also belong to –HH- based HNH endonucleases and the -HH –NL- and –H- are completely conserved in all the sources. The C-terminal regions are more conserved among them and like colicins the active site region is placed at the C-terminal region (Fig. 7). Like in the colicins, in the pyocins also the third H is followed by an invariant Ile and the second N is followed by a branched-chain amino acid, viz. a Leu.

3.6 MSA Analysis of Group II Introns Harbouring Both a Reverse Transcriptase (Rtase) and an HNH Endonuclease

Though both Group I and II introns splice through RNA catalyzed pathways, product of group I introns encode usually a site-specific endonuclease (HNH type) whereas the products of group II introns are usually a reverse transcriptase (Rtase) with an associated endonuclease activity (HNH type) for intron movement. Rtases containing the intronic

CLUSTAL O (1.2.4) MSA of pyocins

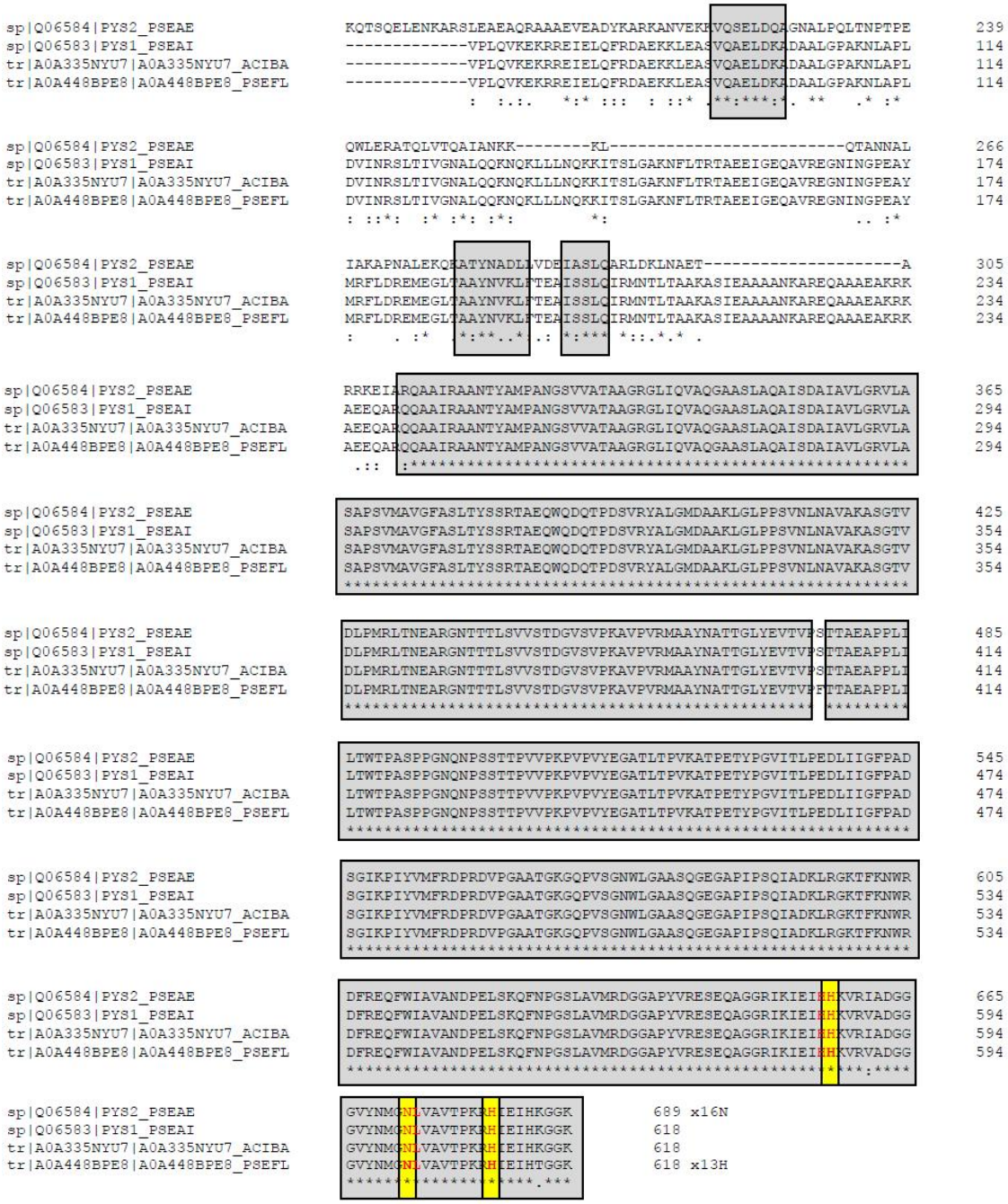


Fig. 7 MSA of pyocins from different organisms

Q06584|PYS2_PSEAE Pyocin-S2, *Pseudomonas aeruginosa*
Q06583|PYS1_PSEAI Pyocin-S1, *Pseudomonas aeruginosa*
A0A335NYU7_ACIBA Pyocin-S1, *Acinetobacter baumannii*
A0A448BPE8_PSEFL Pyocin, *Pseudomonas* sp

proteins are found in both mitochondrial and non-mitochondrial origins [26]. Fig. 8 shows the MSA analysis of group II introns harbouring both an Rtase and an associated HNH endonuclease as discussed above. The Rtase is located in the middle portion of the enzyme (e.g., in *S.*

cerevisiae, an ascomycete fungus, it is placed between 329-613, highlighted in magenta; in *P. anserine*, an ascomycete fungus, it is placed between 228 and 527; in *S. pombe*, it is placed between 281-566; in *K. lactis* it is placed between 261 and 545; etc.) whereas the HNH endonuclease is placed invariably at the end of the C-terminal region. The group II intron HNH endonucleases are typical –HH- type of enzymes and they confer specificity to the group II introns and make a DSB on the recipient DNA and during 'homing' process both are incorporated into the recipient genome. A completely conserved region of probable Rtase catalytic region and a possible primer binding region Y/HG are highlighted in yellow.

The HNH endonuclease appears to be a zinc finger metalloenzyme with an invariant Zn binding motif CxxC about 9 amino acids towards N-terminal of the probable proton acceptor H. The invariant N which confers the nucleotide

CLUSTAL O (1.2.4) MSA of group II Introns harbouring Rtase and HNH endonuclease

```
tr|A0A0S2LQ72|A0A0S2LQ72_9CHLO      KAKFCTN-EDYLNMVDKLSEIVNDPNGYKAIPLDRLYIAKKDGRKRPLSIPSYTDRCLQA      156
tr|Q1KVS2|Q1KVS2_TETOB             KESFRTN-NDYISMMARLEEITADPQKYKATPLSRIYIPKRDGNSRPLSIPSYTDRCLQA      145
sp|P19593|RDPO_TETOB                RESFKTN-KNYVAMMATLEQITSNPHKYKATPLSRIYIPKRDGSARPLSIPSYTDRCLQA      146
tr|A0A249RX17|A0A249RX17_TETOB      RESFKTN-KNYVAMMATLEQITSNPHKYKATPLSRIYIPKRDGSARPLSIPSYTDRCLQA      146
tr|A9IAV6|A9IAV6_BORPD              GSSLDGMSYER---LAGLMAA-VKSGNYRFKPVRRVLIPKSNGKTRPLGIPTGDDKLVQE      111
tr|Q35366|Q35366_PODAS              SETLDGISKEW---FEKISEQ-LKSEQFRFRPTRRVYIPKANGKMRPLGIASPRDKIVQE      269
sp|P05511|YMC6_SCHPO                PDTLDGMSIDV---IDKIIQS-LKSEEFNFTPGRRILIDKASGGKRPLTIGSPRDKLVQE      322
P03875                              LETLDGMNMMY---LNKLSNE-LGTGKFKFKPMRMVNIPKPKGGMRPLSVGNPRDKIVQE      337
P03876.2                            NITLDGINISY---LNKLSKD-INTNMFKFSPVRRVEIPKTSGGFRPLSVGNPREKIVQE      370
tr|A0A0H3WIQ7|A0A0H3WIQ7_YEASX      NITLDGINISY---LNKLSKD-INTNMFKFSPVRRVEIPKTSGGFRPLSVGNPREKIVQE      301
tr|C0J5P2|C0J5P2_SACPW              NITLDGINISY---LNKLSKD-INTNMFKFSPVRRVEIPKTSGGFRPLSVGNPREKIVQE      370
tr|A0A1D8GYT0|A0A1D8GYT0_SACPS      NITLDGINISY---LNKLSKD-INTNMFKFSPVRRVEIPKTSGGFRPLSVGNPREKIVQE      301
X57546                              NITLDGINISY---LNKLCKH-INTNMFKFSPVRRVEIPKTSGGFRPLSVGNPREKIVQE      302
tr|Q34832|Q34832_KLULC              NITLDGINISY---LNKLCKH-INTNMFKFSPVRRVEIPKTSGGFRPLSVGNPREKIVQE      302
                                      .:         .  :       :.  *   : * * .*  *** : .  :: :*

tr|A0A0S2LQ72|A0A0S2LQ72_9CHLO      LYKLALEPIAEEMSDLSSYGFRPIRSTSWAVGRVLNLLANSLANYSFVVEIDILGCFDNI      216
tr|Q1KVS2|Q1KVS2_TETOB             LYKLAIEPMAEEVADVSSYGFRPMRNVSWAVGRVLNGLNNPLANYQYVVEIDIKGCFDNI      205
sp|P19593|RDPO_TETOB                LYKLAIEPMAEEVADLSSYGFRPMRNVSWAVGRVLNGLNNPLANYQYVVEIDIKGCVDNI      206
tr|A0A249RX17|A0A249RX17_TETOB      LYKLAIEPMAEEVADLSSYGFRPMRNVSWAVGRVLNGLNNPLANYQYVVEIDIKGCFDNI      206
tr|A9IAV6|A9IAV6_BORPD              VVRMLLVKIYEPVFSDDSHGFRNGRSCHTALMQVRQK----WTGMKWIVNMDIKGYFDNI      167
tr|Q35366|Q35366_PODAS              VFRAILEQVLEPRFHSSSHGFRPGRGCHSALATI-RY----WNGIKWFIEGDIKGFFDNI      324
sp|P05511|YMC6_SCHPO                ILRIVLEAIYEPLFNTASHGFRPGRSCHSALRSIFTN----FKGCTWWIEGDIKACFDSI      378
P03875                              VMRMILDTIFDKMSTHSHGFRKNMSCQTAIWEVRNM----FGGSNWFIEVDLKKCFDTI      393
P03876.2                            SMRMMLEIIYNNSFSYYSHGFRPNLSCLTAIIQCKNY----MQYCNWFIKVDLNKCFDTI      426
tr|A0A0H3WIQ7|A0A0H3WIQ7_YEASX      SMRMMLEIIYNNSFSYYSHGFRPNLSCLTAIIQCKNY----MQYCNWFIKVDLNKCFDTI      357
tr|C0J5P2|C0J5P2_SACPW              SMRMILEIIYNNSFSNYSHGFRPNLSCLTAIIHCKNY----MQHCNWFIKVDLNKCFDTI      426
tr|A0A1D8GYT0|A0A1D8GYT0_SACPS      SMRMILEIIYNNSFSNYSHGFRPNLSCLTAIIHCKNY----MQHCNWFIKVDLNKCFDTI      357
X57546                              SMRMILEIIYNNSFSNYSHGFRPNLSCLTAIVHCKNY----MKHCNWFIKVDLNKCFDTI      358
tr|Q34832|Q34832_KLULC              SMRMILEIIYNNSFSNYSHGFRPNLSCLTAIVHCKNY----MKHCNWFIKVDLNKCFDTI      358
                                      :  :  : :      *:***   .   *:              : :: *:   .*.*

tr|A0A0S2LQ72|A0A0S2LQ72_9CHLO      DHTFLMQFVPVI-----PKKILWEWLSCGYVERDDNKEVHETLRGVPQGGILSPLLSNLT      271
tr|Q1KVS2|Q1KVS2_TETOB             DHEFVSQVTPFI-----PKSILWSWLKCGYLEQNSE-QLQPTTSGVPQGGIISPLIMNLT      259
sp|P19593|RDPO_TETOB                NHQFISQVTPFI-----PKKILWAWLKCGYIERNSN-TLQPTTTGVPQGGIISPLIMNLT      260
tr|A0A249RX17|A0A249RX17_TETOB      NHQFISQVTPFI-----PKKILWAWLKCGYIERNSN-TLQPTPTGVPQGGIISPLIMNLT      260
tr|A9IAV6|A9IAV6_BORPD              DHEVLVDVLAKRIDDKRFLGLIHSMLKAGYMED-W--KFHDTFSGTPQGGVVSPVLANIY      224
tr|Q35366|Q35366_PODAS              DHHILEKLLVKHFQDQRFIDLYWKMVKAGYVEF-D--KDKSSIIGVPQGGIASPILSNLV      381
sp|P05511|YMC6_SCHPO                PHDKLIALLSSKIKDQRFIQLIRKALNAGYLTE-N--RYKYDIVGTPQGSIVSPILANIY      435
P03875                              SHDLIIKELKRYISDKGFIDLVYKLLRAGYIDEKG--TYHKPMLGLPQGSLISPILCNIV      451
P03876.2                            PHNMLINVLNERIKDKGFMDLLYKLLRAGYVDKNN--NYHNTTLGIPQGSVVSPILCNIF      484
tr|A0A0H3WIQ7|A0A0H3WIQ7_YEASX      PHNMLINVLNERIKDKGFMDLLYKLLRAGYVDKNN--NYHNTTLGIPQGSVVSPILCNIF      415
tr|C0J5P2|C0J5P2_SACPW              PHNMLINVLNERIKDKGFIDLLYKLLRAGYVDKHN--NYHHTTLGIPQGSVVSPILCNIF      484
tr|A0A1D8GYT0|A0A1D8GYT0_SACPS      PHNMLINVLNERIKDKGFIDLLYKLLRAGYVDKHN--NYHHTTLGIPQGSVVSPILCNIF      415
X57546                              PHNMLINVLNERIKDKGFIDLLYKLLRAGYVDKHN--NYHHTTLGIPQGSVVSPILCNIF      416
tr|Q34832|Q34832_KLULC              PHNMLINVLNERIKDKGFIDLLYKLLRAGYVDKHN--NYHHTTLGIPQGSVVSPILCNIF      416
                                     *  :              :    : .**:        :    *  **.: **:: *:

tr|A0A0S2LQ72|A0A0S2LQ72_9CHLO      LDGLEDHIRKRIAES-------------------------------------------      286
tr|Q1KVS2|Q1KVS2_TETOB             LDGLEFHIYKRIQQS-------------------------------------------      274
sp|P19593|RDPO_TETOB                LDGLEFHIYKKIQKS-------------------------------------------      275
tr|A0A249RX17|A0A249RX17_TETOB      LDGLEFHIYKKIQKS-------------------------------------------      275
tr|A9IAV6|A9IAV6_BORPD              LHELDEYVAGLKAEFNRGNR----RASNREYKRISGAIERLMKRIDAYKADGDSPKVEEA      280
tr|Q35366|Q35366_PODAS              LNELDEFVQNIVDEFNEKLKGGKHTSKNPAYVVIDSRIGKITRLERKLKSKGQELDS---      438
sp|P05511|YMC6_SCHPO                LHQLDEFIENLKSEFDYKGPI-A-RKRTSESRHLHYLMAKA-----------------      474
P03875                              MTLVDNWLEDYI-NLYNKGKV---KKQHPTYKKLSRMIAKA-----------------      488
P03876.2                            LDKLDKYLENKFENEFNTGNM-SNRGRNPIYNSLSSKIYRC-----------------      524
tr|A0A0H3WIQ7|A0A0H3WIQ7_YEASX      LDKLDKYLENKFENEFNTGNM-SNRGRNPIYNSLSSKIYRC-----------------      455
tr|C0J5P2|C0J5P2_SACPW              LDKLDKYLENKFENEFNIGSM-SNRSRNPIYNDLSSKIRRC-----------------      524
tr|A0A1D8GYT0|A0A1D8GYT0_SACPS      LDKLDKYLENKFENEFNIGSM-SNRSRNPIYNDLSSKIRRC-----------------      455
X57546                              LDKLDKYLENKFENEFNTGSM-SNRSRNPIYNDLSSKIRRC-----------------      456
tr|Q34832|Q34832_KLULC              LDKLDKYLENKFENEFNTGSM-SNRSRNPIYNDLSSKIRRC-----------------      456
                                    :  ::  :     :
```

Fig. 8. MSA of reverse transcriptase and HNH endonuclease in group II introns

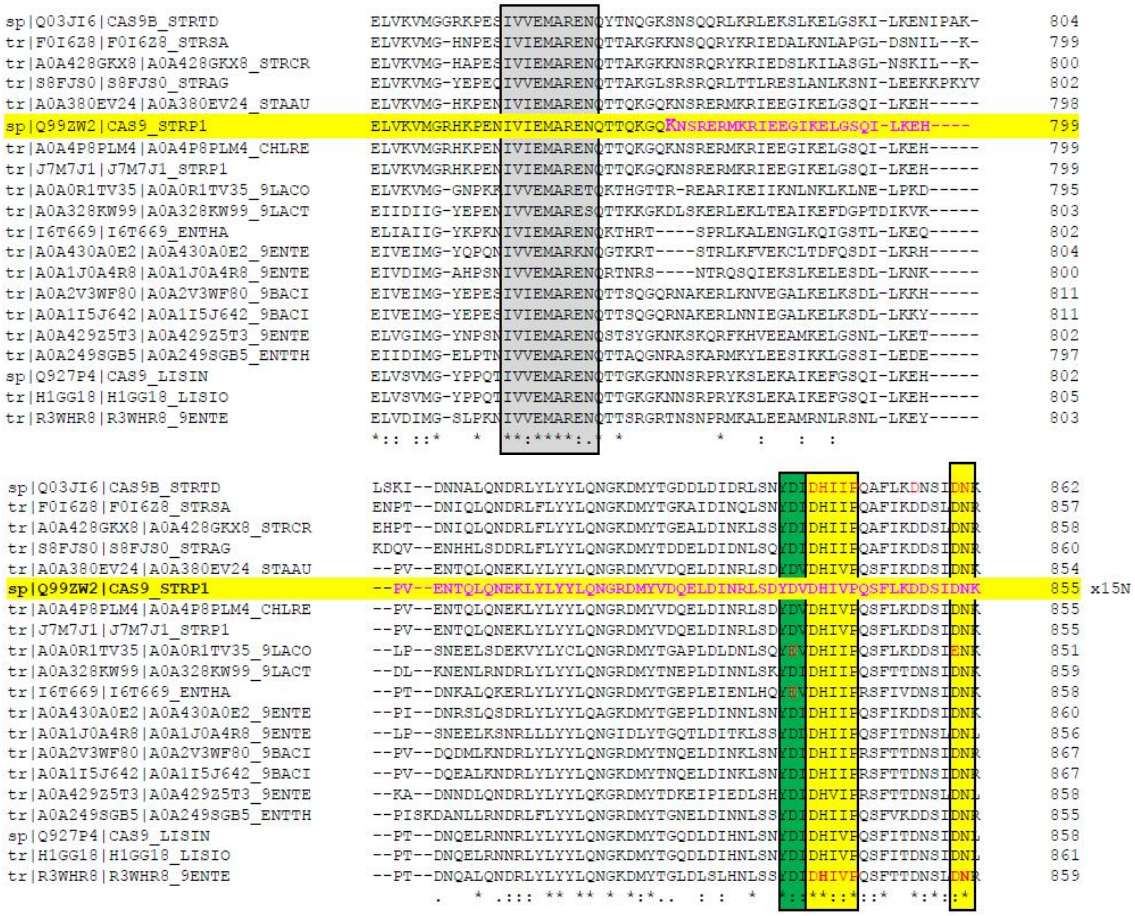
A0A0S2LQ72_9CHLO Putative reverse transcriptase, *Bracteacoccus giganteus*
Q1KVS2_TETOB Intron-encoded reverse transcriptase, *Tetradesmus obliquus*
RDPO_TETOB Group II intron-encoded Probable reverse transcriptase, *Tetradesmus obliquus*
A0A249RX17_TETOB Group II intron-encoded protein LtrA, *Tetradesmus obliquus*
A9IAV6_BORPD Mobile mitochondrial Group II intron of COX1, *Bordetella petrii*
Q35366_PODAS Cytochrome- *c* oxidase subunit I (Fragment), *Podospora anserina*
P05511|YMC6_SCHPO Uncharacterized 91 kDa protein in cob intron, *S. pombe*
P03876.2 Mitochondrial Intron 2 COX1-OXI3 gene, *S. cerevisiae*
A0A0H3WIQ7_YEASX Reverse transcriptase domain-containing protein, *S. cerevisiae*
C0J5P2_SACPW Reverse transcriptase, *S. pastorianus*
A0A1D8GYT0_SACPS Reverse transcriptase (Fragment), *S. pastorianus*
X57546 Group II intron-encoded reverse transcriptase, Mitochondrial COX1 gene, *K. lactis*
Q34832_KLULC Intron ORF, *K. lactis*

specificity is found towards the C-terminal end with respect to the invariant H but not completely aligned in all the sequences. However, the second H of the HNH is completely conserved as a CH diad. Interestingly, one more invariant CxxC (CxxxC in first four sequences) is located in front of the second H of the HNH, which could also participate in Zn binding. Except for *Bordetella petrii*, all show a conserved N at the C-terminal region (marked in red) from the proton acceptor where a conspicuous absence of N is observed, but it possesses a DxD motif (marked in green) a divalent metal-binding site (Fig. 8).

3.7 MSA of CRISPR-Cas9 HNH Endonucleases

Fig. 9 shows the MSA of only the HNH regions of Cas9 endonucleases from various sources. The *S. pyogenes'* enzyme is used as the standard and highlighted in yellow. The HNH endonuclease region of Spycas9 is highlighted in magenta. These Cas9 HNH enzymes are found to be of –DH- type. It is interesting to note that the typical CxxC zinc finger motif, downstream and upstream from the proton acceptor H, is not found in any of these CRISPR-Cas9 HNH endonucleases. The only exception is the HNH endonuclease from of *A. naeslundii*'s HNH-Cas9, i.e., the HNH region of the CRISPR-Cas9 from *A. naeslundii* showed the typical –CxxC- motifs as found in other HNH endonucleases (Fig. 5). However, in all other Cas9 HNH endonucleases a -D/ExD- metal-binding motif very close to the proton acceptor is observed. The first and second invariant Ns are followed by mostly an R/K (Fig. 9). This is in close agreement with other –DH- based enzymes where they show two conserved Ns after the catalytic -DH- diad (Fig. 5).

CLUSTAL O (1.2.4) MSA of CRISPR-Cas9 endonucleases (Shown HNH region 775-908)



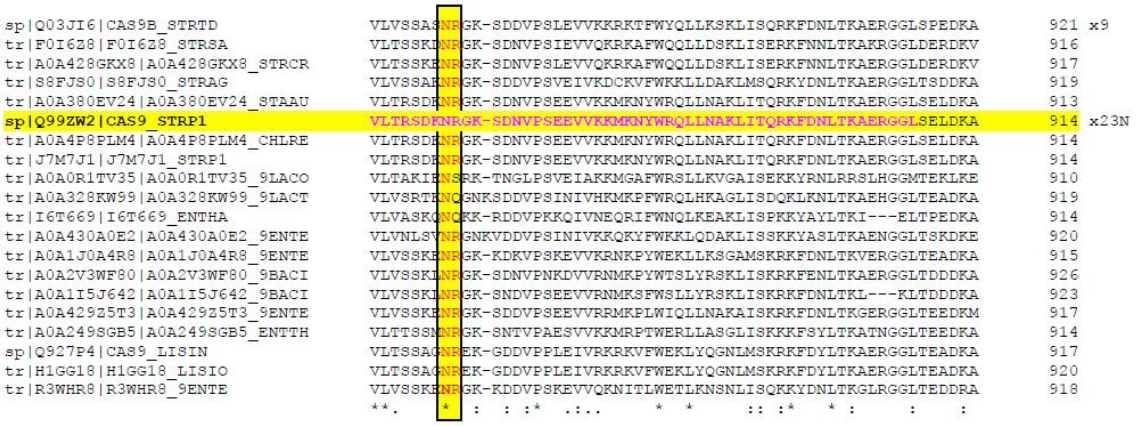


Fig. 9 MSA of CRISPR-Cas9 HNH endonucleases

The HNH endonuclease portion is shown in magenta
Q03JI6|CAS9B_STRTD CRISPR-associated endonuclease Cas9 2, *Streptococcus thermophilus*
F0I6Z8_STRSA CRISPR-associated endonuclease Cas9, *Streptococcus sanguinis* SK115
A0A428GKX8|A0A428GKX8_STRCR CRISPR-associated endonuclease Cas9, *Streptococcus cristatus*
S8FJS0_STRAG CRISPR-associated endonuclease Cas9, *Streptococcus agalactiae* FSL S3-277
Q99ZW2|CAS9_STRP1 CRISPR-associated endonuclease Cas9/Csn1,*Streptococcus pyogenes* serotype M1
A0A4P8PLM4_CHLRE Cas9c, *Chlamydomonas reinhardtii*
J7M7J1_STRP1 CRISPR-associated endonuclease Cas9, *Streptococcus pyogenes*, M1 476
A0A0R1TV35_9LACO CRISPR-associated endonuclease Cas9, *Lactobacillus apodemi*
A0A328KW99_9LACT CRISPR-associated endonuclease Cas9, *Dolosigranulum pigrum*
I6T669|I6T669_ENTHA CRISPR-associated endonuclease Cas9, *Enterococcus hirae*
A0A430A0E2_9ENTE CRISPR-associated endonuclease Cas9, *Vagococcus* sp.
A0A1J0A4R8_9ENTE CRISPR-associated endonuclease Cas9, *Vagococcus teuberi*
A0A2V3WF80_9BACI CRISPR-associated endonuclease Cas9, *Streptohalobacillus salinus*
A0A1I5J642_9BACI CRISPR-associated endonuclease Cas9, *Halolactibacillus alkaliphilus*
A0A429Z5T3_9ENTE CRISPR-associated endonuclease Cas9, *Vagococcus humatus*
A0A249SGB5_ENTTH CRISPR-associated endonuclease Cas9, *Enterococcus thailandicus*
Q927P4|CAS9_LISIN CRISPR-associated endonuclease Cas9, *Listeria innocua* serovar
H1GG18_LISIO CRISPR-associated endonuclease Cas9, *Listeria innocua*
R3WHR8_9ENTE CRISPR-associated endonuclease Cas9, *Enterococcus phoeniculicola*

4. ACTIVE SITE REGIONS OF THE WELL KNOWN HNH ENDONUCLEASES

A summary of all the active site regions of the well known HNH endonucleases is shown in Table 1. The possible metal-binding regions are highlighted in green. The Cs involved in forming the zinc finger-type motif is shown in magenta. The smaller version of CRISPR-Cas9 (1101 amino acids) from *A. naeslundii* shows a zinc finger motif, whereas the larger version of CRISPR-Cas9 from *S. pyogenes* (1368 amino acids) shows DxD metal-binding motifs. Some of the HNH endonucleases shows typical HNH motif whereas others show HNN motif. However, SDM experiments have shown H or N is equally efficient and exhibit similar enzymatic activities. In some the HNH endonucleases, including the CRISPR-Cas9, the conserved second H are not observed, suggesting that without this second H, the enzymes can efficiently cleave the substrate DNAs. From the Table, it is clear that the *S. pyogenes'* HNH deviates from other zinc finger types and shows similarity to phage type homing endonucleases with a DxD metal-binding site but a –DH- based catalytic diad.

5. METAL-BINDING SITES IN HOMING ENDONUCLEASES

HNH endonucleases use different divalent metal ions in catalysis. For example, the metal ions located is a magnesium ion in Vvn, *Serratia* nuclease and I-*Ppo*I; a calcium ion in Endo VII and a zinc ion in ColE7. Based on the MSA analysis three different types of metal-binding sites are observed.

Table 1. Active site regions in different HNH endonucleases

Type	Organism	Active site region
HH- Homing endonucleases	(Bacteriophage Bp7, I-*Tev*III)	-YEIHHKDGNRENNDLDNLMCLSIQEHYDIHLAQKDY[58-]
HH- based group II introns	(*S. cerevisiae*)	-QICGSKHDLEVHHVRTLNNAANKIKDDYLLGRMIKMNRKQITICKTCHF[842]-
HH- based mcrA restriction endonuclease (*E. coli)*		-CENCGKNAPFYLNDGNPYLEVHHVIPLSSGGADTTDNCVALCPNCHRELHYS[258]-
DH- based HNH endonucleases (*E. coli* plasmids)		GGRCAYCGCELPEKGWYADHVQAVLRKSEQCMKAAEKRIFRLK STGDVFRPEADCPENLVPACAPCNLLK[85]
DH- based HNH endonucleases (*E. proavitum)*		-GICQKCKTHFEINEMEADHITPWHEGGKTTSVNCQMLCKDCNRRK[355]
DH- based Endonuclease VII (Resolvase) (T4 Phage)		-GKCLICQRELNPDVQANHLDHDHELNGPKAGKVRGLLCNL CNAAEGQMKHKFNR[74]-
HH based- Colicins endonucleases (Type 9*)	(*E. coli*)	-YELHHDKPISQGGEVYDMDNIRVTTPKRHIDIHRGK[592]
HH- Pyocins endonucleases (Type-S1)	(*P. aeruginosa)*	-KIEIHHKVRVADGGGVYNMGNLVAVTPKRHIEIHKGGK[618]
HH- Pyocins endonucleases (Type-S2)	(*P. aeruginosa)*	-KIEIHHKVRIADGGGVYNMGNLVAVTPKRHIEIHKGGK[689]
DH- basedCRISPR-Cas9 HNH endonucleases (*S. pyogenes)*		-YVDQELDINRLS**DYD**VDH[840]IVPQSFLKDDSIDNKVLTRSDKNRGK[865]
DH- based CRISPR-Cas9 HNH endonucleases *(A. naeslundii)*		-ACLYCGTTIGYHTCQLDH[582]IVPQAGPGSNNRRGNLVAVCERCNR SKSNTPFAVWAQKCGIPHVG[627]-

*NB: *Colicins 2-9 from E coli have identical sequences at the active site region (Fig. 10)*

1) With DxD type Mg^{2+} binding motif(s) using a Mg^{2+} ion-bound water molecule as the nucleophile for initiating the catalysis, e.g., *S. pyogenes* CRISPR-Cas9 HNH endonuclease
2) With CxxC or CxxxC type Zn binding motif(s) using a Zn^{2+} ion-bound water molecule as the nucleophile for initiating the catalysis, e.g., bacteriocin type of HNH endonucleases
3) With no well known conserved metal-binding motif as above.

HNH endonucleases use invariably a divalent cation as a cofactor for catalysis. Usually a Mg^{2+} or Zn^{2+} atom is found to be at the centre of catalytic events in most of the HNH endonucleases. It is suggested that these metal ions play an important role in the activation of a nucleophilic water molecule bound to them. For example,

The homing endonuclease *Ppo*I from the slime mould *Physarum polycephalum* was analyzed by both by X-ray crystallography and SDM experiments. Flick et al. [27] found by X-ray crystallographic analysis of the enzyme that the Mg^{2+} bound water is deprotonated by a general base and His^{98} was found to be in the active site. Furthermore, Mannino et al. [28] on analyzing the same enzyme by SDM experiments found that the three amino acid residues, viz. Arg^{61}, His^{98}, and Asn^{119} are important for efficient DNA cleavage by the enzyme. This finding was consistent with their proposed mechanism in which His^{98} abstracts a proton from an attacking water molecule and Arg^{61} and Asn^{119} stabilize the pentavalent transition state. They also suggested that the Asn^{119} also binds to the essential divalent metal cation, i.e., the Mg^{2+} ion.

Type II restriction endonuclease R.KpnI, a member of the HNH endonuclease with the typical β-β-α-Metal fold, was studied by SDM by Saravanan et al. [29]. They found that D^{148}, H^{149} and Q^{175} in the active site of the enzyme. The mutant enzyme Q^{175}→E fails to bind DNA at the standard conditions, although the DNA binding and cleavage can be rescued at pH 6.0, indicating a role for Q175 in DNA binding. They also found by SDM experiments that the proton acceptor H^{149}→L mutant showed no detectable activity, even at 100-fold excess protein concentrations. Other mutants D^{148}→G and Q^{175}→E showed only traces of the DNA cleavage activity when used in large excess suggesting that the DH----Q are the important amino acids playing a critical role in catalysis. They also suggested that the D148 which is found near the proton acceptor might be also involved in metal-binding, possibly an Mg^{2+} like other Type II restriction endonucleases.

The T_4 endonuclease VII also belongs to HNH endonucleases and is also known as resolvase as it involves in resolving Holliday junctions during recombination events. Raaijmakers et al. [30] by X-ray crystallographic analysis of the enzyme found that the similarly placed amino acids, viz. D^{40}, H^{41} and N^{62} play a similar role in catalysis as in R.KpnI. However, unlike the R.KpnI enzyme, this enzyme contains one Zn atom per molecule coordinated by 4 Cs. Interestingly, Giraud-Panis et al. [31] have shown by SDM experiments that the Cys mutants (C^{23}→S and C^{61}→S) which no longer binds Zn atom was found to be inactive. They found that the SDM analysis also showed that the outer cysteine residues marked in bold $\mathbf{C^{23}}$, C^{25}, C^{58} and $\mathbf{C^{61}}$ are essential for zinc binding, whereas the inner Cs are redundant. Furthermore, they also found that the mutant enzyme N^{62}→D is inactive and therefore, suggested that the D^{40} and N^{62} could act as a ligand for binding the metal ion which participates in catalysis. MSA analysis shows that this is an unusual type of HNH endonuclease as it contains both the highly conserved Zn^{2+} binding motif, coordinated by the conserved 4Cs and also the DxD Mg^{2+} binding motif (-HL$\mathbf{D^{40}HD}$HE- marked bold) in its structure. By X-ray crystallographic analysis, Raaijmakers et al. [30] have shown that D^{40} and N^{62} are essential for the activity of the enzyme, which binds to the metal ion. They suggested that the Zn atom possibly could play a structural role and in maintaining the structural stability of the enzyme.

Li et al. [32] have studied the crystal structure of a periplasmic endonuclease, Vvn from *Vibrio vulnificus,* for its DNA binding and cleavage properties. The overall structure of the enzyme showed no similarity with other HNH enzymes but a ββα-metal motif was identified. When the proton acceptor His was subjected to SDM, the mutant enzyme H^{80}→A did not show any activity. The conserved Asn^{127} was found to be responsible for metal-binding (Mg^{2+}) and is also well conserved among this group of endonucleases, except that it is replaced by a histidine (His^{569}) in ColE7.

It is interesting to note that Wy et al. [23] found in the crystal structure of the ColE7 HNH

endonuclease a water molecule is bound to the Zn atom at the active site suggesting Zn^{2+} bound water molecule at the active site play a crucial role in catalysis. Based on these studies with various HNH endonucleases from different sources, it is clear that these enzymes may make DSB on DNAs by a single-metal ion mechanism.

Palanivelu [33] have shown that the proof-reading activity (an exonuclease) in DNA polymerase I use a metal- bound water molecule.

6. MECHANISM OF ACTION OF SpyCRISPR-Cas9 HNH ENDONUCLEASE

There are reports suggesting either the active site His-bound nucleophilic water molecule initiation of the DSB on the DNA molecule or the divalent metal ion-bound nucleophilic water molecule initiating the DSB on the DNA molecule. Based on the evidences presented in this communication, including the SDM and X-ray crystallographic data, it is clear that the metal-bound nucleophilic water molecule could be a more plausible initiator of reaction.

The proposed mechanism for Cas9 HNH endonucleases is based on the MSA data obtained from various Cas9 HNH endonucleases from this communication along with the supporting data from already published reports from experiments like SDM experiments, X-ray crystallographic analysis, etc.

Unlike most of the HNH endonucleases, the CRISPR-Cas9 enzymes use a completely conserved DXD type of metal-binding (in one or two cases a functionally equivalent ExD motif is found) (Fig. 8). In type II restriction enzymes two acidic residues D and D/E are involved in metal-binding, i.e., binding to a Mg^{2+} ion [34]. In fact, Mg^{2+} binding confers specificity and cleavage at the palindromic sequences only.

Furthermore, the HNH enzyme, Vvn uses a very similar type of metal-binding motif -E^{77}x$E^{79}H^{80}$ where the X-ray crystallographic analysis of the enzyme have shown that the E79 is the second metal-binding residue in addition to the invariant N^{127} where they bind to a Mg^{2+} ion. The Mg^{2+} ion also binds 3 water molecules [32].

Moreover, a Mg^{2+} ion in the HNH endonuclease I-*Ppo*I has been shown biochemically to accelerate the reactions in three ways: positioning and activating a water molecule to donate a proton to the leaving 3'-oxygen; introducing strain into the substrate complex that is relieved in the product complex and stabilizing the phosphoanion transition state [28].

Therefore, it is proposed in this communication, that the completely conserved catalytic amino acid H in CRISPR-Cas9 enzymes acts as the proton acceptor from the metal-bound water molecule resulting in the highly reactive metal hydroxide which attacks the electrophilic centre on the phosphate, cleaving the phosphodiester bond.

SDM experiments paved the way for more or less pinpointing the active site amino acids in SpyCRISPR-Cas9 HNH endonuclease region. For example, the H^{840}→A did not show any activity and shown as the proton acceptor [17]. In the same way, N^{854}→A showed activity but reduced activity [35,17] and the N^{863}→A did not show any activity and *in vivo* experiments have shown loss of Cas9-mediated CRISPR interference in plasmid transformations [35,17,36].

In the structure of the Endo VII, N^{62}→D mutant in complex with a Holliday junction, a Mg^{2+} ion is coordinated by Asp^{40}, Asn^{62}, and the oxygen atoms of the scissile phosphate group of the substrate. X-ray crystallographic studies of the SpyCRISPR-Cas9 HNH domain have found that Asp^{839}, His^{840}, and Asn^{863} corresponded to Asp^{40}, His^{41}, and Asn^{62} of the T_4 Endonuclease VII, respectively, consistent with the observation that His^{840} is critical for the cleavage of the complementary DNA strand [17,37]. Thus, the SDM experiments and X-ray crystallographic studies have shown at least three amino acids, viz. Asp^{839}, His^{840} and Asn^{863} are critical in substrate binding and DNA cleavage in SpyCRISPR-Cas9 HNH endonuclease.

Furthermore, Gasiunas et al. [16] have shown that Mg^{2+} ions are essential for phosphodiester bond cleavage to occur on both strands, 3 nt upstream of the PAM sequence to generate blunt DNA ends. In the absence of Mg^{2+} ions, no cleavage occurred. Based on these observations, a plausible mechanism is proposed for the SpyCRISPR-Cas9 HNH endonuclease domain.

Moreover, in the CRISPR-Cas9 enzyme from *A. naeslundii* (AnaCas9) the corresponding amino acid H^{582} in the active site acts as the proton acceptor. X-ray crystallographic studies revealed

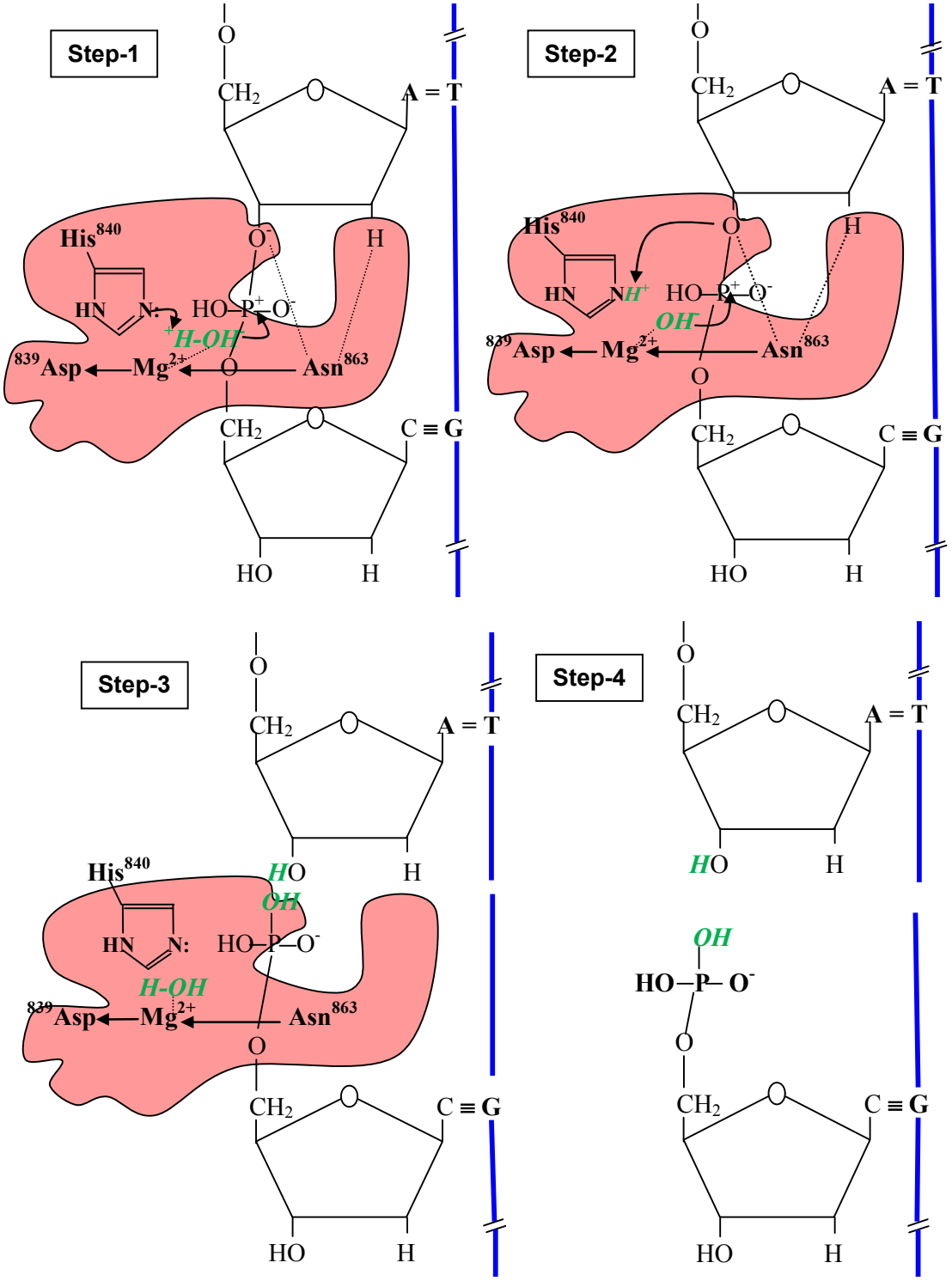


Fig. 10. Proposed mechanism (steps 1-4) for SpyCRISPR-Cas9 HNH endonuclease

NB: The AnaCas9 enzyme though looks different in MSA in having an additional Zn binding motif CxxC (Table 1) could follow the same mechanism as the X-ray crystallographic studies have shown a hydrated magnesium ion in the active site is coordinated by corresponding invariant amino acids D and N [37].

The additional Zn binding site is implicated in structural and conformational stability of this enzyme

the D^{581} and N^{606} in *A. naeslundii* coordinated a hydrated magnesium ion that would be involved in binding the scissile phosphate in the target DNA strand [37] further corroborating the metal-bound water molecule in catalysis.

Based on the above observations, a plausible mechanism for the SpyCRISPR-Cas9 HNH endonuclease is proposed in this communication.

Fig. 10 (steps 1-4) shows the proposed reactions involved in the SpyCRISPR-Cas9 HNH endonuclease.

Step 1. Binding of the enzyme with its active site amino acids at the susceptible phosphodiester bond. The Mg^{2+} ion is coordinated by an invariant D and an N with a bound nucleophilic water molecule.

Step 2. The electronic transition occurs at the active site and the active site His accepts a proton from the metal-bound water molecule resulting in a highly reactive metal hydroxide. The presence of a good general base, like the invariant His in the HNH superfamily is easier to deprotonate than a water molecule.

Step 3. The metal hydroxide now attacks the electrophilic centre on the phosphate, resulting in the cleavage of the phosphodiester bond and the proton transferred to the active site His is now transferred to 3'-Oxygen anion completing the cleavage reaction. Active site of the enzyme is restored.

Step 4. The enzyme moves out for the repair mechanism (homing process) to take over at the DSB. (The non-complementary DNA strand is cleaved by the RuvC endonuclease).

7. CONCLUSIONS

All the HNH endonucleases use an invariant His in the active site. Variations are observed in the metal-binding sites and the metal-binding amino acid residue adjacent to the proton acceptor in different HNH endonucleases from various organisms. For example, some are found to be –HH- type and some are of –DH type. There are marked differences in the metal-binding sites also, e.g., some HNH endonucleases follow the –DxD- type and some follow the typical –CxxC- type and some harbour both. Though both *S. pyogenes* and *A. naeslundii* are Gram-positive organisms, their metal-binding motifs are distinctly different. These variations could be more likely a result of divergent evolution from a common ancestor. From the SDM and X-ray crystallographic studies of these enzymes, it is clear that a metal-bound water molecule could be the nucleophile to initiate the cleavage reactions.

ACKNOWLEDGEMENTS

The author wishes to thank Dr. H. Shakila, Professor & Head, Department of Molecular Microbiology, School of Biotechnology, Madurai Kamaraj University, Madurai, for useful suggestions on the manuscript.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli* and identification of the gene product. Journal of Bacteriology. 1987;169:5429-5433.
2. Deveau H, Garneau JE, Moineau S. CRISPR/Cas system and its role in phage-bacteria interactions. Ann Rev Microbiol. 2010;64:475–493.
3. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. Science. 2010;327:167–170.
4. Marraffini LA, Sontheimer EJ. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet. 2010;11:181–190.
5. Terns MP, Terns RM. CRISPR-based adaptive immune systems. Curr Opin Microbiology. 2011;14:321–327.
6. Pandey VK, Tripathi A, Bhushan R, Ali A, et al. Application of CRISPR/Cas9 genome editing in genetic disorders: A systematic review up to date. J Genet Syndr Gene Ther. 2017;8:2.
7. Selle K, Barrangou R. CRISPR-based technologies and the future of food science. J Food Sci. 2015;80:R2367-R2372.
8. Jansen R, Embden JDAV, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol. 2002;43:1565-1575.

9. Mojica FJM, Díez-Villaseñor C, Soria E, Juez G. Biological significance of a family of regularly spaced repeats in the genomes of archaea, bacteria and mitochondria. Mol Microbiol. 2000;36:244-246.
10. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 2007;315:1709–1712.
11. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science. 2008;321:960–964.
12. Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in Staphylococci by targeting DNA. Science. 2008;322:1843–1845.
13. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV. Evolution and classification of the CRISPR–Cas systems. Nat Rev Microbiol. 2011;9:467–477.
14. Athukoralage JS, Graham S, Grüschow S, Rouillon C, White MF. A type III CRISPR ancillary ribonuclease degrades its cyclic oligoadenylate activator. DOI: http://dx.doi.org/10.1101/582114
15. Ahn WC, Par KH, Bak IS, Song HN, An Y, Lee SJ, Jung M, Yoo KW, Yu DY, Kim YS, Oh BH, Woo EJ. *In vivo* genome editing using the Cpf1 ortholog derived from *Eubacterium eligens.* Nature Scientific Reports. 2019;9:13911.
16. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci (USA). 2012;109:E2579–2586.
17. Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell. 2014;156:935-949.
18. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol. 2013;31:827–832.
19. Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. Cell. 2014;157:1262–1278.
20. Jurica MS, Stoddard BL. Homing endonucleases: Structure, function and evolution. Cell Mol Life Sci. 1999;55:1304-26.
21. Ramalingam R, Prasad R, Shivapriya R, Dharmalingam K. Molecular cloning and sequencing of mcrA locus and identification of mcrA protein in *Escherichia coli.* J Biosci. 1992;17:217–232.
22. Huang H, Yuan HS. The conserved asparagine in the HNH motif serves an important structural role in metal finger endonucleases. J Mol Biol. 2007;368:812-821.
23. Wy KU, Liu YW, Hsu YC, Liao CC, Liang PH, Yuan HS, Chak KF. The zinc ion in the HNH motif of the endonuclease domain of colicin E7 is not required for DNA binding but is essential for DNA hydrolysis. Nucleic Acids Res. 2002;30:1670-1678.
24. Sano Y, Kageyama M. Genetic determinant of Pyocin AP41 as an insert in the *Pseudomonas aeruginosa* Chromosome. Journal of Bacteriology. 1984;158:562-570.
25. Shinomiya T, Sano Y, Kikuchi A, Kageyama M. Mapping of pyocin genes on the chromosome of *Pseudotonis aeruginosa* using plasmid R68.45. In S. Mitsuhashi (Ed.), Drug resistance in bacteria. Japan Scientific Societies Press, Tokyo. 1982;213-217.
26. Gorbalenya AE. Self-splicing group I and group I1 introns encode homologous (putative) DNA endonucleases of a new family. Protein Science. 1994;3:1117-1120.
27. Flick KE, Jurica MS, Monnat RJ Jr, Stoddard BL. DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-*Ppo*I. Nature. 1998;394: 96-101.
28. Mannino SJ, Jenkins CL, Raines RT. Chemical mechanism of DNA cleavage by the homing endonuclease I-*Ppo*I. Biochemistry. 1999;38:16178-86.
29. Saravanan M, Bujnicki JM, Cymerman IA, Rao DN, Nagaraja VN. Type II restriction endonuclease R.KpnI is a member of the HNH nuclease superfamily. Nucleic Acids Res. 2004;32:6129–6135.
30. Raaijmakers H, Toro I, Birkenbihl R, Kemper B, Suck D. Conformational

flexibility in T4 endonuclease VII revealed by crystallography: Implications for substrate binding. J. Mol. Biol. 2001;308: 311–323.

31. Giraud-Panis MJE, Duckett DR, Lilley DMJ. The modular character of a DNA junction resolving enzyme: A zinc binding motif in T4 endonuclease VII. J. Mol. Biol. 1995;252:596–610.
32. Li CL, Hor LI, Chang ZF, Tsai LC, Yang WZ, Yuan HS. DNA binding and cleavage by the periplasmic nuclease Vvn: A novel structure with a known active site. EMBO J. 2003;22:4014-4025.
33. Palanivelu P. DNA polymerases – An insight into their active sites and mechanism of action. In: Recent Advances in Biological Research, SCIENCEDOMAIN International Book Publishers, UK. 2019;1: 1-39. ISBN: 9788193422441. DOI: 10.9734/bpi/rabr/v1
34. Selent U, Rueter T, Koehler E, Liedtke M, Thielking V, Alves J, Oelgeschlaeger T, Wolfes H, Peters F, Pingoud A. A site-directed mutagenesis study to identify amino acid residues involved in the catalytic function of the restriction endonuclease *EcoR*V. Biochemistry. 1992;31:4808-4815.
35. Fonfara I, Le Rhun A, Chylinski K, Makarova KS, Lecrivain AL, Bzdrenga J, Koonin EV, Charpentier E. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. Nucleic Acids Res. 2014;42: 2577–2590.
36. Anders C, Niewoehner O, Duerst A, Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature. 2014;513:569–573.
37. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier EA. Programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337:816–821.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/57357