

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<sup>1\*</sup>

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

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-

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<sup>2+</sup> or Zn<sup>2+</sup> 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<sup>2+</sup> 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 wellcharacterized prokaryotic adaptive immunity mechanism that provides **RNA-mediated** protection from viruses and other mobile genetic CRISPR–Cas9 system elements. (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

cleaving of the host genome. The homing occurs when the endonuclease makes a doublestranded break (DSB) in the genome as the first step, which is followed by the host repair

by

step, which is followed by the host repair mechanism which uses the intron-containing allele as the template and inserts the introncontaining 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

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-

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  $\alpha$ - $\alpha$ - $\beta$ -metal finger

endonucleases. The main function of these

enzymes is to promote the lateral transfer of their

own coding and flanking DNA regions between

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

recombination-dependent

а

resistant plants [6,7].

genomes.

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 sixstranded  $\beta$  sheets surrounded by four  $\alpha$  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 20nucleotide (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. will Cas9 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 predesigned 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 genomeediting 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].



NB: \*Cas genes number, order and direction may be different in different organisms; Pr, promoter region; TracrRNA, Trans-activating CRISPR-RNA

1							1368
1- 59	60-93	94-713	718-769	775-908	909-1098	1099-1368	
RuvC-I	R-rich	α-helical region	RuvC-II	HNH	RuvC-III	PAM-ID	CTD



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 eukarvotes 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 -HHbased 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 metalbinding 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<sup>27</sup> and NTP binding -N<sup>39</sup> 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-TevIII 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 belonas to type IV site-specific kDa. 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 (RgIA = "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

tr G3MUM5 G3MUM5_9CAUD	MNYRKIWIDANGPIPKDSDGRTYEIHHKDGNRENNDLDNLMCLSIQEHYDIHLAQKDYQA	60 x13N
tr A0A159B7B6 A0A159B7B6_9CAUD	MNYRKIWIDANGPIPFDSDGRTYEIHHKDGNRENNDLDULACLSIQEHYDIHLAQKDYQA	60 x10H
tr A0A5C1KA79 A0A5C1KA79_9CAUD	MWIDANGPIPPDSDGRTYEIHHKDGNRENNDLDULACLSIQEHYDIHLAQKDYQA	55
tr A0A482GC49 A0A482GC49 9CAUD	MNYRLIWIDANGPIPFDSDGRTYEMHHKDGNRENNDLDULMCLSIQEHYDIHLAQKDYQA	60
tr AUAI93HUR/ AUAI93HUR/ 9CAUD	MNYRKIWIDANGPIPHDSDGRTYEIHHKDGNRENHDLDHLMCLSIQEHYDIHLAQKDYQA	60
tr AUAI93G204 AUAI93G204 SCAUD	MNYRKIWIDANGPIPPDSDERTYEIHHKDGNRENNDEDILLACLSIGEHIDIHLAQKDYGA	60
tr   A0A462GDA5   A0A462GDA5_9CA0D	MNIRKIWIDANGPIPPDSDERTILIINKOGNRENNDINIKCUSIQETIDIHLAQKDIKA	60
tr   A0A482GMDI   A0A482GMDI _ 9CA0D	MNIRKIWIDANGPIPPOSDERTIEIRRKOGNRENNDERLEGISIQERIDIRLAQKDIQA	60
++1000097569100097569 BDD10	MINING THE DANGET PROSDER. HE FINING ON RENADING THE CLOSED TO THE ACKNER AND THE DANGET PROSDER. HE FINING AND THE ACKNER AND	60
+r1202097.74E21202097.74E2 BPR09	MNYRTIWIDANGPIPPDSDERTJEITHIKDGNRENNDLDULWCLSIGETIDIIIDAOKDYGA	60
trla0a097J1a3la0a097J1a3_BPR03	MNYRT WTDANGPT PROSDER THET HHKDGNRENN DLOUL WCLSTOE HYDTHLAOKDYOA	60
tr1A0A097J2T51A0A097J2T5 BPR06	MNYRKIWIDANGPIPEDSDERTDEIHHKDGNRENNDLDULACLSIOE HYDIHLAOKDYOA	60
tr   A0A097J3N2   A0A097J3N2 9CAUD	MNYRKIWIDANGPIPRDSDERTDEIHHKDGNRENNDLDULMCLSIGEHYDIHLAOKDYGA	60
sp Q38419 TEV3 BPR03	MNYRKIWIDANGPIPPDSDGRTDEIHHKDGNRENNDLDULKCLSIGERYDIHLAGKDYGA	60
tr A0A449C669 A0A449C669 9CAUD	MNYRKIWIDADGPIPPDSDGRTYEIHHKDGNRENNDLDTLMCLSIQEHYDIHLAQKDYQA	60
tr   A0A482GBL4   A0A482GBL4 9CAUD	MNYRKIWIDANGPIPF <mark>DSD</mark> GRTYEIHHKDGNRENNDLDDLUCLSIQEHYDIHLAQKDYQA	60
tr A0A5B9N7B0 A0A5B9N7B0_9CAUD	MNYRKIWIDANGPIPR <mark>DSDGRTYEIHH</mark> KDGNRENN <mark>DLDIL</mark> MCLSIQE <mark>HY</mark> DIHLAQKDYQA	60
tr   A0A482N0E0   A0A482N0E0 9CAUD	MNYRKIWIDANGPIPR <mark>DSDGRTYEIHH</mark> KDGNRENN <mark>DLDTL</mark> ACLSIQE <mark>HY</mark> DIHLAQKDYQA	60
tr 17LHL9 17LHL9_9CAUD	MNYRKIWIDANGPIPR <mark>DSDGRTYEIHH</mark> KDGNRENN <mark>DLDUL</mark> ACLSIQE <mark>HY</mark> DIHLAQKDYQA	60
tr A0A023ZV52 A0A023ZV52_9CAUD	MNYRKIWIDANGPIPP <mark>DSDGRTYEIHH</mark> KDGNRENN <mark>DLDUL</mark> ACLSIQE <mark>HY</mark> DIHLAQKDYQA	60
tr A0A482MSG0 A0A482MSG0_9CAUD	MNYRKIWIDANGPIPPDSDGRTYEIHHKDGNRENNDLDULMCLSIQEHYDIHLAQKDYQA	60
tr A0A482MU60 A0A482MU60_9CAUD	MNYRKIWIDANGPIPP <mark>DSDGRTYEIHHK</mark> DGNRENNDLDNLMCLSIQEHYDIHLAQKDYQA	60 x9H
	**************************************	
+ - COMBRE COMBRE BOATE	OUN THE DARY ODD TO DE A OR A RODD TO TONE DRIVE A OTROPHENT A OTROPHENT TO TO T	120
triacalegadoms_SCAUD	CHAINLEMENSPEEISELASKAAKSREIQIPNIPEVRANNIASIKSKIENGTPHLLDGEI	120
+r 1005501K5791005501K579 9CAUD	CHAIRDRWRISPEEISELASRAARSREIQVENIFEVRARVIASIRSREHGTENDBOGE CHAIRDRWRYSDFFISFLASRAARSREIQVENIFEVRARVIASIRSREHGTENDBOGE	115
+r1202482GC491202482GC49_9CNUD	CHAIREDWIKISPEEISELASKAARSKEIQIPNIPEVRAKUIASIKSKIENSTIHEEDSEL CHAIREDWIKVSDEFISELASKAARSKEIQIPNIPEVRAKUIASIKSKIENSTIHEEDSEL	120
+r1202193H0071202193H007 9CAUD	CHAIRLENERS SEEDELASKAARSKEIQITNIFEVRARUTASIKSKIENSTILLESEI CHAIRLENERSSEESELASKAARSKEIQITNIFEVRARUTASIKSKIENSTILLESEI	120
+r1202193G2U41202193G2U4_9C2UD	CHAIREDWEYSDEFFISELASKAARSKEIQIFNIFEVRARVIRSIKSKIENSTHEEDSEI	120
tr 1000482GDV51000482GDV5_9CNUD	CHAIKLOWKYSDEEISELASKAAKSDEIGITNIFEVRANVASIKSKIENGTTHEEDSEI	120
trla0a482GMD11a0a482GMD1_9CAUD	CHATKLOWKVSDFFTSFLASKAAKSKEIQITNITEVRARVIRSIKSKIENGTHILLDGFT	120
tr   A0A097J243   A0A097J243 9CAUD	CHAIKLBMKYSPEEISELASKAAKSBEIOIFNIPEVBAKNIASIKSKIENGTFHLLDGEI	120
tr A0A097J5691A0A097J569 BPR10	CHAIKLEMKYSPEEISELASKAAKSBEIOIENIPEVRAKNIASIKSKIENGTEHLLDGEI	120
tr A0A097J4E2 A0A097J4E2 BPR09	CHAIKLBMKYSPEEISELASKAAKSREIOIFNIPEVRAKNIASIKSKIENGTFHLLDGEI	120
tr   A0A097J1A3   A0A097J1A3 BPR03	CHAIKLEMKYSPEEISELASKAAKSREIOIENIPEVEAKNIASIKSKIENGTEHLLDGEI	120
tr1A0A097J2T51A0A097J2T5 BPR06	CHAIKLEMKYSPEEISELASKAAKSREIOIFNIPEVRAKNIASIKSKIENGTFHLLDGEI	120
tr   A0A097J3N2   A0A097J3N2 9CAUD	CHAIKLEMKYSPEEISELASKAAKSREIOIFNIPEVRAKNIASIKSKIENGTFHLLDGEI	120
sp Q38419 TEV3 BPR03	CHAIKLRMKYSPEEISELASKAAKSREIQIFNIPEVRAKNIASIKSKIENGTFHLLDGEI	120
tr   A0A449C669   A0A449C669 9CAUD	CHAIKLRMKYSPEEISELASKAAKSREIQIFNIPEVRAKNIASIKSKIENGTFHLLDGEI	120
tr A0A482GBL4 A0A482GBL4 9CAUD	CHAIKLRMKYSPEEISELASKAAKSREIQIFNIPEVRAKNIASIKSKIENGTFHLLDGEI	120
tr A0A5B9N7B0 A0A5B9N7B0 9CAUD	CHAIKLRMKYSPEEISELASKAAKSREIQIFNIPEVRAKNIASIKSKIENGTFHLLDGEI	120
tr A0A482N0E0 A0A482N0E0 9CAUD	CHAIKLRMKYSPEEISELASKAAKSREIQIFNIPEVRAKNIASIKSKIENGTFHLLDGEI	120
tr 17LHL9 17LHL9 9CAUD	CHAIKLRMKYSPEEISELASKAAKSREIQIFNIPEVRAKNIASIKSKIENGTFHLLDGEI	120
tr A0A023ZV52 A0A023ZV52_9CAUD	CHAIKLRMKYSPEEISELASKAAKSREIQIFNIPEVRAKNIASIKSKIENGTFHLLDGEI	120
tr   A0A482MSG0   A0A482MSG0 9CAUD	CHAIKLRMKYSPEEISELASKAAKSREIQIFNIPEVRAKNIASIKSKIENGTFHLLDGEI	120
tr A0A482MU60 A0A482MU60_9CAUD	CHAIKLRMKYSPEEISELASKAAKSREIQIFNIPEVRAKNIASIKSKIENGTFHLLDGEI	120
0 040	***************************************	
tr G3MUM5 G3MUM5 9CAUD	ORKSNINRVALGIHNESEGAH-AKVKERNIAAIKEGHVEGGKMOSETOSKDUNDGSHH	179
trla0a159B7B61a0a159B7B6 9CaUD	ORKSNINRVALGTHNFODAEH TKVKERNIAAIKEGTHAFCGGKMOSETOSKRVNIGSH	180
+ x 1 A 0 A 5 C 1 K A 7 9 1 A 0 A 5 C 1 K A 7 9 9 C A UD	OPESNINDVALGTHNEODAFH TAKVERDNIAATEEGEWVECGGEMOSEDOSEDVND GSHH	175
tr1202482GC491202482GC49 9C200	ORKSNINRVALGI HNFODAEH JAKVKERNI JAATKEGEV FOGGKMOSETOSKOVNI GHH	180
+=1303192808713031928087 9CAUD	OPESNINDAT GTHNEODAELTAEVKEPNIAATKEGEVIEGGEWOOREDOSKEVID GHH	190
tr   2021930007   ACRIJSHOK7_JCROD	ORKONDNKARDOTINE OZADILAKVEKNI ARIKEGILECCENOGEBOGEDUNI COH	100
LI ADAI 93G204 ADAI 93G204_SCAUD	OR SNINKAALGINNE OVALHIAKVKEKNIAAIKEGIN FCGGMQSETQSKKVNUGSH	100
LE AUA402GDAS AUA402GDAS SCAUD	OR ONLY NEW ALGENNE ON THE AKY KERNIAAI KEGT WE COCKNON TO SKEVNI OSH	100
tr A0A482GMDI A0A482GMDI 9CAUD	QRKSNLNRVALGIHNFQQAEHIAKVKERNIAAIKEGTHVFCGGKMQSETQSKRVNDSSHH	180
tr[A0A09/J243[A0A09/J243_9CA0D	QRKSNLNRVALGIHNFQQAEHIAKVKERNIAAIKEGTHVFCGGKMQSETQSKRVNDGSHH	180
tr AUAU9/J569 AUAU9/J569 BPRIU	QRKSNLNRVALGIHNFQQAEHIAKVKERNIAAIKEGTHVFCGGKMQSETQSKRVNDGSHH	180
tr A0A097J4E2 A0A097J4E2_BPR09	QRKSNLNRVALGIHNFQQAEHIAKVKERNIAAIKEGTHVFCGGKMQSETQSKRVNDGSHH	180
tr A0A097J1A3 A0A097J1A3_BPR03	QRKSNLNRVALGIHNFQDAEHIAKVKERNIAAIKEGTHVFCGGKMQSETQSKRVNDGSHH	180
tr A0A097J2T5 A0A097J2T5_BPR06	QRKSNLNRVALGIHNFQQAEHIAKVKERNIAAIKEGTHVFCGGKMQSETQSKRVNDGSHH	180
tr A0A097J3N2 A0A097J3N2_9CAUD	QRKSNLNRVALGIHNFQDAEHIAKVKERNIAAIKEGTHVFCGGKMQSETQSKRVNDGSHH	180
sp Q38419 TEV3_BPR03	QRKSNLNRVALGIHNFQDAEHIAKVKERNIAAIKEGTHVFCGGKMQSETQSKRVNDGSHH	180
tr A0A449C669 A0A449C669_9CAUD	QRKSNLNRVALGIHNFQDAEHIAKVKERNIAAIKEGTEVFCGGKMQSETQSKRVNDGSHH	180
tr A0A482GBL4 A0A482GBL4_9CAUD	QRKSNLNRVALGIHNFODAEHIAKVKERNIAAIKEGTHVFCGGKMQSETQSKRVNDGSHH	180
tr   A0A5B9N7B0   A0A5B9N7B0 9CAUD	QRKSNLNRVALGIHNFQDAEHIAKVKERNIAAIKEGTEVFCGGKMQSETQSKRVNDGSHH	180
tr   A0A482N0E0   A0A482N0E0 9CAUD	QRKSNLNRVALGIHNFQDAEHIAKVKERNIAAIKEGTHVFCGGKMQSETQSKRVNDGSHH	180
tr 17LHL9 17LHL9 9CAUD	QRKSNLNRVALGIHNFQDAEHIAKVKERNIAAIKEGTHVFCGGKMQSETQSKRVNDGSHH	180
tr   A0A023ZV52   A0A023ZV52 9CAUD	QRKSNLNRVALGIHNFQDAEH AKVKERNIAAIKEGTEVFCGGKMQSETOSKRVNDGSHH	180
tr   A0A482MSG0   A0A482MSG0 9CAUD	QRKSNLNRVALGIHNFQDAEH AKVKERNIAAIKEGTHVFCGGKMOSETOSKRVNDGSHH	180
tr   A0A482MU60   A0A482MU60 9CAUD	QRKSNLNRVALGIHNFQDAEH AKVKERNIAAIKEGTHVFCGGKMOSETOSKRVNDGSHH	180
	******** ******** * * *****************	
	· · · · · · · · · · · · · · · · · · ·	

tr G3MUM5 G3MUM5_9CAUD		
	FLSEDHKKRTSAKTLEMVKNGTHPAQKEITCDFCGHIGKGPGFYLKHNDRCKLNPNRIQL	239
tr AUA159B/B6 AUA159B/B6 9CAUD	FLSEDHKKRTSAKTLEMVKNOTHPAQKEIT <mark>CDFCC</mark> HTGKGPGFYLKHNDRCKLNPNRIQI	240
tr A0A5C1KA79 A0A5C1KA79 9CAUD	FLSEDHKKRTSAKTLEMVK <mark>IN</mark> THPAQKEIT <mark>CDFC6</mark> HIGKGPGFYLKHNDRCKLNPNRIQI	235
tr A0A482GC49 A0A482GC49_9CAUD	FLSEDHKKRTSAKTLEMVK <mark>UC</mark> THPAQKEIT <mark>CDFCC</mark> HIGKGPGFYLKHNDRCKLNPNRIQI	240
tr A0A193H0R7 A0A193H0R7_9CAUD	FLSEDHKKRTSAKTLEMVK <mark>UGTHPAQKEIT<mark>CD</mark>FCC</mark> HIGKGPGFYLKHNDRCKLNPNRIQI	240
tr A0A193GZU4 A0A193GZU4_9CAUD	FLSEDHKKRTSAKTLEMVKNOTHPAQKEIT <mark>CDFCC</mark> HIGKGPGFYLKHNDRCKLNPNRIQI	240
tr A0A482GDX5 A0A482GDX5 9CAUD	FLSEDHKKRTSAKTLEMVK <mark>UC</mark> THPAQKEIT <mark>CDFCC</mark> HIGKGPGFYLKHNDRCKLNPNRIQI	240
tr  A0A482GMD1  A0A482GMD1_9CAUD	FLSEDHKKRTSAKTLEMVK <mark>UGTHPAQKEIT<mark>CP</mark>FCC</mark> HIGKGPGFYLKHNDRCKLNPNRIQI	240
tr A0A097J243 A0A097J243_9CAUD	FLSEDHKKRTSAKTLEMVK <mark>UC</mark> THPAQKEIT <mark>CPFCC</mark> HIGKGPGFYLKHNDRCKLNPNRIQI	240
tr   A0A097J569   A0A097J569_BPR10	FLSEDHKKRTSAKTLEMVK <mark>UGTHPAQKEIT<mark>CD</mark>FCG</mark> HIGKGPGFYLKHNDRCKLNPNRIQI	240
tr A0A097J4E2 A0A097J4E2_BPR09	FLSEDHKKRTSAKTLEMVK <mark>UGTHPAQKEIT<mark>CD</mark>FCG</mark> HIGKGPGFYLKHNDRCKLNPNRIQI	240
tr A0A097J1A3 A0A097J1A3_BPR03	FLSEDHKKRTSAKTLEMVK <mark>NGTHPAQKEIT<mark>CD</mark>FCC</mark> HIGKGPGFYLKHNDRCKLNPNRIQI	240
tr A0A097J2T5 A0A097J2T5_BPR06	FLSEDHKKRTSAKTLEMVKNGTHPAQKEIT <mark>CPFCG</mark> HIGKGPGFYLKHNDRCKLNPNRIQI	240
tr A0A097J3N2 A0A097J3N2_9CAUD	FLSEDHKKRTSAKTLEMVK <mark>UGTHPAQKEIT<mark>CD</mark>FCC</mark> HIGKGPGFYLKHNDRCKLNPNRIQI	240
sp Q38419 TEV3_BPR03	FLSEDHKKRTSAKTLEMVK <mark>UGTHPAQKEIT<mark>CD</mark>FCC</mark> HIGKGPGFYLKHNDRCKLNPNRIQI	240
tr A0A449C669 A0A449C669_9CAUD	FLSEDHKKRTSAKTLEMVK <mark>NGTHPAQKEIT<mark>CD</mark>FCC</mark> HIGKGPGFYLKHNDRCKLNPNRIQI	240
tr   A0A482GBL4   A0A482GBL4 _ 9CAUD	FLSEDHKKRTSAKTLEMVK <mark>UGTHPAQKEIT<mark>CPFCG</mark>HIGKGPGFYLKHNDRCKLNPNRIQI</mark>	240
tr  A0A5B9N7B0  A0A5B9N7B0_9CAUD	FLSEDHKKRTSAKTLEMVK <mark>NGTHPAQKEIT<mark>CD</mark>FCC</mark> HIGKGPGFYLKHNDRCKLNPNRIQI	240
tr A0A482N0E0 A0A482N0E0 9CAUD	FLSEDHKKRTSAKTLEMVKNOTHPAQKEIT <mark>CPFCC</mark> HIGKGPGFYLKHNDRCKLNPNRIQI	240
tr I7LHL9 I7LHL9_9CAUD	FLSEDHKKRTSAKTLEMVK <mark>U</mark> GTHPAQKEIT <mark>CPFC6</mark> HIGKGPGFYLKHNDRCKLNPNRIQI	240
tr   A0A023ZV52   A0A023ZV52_9CAUD	FLSEDHKKRTSAKTLEMVK <mark>UGTHPAQKEIT<mark>CDFCG</mark>HIGKGPGFYLKHNDRCKLNPNRIQI</mark>	240
tr A0A482MSG0 A0A482MSG0_9CAUD	FLSEDHKKRTSAKTLEMVK <mark>UCTHPAQKEIT<mark>CD</mark>FCC</mark> HIGKGPGFYLKHNDRCKLNPNRIQI	240
tr A0A482MU60 A0A482MU60 9CAUD	FLSEDHKKRTSAKTLEMVK <mark>NGTHPAQKEIT<mark>CP</mark>FCC</mark> HIGKGPGFYLKHNDRCKLNPNRIQI	240
	**************************************	
+ + COMINS   COMINS GODIE		
+= 10015987861001598786 900UD	NOR CONTRACTOR OF CONTRACTOR OF CONTRACTOR	
CI AUAIJJB/B0 AUAIJJB/B0_JCAUD	NGPICEREDESPSTIKKWIGENCKIKINE 205	
5-1303E01W3701303E01W370 003UD		
tr   A0A5C1KA79   A0A5C1KA79_9CAUD	NGPYCDKKDLSPSTYKRWHGDNCKTRFNE 264	
tr A0A5C1KA79 A0A5C1KA79_9CAUD tr A0A482GC49 A0A482GC49_9CAUD	NCPYCOKKDLSPSTYKRWHGDNCKTRFND 264 NGPYCOKKDLSPSTYRRWHGDNCKTRFND 269	
tr A0A5C1KA79 A0A5C1KA79_9CAUD tr A0A482GC49 A0A482GC49_9CAUD tr A0A193H0R7 A0A193H0R7_9CAUD	NCPYCOKKDLSPSTYKRWHGDNCKTRFND 264 NCPYCOKKDLSPSTYKRWHGDNCKTRFND 269 NCPYCOKKDLSPSTYKRWHGDNCKTRFND 269	
tr A0A5C1KA79 A0A5C1KA79_9CAUD tr A0A482GC49 A0A482GC49_9CAUD tr A0A193H0R7 A0A193H0R7_9CAUD tr A0A193GZU4 A0A193GZU4_9CAUD	NGPYGDKKDLSPSTYKRWHGDNCKTRFND 264 NGPYGDKKDLSPSTYKRWHGDNCKTRFND 269 NGPYGDKKDLSPSTYKRWHGDNCKTRFND 269 NGPYGDKKDLSPSTYKRWHGDNCKTRFND 269	
tr  A0A5C1KA79 A0A5C1KA79_9CAUD tr  A0A482GC49 A0A482GC49_9CAUD tr  A0A193H0R7 A0A193H0R7_9CAUD tr  A0A193GZU4 A0A193GZU4_9CAUD tr  A0A482GDX5 A0A482GDX5_9CAUD	NGPYCOKKDLSPSTYKRWHGDNCKTRFNE 264 NGPYCOKKDLSPSTYRRWHGDNCKTRFNE 269 NGPYCOKKDLSPSTYKRWHGDNCKTRFNE 269 NGPYCOKKDLSPSTYKRWHGDNCKTRFNE 269 NGPYCOKKDLSPSTYKRWHGDNCKTRFNE 269	
tr  A0A5C1KA79 A0A5C1KA79_9CAUD tr  A0A482GC49 A0A482GC49_9CAUD tr  A0A193H0R7 A0A193H0R7_9CAUD tr  A0A193GZU4 A0A193GZU4_9CAUD tr  A0A482GDX5 A0A482GDX5_9CAUD tr  A0A482GMD1 A0A482GMD1_9CAUD	NCPYCDKKDLSPSTYKRWHGDNCKTRFNE 264 NCPYCDKKDLSPSTYRRWHGDNCKTRFNE 269 NCPYCDKKDLSPSTYRRWHGDNCKTRFNE 269 NCPYCDKKDLSPSTYKRWHGDNCKTRFNE 269 NCPYCDKKDLSPSTYKRWHGDNCKTRFNE 269 NCPYCDKKDLSPSTYKRWHGDNCKARFNE 269	
tr  A0A5C1KA79  A0A5C1KA79_9CAUD tr  A0A482GC49  A0A482GC49_9CAUD tr  A0A193H0R7  A0A193H0R7_9CAUD tr  A0A193GZU4  A0A193GZU4_9CAUD tr  A0A482GDX5  A0A482GDX5_9CAUD tr  A0A482GMD1  A0A482GMD1_9CAUD tr  A0A097J243  A0A097J243_9CAUD	NCPYCDKKDLSPSTYKRWHGDNCKTRFNE 264 NCPYCDKKDLSPSTYRRWHGDNCKTRFNE 269 NCPYCDKKDLSPSTYKRWHGDNCKTRFNE 269 NCPYCDKKDLSPSTYKRWHGDNCKTRFNE 269 NCPYCDKKDLSPSTYKRWHGDNCKARFNE 269 NCPYCDKKDLSPSTYKRWHGDNCKARFNE 269	
tr  A0A5C1KA79  A0A5C1KA79_9CAUD tr  A0A482GC49  A0A482GC49_9CAUD tr  A0A193H0R7  A0A193H0R7_9CAUD tr  A0A193GZU4  A0A193GZU4_9CAUD tr  A0A482GDX5  A0A482GDX5_9CAUD tr  A0A482GMD1  A0A482GMD1_9CAUD tr  A0A097J243  A0A097J243_9CAUD tr  A0A097J569  A0A097J569_BFR10	NGPYCDKRDLSPSTYKRWHGDNCKTRFNE 264 NGPYCDKRDLSPSTYRRWHGDNCKTRFNE 269 NGPYCDKRDLSPSTYRRWHGDNCKTRFNE 269 NGPYCDKRDLSPSTYRRWHGDNCKTRFNE 269 NGPYCDKRDLSPSTYRRWHGDNCKARFNE 269 NGPYCDKRDLSPSTYRRWHGDNCKARFNE 269 NGPYCDKRDLSPSTYRRWHGDNCKARFNE 269 NGPYCDKRDLSPSTYRRWHGDNCKARFNE 269	
tr  A0A5C1KA79  A0A5C1KA79_9CAUD tr  A0A482GC49  A0A482GC49_9CAUD tr  A0A193H0R7  A0A193H0R7_9CAUD tr  A0A193GZU4  A0A193GZU4_9CAUD tr  A0A482GDX5  A0A482GDX5_9CAUD tr  A0A482GMD1  A0A482GMD1_9CAUD tr  A0A097J243  A0A097J243_9CAUD tr  A0A097J569  A0A097J569_BPR10 tr  A0A097J42  A0A097J422 BPR09	NGPYCDKKDLSPSTYKRWHGDNCKTRFNE 264 NGPYCDKKDLSPSTYRRWHGDNCKTRFNE 269 NGPYCDKKDLSPSTYRRWHGDNCKTRFNE 269 NGPYCDKKDLSPSTYRRWHGDNCKTRFNE 269 NGPYCDKKDLSPSTYRRWHGDNCKARFNE 269 NGPYCDKKDLSPSTYRRWHGDNCKARFNE 269 NGPYCDKKDLSPSTYRRWHGDNCKARFNE 269 NGPYCDKKDLSPSTYKRWHGDNCKARFNE 269	
tr   A0A5C1KA79   A0A5C1KA79 _9CAUD tr   A0A482GC49   A0A482GC49 _9CAUD tr   A0A493H0R7   A0A193H0R7 _9CAUD tr   A0A193H0R7   A0A193G2U4 _9CAUD tr   A0A482GDX5   A0A482GDX5 _9CAUD tr   A0A482GMD1   A0A482GMD1 _9CAUD tr   A0A482GMD1   A0A482GMD1 _9CAUD tr   A0A097J243   A0A097J243 _9CAUD tr   A0A097J269   A0A097J269 _BPR10 tr   A0A097J4E2   A0A097J4E2 _BPR09 tr   A0A097J1A3   A0A097J1A3 _BPR03	NCPYCDKKDLSPSTYKRWHGDNCKTRFNE 264 NCPYCDKKDLSPSTYRRWHGDNCKTRFNE 269 NCPYCDKKDLSPSTYRRWHGDNCKTRFNE 269 NCPYCDKKDLSPSTYRRWHGDNCKTRFNE 269 NCPYCDKKDLSPSTYRRWHGDNCKARFNE 269 NCPYCDKKDLSPSTYRRWHGDNCKARFNE 269 NCPYCDKKDLSPSTYRRWHGDNCKARFNE 269 NCPYCDKKDLSPSTYRRWHGDNCKARFNE 269 NCPYCDKKDLSPSTYRRWHGDNCKARFNE 269 NCPYCDKKDLSPSTYRRWHGDNCKARFNE 269 NCPYCDKKDLSPSTYRRWHGDNCKARFNE 269	
tr  A0A5C1KA79  A0A5C1KA79_9CAUD tr  A0A482GC49  A0A482GC49_9CAUD tr  A0A193H0R7  A0A193H0R7_9CAUD tr  A0A193GZU4  A0A193GZU4_9CAUD tr  A0A482GM5  A0A482GM5_9CAUD tr  A0A482GMD1  A0A482GMD1_9CAUD tr  A0A482GMD1  A0A482GMD1_9CAUD tr  A0A097J243  A0A097J243_9CAUD tr  A0A097J569  A0A097J569_BFR10 tr  A0A097J4E2  A0A097J4E2_BFR03 tr  A0A097J25  A0A097J2T5_BFR06	NCPYCDKKDLSPSTYKRWHGDNCKTRFNE 264 NCPYCDKKDLSPSTYRRWHGDNCKTRFNE 269 NCPYCDKKDLSPSTYRRWHGDNCKTRFNE 269 NCPYCDKKDLSPSTYKRWHGDNCKTRFNE 269 NCPYCDKKDLSPSTYKRWHGDNCKARFNE 269 NCPYCDKKDLSPSTYRRWHGDNCKARFNE 269 NCPYCDKKDLSPSTYRRWHGDNCKARFNE 269 NCPYCDKKDLSPSTYRRWHGDNCKARFNE 269 NCPYCDKKDLSPSTYKRWHGDNCKARFNE 269 NCPYCDKKDLSPSTYKRWHGDNCKARFNE 269 NCPYCDKKDLSPSTYKRWHGDNCKARFNE 269 NCPYCDKKDLSPSTYKRWHGDNCKARFNE 269	
tr  A0A5C1KA79  A0A5C1KA79 9CAUD tr  A0A482GC49  A0A482GC49 9CAUD tr  A0A493H0R7  A0A193H0R7 9CAUD tr  A0A193GZU4  A0A193GZU4 9CAUD tr  A0A482GDX5  A0A482GDX5 9CAUD tr  A0A482GMD1  A0A482GMD1 9CAUD tr  A0A097J243  A0A097J243 9CAUD tr  A0A097J569  A0A097J569 BPR10 tr  A0A097J1A2  A0A097J1A3 BPR03 tr  A0A097J1A3  A0A097J1A3 BPR03 tr  A0A097J375  A0A097J275 BPR06 tr  A0A097J382  A0A097J3N2 9CAUD	NGPYCDKKDLSPSTYKRWHGDNCKTRFNE 264 NGPYCDKKDLSPSTYKRWHGDNCKTRFNE 269 NGPYCDKKDLSPSTYKRWHGDNCKTRFNE 269 NGPYCDKKDLSPSTYKRWHGDNCKTRFNE 269 NGPYCDKKDLSPSTYKRWHGDNCKARFNE 269	
tr   A0A5C1KA79   A0A5C1KA79 _9CAUD tr   A0A482GC49   A0A482GC49 _9CAUD tr   A0A493H0R7   A0A193H0R7 _9CAUD tr   A0A193GZV4   A0A193GZV4 _9CAUD tr   A0A482GDX5   A0A482GDX5 _9CAUD tr   A0A482GMD1   A0A482GMD1 _9CAUD tr   A0A097J243   A0A097J243 _9CAUD tr   A0A097J243   A0A097J243 _9CAUD tr   A0A097J42   A0A097J4E2 _BPR09 tr   A0A097J1A3   A0A097J4E2 _BPR09 tr   A0A097J2T5   A0A097J2T5 _BPR06 tr   A0A097J2T5   A0A097J2T5 _BPR06 tr   A0A097J3N2   A0A097J3N2 _9CAUD an   038419   TFV3 _BPR03	NGPYCDKKDLSPSTYKRWHGDNCKTRFNE 264 NGPYCDKKDLSPSTYRRWHGDNCKTRFNE 269 NGPYCDKKDLSPSTYRRWHGDNCKTRFNE 269 NGPYCDKKDLSPSTYRRWHGDNCKTRFNE 269 NGPYCDKKDLSPSTYRRWHGDNCKARFNE 269	
tr   A0A5C1KA79   A0A5C1KA79 _9CAUD tr   A0A482GC49   A0A482GC49 _9CAUD tr   A0A493H0R7   A0A193H0R7 _9CAUD tr   A0A193H0R7   A0A193G2U4 _9CAUD tr   A0A482GDX5   A0A482GDX5 _9CAUD tr   A0A482GMD1   A0A482GMD1 _9CAUD tr   A0A482GMD1   A0A482GMD1 _9CAUD tr   A0A097J243   A0A097J243 _9CAUD tr   A0A097J269   A0A097J269 _BPR10 tr   A0A097J269   A0A097J462 _BPR09 tr   A0A097J215   A0A097J1A3 _BPR03 tr   A0A097J275   A0A097J3N2 _9CAUD sp   Q38419   TEV3_BPR03	NCPYCDKKDLSPSTYKRWHGDNCKTRFNE 264 NCPYCDKKDLSPSTYKRWHGDNCKTRFNE 269 NCPYCDKKDLSPSTYKRWHGDNCKTRFNE 269 NCPYCDKKDLSPSTYKRWHGDNCKTRFNE 269 NCPYCDKKDLSPSTYKRWHGDNCKARFNE 269 NCPYCDKKDLSPSTYRWHGDNCKARFNE 269 NCPYCDKKDLSPSTYRWHGDNCKAFFNE 269 NCPYCDKKDLSPSTYRWHGDNCKAFFNE 269 NCPYCDKKDLSPSTYRWHGDNCKAFFNE 269 NCPYCDKKDLSPSTYRWHGDNCKAFFNE 269 NCPYCDKKDLSPSTYRWHGDNCKAFFNE 269 NCPYCDKKDLSPSTYRWHGDNCKAFFNE 269 NCPYCDKKDLSPSTYRWHGDNCKAFFNE 269 NCPYCDKKDLSPSTYRWHGDNCKAFFNE 269 NCPYCDKRDLSPSTYRWHGDNCKAFFNE 269	
tr  A0A5C1KA79  A0A5C1KA79 9CAUD tr  A0A482GC49  A0A482GC49 9CAUD tr  A0A493GC49  A0A482GC49 9CAUD tr  A0A193GZU4  A0A193H0R7 9CAUD tr  A0A482GDX5  A0A482GDX5 9CAUD tr  A0A482GMD1  A0A482GDX5 9CAUD tr  A0A097J243  A0A097J243 9CAUD tr  A0A097J569  A0A097J569 BER10 tr  A0A097J1A3  A0A097J42 BER09 tr  A0A097J1A3  A0A097J1A3 BER03 tr  A0A097J3X  A0A097J3N2 9CAUD sp  Q38419  TEV3 BER03 tr  A0A449C669  A0A449C669 9CAUD	NGPYCDKRDLSPSTYKRWHGDNCKTRFNE 264 NGPYCDKRDLSPSTYRRWHGDNCKTRFNE 269 NGPYCDKRDLSPSTYRRWHGDNCKTRFNE 269 NGPYCDKRDLSPSTYRRWHGDNCKTRFNE 269 NGPYCDKRDLSPSTYRRWHGDNCKARFNE 269	
tr   A0A5C1KA79   A0A5C1KA79 _9CAUD tr   A0A482GC49   A0A482GC49 _9CAUD tr   A0A493H0R7   A0A193H0R7 _9CAUD tr   A0A193GZU4   A0A193GZU4 _9CAUD tr   A0A482GDX5   A0A482GDX5 _9CAUD tr   A0A482GMD1   A0A482GMD1 _9CAUD tr   A0A097J243   A0A097J243 _9CAUD tr   A0A097J243   A0A097J243 _9CAUD tr   A0A097J242   A0A097J243 _9CAUD tr   A0A097J422   A0A097J242 _ BPR09 tr   A0A097J1A3   A0A097J1A3 _ BPR03 tr   A0A097J2T5   A0A097J3T5 _ BPR06 tr   A0A097J3N2   A0A097J3N2 _9CAUD sp   G38419   TEV3 _BPR03 tr   A0A449C669   A0A482GBL4 _9CAUD tr   A0A482GBL4   A0A482GBL4 _9CAUD	NGPYCDKKDLSPSTYKRWHGDNCKTRFNE 264 NGPYCDKKDLSPSTYKRWHGDNCKTRFNE 269 NGPYCDKKDLSPSTYKRWHGDNCKTRFNE 269 NGPYCDKKDLSPSTYKRWHGDNCKTRFNE 269 NGPYCDKKDLSPSTYRRWHGDNCKARFNE 269 NGPYCDKKDLSPSTYKRWHGDNCKARFNE 269	
tr   A0A5C1KA79   A0A5C1KA79 _9CAUD tr   A0A482GC49   A0A482GC49 _9CAUD tr   A0A493H0R7   A0A193H0R7 _9CAUD tr   A0A193GZV4   A0A193GZV4 _9CAUD tr   A0A482GDX5   A0A482GDX5 _9CAUD tr   A0A482GMD1   A0A482GMD1 _9CAUD tr   A0A97J243   A0A097J243 _9CAUD tr   A0A097J243   A0A097J243 _9CAUD tr   A0A097J243   A0A097J4E2 BFR09 tr   A0A097J1A3   A0A097J4E2 BFR09 tr   A0A097J1A3   A0A097J1A3 _BFR03 tr   A0A097J2T5   A0A097J2T5 _BFR06 tr   A0A097J2T5   A0A097J2T5 _BFR06 tr   A0A097J3N2   A0A097J3N2 _9CAUD sp   Q38419   TEV3 _BFR03 tr   A0A449C669   A0A449C669 _9CAUD tr   A0A482GBL4   A0A482GBL4 _9CAUD tr   A0A5B9N7B0   A0A5BSN7B0 _9CAUD	NGPYCDKKDLSPSTYKRWHGDNCKTRFNE 264 NGPYCDKKDLSPSTYRRWHGDNCKTRFNE 269 NGPYCDKKDLSPSTYRRWHGDNCKTRFNE 269 NGPYCDKKDLSPSTYRRWHGDNCKTRFNE 269 NGPYCDKKDLSPSTYRRWHGDNCKARFNE 269	
tr   A0A5C1KA79   A0A5C1KA79 _9CAUD tr   A0A482GC49   A0A482GC49 _9CAUD tr   A0A193H0R7   A0A193H0R7 _9CAUD tr   A0A193GZV4   A0A193GZV4 _9CAUD tr   A0A482GDX5   A0A482GDX5 _9CAUD tr   A0A482GMD1   A0A482GMD1 _9CAUD tr   A0A97J243   A0A097J243 _9CAUD tr   A0A097J243   A0A097J245 _9CAUD tr   A0A097J249   A0A097J462 _BPR09 tr   A0A097J462   A0A097J462 _BPR09 tr   A0A097J312   A0A097J462 _BPR09 tr   A0A097J312   A0A097J313 _ BPR03 tr   A0A097J312   A0A097J312 _ 9CAUD sp   Q38419   TEV3 _BPR03 tr   A0A482GE4   A0A482GBL4 _9CAUD tr   A0A482GE4   A0A482GBL4 _9CAUD tr   A0A5B9N7B0   A0A5B9N7B0 _9CAUD tr   A0A482NED   A0A482NE0 _9CAUD	NCPYCDKKDLSPSTYKRWHGDNCKTRFNE 269 NCPYCDKKDLSPSTYKRWHGDNCKTRFNE 269 NCPYCDKKDLSPSTYKRWHGDNCKTRFNE 269 NCPYCDKKDLSPSTYKRWHGDNCKTRFNE 269 NCPYCDKKDLSPSTYKRWHGDNCKARFNE 269	
tr   A0A5C1KA79   A0A5C1KA79 9CAUD tr   A0A482GC49   A0A482GC49 9CAUD tr   A0A493H0R7   A0A193H0R7 9CAUD tr   A0A193GZU4   A0A193GZU4 9CAUD tr   A0A482GDX5   A0A482GDX5 9CAUD tr   A0A482GMD1   A0A482GMD1 9CAUD tr   A0A097J243   A0A097J243 9CAUD tr   A0A097J243   A0A097J243 9CAUD tr   A0A097J42   A0A097J42 BPR09 tr   A0A097J1A3   A0A097J1A3 BPR03 tr   A0A097J1A3   A0A097J1A3 BPR03 tr   A0A097J3X2   A0A097J3N2 9CAUD sp   Q38419   TEV3 BPR03 tr   A0A492GE14   A0A482GBL4 9CAUD tr   A0A482GBL4   A0A482GBL4 9CAUD tr   A0A482N0E0   A0A482N0E0 9CAUD tr   A0A482N0E0   A0A482N0E0 9CAUD tr   A0A482N0E0   A0A482N0E0 9CAUD tr   A0A482N0E0   A0A482N0E0 9CAUD	NGPY CDKKDLSPSTYKRWHGDNCKTRFNE 264 NGPY CDKKDLSPSTYKRWHGDNCKTRFNE 269 NGPY CDKKDLSPSTYKRWHGDNCKTRFNE 269 NGPY CDKKDLSPSTYKRWHGDNCKTRFNE 269 NGPY CDKKDLSPSTYRRWHGDNCKARFNE 269 NGPY CDKKDLSPSTYKRWHGDNCKARFNE 269	
tr  A0A5C1KA79  A0A5C1KA79_9CAUD tr  A0A482GC49  A0A482GC49_9CAUD tr  A0A493H0R7  A0A193H0R7_9CAUD tr  A0A193GZU4  A0A193GZU4_9CAUD tr  A0A482GDX5  A0A482GDX5_9CAUD tr  A0A482GMD1  A0A482GMD1_9CAUD tr  A0A097J243  A0A097J243_9CAUD tr  A0A097J243  A0A097J243_9CAUD tr  A0A097J4E2  A0A097J42_BFR09 tr  A0A097J1A3  A0A097J4E2_BFR09 tr  A0A097J1A3  A0A097J1A3_BFR03 tr  A0A097J275  A0A097J2T5_BFR06 tr  A0A097J275  A0A097J2T5_BFR06 tr  A0A097J275  A0A097J2T5_BFR06 tr  A0A497C669  A0A49C669_9CAUD tr  A0A49C669  A0A449C669_9CAUD tr  A0A482GBL4  A0A482GBL4_9CAUD tr  A0A482NDE0  A0A482NDE0_9CAUD tr  A0A482NDE0  A0A482NDE0_9CAUD tr  A0A482NDE0  A0A482NDE0_9CAUD tr  A0A023ZV52  A0A023ZV52_9CAUD	NGPY CDKKDLSPSTYKRWHGDNCKTRFNE 264 NGPY CDKKDLSPSTYKRWHGDNCKTRFNE 269 NGPY CDKKDLSPSTYKRWHGDNCKTRFNE 269 NGPY CDKKDLSPSTYKRWHGDNCKTRFNE 269 NGPY CDKKDLSPSTYRRWHGDNCKARFNE 269 NGPY CDKKDLSPSTYRWHGDNCKARFNE 269 NGPY CDKKDLSPSTYRWHGDNCKARFNE 269	
tr   A0A5C1KA79   A0A5C1KA79 _9CAUD tr   A0A482GC49   A0A482GC49 _9CAUD tr   A0A193H0R7   A0A193H0R7 _9CAUD tr   A0A193GZV4   A0A193GZV4 _9CAUD tr   A0A482GDX5   A0A482GDX5 _9CAUD tr   A0A482GDX5   A0A482GDX5 _9CAUD tr   A0A97J243   A0A097J243 _9CAUD tr   A0A097J243   A0A097J243 _9CAUD tr   A0A097J243   A0A097J245 _BPR10 tr   A0A097J245   A0A097J4E2 _BPR09 tr   A0A097J275   A0A097J4E2 _BPR09 tr   A0A097J275   A0A097J275 _BPR06 tr   A0A097J275   A0A097J275 _BPR06 tr   A0A097J321   A0A097J372 _9CAUD sp   Q38419   TEV3 _BPR03 tr   A0A492GEL4   A0A492GEL4 _9CAUD tr   A0A482GEL4   A0A482GBL4 _9CAUD tr   A0A482NDE0   A0A482NDE0 _9CAUD tr   A0A482NDE0   A0A482NDE0 _9CAUD tr   A0A482NSG0   A0A482MSG0 _9CAUD	NGPYCDKKDLSPSTYKRWHGDNCKTRFNE 269 NGPYCDKKDLSPSTYRRWHGDNCKTRFNE 269 NGPYCDKKDLSPSTYRRWHGDNCKTRFNE 269 NGPYCDKKDLSPSTYRRWHGDNCKTRFNE 269 NGPYCDKKDLSPSTYRRWHGDNCKARFNE 269	
tr  A0A5C1KA79  A0A5C1KA79 9CAUD tr  A0A482GC49  A0A482GC49 9CAUD tr  A0A193H0R7  A0A193H0R7 9CAUD tr  A0A193GZV4  A0A193GZU4 9CAUD tr  A0A482GDX5  A0A482GDX5 9CAUD tr  A0A482GDX1  A0A482GDX5 9CAUD tr  A0A097J243  A0A097J243 9CAUD tr  A0A097J569  A0A097J569 BER10 tr  A0A097J1A3  A0A097J1A3 BER03 tr  A0A097J1A3  A0A097J1A3 BER03 tr  A0A097J3X2  A0A097J3N2 9CAUD tr  A0A097J3N2  A0A097J3N2 9CAUD sp  Q38419 TEV3 BER03 tr  A0A49C669  A0A48CB14 9CAUD tr  A0A482GE14  A0A482GB14 9CAUD tr  A0A482N0E0  A0A48CN0E0 9CAUD tr  A0A482N0E0  A0A482N0E0 9CAUD tr  A0A482N52  A0A0232V52 9CAUD tr  A0A482M260  A0A482M309 9CAUD tr  A0A482M260  A0A482M309 9CAUD tr  A0A482N52  A0A0232V52 9CAUD tr  A0A482M260  A0A482M309 9CAUD tr  A0A482M400  A0A482M309 9CAUD	NGPY OKKDLSPSTYKRWHGDNCKTRFNE 264 NGPY OKKDLSPSTYKRWHGDNCKTRFNE 269 NGPY OKKDLSPSTYKRWHGDNCKTRFNE 269 NGPY OKKDLSPSTYKRWHGDNCKTRFNE 269 NGPY OKKDLSPSTYKRWHGDNCKARFNE 269	

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

G3MUM5\_9CAUD I-TevIII 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 17LHL9 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  $\beta\beta\alpha$ -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 conserved Ns towards completely 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

sp P24200 MCRA_ECOLI	MHVFDNNGIELKAECSIGEEDGVYGLILESWGPGDRNKDYNIALDYIIERLVDSGVSQVV
tr   A0A5H9M2S7   A0A5H9M2S7 SALET	MHVFDNNGIELKAE¢SIGEEDGVYGLILESWGPGDRNKDYNIALDYIIERLVDSGVSQVV
tr   A0A3A3M3R2   A0A3A3M3R2 SALMO	MHVFDNNGIELKAECSIGEEDGVYGLILESWGPGDRNKDYNIALDYIIERLVDSGVSQVV
tr A0A1S9J5H1 A0A1S9J5H1 SHIBO	MHVFDNNGIELKAECIIGEEDGVYGLILESWGPGDRNKDYNIALDYIIERLVDSGVSQVV
tr A0A403M222 A0A403M222 SHIDY	MHVFDNNGIELKAECIIGEEDGVYGLILESWGPGDRNKDYNIALDYIIERLVDSGVSQVV
	***********
ID04000 Huges month	
SP/F24200/MCRA_ECOLI	VILASSVRRMASLDERATAPGETFTLIGNSPRJIKLEARGIQATSKTGRETPSGRR
tr AUASH9M2S / AUASH9M2S / SALET	VYLASSSVRKHMHSLDERKIH PGEYFTLIGNSPHIIRLKMCGYQAYFSRTGRKEIPSGNR
tr   AUA3A3M3R2   AUA3A3M3R2_SALMO	VYLASSSVRKHMHSLDERKIH GEYFTLIGNSPHIIRLKMCGYQAYFSRTGRKEIPSGNR
tr A0A1S9J5H1 A0A1S9J5H1_SHIBO	VYLASSSVRKHMHSLDERKIH GEYFTLIGNSPFVIRLKMCGYQAYFSRTGRKEIPSGNR
tr A0A403M222 A0A403M222_SHIDY	VYLASSSVRKHMHSLDERKIHIGEYFTLIGNSPHVIRLKMCGYQAYFSRTGRKEIPSGNR
	***************************************
sp P24200 MCRA ECOLI	TKRILINVPGIYSDSFWASIIRGELSELSQPTDDESLLNMRVSKLIKKTLSQPEGSRKPV
tr A0A5H9M2S7 A0A5H9M2S7 SALET	TKRILINVPGIYSDSFWASIIRGELSELSOPTDDESLLNMRVSKLIKKTLSOPEGSRKPV
tr A0A3A3M3R2 A0A3A3M3R2 SALMO	TKRILINVPGIYSDSFWASIIRGELSELSOPTDDESLLNMRVSKLIKKTLSOPEGSRKPV
tr A0A1S9J5H1 A0A1S9J5H1 SHIBO	TKRILINVPGIYSDSFWYSIIRGELSELSOPTDDESLLNMRVSKLIKKTLSOPEGSRKPV
+r1202403M2221202403M222 SHTDY	TWO TI THYDGT VSDSFWYST TOGET, SFI SODTODEST, INMOVSKI, TWWT, SODEGSDWDV
	Di ante de la companya
sp P24200 MCRA_ECOLI	EVERLOKVYVRDPMVKAWILQOSKGICENCGKNAPFYLNDGNPYLEVHHVIPLSSGGADT 240
tr A0A5H9M2S7 A0A5H9M2S7 SALET	EVERLOKVYVRDPMVKAWILQOSKGICENOSKNAPFYLNDSPYLEVENVIPLSSGGADT 240
tr A0A3A3M3R2 A0A3A3M3R2_SALMO	EVERLOKVYVRDPMVKAWILQOSKGICENOSKNAPFYLNDSPYLEVENVIPLSSGGADT 240
tr A0A1S9J5H1 A0A1S9J5H1 SHIBO	EVERLQKVYVRDPMVKAWILQQSKGICENCKNAPFYLNDGSPYLEVEHVIPLSSGGADT 240
tr A0A403M222 A0A403M222 sHIDY	EVERLOKVYVRDPMVKAWILOOSKGICENOGKNAPFYLNDGSPYLEVEHVIPLSSGGADT 240
	**************************************
sp P24200 MCRA_ECOLI	TDNCVALCPNCHRELHYSKNAKELIEMLYVNINRLOK 277 X15/X14
tr   A0A5H9M2S7   A0A5H9M2S7_SALET	TINCVALCPICHREIHYSKNAKELIEMLYVNINRLOK 277
tr A0A3A3M3R2 A0A3A3M3R2_SALMO	TINCVALCPUCHRELHYSKNAKELIEMLYVNINRLOK 277
tr A0A1S9J5H1 A0A1S9J5H1 SHIBO	TINCVAICPECHRELHYSKNAKELIEMLYVNINRLOK 277
tr A0A403M222 A0A403M222 SHIDY	TINCVALCPICHREIHYSKNAKELIEMLYVNINRLOK 277
	** <b>*</b> *** <b>*</b> *** <b>*</b> *********************

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

QBB00100.1			MVKVININGNLVELPEPSAKLSKAES	26
AVX35624.1				0
ARX61598.1				0
ARX61672.1				0
sp Q6QGL2 TFLIV_BPT5	-RMLSCS	SSN	TLEAYLGKLGIVVIPTNR-QYD	55
sp Q38419 TEV3 BPR03	AKVKERNIAAIKEGTHVFC-		Q-SGGKMQSETQ-S	17
tr Q4TZV1 Q4TZV1 9CAUD	MITK	N	WDDEFQAEIRRRFYHVNGVLYHKTE-SKQTAFS	37
AXY85329.1	MAHCDTPF	HSP	ASEELR-TIHKRKYKLYREA	29
AKL98006.1	VKIIFPKYRKEMKC	GVA	WGELYNNYKDKKLDANELE-KEIVKLM	26
ACL33437.1	VKTTFPNYRKEMKG	GLE	WGLFYNTHKDRDLNPTTLE-AKIKTLM	28
ST081700.1	VKTTFPNYRKEMKO	GLE	WGLFYNTHKERDLNPTTLE-AKIKTLM	28
pdb 5H0M A			TRRND-KEYDKH-	46
tr10087241008724 9CYAN	GKRRHPDKSNVWVTKKYWHTVGC	GDN	WVFAATKNGEITMRLFKHSO-KEIVRH-	47
EFX42782.1	AHPE-IVDWLTOEYTKR		D-RGIRKLH	53
AFZ55703.1				0
tr 038112 038112 BPR1T	IALRADRTGA		HRVAFDKNR-KI	21
HNH Ana-Cas9	SERMADERDKA		Q-EA	22
UNU SpyCaco			VI VI VVI ONCO	11
OBB00100 1	PDGRESKP		KNKISKIORAELEMKEGGREAVECKIPE	63
AVX35624 1			MALTKKOBEKLEMKEGGECAVCCCELPE	28
AVA55024.1			MALTRINGRENI DMKEGGDCAVOCCEI DE	20
ARX61550.1			MALIKKQKEKERIKIKI GGRERI DEL-	20
SOLOGOGI 2 LTELIV BDT5	NK_VKSATEVI VNGSD		TSSVELENETINECIEDERSON FSWI	00
cp103841817EV3 BBD03	KEVNIDGSHHEI SEDHKK-			22
+»104mgy1104mgy1 0chup	DACAAER DC		-KISAKILENVKWGINFAQKEITEDI CBNIGKG	70
AVV05220 1	PAGITINPG		VDICY VDVDCYK ISVCC V	57
AKI03323.1	MDDDWTW_WSGIVDVUISCEEV_		_VINTESETDNORDEUVERORGICORCETEL_	20
ACL 22/27 1	FDDFUGK_KGCIVAVII TCFFD_		-ILNIKSFIDNQKKEVIEKQKGIDQKCKIHFEI	22
CE001700 1	EDDEVSK-KSGIIAIILIGEEK-	-	VICIDAE TORDRORMERDODGICE TOURICKERT	22
SIUGI/UU.I	EDDEVSK-KSGITATILIGEER-		-ILSLKAFIDKDKKIMFERQUGICFHUNGKFKI	05
+=10097241009724 0CVIN	QQARAF	, inc	SEWEKIKLA-VLANDNILLOHULKERNI	00
EEV/2702 1		IWDA	SKAGENFLVFARVAILLAAALGALDAUUTCACEVE	10
NE755703 1	QQGMLQKIAKGVIKIDFMLVL	2011	V TN UDI DDI WEDACNICEVOI I CAUD	21
Ar233703.1	M5K1		-IINVDLKKLVVERAGNIESICHISAVD	20
UPIQ38112 Q38112 BPRIT	MUNTORD N. CV	-	LLKTONTCHCGKPIDK	38
IINII Alia-Cass	MARIQRD-1GR		JIIS-KODIVKEDALELQGCARDITEIIIGI	03
HNH SpyCas9	-DMYVDOELDINRLSDYDVT	DHI	VPOSFLKDDS	4
OBB00100.1	KGWHZ	DHV	EPVRRDFELVRAPVGSGVTHVARSTGKVMHPELHA	10
ÃVX35624.1	KGWYAT	DHV	OAVLRKSEOCMKAAEK-RIFRLKSTGDVFRPEADC	7
ARX61598.1	KGWY2T	DHV	OAVLRKSEOCMKAAEK-RIFRLKSTGDVFRPEADC	70
ARX61672.1	KGWHA	DHV	OAVLRKSEOCMKAAEK-RIFRLKSTGDVFRPEADC	70
SD1060GL21TELTV BPT5		EHK	DGNHYNNF	1
splo384191TEV3 BPR03	PGFVI	KHM	DRCKI.N	2
trio4TZV1104TZV1 9CAUD	YLYVHRVVWFLEVGYOVDTIT	DHT	DI.NPINND	1
AVV85320 1	OMI A HOW WITH DIGIOUVE III	DHT	NNDSDND	90
AK103323.1	VENENT NEWEN	DUT		2
ACL30000.1	FEMERAL FEMERAL	DUT	TFWRDGGK-11	25
RCL33437.1			TEMSQGGK-ID	3.
ndh 5H0MIN	mparrent for the second	DUT	TIWSYGGK-TD	3:
trio09724 009724 009724	TRAVIVI	PHL:	IPLLVDWSKR-LD	L .
LT 10087241008724_9CYAN	DDL1E1		TERPAGAGGK-DA	24
EFX42/82.1	GVELHVI	DHI	KFKDLGGKA-T	1
AF255/03.1	RSSGCQVI	DHL	ISVKHGGA-ST	5.
TT12381121238112_BPR1T	RLKAPDPLSPVVI	DHI	IPINKGGHPSA	6.
HNH Ana-Cas9	CQL	DHI	VPQAGPGSNNR	82

. \*

IKVLTRSDR HNH SpyCas9 NLFPSCAF NLVPACAF LFKGA----FSVE-----GMRNEITKQVERARAYSVNFRTAER 152 TF OBB00100.1 LLKTT----YSLE-----MFRTQVSLQVERGRRSSVNFRTAER AVX35624.1 PF 116 LLKTT----YSLE-----MFRTQVSLQVERGRRSSVNFRTAER 116 ARX61598.1 PENLVPA NLVPA NLALL AF LLKTT----YSLE-----MFRTQVSLQVERGRRSSVNFRTAER ARX61672.1 PF 116 sp|Q6QGL2|TFLIV BPT5 WI PN ALTPTHAGKNIGRYTERTVNTCAICHCEISSRATHCK---SCTPKGI 174 RIQLN PY KKDLS--PSTYKRWHGDNCKA-RFND-----sp|Q38419|TEV3 BPR03 PN 269 TQ QA tr|Q4TZV1|Q4TZV1 9CAUD PS LREA HYNKNGKPNTSSTYKGVSWDK-STCKWKVGIQHNKKKIHL-GYFDDEV 161 AXY85329.1 IE LRCV LSNVDVRSHCLSGEKYIALDK-RTGRFAVRIRRKSHG-TY-GTLEEAI 147 AKL98006.1 SV ICQML ÑІ КЗ КЗ RRK----SGI-----358 RKK----SDK------16QML ACL33437.1 LN 370 RKK----SDK------ 370 ST081700.1 ln<mark>n</mark>goml NRK----TAEDKRRYG----pdb | 5H0M | A MI LQSL QA 130 YE LQAL DVK----TATDNSYNQPKSDT-EINVMW----tr|Q08724|Q08724\_9CYAN RH 584 SF 1GQTL EFX42782.1 LE FLK----KNFROTETG---KK-MFIRMLESAR-KAGELELVAFLEEVL 176 NLQK----GTDL-----GSINW---R-N-GEL-----VF AFZ55703.1 DE ILCYA WICHRQK----SDKL------P 91 EFINRSK----SNTP------FNVKQ---E----E-----P 91 80 tr|Q38112|Q38112 BPR1T MINLQLA RSK----SNTP-----PAVWA---Q-K-CGI-----P HNH Ana-Cas9 R VAT 112 HNH SpyCas9 -----YWRQLLNAKLI---TQRKFD---NLTKAERGGLS---ELDKAGFIK-109 FGLLHIVVKPVVFWFEQYNEQKQNE-----QBB00100.1 177 FGLIQVVNKPVVFWFEQYQRESEV------140 AVX35624.1 FGLIQVVNKPVVFWFEQYQRESEV-----ARX61598.1 140 FGLIQVVNKPVVFWFEQYQRESEV-----ARX61672.1 140 sp|Q6QGL2|TFLIV BPT5 TINPDITVEQIEYWVSKYSWIRASKE-LGLSDTGLRKRYKS-----LTGKD 219 sp|038419|TEV3 BPR03 269 EAAKAYDAKAIEIGGQYAR-----LNF----HDYVRTRI------ 191 tr|Q4TZV1|Q4TZV1 9CAUD SVR-----DAA-----INP----H-----157 AXY85329.1 -----AKL98006.1 358 ----- 370 ACL33437.1 ST081700.1 ----- 370 5HOM A 130 tr|Q08724|Q08724\_9CYAN ------584 SVYEKHGINGHIVWRKD----er mili EFX42782.1 193 AFZ55703.1 ---VRFFNPRRDLWGDHFRLDEAVIQPLTDI-GEVTTRILDFNNDERIIERLLLIEVGKY 136 KVLGNRNLPQSRDWSSYVS-----Q38112|Q38112 BPR1T 110 HNH Ana-Cas9 HVGVKEAIGRVRGWRKQTPNTS--SEDLTRLKKEVIARLRRTQEDPEIDERSM------163

#### 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-TflIV, 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 (CoIE7), 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 CoIE7 (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<sup>545</sup>→N-ColE7 mutant was completely abolished while activities of N<sup>560</sup> and H<sup>573</sup> 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

SGVHWGGSGHGNGGGNGNSGGGSNSSVAAVAFGPPA SGVHWGGSGHONGGGNGNSGGSGSGSGGSNSSVAAVAFGPPA SGIHWGGSGHONGGGNGNSGGSGSGTGGNLSAVAAPVAFGPPA SGIHWGGSGHGNGGGNGNSGGSGTGGNLSAVAAPVAFGPPA SGIHWGGSGHGNGGGNGNSGGSGTGGNLSAVAAPVAFGPPA SGIHWGGSGHGNGGGNGNSGGSGTGGNLSAVAAPVAFGPPA SGIHWGGSGHGNGGGNGNSGGSGTGGNLSAVAAPVAFGPPA SGIHWGGSGHGNGGGNGNSGGSGTGGNLSAVAAPVAFGPPA SGIHWGGSGGHGNGGGNGNSGGSGTGGNLSAVAAPVAFGPPA SGIHWGGSGGHGNGGGNGNSGGSGTGGNLSAVAAPVAFGPPA SGIHWGGSGGHGNGGGNGNSGGSGTGGNLSAVAAPVAFGPPA SGIHWGGSGGHKPGNSGNSGGSGTGGNLSAVAAPVAFGPPA SGIHWGGSGGHKPGNSGNSGGSGTGGNLSAVAAPVAFGPPA SGIHWGGSGGHKPGNSGNSNSGGSSTGGNLSAVAAPVAFGPPA SGIHWGGSGHKPGNSGNSSGGSGTGGNLSAVAAPVAFGPPA SGIHWGGSGHKPGNSGNSSGGSSTGDQSVAAPNAFGPPA SGIHWGGSGHKPGNSGNSGGSNSGGSNSSVAAPNAFGPPA SGIHWGGSGHKPGNSGNSGGSNSGGSNSSVAAPNAFGPPA SGIHWGGSGHKPGNSGNSGGSNSGGSNSSVAAPNAFGPPA SGIHWGGSGHKPGNSGNSGGSNSGGSNSSVAAPNAFGPPA SGIHWGGSGHGNGGGNSNSGGSNSGGSNSSVAAPNAFGPPA SGIHWGGSGHKPGNGSGNSGGSNSGGSNSSVAAPNAFGPPA SGIHWGGSGHKPGNGGNSNSGGSNSGGSNSSVAAPNAFGPPA SGIHWGGSGHKPGNGNSGNSGGSNSGGSNSSVAAPNAFGPPA SGIHWGGSGHKPGNSGNSGGSNSGGSNSGGSNSSVAAPNAFGPPA SGIHWGGSGNGNGGNGNSGGSNSGGSNSSVAAPNAFGPPA SGIHWGGSGHKPGNSGNSGGSNSGGSNSSVAAPNAFGPPA SGIHWGGSGNGGNGGNGGNSNSGGSNSSVAAPNAFGPPA SGIHWGGSGHKPGNSGNSGGSNSGGSNS LSAAIPDIMALK-GPFKFGJWGVALYGVLPSQIAKDDPNNMS LSAAIPDIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPNNMS LSAAIPDIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPNNMS LSAAIPDIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPNNMS LSAAIPDIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPNNMS LSAAIPDIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPNNMS LSAAIPDIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPNNMS LSAAIPDIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPNNMS LSAAIPDIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPNNMS LSAAIPDIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPNMS LSAAIPDIMALK-GPFKFGJWGVALYGVLPSQIAKDDPNNMS	AABGA LATPGA LAT	
SGVHWGGSGNGNGGGNGNSGGGSNSSVAA AFGPPA SGIHWGGSGHONGGGNGNSGGSGTGGNLSAVAAPAFGPPA SGIHWGGSGHONGGGNGNSGGSGTGGNLSAVAAPAFGPPA SGIHWGGSGHONGGGNGNSGGSGTGGNLSAVAAPAFGPPA SGIHWGGSGHONGGGNGNSGGSGTGGNLSAVAAPAFGPPA SGIHWGGSGRGNGGGNGNSGGSGTGGNLSAVAAPAFGPPA SGIHWGGSGRGNGGGNGNSGGSGTGGNLSAVAAPAFGPPA SGIHWGGSGRGNGGGNGNSGGSGTGGNLSAVAAPAFGPPA SGIHWGGSGRGNGGGNGSGNSGGSGTGGNLSAVAAPAFGPPA SGIHWGGSGRGNGGGNGSGNSGGSGTGGNLSAVAAPAFGPPA SGIHWGGSGRGNGGGNGSGNSGGSGTGGNLSAVAAPAFGPPA SGIHWGGSGRGNGGGNGSGNSGGSGTGGNLSAVAAPAFGPPA SGIHWGGSGRGNGGGNSSGGSGTGGNLSAVAAPAFGPPA SGUHWGGSGRGNGGGNSSGGSGTGGNLSAVAAPAFGPPA SGVHWGGSGRGNGGGNSSGGSGTGGNLSAVAAPAFGPPA SGVHWGGSGRGNGGGNSSGGSGTGGNLSAVAAPAFGPPA SGVHWGGSGRGNGGGNSSGGSGTGGNLSAVAAPAFAFGPPA SGVHWGGSGRGNGGGNSSGGSGTGGNLSAVAAPAFAFGPPA SGVHWGGSGRGNGGGNSSGGSGTGGNLSAVAAPAFAFGPPA SGVHWGGSGGHKPGNSGHSGTGDQ SGVHWGGSGGHKPGNSGHSGTGDQ SGVHWGGSGGHKPGNSGGSGSTGGNLSAVAAPAFAFGPPA SGVHWGGSGGHKPGNSGNSGGSSNSSVAAPAFAFGPPA SGVHWGGSGGHKPGNSGGSGSGTGGNLSAVAAPAFAFGPPA SGVHWGGSGGHKPGNSGGSGSGTGGNLSAVAAPAFAFGPPA SGVHWGGSGGHKPGNSGGSGSGTGGNLSAVAAPAFAFGPPA SGVHWGGSGGHKPGNSGGSGSGTGGNLSAVAAPAFAFGPPA SGVHWGGSGGHKPGNSGGSGSGTGGNLSAVAAPAFAFGPPA SGVHWGGSGGHKPGNSGGSGSGTGGNLSAVAAPAFAFGPA SGVHWGGSGGHKPGNSGGSGGGGGGSGNS SGVHWGGSGGHKPGNSG SGVHWGGSGGHKPGNGGNSSGGSGNSSVAAPAFAFGPPA SGVHWGGSGGHKPGNSGGSGGSGTGGNLSAVAAPAFAFGPNMSS LSAAIADIMAALK-GPFKFGLWGALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGLWGALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGLWGALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGLWGALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGLWGALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGLWGALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGLWGALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGLWGALYGVLPSQIAKDDPNMSS LSAAIADIMAALK-GPFKFGLWGALYGVLPSQIAKDDPNMSS LSAAIADIMAALK-GPFKFGLWGALYGVLPSQIAKDDPNMSS LSAAIADIMAALK-GPFKFGLWGALYGVLPSQIAKDDPNMSS LSAAIADIMAALK-GPFKFGLWGALYGVLPSQIAKDDPNMSS LSAAIADIMAALK-GPFKFGLWGALYGVLPSQIAKDDPNMSS	AT SGA LST SGAA LST SGA LST SG	
SGIHWGGSGHONGGNONSGGSGTGGNLSAVAPTAFGPPA SGIHWGGSGHONGGONGNSGGSGTGGNLSAVAAPTAFGPPA SGIHWGGSGHONGGGNGNSGGSGTGGNLSAVAAPTAFGPPA SGIHWGGSGHONGGGNGNSGGSGTGGNLSAVAAPTAFGPPA SGIHWGGSGGNONGGGNGNSGGSGTGGNLSAVAAPTAFGPPA SGIHWGGSGRONGGGNGNSGGSGTGGNLSAVAAPTAFGPPA SGIHWGGSGRONGGGNGNSGGSGTGGNLSAVAAPTAFGPPA SGIHWGGSGRONGGGNSNSGGSGTGGNLSAVAAPTAFGPPA SGIHWGGSGRONGGGNSNSGGSGTGGNLSAVAAPTAFGPPA SGIHWGGSGRONGGGNSNSGGSGTGGNLSAVAAPTAFGPPA SGIHWGGSGRONGGGNSNSGGSGTGGNLSAVAAPTAFGPPA SGIHWGGSGHONGGNSNSGGSSTGGNLSAVAAPTAFGPPA SGIHWGGSGHONGGNSNSGGSSTGNLSAVAAPTAFGPPA SGIHWGGSGHONGGNSNSGGSNSSVAAPIAFGPPA SGIHWGGSGHONGGNSNSGGSNSSVAAPIAFGPPA SGIHWGGSGHONGGNSNSGGSNSNAFGFPA SGIHWGGSGHONGGNSNSGGSNSNAFGFPA SGIHWGGSGHONGGNSNSGGSNSNAFGFPA SGIHWGGSGHONGGNSNSGGSNS	ST PGA LST	
SGIHWGGSSGHONGGONONSGGSSGTGGNLSAVAAPTAFGFPA SGIHWGGSSGHONGGONONSGGSSGTGGNLSAVAAPTAFGFPA SGIHWGGSSGHONGGONONSGGSSGTGGNLSAVAAPTAFGFPA SGIHWGGSSGRONGGONONSGGSSGTGGNLSAVAAPTAFGFPA SGIHWGGSSGRONGGGNONSGGSSGTGGNLSAVAAPTAFGFPA SGIHWGGSSGRONGGGNONSGGSSGTGGNLSAVAAPTAFGFPA SGIHWGGSSGRONGGGNONSGGSSGTGGNLSAVAAPTAFGFPA SGIHWGGSSGRONGGGNONSGGSSGTGGNLSAVAAPTAFGFPA SGIHWGGSSGRONGGGNSNSGGSSGTGGNLSAVAAPTAFGFPA SGIHWGGSSGRONGGGNSNSGGSSGTGDQSVAAPTAFGFPA SGIHWGGSSGRONGGGNSNSGGSSTGDQSVAAPTAFGFPA SGIHWGGSSGRONGGGNSNSGGSNSSVAAPTAFGFPA SGIHWGGSSGRONGGGNSNSGGSNSSVAAPTAFGFPA SGIHWGGSSGRONGGGNSNSGGSNSSVAAPTAFGFPA SGIHWGGSSGRONGGGNSNSGGSNSSVAAPTAFGFPA SGIHWGGSSGRONGGNSNSGGSNSSVAAPTAFGFPA SGIHWGGSSGRONGGNSNSGGSNSSVAAPTAFA SGIHWGGSSGRONGGNSNSGGSNSSVAAPTAFAFGFPA SGIHWGGSSGRONGGNSNSGGSNSSVAAPTAFA SGIHWGGSSGRONGGNSNSGGSNSSVAAPTAFAFGFPA SGIHWGGSSGRONGGNSNSGGSNSSVAAPTAFAFGFPA SGIHWGGSSGRONGGNSSGGNSSGGSNSSVAAPTAFAFGFPA SGIHWGGSSGRONGGNSNSGGSNSSVAAPTAFAFGFPA SGIHWGGSSGRONGGNSNSGGSNSGGSNSSVAAPTAFA SGIHWGGSSGRONGGNSNSGGSNSSVAAPTAFAFGFPA SGIHWGGSSGRONGGNSSGGSNSGGSNSSVAAPTAFAFA SGIHWGGSSGRONGGNSNSGGSNSSVAAPTAFAFGFPA SGIHWGGSSGRONGGNSSGGSNSSVAAPTAFAFGFPA SGIHWGGSSGRONGGNSSGGSNSSVAAPTAFAFTAF SGIHWGGSSGNGGNSSGGSNSSVAAPTAFAFGFPA SGIHWGGSSGRONGGNSSGGSNSSVAAPTAFAFTAF SGIHWGGSSGNSGNSGGNSSGGSNSSVAAPTAFAFTAF SGIHWGGSSGNSGNSGGNSSGGSNSSVAAPTAFAFTAFA SGIHWGGSSGNSGNSGNSGGNSSGGNS	STPGA STPGA STPGA STPGA STPGA STPGA STPGA STPGA STPGA STPGA STPGA STPGA STPGA STPGA STPGA STPGA STPGA STPGA STGA STGA STGA STPGA STGA STGA STGA STPGA STGA STGA STGA STGA STGA STPGA STGA STGA STGA STGA STGA STGA STGA ST	
SGTHWGGSSGHONGGNONSGGSSGTGGNLSAVAPVAFGPPA SGIHWGGSSGHONGGONGNSGGSSGTGGNLSAVAPVAFGPPA SGIHWGGSSGHONGGGNONSGGSSGTGGNLSAVAPVAFGPPA SGIHWGGSSGRONGGGNONSGGSSGTGGNLSAVAPVAFGPPA SGIHWGGSSGRONGGNONSGGSSGTGGNLSAVAPVAFGPPA SGUHWGGSSGRONGGNONSGGSSGTGGNLSAVAPVAFGPPA SGUHWGGSSGRONGGNONSGGSSGTGGNLSAVAPVAFGPPA SGUHWGGSSGRONGGNONSGGSSGTGGNLSAVAPVAFGPPA SGUHWGGSSGRONGGNONSGGSSGTGGNLSAVAPVAFGPPA SGUHWGGSSGRONGGNONSGGSSGTGGNLSAVAPVAFGPPA SGUHWGGSSGRONGGNONSGGSSGTGGNLSAVAPVAFGPPA SGUHWGGSSGRONGGNONSGGSSGTGGNLSAVAPVAFGPPA SGUHWGGSSGRONGGNONSGGSSTGGNLSAVAPVAFGPPA SGUHWGGSSGHSNGGNSNSGGSSNSSUAPHLAFGPPA SGUHWGGSSGHGNGGNSNSGGSNSSUAPHLAFGPPA SGUHWGGSSGHONGGNSNSGGSNSSUAPHLAFGPPA SGUHWGGSSGHONGGNSNSGGSNSSUAPHLAFGPPA SGUHWGGSSGHONGGNSNSGGSNSSUAPHLAFGPPA SGUHWGGSSGHONGGNSNSGGSNSSUAPHLAFGPPA SGUHWGGSSGHONGGNSNSGGSNSSUAPHLAFGPPA SGUHWGGSSGHONGGNSNSGGSSNSSUAPHLAFGPPA SGUHWGGSSGHONGGNSNSGGSNSSUAPHLAFGPPA SGUHWGGSSGHONGGNSNSGGSSNSSUAPHLAFGPPA SGUHWGGSSGHONGSGNSSSGSSNSSUAPHLAFGPPA SGUHWGGSSGHONGSGNSSSGSSNSSUAPHLAFGPPA SGUHWGGSSGHONGSGNSSSGSSNSSUAPHLAFGPPA SGUHWGGSGGNGGNSSSGGSNSSUAPHLAFGPPA SGUHWGGSGHONGSGNSSGGSNSSUAPHLAFGPPA SGUHWGGSGHONGSGNSSGGSSNSSUAPHLAFGPPA SGUHWGGSGHONGSGNSSGGSNSSUAPHLAFGPPA SGUHWGGSGHONGSGNSSGGSNSSUAPHLAFGPPA LSAAILDIFAALK-GPFKFGWGVALYGULPSQIAKDDPNMMS LSAAILDIMAALK-GPFKFGWGVALYGULPSQIAKDDPNMMS LSAAILDIAALK-GPFKFGWGVALYGULPSQIAKDDPNMMS LSAAILDIAALK-GPFKFGWGVALYGULPSQIAKDDPNMMS LSAAILDIAALK-GPFKFGWGVALYGULPSQIAKDDPNMMS LSAAILDIAALK-GPFKFGWGVALYGULPSQIAKDDPNMMS LSAAILDIAALK-GPFKFGWGVALYGULPSQIAKDDPNMMS LSAAILDIAALK-GPFKFGWGVALYGULPSQIAKDDPNMSS LSAAILDIAALK-GPFKFGWGVALYGULPSQIAKDDPNMSS LSAAILDIAALKGNUKFTFFGVUSSLIPSEIAKDDPNMSS	ST PGA LST	
SGIHWGGSSGHONGGONGNSGGSSGTGGNLSAVAAP <sup>T</sup> AFGFPA SGIHWGGSSGHONGGGNGNSGGSSGTGGNLSAVAAP <sup>T</sup> AFGFPA SGIHWGGSSGRONGGGNGNSGGGSGTGGNLSAVAAP <sup>T</sup> AFGFPA SGIHWGGSSGRONGGNSSGGSSGTGGNLSAVAAP <sup>T</sup> AFGFPA SGUHWGGSSGRONGGNSSGGSGTGGNLSAVAAP <sup>T</sup> AFGFPA SGVHWGGSSGHGGNSGNSSGGSSNSSVAAPIAFGFPA 	STEGA STEGA	
SGIHWGGSGHONGGONGSGGSGTGGNLSAVAP <sup>T</sup> AFGFPA SGIHWGGSGRGNGGGNGGNGSGGSGTGGNLSAVAAP <sup>T</sup> AFGFPA SGIHWGGSGRGNGGGNGNSGGGSGTGGNLSAVAAP <sup>T</sup> AFGFPA SGIHWGGSGRONGGGNSNSGGSSTGGNLSAVAAP <sup>T</sup> AFGFPA SGYHWGGSGHONGGGNSNSGGGSNS	LST PGA LST PGA LST PGA LST PGA LAP PGA LAP PGA LAP PGA LAP PGA 	
SGIHWGGSGRONGGONGSGGSGTGGNLSAVAAP\AFGPA SGIHWGGSGRGNGGGNGGNGSGGSGTGGDLSAVAAP\AFGPFA GEFGGGGHKPGGNSGHSGGGSGTGDGQSGAANAFGFPA SGVHWGGSGHGNGGGNSNSGGGSNSSVAAPNAFGFPA 	LSTPGA LSTPGA LSAPGA LA	
SGIHWGGGSGRGNGGGNGNSGGSGTGGNLSAVAAP AFGPPA PGEFGGGHKPGGNSGNHSGTGDGQPSGAANAFGPPA SGVHWGGSGHKPGGNSNNSGGSNSSVAAPHAFGPPA MAFGPPA 	LST BGA LVPAGA LAA BGA LAA BGA LAA PGA LAT PGA 	
PGEFGGGGHKPGGNSGNHSGGTGDGPSGAALAFGFPA SGTHWGGSGHONGGGNSNSGGGSNSSVAAPAFGFPA 	LVPAGA LAAPGA LAAPGA LAAPGA LATPGA (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL	
SGVHWGGGSGHGNGGGNSNSGGGSNSSVAAPHAFGPPA HAFGPPA LSAAIADIFAALK-GPFKFS.WGIALYSILPSEIAKDDPKMMS LSAAIADIFAALK-GPFKFS.WGIALYGILPSEIAKDDPKMMS LSAAIADIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPKMMS LSAAIADIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPKMMS LSAAIADIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPKMMS LSAAIADIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPKMMS LSAAIADIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPKMMS LSAAIADIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPKMMS LSAAIADIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPKMMS LSAAIADIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPKMMS LSAAIADIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPKMMS LSAAIADIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPKMMS	LAABGA LAAPGA LATPGA 	
LSAAIADIMAALK-GPFKFS.WGIALYSILPSEIAKDDPMMMS LSAAIADIFAALK-GPFKFS.WGIALYGILPSEIAKDDPMMMS LSAAIADIFAALK-GPFKFS.WGVALYGVLPSQIAKDDPMMMS LSAAIADIMAALK-GPFKFG.WGVALYGVLPSQIAKDDPMMMS LSAAIADIMAALK-GPFKFG.WGVALYGVLPSQIAKDDPMMMS LSAAIADIMAALK-GPFKFG.WGVALYGVLPSQIAKDDPMMMS LSAAIADIMAALK-GPFKFG.WGVALYGVLPSQIAKDDPMMMS LSAAIADIMAALK-GPFKFG.WGVALYGVLPSQIAKDDPMMMS LSAAIADIMAALK-GPFKFG.WGVALYGVLPSQIAKDDPMMMS LSAAIADIAKLKVNLKFTPFGVULSSLIPSEIAKDDPMMMS	LAAPGA LATPGA *** (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL	
LSAAIADIFAALK-GPFKFSAWGIALYSILPSEIAKDDPKMMS LSAAIADIFAALK-GPFKFSAWGIALYGILPSEIAKDDPKMMS LSAAIADIFAALK-GPFKFGWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGWGVALYGVLPSQIAKDDPNMMS LSAAIADIAALK-GPFKFGWGVALYGVLPSQIAKDDPNMMS LSAAIADIAKLKVNLKFTPFGVUSSLIPSEIAKDDPNMMS	ATPGA (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL (IVTSL	
****** LSAAIADIFAALK-GPFKFSAWGIALYSILPSEIAKDDPKMMS LSAAIADIFAALK-GPFKFSWGIALYGULPSEIAKDDPNMMS LSAAIADIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGJWGVALYGVLPSQIAKDDPNMMS	XIVTSL XIVTSL	
LSAAIADIFAALK-GPFKFSAWGIALYSILPSEIAKDDPKMMS LSAAIADIFAALK-GPFKFSAWGIALYGILPSEIAKDDPNMMS LSAAIADIMAALK-GPFKFGLWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGLWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGLWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGLWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGLWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGLWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGLWGVALYSVLPSQIAKDDPNMMS LSAAIADIMAALK-WNLKFTFFGVUSSLIPSEIAKDDPNMMS	KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL	
LSAAIADIFAALK-GPFKFSAWGIALYSILESEIAKDDFKMMS. LSAAIADIFAALK-GPFKFSAWGIALYGILESEIAKDDFKMMS. LSAAIADIMAALK-GPFKFGLWGVALYGVLESQIAKDDFNMMS. LSAAIADIMAALK-GPFKFGLWGVALYGVLESQIAKDDFNMMS. LSAAIADIMAALK-GPFKFGLWGVALYGVLESQIAKDDFNMMS. LSAAIADIMAALK-GPFKFGLWGVALYGVLESQIAKDDFNMMS. LSAAIADIAALK-GPFKFGLWGVALYGVLESQIAKDDFNMMS. LSAAIADIAALKYNLKFTFFGVUSSLIESEIAKDDFNMMS.	KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL	P1 P1 P1 P1 P1 P1 P1 P1 P1
LSAAIADIFAALK-GPFKFSAWGIADIGIDESEIAKDDPNMMS LSAAIADIMAALK-GPFKFGWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGWGVALYGVLPSQIAKDDPNMMS LSAAIADIMAALK-GPFKFGWGVALYGVLPSQIAKDDPNMMS LSAAIADIAALK-GPFKFGWGVALYGVLPSQIAKDDPNMMS LSAAIADIAKLKKVNLKFFFGVULSSLIPSEIAKDDPNMMS	KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL	
LSAAIADIMAALK-GFFKFGLWGVADYGVLFSQIAKDDFNMMS; LSAAIADIMAALK-GFFKFGLWGVALYGVLFSQIAKDDFNMMS; LSAAIADIMAALK-GFFKFGLWGVALYGVLFSQIAKDDFNMMS; LSAAIADIMAALK-GFFKFGLWGVALYGVLFSQIAKDDFNMMS; LSAAIADIMAALK-GFFKFGLWGVALYGVLFSQIAKDDFNMMS; LSAAIADIAKLKKVNLKFFFFGVUSSLIFSEIAKDDFNMMS; LSAAIAGIIAKLKKVNLKFFFFGVUSSLIFSEIAKDDFNMMS;	KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL	P1 P1 P1 P1 P1 P1 P1
LSAAIADIMAALK-GPFKFGLWGVALYQVLSQIAKDDPNMMS: LSAAIADIMAALK-GPFKFGLWGVALYQVLSQIAKDDPNMMS: LSAAIADIMAALK-GPFKFGLWGVALYQVLSQIAKDDPNMMS: LSAAIADIMAALK-GPFKFGLWGVALYQVLSQIAKDDPNMMS: LSAAIAGIIAKLKKVNLKFTPFGVVLSSLIPSEIAKDDPNMMS:	KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL	
LSAAIADIMAALK-GPFKFGLWGVALYQVLBSQIAKDDPNMMS: LSAAIADIMAALK-GPFKFGLWGVALYQVLBSQIAKDDPNMMS: LSAAIADIMAALK-GPFKFGLWGVALYQVLBSQIAKDDPNMMS; LSAAIAGIIAKLKKVNLKFTPFGVLSSLIPSEIAKDDPNMMS; LSAAIAGIIAKLKKVNLKFTPFGVLSSLIPSEIAKDDPNMMS;	KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL	PI PI PI PI
LSAAIADIMAALK-GFFKFGLWGVALYQVLFSQIAKDDPNMMS: LSAAIADIMAALK-GFFKFGLWGVALYQVLFSQIAKDDPNMMS: LSAAIAGIIAKLKKVNLKFFFGVVLSSLIFSEIAKDDPNMMS:	KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL	PJ PJ PJ PJ
lsaaiadimaalk-gpfkfglwgvalygvlpsqiakddpnmms lsaaiagiiaklkkvnlkftpfgvlsslipseiakddpnmms lsaaiagiiaklkkvnlkftpfgvlsslipseiakddpnmms	KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL	PI PI PI
LSAAIAGIIAKLKKVNLKFTPFGVVLSSLIPSEIAKDDPNMMS: LSAAIAGIIAKLKKVNLKFTPFGVVLSSLIPSEIAKDDPNMMS:	KIVTSL KIVTSL KIVTSL KIVTSL KIVTSL	P2 P2 P2
LSAAIAGIIAKLKKVNLKFTPFGVVLSSLIPSEIAKDDPNMMS:	KIVTSL KIVTSL KIVTSL KIVTSL	P2 P2
	KIVTSL KIVTSL KIVTSL	$\mathbf{P}$
LAAAIADVLAVLK-GPFKFGVWGIALYGILPTEIAKDDPRMMS	KIVTSL KIVTSL	-
LSAAIADIFAALK-GPFKFSAWGIALYGILPSEIAKDDPNMMS	XIVTSL	P1
LSAAIADIFAALK-GPFKFSAWGIALYGILPSEIAKDDPNMMS	T T T T T T C T	P
LSAAIADIFAALK-GPFKFSAWGIALYGILPSEIAKDDPNMMS	TALST	P2
*:****	*****	* 3
		1
TLPLDQATTRVTKRVTDVVKPHRQHIAVVAGVPMSVPVVNAKP	CRTPGV	FF
TLPLDQATVSVTKRVTDVVKDTRQHIAVVAGVPNSVPVVNAKP	TRTPGV	E l
SLPLDKATVNVNVRVVDIVKDERQNISVVSGVPNSVPVVDAKP	TERPGV	E 1
SLPLDKATVNVNVRVVDDVKPERQNISVVSGVENSVPVVDAKP	TERPGV	F 1
SLPLDKATVNVNVRVVDDVKDERQNISVVSGVENSVPVVDAKP	FERPGV	E 1
SLPLDKATVNVNVRVVDLVKDERQN1SVVSGVPNSVPVVDAKP	TERPGV	1
STREIKATVNVNVRVVD VKDBRONTSVVSGVPNSVPVVDAKP	TERPGV	1
	TERPGV	1
SLPLDKATYNVNVRVVDIVKDERQNISVVSGVINSVPVVDAKP	TENEGA	1
SLPLDKATYNVNVRVVDIVKÞERQNISVVSGVENSVPVDAKP SLPLDKATYNVNVRVVDIVKÞERQNISVVSGVENSVPVDAKP	THE DOTT	<b>10</b> / 2
SLELDKATYNVNVRVVDDVKDERQNISVVSGVENSVPVVDAKP SLELDKATYNVNVVVDIVKDERQNISVVSGVENSVPVVDAKP SLELGQATISVVRVVDVKDERQHISVVSGVENSVPVVDAKP	THPGV	E S
SLELDKATYNVNVRVVDIVKPERQNISVVSGVENSVPVVDAKP SLELDKATYNVNVRVVDIVKPERQNISVVSGVENSVPVVDAKP SLELGQATISVTKRVTDVKPERQHIAVVAGVENSIFVVDAKP TLELDQATYSVURRVTDVKPTRQHIAVVAGVENSVPVVNAKP	THPGV TRTPGV	E E E
SLFLDRATYNVNVRVVDIVKDERQNISVVSGVENSVPVVDAKP SLFLDRATYNVNVRVDDVKDFRQNISVVSGVENSVPVVDAKP SLFLGQATSVTKRVTDVKDFRQHIAVVAGVENSIPVVDAKP TLFLDQATYSVTRRVTDVVKDFRQHIAVVAGVENSVPVVNAKP TLFLDQATYSVTRRVTDVKDFRQHIAVVAGVENSVPVVNAKP	THPGV TRTPGV TRTPGV	EF
S	SSLPLDKATYNVNVRVVDIVKPERQNISVVSGVPMSVPVVDAKP	SSLPLDKATVNVNVRVVDIVKPERQNISVVSGVPNSVPVVDAKPTERPGV

+~   3   3 V 211 V 10   3   3 V 211 V 10   3 V 21 V 21 V 21 V 20   3 V 21 V 21 V 20   3 V 21 V 21 V 20   3 V		275
tr AUASISUKTS AUASISUKTS SALEN	ASPPGVPSLTLSTVKGLPASTTLPRGITEDKGRTASPAGFTFGGGSHEAVIRFPKESGQK	275
CI AUAUP 672M9 AUAUP 672M9_SHISO	ASE POVESUTVSTVRGLEVSTTLERGITEDRGRTAVEAGETEGGGSBEAVIREPRESON	273
SPIPI/SSSICEA6_ECOLA	ASIPGAPVLNISVNNSTPAVQTLSPGVINNTDKDVRPAGF1QGGHTRDAVIRFPKDSUHN	201
+~1000402055410004020554 9HTDV	ASIPGAPULNISVNNSTPLVQTLSPSVINNTDKDVRPAGPIQGGHTRDAVIRPRDSGIN	290
ty FGPmp4 FGPmp4 FCOLV	ASTPGAPVENTSVNNSTPAVQTESPSVINNTDADVRAGPTQGGTARDAVTAPRDADAUT	201
tyla0a2v0kBu7la0a2v0kBu7 surso	ASTEGAEVINISTERVQIISESVINITERSTRAVIKERSEIQOGITKDAVIKERDAUTERDAUTERDAUTER	201
+= IBGINGG IBGINGG FOOLV	ASTPGAPVENTSVNNSTPAVQTESPSVINNTDKDVRPASPIQSSHTKDAVTRPROSONN	201
CI   B3VM35   B3VM35 _ ECOLA	ASIPGAPVINISVNDSTPAVQTLSPGVINNTDKDVRPAGPIQGGFTKDAVIRPPKDSOHN	292
+~12022150VV712022150VV7 02100	ASIPGAPVINISVANSTPAVQTISPOVINITDRDVKPAGPIQGGITRDAVIRPROSON	202
ap 10471121CP27 PCOLV	A SECURAT MUSEURCE DUSING DECTREDUCED AVENUE AND A CREECE OF SUCH AVENUE AND A CREECE OF SUCH AVENUE AND A CREECE OF AVENUE AND A CREECE AVENUE AVENUE AND A CREECE AVENUE AVENUE AND A CREECE AVENUE	200
+=1303270W4T01303270W4T0 0P20H	AST FOVESLIVSIVASLEVSIILFRSIIEDKORIAVERGE FOGOSNEAVIREFRESOOR	10/
+=1202100 T0V410 A02100 T0V410_555CH	ASP FOVESHIVSTVRSLEVSTTLERGITEDRORTAVERGETEGGGSDERVIREPRESOUR	104
CI AUAIS 50 5A4 AUAIS 50 5A4_SHIBO	ASP PGVPSLTVSTVRGLPVSTTLPRSITEDRGRTAVPAGPTPGGGSREAVIRPRESGOR	154
TT AUASYSUATS AUASYSUATS SALEN	PVYVSVTDVLTSAQVKQRQDEENKLQQEWNDTHPVEVAERNEEQARAELNQAIHDVARNQ	335
TT AUAUF6T2M9 AUAUF6T2M9_SHISO	EVYVSVTDVLTPAQVKQRQDEEKKIQQEWNDAHPVLVAERNKEQARAELNQADHDVARNQ	335
SPIPI/999ICEA6_ECOLX	AVYVSVSDVLSPLQVKQRQDEENRHQQEWDATHPVLAAERNEERARAELNQAIHDVARNQ	341
SP P04419 CEA2_ECOLX	AVYVSVSDVLSPLQVKQRQDEENRHQQEWDATHPVLAAERNYERARAELNQAIHDVARNQ	350
tr AUA403A554 AUA403A554 SHIDY	AVYVSVSDVLSPDQVKORODEENRHQQEWDATHPVLAAERNYERTRAELNQAIHDVARNQ	341
tr[E6BTB4[E6BTB4_ECOLX	AVYVSVSDVLSPDQVKQRQDEENKHQQEWDATHPVLAAERNYERARAELNQAIHDVARNQ	341
tr AUA3YUKRU/ AUA3YUKRU/_SHISO	AVYVSVSDVLSPDQVKQRQDEENRRQQEWDATHPVEAAERNYERARAELNQADHDVARNQ	341
tr B9VM99 B9VM99_ECOLX	AVYVSVSDVLSPDQVKQRQDEENRRQQEWDATHPVEAAERNYERARAELNQATHDVARNQ	352
tr AUA3T6S430 AUA3T6S430 SALET	AVYVSVSDVLSPDQVKQRQDEENRRQQEWDATHPVEAAERNYERARAELNQATHDVARNQ	342
tr A0A315FXY/ A0A315FXY/_SALET	PVYVSVTDVLTPEQVKQRQDEENRFQQEWDATHPVEVAERNYRLASDELNRAIVDVAGKQ	343
sp Q47112 CEA7_ECOLX	PVYVSVTDVLTPAQVKQRQDEEKRIQQEWNDAHPVIVAERNYEQARAELNQANHDVARNQ	336
tr A0A370V4L8 A0A370V4L8_9ESCH	FVYVSVTDVLTPAQVKQRQDEEKRIQQEWNDAHFVEVAERNYEQARAELNQATHDVARNQ	254
tr A0A1S9J9X4 A0A1S9J9X4_SHIBO	FIYVSVTDVLTPAQVKQRQDEENRIQQEWNDTHPVEVAERNYEQARAELDQATHDVARNQ	254
	······································	
tr AUA3Y3UXT9 AUA3Y3UXT9 SALEN	EROHKAVH VYNSRKHELDAANKTFADAKAEIKRFERFAREPMAHGHRMWOMAGLKAORAO	395
LE AUAUPOTZMS AUAUPOTZMS_SHISO	EROHAVUVINSRIAELDAANKTLADAKAEIKUPERFAREPMA GIRMWQMAGLIKAQRAO	390
SPIPI/999 CEA6_ECOLX	ERQHKAVQVYNSRKHELDAANKTLADATAETKQFNRFAHDFMAHGHRMWQMAGLKAQRAO	401
Sp[P04419]CEA2_ECOLX	ERQHKAVOVYNSRKHELDAANKTLADAIAEIKQFNRFAHDPMAHGHRMWQMAGLKAQRAO	410
tr AUA403A554 AUA403A554 SHIDY	EROHKAVOVYNSRKEELDAANKTLADAIAEIKOFNRFAHDPMAIGHRMWOMAGLKAORAO	401
tr   E6BTB4   E6BTB4 ECOLX	EROHKAVOVYNSEKSELDAANKTLADATAETKOFDEFAHDPMAIGHEMWOMAGLKAORAO	401
LE ADASIORRO/ ADASIORRO/_SHISO	EROHAAVU VIIISRAELIDAANATLADAIAEIROFERFARDPMA GRAMWQMAGLAAQAAO	401
TLIBANDALBANDA ECOTY	ERQHKAVUVYNSRKHELDAANKTLADAIAKIKQFNRFAHDPMAHGHRMWQMAGLKAQRAO	412
tr AUA3T6S430 AUA3T6S430 SALET	EROHKAVOVYNSRKHELDAANKTLADAIAEIKOFNRFAHDPMAHGHRMWQMAGLKAQRAO	402
tr AUA315FXY/ AUA315FXY/ SALET	ERQLQAAQAVAARKHELDAANKTFADAKEEIKKFERFAHDPMAHGHRMWQMAGLKAQRAQ	403
sp/Q4/112/CEA/_ECOLX	EROHKAVOVYNSRKHELDAANKTLADAKAEIKOFERFAREPMARGHRMWQMAGLKAQRAO	396
tr AUA370V4L8 AUA370V4L8_9ESCH	EROMKAVQVYNSRKHELDAANKTLADAKAEIKQFERFAREPMANGHRMWQMAGLKAQRAQ	314
tr AUA1S9J9X4 AUA1S9J9X4_SHIBO	EROHKAVQVYNSRKEELDAANKTLADAKAEIKQFERFAREPMAHGHRMWQMAGLKAQRAQ	314
tr A0A3Y3UXT9 A0A3Y3UXT9 SALEN	TOVNNKKAAFDAAAKEKSAADAALSSAMESRKKKEDEKRSAENKLNEEKNKPRKGVK	452
tr A0A0F6T2M9 A0A0F6T2M9 SHISO	TDVNNKKAAFDAAAKEKSAADAALSSAMEBRKKKEDEKRSAENKLNEEKNKRRKGTK	452
sp P17999 CEA6 ECOLX	TDVNNKQAAFDAAAKEKSDADAALSSAMESRKKKEDEKRSAENKLNEEKNKPRKGVK	458
sp P04419 CEA2 ECOLX	TDVNNKOAAFDAAAKEKSDADAALSAAQERRKOKENHEKDAKDKLDKESKRNKPGKATGK	470
tr A0A403A554 A0A403A554 SHIDY	TDVNNKQAAFDAAAKEKSDADAALSAAQERRKQKENHEKDAKDKLDKESKRNKPGKATGK	461
tr   E6BTB4   E6BTB4 ECOLX	TDVNNKQAAFDAAAKEKSDADAALSAAQERRKQKENFEKDAKDKLDKESKRNKPGKATGK	461
tr   A0A3Y0KRU7   A0A3Y0KRU7 SHISO	TDVNNKQAAFDAAAKEKADADAALSAAQERRKQKENHEKDAKDKLDKESKRNKPGKATGK	461
tr B9VM99 B9VM99 ECOLX	TDVNNKQAAFDAAAKEKSDADAALSAAQERRKOKENHEKDAKIKLDKESKRNKPGKATGK	472
tr A0A3T6S430 A0A3T6S430 SALET	TDVNNKQAAFDAAAKEKSDADAALSAAQERRKOKENHEKDAKDKLDKESKRNKPGKATGK	462
tr   A0A315FXY7   A0A315FXY7 SALET	NEVNOKOAE FNAAEKEKADADAALNVALE SRKOKEORAKDASDKLDKENKRNHPGKATGK	463
sp Q47112 CEA7 ECOLX	TDVNNKKAAFDAAAKEKSDADVALSSALERRKOKENFEKDAKAKLDKESKRNKPGKATGK	456
tr A0A370V4L8 A0A370V4L8 9ESCH	TDVNNKKAAFDAAAKEKSDADVALSSALERRKOKENFEKDAKAKLDKESKRNKPGKATGK	374
tr A0A1S9J9X4 A0A1S9J9X4 SHIBO	TDVNNKKAAFDAAAKDKSDADAALGAALERRKOKENFEKDSKDKLDKESKRNKPGKATGK	374
	.:**:*:* *:** *:*: **.**. * * **:**:* :.:. **::*	
tr A0A3Y3UXT9 A0A3Y3UXT9 SALEN	ELEG-YRASDGEHIGAFDPKTGKOIKGPDPK	538
tr A0A0F6T2M9 A0A0F6T2M9 SHISO	ELEG-YRASDGQHLGSFDPKTGKQLKGPDPK	538
sp P17999 CEA6 ECOLX -	ELEG-YRASDG <mark>DHL</mark> GSFEPKTG	544
sp P04419 CEA2_ECOLX	DLSKQFKGSNKTNIQKGKAPFARKKDQVGGRERFEL <mark>HHD</mark> KPISQDGGVYDMN <mark>NIR</mark> VTTPK	581
tr A0A403A554 A0A403A554 SHIDY	DLSKQFKGSNKTNIQKGKAPFARKKDQVGGRERFEL <mark>HHD</mark> KPISQDGGVYDMN <mark>NIR</mark> VTTPK	572
tr E6BTB4 E6BTB4_ECOLX	DLSKQFKGSNKTNIQKGKAPFARKKDQVGGRERFEL <mark>HHD</mark> KPISQDGGVYDMN <mark>NIR</mark> VTTPK	572
tr A0A3Y0KRU7 A0A3Y0KRU7_SHISO	DLSKQFKGSNKTNIQKGKAPFARKKDQVGGRERFEL <mark>HHD</mark> KPISQDGGVYDMN <mark>NIR</mark> VTTPK	572
tr B9VM99 B9VM99_ECOLX	ELSKNLNPSNKSSVSKGYSPFTPKNQQVGGRKVYEL <mark>HH</mark> DKPISQGGEVYDMD <mark>WIR</mark> VTTPK	583
tr A0A3T6S430 A0A3T6S430_SALET	ELSKQFNPGNKKRLSQGLAPRARNKDTVGGRRSFEL <mark>HH</mark> KPISQDGGVYDMD <mark>ULR</mark> ITTPK	573
tr A0A315FXY7 A0A315FXY7_SALET	ELSKQFIKGNRDRMQVGKAPKSRKSDAAGKRTSFEL <mark>HHD</mark> KPISQDGGVYDMD <mark>HIR</mark> VTTPK	574
sp Q47112 CEA7_ECOLX	ELSKQFSRNNNDRMKVGKAPKTRTQDVSGKRTSFEL <mark>HHE</mark> KPISQNGGVYDMD <mark>NIS</mark> VVTPK	567
tr A0A370V4L8 A0A370V4L8_9ESCH	ELSKQFSRNNNDRMKVGKAPKTRTQDVSGKRTSFEL <mark>HHE</mark> KPISQNGGVYDMD <mark>HIS</mark> VVTPK	485
tr A0A1S9J9X4 A0A1S9J9X4_SHIBO	ELSKQFKDSNKTNIQKGKAPFARKKDQVGGRERFEL <mark>HHD</mark> KPISQDGGVYDMD <mark>NIR</mark> VTTPK	485 x16
	:*:	

tr A0A3Y3UXT9 A0A3Y3UXT9 SALEN	GRNIKKYL	546 x10N
tr A0A0F6T2M9 A0A0F6T2M9 SHISO	-RSIKKYL	545
sp P17999 CEA6 ECOLX	-RNIKKYL	551 x9N
sp   P04419   CEA2 ECOLX	-RHIDIHRGK	590
tr A0A403A554 A0A403A554 SHIDY	-RHIDIHRGK	581
tr E6BTB4 E6BTB4 ECOLX	-RHIDIHRGK	581
tr A0A3Y0KRU7 A0A3Y0KRU7 SHISO	-RHIDIHRGK	581
tr B9VM99 B9VM99_ECOLX	-RHIDIHRGK	592
tr A0A3T6S430 A0A3T6S430 SALET	-RHIDIHRGQ	582 x9H
tr   A0A315FXY7   A0A315FXY7 SALET	-HHIDIHRGK	583
sp Q47112 CEA7 ECOLX	-RHIDIHRGK	576
tr A0A370V4L8 A0A370V4L8 9ESCH	-RHIDIHRGK	494
tr A0A1S9J9X4 A0A1S9J9X4_SHIBO	-RHIDIHRGK	494
	*. :	

#### 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 may synthesize several pyocins. strain Structurally the pyocins contain three domains, N-terminal receptor-binding viz 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.

- R-type pyocins resemble non-flexible and contractile tails of bacteriophages. They depolarize the cytoplasmic membrane in relation with pore formation.
- F-type pyocins also resemble phage tails, but with a flexible and non-contractile rodlike structure.
- S-type pyocins are colicin-like, proteasesensitive 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

sp Q06584 PYS2_PSEAE	KQTSQELENKARSLEAEAQRAAAEVEADYKARKANVEKFVQSELDQAGNALPQLTNPTPE	239
sp Q06583 PYS1_PSEAI	VPLQVKEKRREIELQFRDAEKKLEASVQAELDKADAALGPAKNLAPL	114
tr A0A335NYU7 A0A335NYU7_ACIBA	VPLQVKEKRREIELQFRDAEKKLEASVQAELDKADAALGPAKNLAPL	114
tr A0A448BPE8 A0A448BPE8_PSEFL	VPLQVKEKRREIELQFRDAEKKLEASVQAELDKADAALGPAKNLAPL	114
	: :.:. *:* ::: : ::* <mark>**:***:4</mark> . ** .* :*	
spl006584/PYS2 PSEAE	OWLERATOLVTOAIANKKKLOTANNAL	266
sp  006583 PYS1 PSEAI	DVINRSLTIVGNALOOKNOKLLLNOKKITSLGAKNFLTRTAEEIGEOAVREGNINGPEAY	174
tr A0A335NYU7 A0A335NYU7 ACIBA	DVINRSLTIVGNALOOKNOKLLLNOKKITSLGAKNFLTRTAEEIGEOAVREGNINGPEAY	174
tr A0A448BPE8 A0A448BPE8 PSEFL	DVINRSLTIVGNALOOKNOKLLLNOKKITSLGAKNFLTRTAEEIGEOAVREGNINGPEAY	174
	: ::*::*	
sp Q06584 PYS2 PSEAE	IAKAPNALEKQEATYNADLIVDEIASLQARLDKLNAETA	305
sp Q06583 PYS1 PSEAI	MRFLDREMEGLTAAYNVKLTTEAISSLCIRMNTLTAAKASIEAAAANKAREQAAAEAKRK	234
tr A0A335NYU7 A0A335NYU7 ACIBA	MRFLDREMEGLTAAYNVKLTTEAISSLCIRMNTLTAAKASIEAAAANKAREQAAAEAKRK	234
tr A0A448BPE8 A0A448BPE8 PSEFL	MRFLDREMEGLTAAYNVKLTTEAISSLOIRMNTLTAAKASIEAAAANKAREQAAAEAKRK	234
	· · · * *·**··* ·· *···* ·	
		672.2
splQ06584 PYS2_PSEAE	RRKEIARQAAIRAANTYAMPANGSVVATAAGRGLIQVAQGAASLAQAISDAIAVLGRVLA	365
sp QU6583 PYS1_PSEAI	AEEQARQQAAIRAANTYAMPANGSVVATAAGRGLIQVAQGAASLAQAISDAIAVLGRVLA	294
tr A0A335NYU7 A0A335NYU7_ACIBA	AEEQARQQAAIRAANTYAMPANGSVVATAAGRGLIQVAQGAASLAQAISDAIAVLGRVLA	294
tr A0A448BPE8 A0A448BPE8_PSEFL	AEEQARQQAAIRAANTYAMPANGSVVATAAGRGLIQVAQGAASLAQAISDAIAVLGRVLA	294
	.:: :**********************************	
		405
splQ06584 PISZ_PSEAE	SAPSVMAVGFASLTISSKTAEQWQDQTPDSVRIALGMDAAKLGLPPSVNLNAVAKASGTV	420
spiQ06583 PISI_PSEAL	SAPSVMAVGFASLTYSSRTAEQWQDQTPDSVRYALGMDAAKLGLPPSVNLNAVAKASGTV	354
TT AUASSSNIU/ AUASSSNIU/ ACIBA	SAPSVMAVGFASLTISSRTAEQWQDQTPDSVRIALGMDAAKLGLPPSVNLNAVAKASGTV	304
tr AUA448BPE8 AUA448BPE8_PSEFL	SAFSVMAVGFASLTISSRTAEQWQDQTFDSVRIALGMDAAKLGLPFSVNLNAVAKASGTV	354
spic065841PVS2 PSFAF	DT.PMRT.MNRARGNMMMT.SVV.SMDGVSVPKAVPVRMAAVNAMGT.YRVMVPSPMARAPPT.T	485
spl006583 PVS1 PSEAT	DI. PMRI. TNEARGNTTTT. SVVSTDGVSVPKAVPVRMAAVNATTGI. YEVTVPSTTARA PPI. I	414
tr   A0A335NYU7   A0A335NYU7 ACTBA	DI.PMRT.TNEARGNTTTT.SVVSTDGVSVPKAVPVRMAAYNATTGI.YEVTVPSTTAEAPPI.T	414
trla02448BPE81202448BPE8 PSEFL	DI PMRI TNEARGNTTTI SVVSTDGVSVPRAVPVRMAAYNATTGITEVTV PTTAEAPPI.T	414
	*****	
sp Q06584 PYS2_PSEAE	LTWTPASPPGNQNPSSTTPVVPKPVPVYEGATLTPVKATPETYPGVITLPEDLIIGFPAD	545
sp Q06583 PYS1_PSEAI	$\verb"LTWTPASPPGNQNPSSTTPVVPKPVPVYEGATLTPVKATPETYPGVITLPEDLIIGFPAD"$	474
tr A0A335NYU7 A0A335NYU7_ACIBA	$\verb"LTWTPASPPGNQNPSSTTPVVPKPVPVYEGATLTPVKATPETYPGVITLPEDLIIGFPAD"$	474
tr A0A448BPE8 A0A448BPE8_PSEFL	LTWTPASPPGNQNPSSTTPVVPKPVPVYEGATLTPVKATPETYPGVITLPEDLIIGFPAD	474
	***************************************	
SD10065841PYS2 PSEAE	SGIKPIYVMFRDPRDVPGAATGKGOPVSGNWLGAASOGEGAPIPSOIADKLRGK#FKNWR	605
sp 006583 PYS1 PSEAL	SGIKPIYVMFRDPRDVPGAATGKGOPVSGNWLGAASOGEGAPIPSOIADKLRGKTFKNWR	534
tr A0A335NYU7 A0A335NYU7 ACIBA	SGIKPIYVMFRDPRDVPGAATGKGOPVSGNWLGAASOGEGAPIPSOIADKLRGKTFKNWR	534
tr A0A448BPE8 A0A448BPE8 PSEFL	SGIKPIYVMFRDPRDVPGAATGKGOPVSGNWLGAASOGEGAPIPSOIADKLRGKTFKNWR	534
	************************	
		CCE
ap 1006592   DV91 DAPAT	DEDECEMENTATIONEDERE SKORNEGEL VINDEGER I VEDECERGEKTETETER INDEGE	600
SPIQUOJOJIPISI_PSEAL	DEREQUINTAVANDEELSKOFNEGSLAVMKDGGAFIVKESEQAGGKIKIEIHKVKVADGG	594
triadadadadadadadadadadadadada	DEREGENTAVANDERE SKORNEGET XXXXBCGAFT VKESEQAGGKIKISIIINKVKIADGG	554
CT LANA 4 4 0 DE DO LANA 4	**************************************	594
apiQuoJO4[FIJZ_FJEAE	GVINNON VAVIENCE 610	
++1202335NVII7120225NVII7 20182	GVVNMON VAVTECCE 610	
tr A0A448BPE8 A0A448BPE8 PSEFL	GVYNMEN VAVTPKRHIEIHTGGK 618 x13H	
	**************	

### 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 nonmitochondrial 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 tr Q1KVS2 Q1KVS2_TETOB sp P19593 RDPO_TETOB tr A0A249RX17 A0A249RX17_TETOB tr A9IAV6 A9IAV6_BORPD tr Q35366 Q35366_PODAS sp P05511 YMC6_SCHPO P03875_Q	KAKFCTN-EDYLNMVDKLSEIVNDPNGYKAIPLDRLYIAKKDGRKRPLSIPSYTDRCLQA KESFRTN-NDYISMMARLEEITADPQKYKATPLSRIYIPKRDGNSRPLSIPSYTDRCLQA RESFKTN-KNYVAMMATLEQITSNPHKYKATPLSRIYIPKRDGSARPLSIPSYTDRCLQA GSSLDGMSYERLAGLMAA-VKSGNYRFKPVRVLIPKSNGKTRPLGIPTGDDKLVQE SETLDGISKEWFEKISEQ-LKSEQFFFPTRRVVIPKANGKMRPLGIASPRDKIVQE PDTLDGMSIDVIDKIIQS-LKSEEFNFTPGRRILIDKASGGKRPLIGSPRDKLVQE LETLDGNNMMYIMKLSNE-LGTGKFKFKPMRVNIPKEGGMRPLSVGNPERDKIVQE	156 145 146 146 111 269 322 337
P03876.2	NITLDGINISYLNKLSKD-INJNMEKESPVRRVEIPKISGGERPLSVGNPREKIVQE	370
tr AOAOH3WIQ7 AOAOH3WIQ7 YEASX	NITLDGINISYLNKLSKD-INTNMFKFSPVRRVEIPKTSGGFRPLSVGNPREKIVQE	301
tr C0J5P2 C0J5P2_SACPW	NITLDGINISYLNKLSKD-INTNMFKFSPVRRVEIPKTSGGFRPLSVGNPREKIVQE	370
tr A0A1D8GYT0 A0A1D8GYT0 SACPS	NITLDGINISYLNKLSKD-INTNMFKFSPVRRVEIPKTSGGFRPLSVGNPREKIVQE	301
X57546	NITLDGINISYLNKLCKH-INTNMFKFSPVRRVEIPKTSGGFRPLSVGNPREKIVQE	302
tr Q34832 Q34832_KLULC	NITLDGINISYLNKLCKH-INTNMFKFSPVRRVEIPKTSGGFRPLSVGNPREKIVQE	302
+		216
tr AUAUS2LQ/2 AUAUS2LQ/2_SCHLO	LIKLALEPIAEEMSDLSSIGFPIKSTSWAVGRVLNLLANSLANISFVVEIDILGCFDNI	216
TETOB	LIKLAIEPMAEEVADVSSIGFHPMRNVSWAVGRVLNGLNNPLANIQIVVEIDIRGCPDNI	205
5012133331KDF0_1E10B	LINLAIEPMAEEVADLOSIGPHPMENVOWAVGEVENGENNPEANIQIVVEIDIEGEVENI	206
trlagtavelagtave BODDD	UNDMILUTET VEDUESDIGEREMICE SCHEMINGERVENGENNEDEN IGT VEDUIGGEDNI	167
+=10252661025266 BODAR	VVRIDDVRITEPVISDIGIANGRSCHIALMOVROR WIGMAWIVMMDIRGIPDNI	224
en 19055111 VMC6 8CHP0	TLDIVLEATVED.ENTASHGEDCDSCHSALDSIETNEKCCTWWIEGDIKACEDSI	378
P03875	VMRMILDTIFDKKMSTHSHGFFKNMSCOTAIWEVRNMFGGSNWFIEVDLKKCFDTI	393
P03876.2	SMRMMLEIIVNNSFSYYSHGERENLSCLTAIIOCKNYMOVCNWFIKVDLNKCFDTI	426
tr A0A0H3WIQ7 A0A0H3WIQ7 YEASX	SMRMMLEIIYNNSFSY <mark>SHGFR</mark> PNLSCLTAIIOCKNYMOYCNWFIKVDLNKCFDTI	357
tr C0J5P2 C0J5P2 SACPW	SMRMILEIIYNNSFSNY <mark>SHGFF</mark> PNLSCLTAIIHCKNYMQHCNWFIKVDLNKCFDTI	426
tr A0A1D8GYT0 A0A1D8GYT0 SACPS	SMRMILEIIYNNSFSNY <mark>SHGFR</mark> PNLSCLTAIIHCKNYMQHCNWFIKVDLNKCFDTI	357
x57546	SMRMILEIIYNNSFSNY <mark>SHGFF</mark> PNLSCLTAIVHCKNYMKHCNWFIKVDLNKCFDTI	358
tr Q34832 Q34832_KLULC	SMRMILEIIYNNSFSNY <mark>SHGFF</mark> PNLSCLTAIVHCKNYMKHCNWFIKVDLNKCFDTI	358
	::::: *: *: :::*: *:***	
tr  A0A0S2LQ72  A0A0S2LQ72_9CHLO tr  Q1KVS2 Q1KVS2_TETOB sp  P19593 RDPO_TETOB tr  A0A249RX17 A0A249RX17_TETOB tr  A0A249RX17 A0A249RX17_TETOB tr  Q35366 Q35366_PODAS sp  P05511 YMC6_SCHPO P03876.2 tr  A0A0H3WIQ7 A0A0H3WIQ7_YEASX tr  C055P2 C0J5P2_SACPW tr  A0A1D8GYT0 A0A1D8GYT0_SACPS	DHTFLMQFVPVIPKKILWEWLSCGYVERDDNKEVHETLRGVPQGGILSPLLSNLT DHEFVSQVTPFIPKSILWSWLKCGYLEQNSE-QLQPTTSCVPQGGILSPLINNLT NHQFISQVTPFIPKKILWAWLKCGYLERNSN-TLQPTTGVPQGGISPLINNLT DHEVLVDVLAKRIDDKRFLGLHSMLKAGYMED-WKFHDTFSGTPQGGVSPVLANIY DHHILEKLLVKHFQQQFFIDLYKKNKAGYVEF-DKDKSSIIGVPQGISPLINNLT HHILEKLLVKHFQQQFFIDLYKLKRAGYVEF-DKVKJVUVGGISPLINNIY SHDLIIKELKRYISDKGFIDLVKLKRAGYIDEKGTYHKPMLGLPQGSVSPVLANIY SHDLIIKELKRYISDKGFIDLVKLKRAGYUDEKMNYHNTIGIFQGSVSPILONIF PHNKLINVLNERIKDKGFHDLLYKLLRAGYUDKNNNYHNTIGIFQGSVSPILONIF PHNLINVLNERIKDKGFIDLYKLLRAGYUDKNNNYHNTIGIQQSVSPILCNIF	271 259 260 224 381 435 451 484 415 484 415
X57546	PHNMLINVLNERIKDKGFIDLLYKLLRAGYVDKHNNYHHTTLGI <mark>PQG</mark> SVVSPILCNIF	416
tr Q34832 Q34832_KLULC	PHNMLINVLNERIKDKGFIDLLYKLLRAGYVDKHNNYHHTTLGIPOGSVVSPILCNIF	416
tr  A0A0S2LQ72  A0A0S2LQ72_9CHLO tr  Q1KVS2 Q1KVS2_TETOB sp  P19593 RDP0_TETOB tr  A0A249RX17  A0A249RX17_TETOB tr  A91AV6 A91AV6_BORPD tr  Q35366 Q35366_PODAS sp  P05511  YMC6_SCHP0 P03875	C : : : : : : : : : : : : : : : : : : :	286 274 275 275 280 438 474 488
P03876.2	LDKLDKYLENKFENEFNIGNM-SNRGRNPIYNSLSSKIYRC	524
tr A0A0H3WIQ7 A0A0H3WIQ7_YEASX tr C0J5P2 C0J5P2_SACFW tr A0A1D8GYT0 A0A1D8GYT0_SACPS X57546 tr Q34832 Q34832_KLULC	LDRLDKVLENKFENEFNTGSM-SNRGRNFIYNSLSSKIYRC LDRLDKVLENKFENEFNIGSM-SNRSRNFIYNDLSSKIRRC LDRLDKVLENKFENEFNIGSM-SNRSRNFIYNDLSSKIRRC	455 524 455 456 456

tr A0A0S2LQ72 A0A0S2LQ72_9CHL0 tr Q1KVS2 Q1KVS2_TETOB sp P19593 RDP0_TETOB tr A0A249RX17 A0A249RX17_TETOB	PSKSKGAGFC SSKSKGCTYC SSQSKGNTYC SSQSKGNTYC SSKSKGNTYC	RYADDWYCF-TTTRENADTAVEAIK RYADDWYIL-TTSQETAQTALEAVK RYADDWYIL-TTTEETALIALPAVK RYADDWYIL-TTTEETALIALEAVK	320 308 309 309
tr ASIAV6 ASIAV6_BORPD	KRELAELY-LRRKALSSSDPMDANYRRLVYV	RYADDFLIGIIGSRDEAVTVMQRVA	335
en 10055111VMC6 SCHDO	GRADERMADIAVRATMPSMIPNPDLAEIIIV KDENADSKUTOKIAIEMONVONKIHGIOSNKI.MVV	KIADDWLIGVAGSSETARAIKERIA	534
P03875	KMESTRIKIHKERAKGPTEIYNDPNFKRMKYV	RYADDILIGVLGSKNDCKMIKEDLN	545
P03876.2	KLLSEKIKLIRLRDHYORNMGSEKSFKRAYFV	RYADDIIIGVMGSHNDCKNILNDIN	581
tr A0A0H3WIQ7 A0A0H3WIQ7 YEASX	KLLSEKLKLIRLRDHYQRNMGSDKSFKRAYFV	RYADDIIIGVMGSHNDCKNILKDIN	512
tr C0J5P2 C0J5P2_SACPW	KLLSDKLKLIRLRDHYQRNLGSDKSFKRAYFV	RYADDIIIGVMGSHDDCKNILNDMN	581
tr A0A1D8GYT0 A0A1D8GYT0_SACPS	KLLSDKLKLIRLRDHYQRNLGSDKSFKRAYFV	RYADDIIIGVMGSHDDCKNILNDMN	512
X57546	KLLSNKLKLIRLRDHYQRNLGSDKNFKRACFV	RYADDIIIGVMGSHDDCKNILNDMN	513
tr Q34832 Q34832_KLULC	KLLSNKLKLIRLRDHYQRNLGSDKNFKRACFV	RYADDIIIGVMGSHDDCKNILNDMN	513
tr A0A0S2L072 A0A0S2L072 9CHL0	EFLV-ARGLEIKEAKSRITNVNSESFEFLGYKF	SKVYRHNRKR	362
tr Q1KVS2 Q1KVS2 TETOB	EFLA-VRKLEVKLAKTAIKDIINDRNGFEFLSFRF	RKVYRRNRKR	352
sp P19593 RDFO_TETOB	EFLA-VRGLEVKLAKTTIKNIINDRNGFEFLSFRF	RKVYRRNRKR	353
tr A0A249RX17 A0A249RX17_TETOB	EFLA-VRGLEVKLAKTTIKNIINDRNGFEFLSFRF	RKVYRRNRKR	353
tr A9IAV6 A9IAV6_BORPD	GFISDKLHLEIAEEKSGVVHASDG-VRFLGYDV	RTYSGDRN-VRTVRSGRSITARSVS	391
tr   Q35366   Q35366_PODAS	AYLKDILKLELSMEKTLITNASEDKAYFLGTEI	QRISSVKGEIKRFKNIKGHPQRIPT	552
BD2075	CFCS-SIGLTVSPTKTKITNSITDKILFLGTNI	SHSKNVTFSRHFGILQKNS	585
P03876-2	NELKENLOWSINNDESVIENS-KEGVSELGYEV	KVTPWEKRPYRMIKKGDNFIRVRHH	638
tr A0A0H3WIO7 A0A0H3WIO7 YEASX	NFLKENLGMSINMDKSVIKHS-KEGVSFLGYDV	KVTPWEKRPYRMIKKGDNFIRVRHH	569
tr C0J5P2 C0J5P2 SACPW	NFLTENLGMSINMDKSIIKHS-KEGVSFLGYDV	KVTPWEMRPYRMIKKGDKFIRVRHH	638
tr  A0A1D8GYT0  A0A1D8GYT0_SACPS	NFLTENLGMSINMDKSIIKHS-KEGVSFLGYDV	KVTPWEMRPYRMIKKGDKFIRVRHH	569
X57546	NFLKENLGMSINMDKSIIKHS-KEGVSFLGYDV	KVTPWEMRPYRMIKRGDKFIRVRHH	570
tr Q34832 Q34832_KLULC	NFLKENLGMSINMDKSIIKHS-KEGVSFLGYDV	KVTPWEMRPYRMIKRGDKFIRVRHH	570
tr A0A0S2LQ72 A0A0S2LQ72 9CHLO	KPTWRDVVMDKWLGCCGLCRKRLDINYVPYELHH	LPRRFGGKDQPS	566
tr Q1KVS2 Q1KVS2 TETOB	KSSWKRVILEKWGPC <mark>C</mark> GL <mark>C</mark> RKNLEINSTPYEL <mark>HH</mark>	LPKRFGGKNTPN	555
sp P19593 RDPO_TETOB	KSSWKRVILEKWGPC <mark>C</mark> GL <mark>C</mark> RKNLEINSIPYEL <mark>HH</mark>	LPKRFGGKDTPN	556
tr A0A249RX17 A0A249RX17_TETOB	KSSWKRVILEKWGPC <mark>C</mark> GL <mark>C</mark> RKNLEINSIPYEL <mark>HH</mark>	LPKRFGGKDTPN	556
tr A9IAV6 A9IAV6 BORPD	THSRSELIQRLEARKCEYCGTTEGSFQVHH.	RKLKDVDKGKYLWPQVMSWRRR	588
TT   Q35366   Q35366_PODAS	RNLQVHH	WHIRTIDVKLSGFDKQLAAINR	701
P03875 S carevisiae	KVMLDT-AKANENKDOBIOUSTIRVEMAN	KOTHECH-TKATKDITETTMAKANK	809
P03876.2 S. cerevisiae	KYMLPR-SLSLFSGICQICGSKHDLEVHH	VRTLNNAA-NKIKDDYLLGRMIKMNR	831
tr A0A0H3WIQ7 A0A0H3WIQ7_YEASX	KYMLPR-SLSLFSGICDICSSKHDLEVHH	RTLNNAA-NKIKDDYLLGRMIKMNR	762
tr C0J5P2 C0J5P2_SACPW	KYMLPR-SLSLFSGICDVCGSKQNLEVH	KMLNNAA-NKIKNDYLLGRMIKMNR	831
tr   A0A1D8GYT0   A0A1D8GYT0_SACPS	KYMLPR-SLSLFSGICDVCSSKQNLEVH	7KMLNNAA-NKIKNDYLLGRMIKMNR	762
X57546	KYMLPR-SLSLFSGICDVCGSKQNLEVHH	WMLNNAA-NKIKNDYLLGRMIKMNR	763
CI  Q34032 Q34032_KD0DC	KIMDER-STSPLSCIPSACNPEAU	KMENNAA-NKIKNDILLGRMIRMIK	/03
tr A0A0S2LQ72 A0A0S2LQ72 9CHLO	NIVPICKSECHNQVSSAISSKDVENILQFMALGI	600	
tr  Q1KVS2  Q1KVS2_TETOB	NIVPICKSECHQLVSSSIQKSNVSEIKKYISLGI:	LEIPSDYLDNLQSSPSSY 607	
sp P19593 RDPO_TETOB	NMVLICKSECHQLVSSSIQKADVSEIQNYISLGI	LEIPMDYLENLKTSQSSY 608	
tr A0A249RX17 A0A249RX17_TETOB	NMVPICKSFCHQLVSSSIQKADVSEIQNYISLGI:	LEIPMDYLENLKTSQSSY 608	
tr A9IAV6 A9IAV6 BORPD	KTLVIOV-FCHQKLHAGVL	606 NO0 797	
splP055111YMC6_SCHP0	KOIPICR-SCHMKLHANKLTINEDKKV	807	
P03875 S. cerevisiae	KOIPLCK-OCHIKTHKNKFKNMGPGM	834	
P03876.2 S. cerevisiae	KOITICK-TCHFKVHOGKYNGPGL	854	
tr A0A0H3WIQ7 A0A0H3WIQ7_YEASX	KQITICK-TCHFKVHQGKYNGPGL	785	
tr C0J5P2 C0J5P2_SACPW	KQITICK-ICHYKIHQGKYNGPGL	854	
tr A0A1D8GYT0 A0A1D8GYT0_SACPS	KQITICK-TCHYKIHQGKYNGPGL	785	
X5/546 K. Lactis	KQITICK-TCHFKIHQGKYNGPGL	786	
CT16240351634037_VT0TC	: : : * **	/86	

### 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* P055111YMC6\_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* 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)

sp Q03JI6 CAS9B STRTD	ELVKVMGGRKPESIVVEMARENQYTNQGKSNSQQRLKRLEKSLKELGSKI-LKENIPAK-	804
tr F0I628 F0I628 STRSA	ELVKVMG-HNPESIVIEMARENQTTAKGKKNSQQRYKRIEDALKNLAPGL-DSNILK-	799
tr A0A428GKX8 A0A428GKX8 STRCR	ELVKVMG-HAPESIVIEMARENQTTAKGKKNSRQRYKRIEDSLKILASGL-NSKILK-	800
tr S8FJS0 S8FJS0 STRAG	ELVKVMG-YEPECIVVEMARENOTTAKGLSRSRORLTTLRESLANLKSNI-LEEKKPKYV	802
tr A0A380EV24 A0A380EV24 STAAU	ELVKVMG-HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI-LKEH	798
sp Q99ZW2 CAS9_STRP1	ELVKVMGRHKPENIVIEMARENQTTOKGQKNSRERMKRIEEGIKELGSQI-LKEH	799
tr A0A4P8PLM4 A0A4P8PLM4_CHLRE	ELVKVMGRHKPENIVIEMARENOTTOKGQKNSRERMKRIEEGIKELGSQI-LKEH	799
tr J7M7J1 J7M7J1_STRP1	ELVKVMGRHKPENIVIEMARENOTTOKGOKNSRERMKRIEEGIKELGSQI-LKEH	799
tr   A0A0R1TV35   A0A0R1TV35 9LACO	ELVKVMG-GNPKRIVVEMARETOKTHGTTR-REARIKEIIKNLNKLKLNE-LPKD	795
tr A0A328KW99 A0A328KW99 9LACT	EIIDIIG-YEPENIVVEMARESOTTKKGKDLSKERLEKLTEAIKEFDGPTDIKVK	803
tr 16T669 16T669 ENTHA	ELIAIIG-YKPKNIVIEMARENQKTHRTSPRLKALENGLKQIGSTL-LKEQ	802
tr A0A430A0E2 A0A430A0E2_9ENTE	EIVEIMG-YQPQNIVVEMARKNQGTKRTSTRLKFVEKCLTDFQSDI-LKRH	804
tr A0A1J0A4R8 A0A1J0A4R8 9ENTE	EIVDIMG-AHPSNIVVEMARENORTNRSNTROSQIEKSLKELESDL-LKNK	800
tr A0A2V3WF80 A0A2V3WF80 9BACI	EIVEIMG-YEPESIVIEMARENQTTSQGQRNAKERLKNVEGALKELKSDL-LKKH	811
tr A0A115J642 A0A115J642 9BACI	EIVEIMG-YEPESIVIEMARENQTTSQGQRNAKERLNNIEGALKELKSDL-LKKY	811
tr   A0A429Z5T3   A0A429Z5T3 9ENTE	ELVGIMG-YNPSNIVIEMARENQSTSYGKNKSKQRFKHVEEAMKELGSNL-LKET	802
tr   A0A2498GB5   A0A2498GB5 ENTTH	EIIDIMG-ELPTNIVVEMARENQTTAQGNRASKARMKYLEESIKKLGSSI-LEDE	797
sp Q927P4 CAS9 LISIN	ELVSVMG-YPPQTIVVEMARENQTTGKGKNNSRPRYKSLEKAIKEFGSQI-LKEH	802
tr H1GG18 H1GG18 LISIO	ELVSVMG-YPPQTIVVEMARENQTTGKGKNNSRPRYKSLEKAIKEFGSQI-LKEH	805
tr R3WHR8 R3WHR89ENTE	ELVDIMG-SLPKNIVVEMARENQTTSRGRTNSNPRMKALEEAMRNLRSNL-LKEY	803
	*:: ::* * *:****:.* * * : : :	
ep1003.TT61C3998_STPTD	LSKT DNNALONDELVLVVLONGKDMYTGDDLDTDELSMUDTDHTTEOAFLKDNSTDNE	862
+rIF01628/F01628 STRSA	ENETDNIOLONDRIFILYVLONGKDMYTGKAIDINOLSNKDIDHIIFOAFIKDDSIDNA	857
trla02428GKX81202428GKX8 STRCR	EHPTDNIGLONDBLELVYLONGKDMYTGEALDINKLSSYDIDHITEOAFIKDDSLDN	858
trissEIS0188EIS0 STEAG	KDOVENHHLSDDRLELVVLONGKDMYTDDELDIDNLSOKDIDHILEOAEIKDDSIDN	860
+r 101380EV24 202380EV24 STALL	PVENTOLONEKLYLYLONGEDWYVDOELDINELSDYDUDHIVEOSEIKDDSIDNY	854
spl0992W21CAS9_STRP1	PVENTOLONEKLYLYLONGROMYVDOELDINRLSDYDVDHIVPOSELKDDSIDNK	855 x15N
tr A0A4P8PLM4 A0A4P8PLM4 CHLRE	PVENTOLONEKLYLYYLONGROMYVDOELDINRLSD TOUDHIVEOSELKDDSTONK	855
tr J7M7J1 J7M7J1 STRP1	FVENTOLONEKLYLYYLONGROMYVDOELDINRLSDVDVDHIVEOSFLKDDSIDNK	855
tr A0A0R1TV35 A0A0R1TV35 9LACO	LPSNEELSDEKVYLYCLONGRDMYTGAPLDLDNLSOVWDHIIFOSFLKDDSIENK	851
tr A0A328KW99 A0A328KW99 9LACT	DLKNENLRNDRLYLYYLONGRDMYTNEPLDINNLSKYDIDHIIFOSFTTDNSIDNK	859
tr I6T669 I6T669 ENTHA	PTDNKALOKERLYLYYLONGRDMYTGEPLEIENLHOVWDHIIERSFIVDNSIDNK	858
tr A0A430A0E2 A0A430A0E2 9ENTE	PIDNRSLOSDRLYLYYLOAGKDMYTGEPLDINNLSNVDIDHIIFOSFIKDDSIDNK	860
tr A0A1J0A4R8 A0A1J0A4R8 9ENTE	LPSNEELKSNRLLLYYLONGIDLYTGOTLDITKLSSYDUDHIIFOSFITDNSLDNL	856
tr A0A2V3WF80 A0A2V3WF80 9BACI	PV DODMLKNDRLYLYYLONGRDMYTNOELDINKLSN DI DHIIFRSFTTDNSIDNR	867
tr A0A115J642 A0A115J642 9BACI	PV DOEALKNDRLYLYYLONGKDMYTNOELDINKLSN/DIDHIIERSFTTDNSIDNR	867
tr A0A42925T3 A0A42925T3 9ENTE	KADNNDLQNDRLYLYYLQKGRDMYTDKEIPIEDLSHYDIDHVIHRSFTTDNSIDNI	858
tr A0A249SGB5 A0A249SGB5 ENTTH	PISKDANLLRNDRLFLYYLONGRDMYTGNELDINNLSSYDIDHIIF	855
sp 0927P4 CAS9 LISIN	PTDNOELRNNRLYLYYLONGKDMYTGODLDIHNLSNYDIDHIVFOSFITDNSIDNL	858
tr H1GG18 H1GG18 LISIO	PTDNQELRNNRLYLYYLQNGKDMYTGQDLDIHNLSNYDIDHIVEQSFITDNSIDNL	861
tr   R3WHR8   R3WHR8 9ENTE	PTDNQALQNDRLYLYYLQNGKDMYTGLDLSLHNLSS <mark>/DIDHIVF</mark> QSFTTDNSL <mark>DNR</mark>	859
na n	. * ** ** * *:* : : * <mark>*</mark> ***********	

sp Q03JI6 CAS9B_STRTD	VLVSSA <mark>SNR</mark> GK-SDDVPSLEVVKKRKTFWYQLLKSKLISQRKFDNLTKAERGGLSPEDKA	921 x9
tr F0I6Z8 F0I6Z8_STRSA	VLTSSKI <mark>NR</mark> GK-SDNVPSIEVVQKRKAFWQQLLDSKLISERKFNNLTKAKRGGLDERDKV	916
tr A0A428GKX8 A0A428GKX8 STRCR	VLTSSKE <mark>NR</mark> GK-SDNVPSLEVVQKRKAFWQQLLDSKLISERKFNNLTKAERGGLDERDKV	917
tr S8FJS0 S8FJS0 STRAG	VLVSSAHNRGK-SDDVPSVEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKA	919
tr A0A380EV24 A0A380EV24 STAAU	VLTRSDENRGK-SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA	913
sp Q99ZW2 CAS9 STRP1	VLTRSDKNRGK-SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA	914 x23N
tr A0A4P8PLM4 A0A4P8PLM4_CHLRE	VLTRSDE <mark>NR</mark> GK-SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA	914
tr J7M7J1 J7M7J1_STRP1	VLTRSDE <mark>NR</mark> GK-SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA	914
tr A0A0R1TV35 A0A0R1TV35 9LACO	VLTAKIENSRK-TNGLPSVEIAKKMGAFWRSLLKVGAISEKKYRNLRRSLHGGMTEKLKE	910
tr A0A328KW99 A0A328KW99 9LACT	VLVSRTE <mark>NO</mark> GNKSDDVPSINIVHKMKPFWRQLHKAGLISDQKLKNLTKAEHGGLTEADKA	919
tr 16T669 16T669 ENTHA	VLVASKONOKK-RDDVPKKQIVNEQRIFWNQLKEAKLISPKKYAYLTKIELTPEDKA	914
tr A0A430A0E2 A0A430A0E2 9ENTE	VLVNLSV <mark>NR</mark> GNKVDDVPSINIVKKOKYFWKKLODAKLISSKKYASLTKAENGGLTSKDKE	920
tr A0A1J0A4R8 A0A1J0A4R8 9ENTE	VLVSSKENRGK-KDKVPSKEVVKRNKPYWEKLLKSGAMSKRKFDNLTKVERGGLTEADKA	915
tr   A0A2V3WF80   A0A2V3WF80 9BACI	VLVSSKI <mark>NR</mark> GK-SDNVPNKDVVRNMKPYWTSLYRSKLISKRKFENLTKAERGGLTDDDKA	926
tr A0A115J642 A0A115J642 9BAC1	VLVSSKI <mark>NR</mark> GK-SNDVPSEEVVRNMKSFWSLLYRSKLISKRKFDNLTKLKLTDDDKA	923
tr A0A429Z5T3 A0A429Z5T3_9ENTE	VLVSSKENRGK-SDDVPSEEVVRRMKPLWIQLLNAKAISKRKFDNLTKGERGGLTEEDKM	917
tr   A0A249SGB5   A0A249SGB5 ENTTH	VLTTSSN <mark>NR</mark> GK-SNTVPAESVVKKMRPTWERLLASGLISKKKFSYLTKATNGGLTEEDKA	914
sp Q927P4 CAS9 LISIN	VLTSSAG <mark>NR</mark> EK-GDDVPPLEIVRKRKVFWEKLYQGNLMSKRKFDYLTKAERGGLTEADKA	917
tr H1GG18 H1GG18 LISIO	VLTSSAG <mark>NR</mark> EK-GDDVPPLEIVRKRKVFWEKLYQGNLMSKRKFDYLTKAERGGLTEADKA	920
tr R3WHR8 R3WHR8_9ENTE	VLVSSKE <mark>NR</mark> GK-KDDVPSKEVVQKNITLWETLKNSNLISQKKYDNLTKGLRGGLTEDDRA	918
na se-re la anerodo e en conserva eserva eserva eserva — 1987 1800/180	**. * : : : * * : : : * * : : :	

#### 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 I6T669II6T669 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 Q927P4ICAS9 LISIN CRISPR-associated endonuclease Cas9, Listeria innocua serovar H1GG18\_LISIO CRISPR-associated endonuclease Cas9, Listeria innocua R3WHR8\_9ENTE CRISPR-associated endonuclease Cas9, Enterococcus phoeniculicola

# **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

4. ACTIVE SITE REGIONS OF THE WELL 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-Ppol; 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.

Type Organism		Active site region
HH- Homing endonucleases (Bacteriophage Bp7, I-TevIII)		-YEIHHKDGNRENNDLDNLMCLSIQEHYDIHLAQKDY <sup>58-</sup>
HH- based group II introns (S. cerevisiae)		<sup>-</sup> QICGSKHDLEV <mark>HH</mark> VRTLNNAANKIKDDYLLGRMIKM <mark>NR</mark> KQITICKTCHF <sup>842</sup> -
HH- based mcrA restriction endonuclease (E. coli)		-CENCGKNAPFYLNDGNPYLEVHHVIPLSSGGADTTDNCVALCPNCHRELHYS <sup>258</sup> -
DH- based HNH endonucleases (E. coli plasmids)		GGRCAYCGCELPEKGWYADHVQAVLRKSEQCMKAAEKRIFRLK
		STGDVFRPEAD <mark>C</mark> PE <mark>NL</mark> VPACAPCNLLK <sup>85</sup>
DH- based HNH endonucleases (E. proavitum)		-GICQKCKTHFEINEMEADHITPWHEGGKTTSVNCQMLCKDCNRRK <sup>355</sup>
DH- based Endonuclease VII (Resolvase) (T4 Phage)		-GKCLICQRELNPDVQANHLDH <mark>DHE</mark> LNGPKAGKVRGLLCNL
		CNAAEGQMKHKFNR <sup>74</sup> -
HH based- Colicins endonucleases (Type 9*)	(E. coli)	-YEL <mark>HH</mark> DKPISQGGEVYDMD <mark>NI</mark> RVTTPKRHIDIHRGK <sup>592</sup>
HH- Pyocins endonucleases (Type-S1)	(P. aeruginosa)	-KIEIHHKVRVADGGGVYNMGNLVAVTPKRHIEIHKGGK <sup>618</sup>
HH- Pyocins endonucleases (Type-S2)	(P. aeruginosa)	-KIEI <mark>HH</mark> KVRIADGGGVYNMG <mark>NL</mark> VAVTPKRHIEI <mark>HK</mark> GGK <sup>689</sup>
DH- basedCRISPR-Cas9 HNH endonucleases (S. pyogenes)		-YVDQELDINRLS <b>DYD</b> VDH <sup>840</sup> IVPQSFLKDDSIDNKVLTRSDKNRGK <sup>865</sup>
DH- based CRISPR-Cas9 HNH endonucleases (A. naeslundii)		-ACLYCGTTIGYHTCQLDH <sup>582</sup> IVPQAGPGSNNRRGNLVAVCERCNR
	. ,	SKSNTPFAVWAQKCGIP <mark>HV</mark> G <sup>627</sup> -

## Table 1. Active site regions in different HNH endonucleases

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

- With DxD type Mg<sup>2+</sup> binding motif(s) using a Mg<sup>2+</sup> ion-bound water molecule as the nucleophile for initiating the catalysis, e.g., *S. pyogenes* CRISPR-Cas9 HNH endonuclease
- With CxxC or CxxxC type Zn binding motif(s) using a Zn<sup>2+</sup> ion-bound water molecule as the nucleophile for initiating the catalysis, e.g., bacteriocin type of HNH endonucleases
- With no well known conserved metalbinding 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 *Ppol* 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<sup>2+</sup> bound water is deprotonated by a general base and His<sup>98</sup> 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<sup>61</sup>, His<sup>98</sup>, and Asn<sup>119</sup> are important for efficient DNA cleavage by the enzyme. This finding was consistent with their proposed mechanism in which His<sup>98</sup> abstracts a proton from an attacking water molecule and Arg<sup>61</sup> and Asn<sup>119</sup> stabilize the pentavalent transition state. They also suggested that the Asn<sup>119</sup> also binds to the essential divalent metal cation, i.e., the Mg<sup>2+</sup> ion.

Type II restriction endonuclease R.KpnI, a member of the HNH endonuclease with the typical  $\beta$ - $\beta$ - $\alpha$ -Metal fold, was studied by SDM by Saravanan et al. [29]. They found that D<sup>148</sup>, H<sup>149</sup> and Q<sup>175</sup> in the active site of the enzyme. The mutant enzyme Q<sup>175</sup> $\rightarrow$ 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<sup>149</sup> $\rightarrow$ L mutant showed no detectable activity, even at 100-fold excess protein concentrations. Other mutants D<sup>148</sup> $\rightarrow$ G and Q<sup>175</sup> $\rightarrow$ 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 metalbinding, possibly an  $Mg^{2+}$  like other Type II restriction endonucleases.

The T<sub>4</sub> 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<sup>40</sup>, H<sup>41</sup> and N<sup>62</sup> play a similar role in catalysis as in R.Kpnl. 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  $(C^{23} \rightarrow S \text{ and } C^{61} \rightarrow S)$  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  $C^{23}$ ,  $C^{25}$ ,  $C^{58}$ and **C<sup>61</sup>** are essential for zinc binding, whereas the inner Cs are redundant. Furthermore, they also found that the mutant enzyme  $N^{62} \rightarrow D$  is inactive and therefore, suggested that the D<sup>40</sup> and N<sup>62</sup> 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<sup>2+</sup> binding motif, coordinated by the conserved 4Cs and also the DxD Mg<sup>2+</sup> binding motif (-HLD<sup>40</sup>HDHE- marked bold) in its structure. By X-ray crystallographic analysis, Raaijmakers et al. [30] have shown that D<sup>40</sup> and N<sup>62</sup> 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  $\beta\beta\alpha$ -metal motif was identified. When the proton acceptor His was subjected to SDM, the mutant enzyme H<sup>80</sup> $\rightarrow$ A did not show any activity. The conserved Asn<sup>127</sup> was found to be responsible for metal-binding (Mg<sup>2+</sup>) and is also well conserved among this group of endonucleases, except that it is replaced by a histidine (His<sup>569</sup>) in CoIE7.

It is interesting to note that Wy et al. [23] found in the crystal structure of the CoIE7 HNH

endonuclease a water molecule is bound to the Zn atom at the active site suggesting Zn<sup>2+</sup> 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 proofreading 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 metalbound 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  $-E^{77}xE^{79}H^{80}$  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<sup>127</sup> where they bind to a Mg<sup>2+</sup> ion. The Mg<sup>2+</sup> ion also binds 3 water molecules [32].

Moreover, a Mg<sup>2+</sup> ion in the HNH endonuclease I-*Ppol* 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  $H^{840} \rightarrow A$  did not show any activity and shown as the proton acceptor [17]. In the same way,  $N^{854} \rightarrow A$  showed activity but reduced activity [35,17] and the  $N^{863} \rightarrow A$  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,  $N^{62} \rightarrow D$  mutant in complex with a Holliday junction, a  $Mg^{2^+}$  ion is coordinated by Asp<sup>40</sup>, Asn<sup>62</sup>, 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<sup>839</sup>, His<sup>840</sup>, and Asn<sup>863</sup> corresponded to Asp<sup>40</sup>, His<sup>41</sup>, and Asn<sup>62</sup> of the T<sub>4</sub> Endonuclease VII, respectively, consistent with the observation that His<sup>840</sup> 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<sup>839</sup>, His<sup>840</sup> and Asn<sup>863</sup> are critical in substrate binding and DNA cleavage in SpyCRISPR-Cas9 HNH endonuclease.

Furthermore, Gasiunas et al. [16] have shown that Mg<sup>2+</sup> 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<sup>2+</sup> 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<sup>582</sup> 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 –CxxCtype 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

- 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.
- Deveau H, Garneau JE, Moineau S. CRISPR/Cas system and its role in phagebacteria 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.
- 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.
- 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.

- 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.
- 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.
- 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.
- 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
- 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.
- 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.
- 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.
- 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.

- 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.
- 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.
- 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-*Ppol*. 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.

- 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.
- 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.
- 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 sitedirected mutagenesis study to identify amino acid residues involved in the catalytic function of the restriction endonuclease *EcoRV*. 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.
- 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.
- 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.

© 2020 Palanivelu; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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