Natural zinc ribbon HNH endonucleases and engineered zinc finger nicking endonuclease

Shuang-yong Xu1,* and Yogesh K. Gupta2

1New England Biolabs, Inc., Research Department, 240 County Road, Ipswich, MA 01938 and 2Department of Structural and Chemical Biology, Mount Sinai School of Medicine, Box 1677, 1425 Madison Avenue, New York, NY 10029, USA

Received July 16, 2012; Revised October 4, 2012; Accepted October 8, 2012

ABSTRACT

Many bacteriophage and prophage genomes encode an HNH endonuclease (HNHE) next to their cohesive end site and terminase genes. The HNH catalytic domain contains the conserved catalytic residues His-Asn-His and a zinc-binding site [CxxC]2. An additional zinc ribbon (ZR) domain with one to two zinc-binding sites ([CxxxxC], [CxxxxH], [CxxxC], [HxxxxH], [CxxC] or [CxxH]) is frequently found at the N-terminus or C-terminus of the HNHE or a ZR domain protein (ZRP) located adjacent to the HNHE. We expressed and purified 10 such HNHEs and characterized their cleavage sites. These HNHEs are site-specific and strand-specific nicking endonucleases (NEase or nickase) with 3- to 7-bp specificities. A minimal HNH nicking domain of 76 amino acid residues was identified from Bacillus phage γ HNHE and subsequently fused to a zinc finger protein to generate a chimeric NEase with a new specificity (12–13 bp). The identification of a large pool of previously unknown natural NEases and engineered NEases provides more ‘tools’ for DNA manipulation and molecular diagnostics. The small modular HNH nicking domain can be used to generate rare NEases applicable to targeted genome editing. In addition, the engineered ZF nickase is useful for evaluation of off-target sites in vitro before performing cell-based gene modification.

INTRODUCTION

A large number of restriction endonucleases (REases) have been found in bacteria and viruses and are widely used in recombinant DNA technology (1–5). In contrast, relatively few natural nicking endonucleases (NEases or nickase), which introduce single-strand breaks in double-stranded DNA (dsDNA), have been discovered (3,6–8). NEases were engineered from type IIS REases by mutating one of the two catalytic sites or disrupting the dimerization domain (9–11), by alteration of binding specificity of a type IIP REase (12) or by combining a DNA binding-deficient and catalytic-proficient FokI subunit with a DNA binding-proficient and catalytically inert FokI subunit (13). Naturally occurring nicking enzymes may be found as part of heterodimeric REases or as stand-alone enzymes. One important group of natural NEases are replication initiation proteins that specifically nick the viral genome or conjugative plasmid during phage DNA replication or conjugative plasmid transfer such as gpII of f1 phage (14) and Salmonella typhimurium relaxase encoded by plasmid pCU1 (15). Another group of natural NEases are homing nicking endonucleases that play a role in gene conversion (insertion of intron and intron-encoded homing endonuclease into an intron-less allele) [reviewed in (16–18)]. Artificial zinc finger nuclease (ZFN)-based nicking enzymes (ZF nickases) have been constructed by fusion of three to four ZF arrays with FokI cleavage domains, one catalytic proficient (FokIR+) and the other catalytic deficient (FokIR−). Transient heterodimer formation by R+/R− cleavage domains upon binding to the target sites with a 5- to 6-bp spacer leads to DNA nicking which stimulates targeted gene disruption or gene addition. Such nick-induced gene targeting displays a low frequency of insertion or deletion, in sharp contrast to the non-homologous end-joining process that follows a dsDNA break and repair response (19–21). Thus, it is desirable to use ZF nickases to introduce nicks to initiate a DNA repair response that results in homology-directed recombination (HDR) for genome editing and targeted gene correction. In addition to the ZF nickases used in genome editing, a nicking variant was engineered from I-AniI homing endonuclease (HEase) and its utility was demonstrated in nick-induced gene correction via HDR in human cells (22). Furthermore, it was shown that an I-AniI nicking variant induces homologous recombination with both plasmid and adeno-associated virus (AAV) vector templates and the nicking variant alleviates the in vivo toxicity problem conferred by the dsDNA cleaving

*To whom correspondence should be addressed. Tel: +1 978 380 7287; Fax: +1 978 921 1350; Email: xus@neb.com

© The Author(s) 2012. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/3.0/), which permits non-commercial reuse, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com.
enzyme (23). A strand-specific nicking variant has also been isolated from I-SceI HEase (24). The inherent disadvantage of using HEases in genome editing is that most of the HEases studied so far tolerate degenerate sequences, and off-target sites with a few base pair mismatches are also cleaved (25,26). ZF recombinases (ZFRs) have also been constructed by fusion of ZF arrays with Tyr or Ser recombinase for ZFR-mediated integration of the donor plasmid in mammalian genomes and site-specific recombination in a bacterial genome (27,28).

The HNH superfamily nucleases include HEases, REases, structure-specific endonucleases, non-specific nucleases, CRISPR-associated protein Cas9 (CRISPR, clustered regularly interspaced short palindromic repeat) and DNA repair enzymes. The invariant His residue in the conserved motif HNH (sometimes found as HNK or HNN) serves as the general base that activates a water molecule for a nucleophilic attack on the sugar phosphate backbone of nucleic acids. We recently investigated the restriction–modification (R–M) systems in the sequenced genome Bacillus cereus ATCC 10978 (GenBank: AE017194) and found a prophage-encoded multi-specificity C5 methyltransferase (MTase) and a ParB-MTase fusion protein with non-specific DNA nicking activity (29). In the vicinity of the prophage-encoded C5 MTase, three genes were predicted to encode DNA-cleaving enzymes (genes in the order of Vrr_Nuclease, DNA Integrase, HNH endonuclease (HNHE), Terminase S subunit, ParB-MTase, C5 MTase). Here, we report the DNA nicking activity of the Bacillus cereus HNHE. In addition, we identified and characterized nine HNHE homologs and their nicking sites. We defined a minimal nicking domain of 76 amino acids (aa) from phage HEase (F5, 76 aa) found in B. cereus (Supplementary Figure S1A) was expressed in E. coli using the IMPACT protein expression system (Supplementary Figure S1B). After purification, the protein displayed low DNA nicking activity in Mg++ buffer and robust activity in Mn++ buffer (data not shown) and was subsequently named N.BceSVIII. It is partially active in reaction buffers with Co++ and is completely inhibited by addition of ethylenediaminetetraacetic acid (EDTA) (data not shown). Run-off DNA sequencing of the cleavage products indicates that N.BceSVIII cuts frequently with the recognition sequence of 5′ S↓RT 3′ (S = C/G, R = A/G) (Supplementary Figure S1C and D). We will use a downward pointing arrow to indicate nicking of the strand shown and an upward pointing arrow to indicate nicking of the complementary strand, following accepted nomenclature (32).

A BlastP search of the GenBank NR database using the N.BceSVIII aa sequence as a query revealed over 200 related HNHEs, all similarly small (95 to ~300 aa) and predominantly encoded by phage and prophage. These HNHE homologs can be divided into four major groups. Group 1 contains one to two zinc-binding sites with two to four zinc ribbons (ZRs) each with the motif [CxxxxC], [CxxxxH], [CxxxxX], [HxxxxH], [CxxH] or [CxC] at the N-terminus, and the HNH catalytic domain with one zinc-binding site with the motif [CxC] located at the C-terminus (each zinc ion is presumably coordinated by a tetrahedral geometry and thus requires two ZRs). Group 2 reverses the domain organization of the Group 1 enzymes. Group 2 lacks the N-terminal zinc-binding region of Group 1 and contains only one zinc-binding site in the HNH catalytic domain. A non-specific HNH (gp74) with the Group 3 domain architecture has been characterized recently from a λ-related phage HK97 (33). Group 4 HNHEs are similar to Groups 1 and 2, but located side-by-side with another ZR domain protein (ZRP) either immediate upstream or downstream. Some Group 4 enzymes are fused to the adjacent ZRP (see, e.g. Bth193 HNH below, which has four putative zinc-binding sites). Some putative HNHEs carry additional functional domains such as a...
Characterization of Bacillus anthracis phage γ HNH nickase

One HNHE homolog (accession YP_338236, 127 aa) from Group 1 was found in the genome of B. anthracis phage γ (accession NC_007458) (35), and which we will refer to as phage γ HNHE (N.φGamma or N. phiGamma). Its secondary structure as predicted by the Phyre server is shown in Supplementary Figure S2A (36). A synthetic gene was expressed in E. coli and the gene product purified by chromatography through two columns (Figure 1A). N.φGamma is active in the presence of Mg++ or Mn++, partially active in the presence of Ca++, Ni++ or Zn++ and inactive in the presence of EDTA (Figure 1B). Run-off sequencing of the cleavage products indicates that N.φGamma nicks predominantly at CG|GT sites in the presence of Mg++ (Figure 1C and D). A small amount of linear DNA was also detected in Mg++ buffer, most likely due to the nicking of two closely positioned sites on opposite strands. The sequence specificity is reduced to 2–3 bp recognition in Mn++ buffer as determined by run-off sequencing of the cleavage products (data not shown). Using an independent method, we digested bacteriophage λ DNA in the presence of Mn++ and cloned digestion products after performing a blunting reaction. Sequencing of the cloned fragments also revealed a relaxed specificity with nicking occurring at sites with one to two mismatches relative to the cognate site (data not shown).

Heat treatment of ~5 U of N.φGamma at 80 °C for 20 min completely abolished its activity, whereas treatment at lower temperatures (55–75 °C) resulted in only partial loss of nicking activity. Furthermore, treatment at 80 °C for only 10 min was not sufficient for complete inactivation (data not shown).

Mapping the minimally functional nicking domain in N.φGamma

Catalytic domains of several type IIS REases including FokI, BmrI and BpuJI when cloned and purified in absence of the DNA binding domains have been shown to display non-specific nuclease activity (37–39). To test whether the N.φGamma catalytic domain (the HNH domain) displays a similar non-specific activity in absence of the ZR domain, we deleted 66 aa (Δ66 aa) residues from the N-terminus of the wild-type (wt) enzyme, resulting in a mutant containing only the 61 C-terminal aa residues (F6). Additional deletion mutants were also constructed, removing 22 aa (F3), 43 aa (F4) and 51 aa (F5) from the N-terminus, respectively (Supplementary Figure S2A). The deletion variants were purified and their nicking activity was assayed in Mg++ or Mn++ buffer. Truncated mutants F3 and F4 displayed slightly lower activity in Mn++ buffers compared with the full-length enzyme. Mutant F5 showed nicking activity in Mn++ buffer and low activity in Mg++ buffer (Supplementary Figure S2B and C), and mutant F6 has minimal nicking activity in Mg++ or Mn++ buffers. The nicking specificities of F4 and F5 are nearly identical: both deletion variants nick the cognate sites (CG|GT) or variations thereof with one base mismatch (F5 ‘star’ sites CG|GG, CG|GC, CA|GT or CG|AT; data not shown).

Enzyme concentration dependence in nicking reactions, DNA nicking kinetics in a time course by N.φGamma and flanking sequence effect on nicking efficiency

Partially purified N.φGamma and a deletion variant F3 were used to nick pBR322 DNA. At low enzyme concentration, supercoiled DNA was converted to nicked circular form. At high enzyme concentration (2–4×), however, a small fraction of DNA was converted to linear form, probably generated from nicks introduced on the opposite strands at adjacent sites (Supplementary Figure S2D). Supplementary Figure S2E shows a time course in N.φGamma-mediated nicking reactions. The substrate DNA (pBR322) was incubated with diluted N.φGamma from 2 min to 2 h in a limited digestion. Supercoiled DNA was partially converted to nicked circular DNA in the time period. No linear DNA was detected due to the limited digestion. In order to examine the nicking efficiency of various ACCG sites in pBR322, the DNA substrate was incubated with purified N.φGamma at 37 °C for 2 h and then subjected to run-off sequencing. Supplementary Figure S2F shows a few examples of the nicked ACC|CG sites followed by A, G, C or T. Overall, the DNA sites ACCGR appeared to be nicked more efficiently than ACCGY (ACCGR > ACCGC > ACCGT) by comparing the ‘A’ peak in double peaks at the run-off sites. A quantitative nicking assay to evaluate the nicking efficiency of all possible flanking sequences (N’3N’2N’1-ACC-G-N1N2N3) remains to be developed.

Nicking short DNA duplex oligos by N.φGamma

DNA duplex oligos with one ACCGG site (28-mer) or two ACCGG sites (32-mer) were digested by N.φGamma at 37°C for 1 h in either Mg++ or Mn++ buffer and the cleavage products were resolved in a denaturing gel. Both substrates were nicked and new products were detected (data not shown). In support of this, 3’ FAM-labeled duplex oligos with one ACCGN site (N = A, G, C or T) were also nicked by N.φGamma. Supplementary Figure S2G shows the nicking results of ACCGA, ACCGG, ACCGC and ACCGT duplex oligos at four enzyme concentrations. Consistent with the nicking site efficiency in pBR322, the 28-bp short oligos with ACCGR sites were nicked more readily than ACCGY sites. The nicking reactions nearly reached completion at high enzyme concentration for ACCGR oligos, whereas partial nicking was achieved using the ACCGY.
substrates. Although the exact size of the nicked products remains to be determined by running a sequencing gel at the single nucleotide resolution, it suffices to say that short DNA duplex oligos can serve as the substrates for nicking reactions by N.\(\gamma\)Gamma.

Homologs of phage gamma HNHE (N.\(\varphi\)Gamma)

A BlastP search in GenBank revealed that Bacillus phage Cherry and WBeta genomes encode a putative HNHE with identical aa sequence to N.\(\varphi\)Gamma (data not shown). N.\(\varphi\)Gamma also shares 99% aa sequence identity to a putative HNHE (protein ID: EJR41403, 129 aa, probably encoded by a prophage belonging to HK97 family phage) in the genome of B. cereus VD045 strain, and 91% aa sequence identity to a putative HNHE (protein ID: YP_001375827, 127 aa, probably encoded by a prophage similar to P27 family phage, and located next to open reading frames (ORFs) coding for terminase small and large subunits) in the genome of Bacillus cytotoxicus NVH 391-98 strain. The next group of putative HNHEs from Geobacillus phage or prophage shares 51–56% aa sequence identity to that of N.\(\varphi\)Gamma. Because of the high aa sequence identity (similarity) among these HNHEs, it is highly likely that they are nicking enzymes with nicking site CG\(\downarrow\)GT (AC\(\uparrow\)CG) or minor variations of such a sequence. Consistent with this prediction, the HNHE encoded by Geobacillus virus E2 (N.\(\varphi\)E2), which shares 51% aa sequence identity to N.\(\varphi\)Gamma, displays the nicking specificity of CG\(\downarrow\)GT (data not shown) (Table 1).

Lactobacillus phage Sal2 HNH nicking endonuclease

An HNHE homolog was found in the Lactobacillus phage Sal2 with additional aa blocks located upstream and
downstream of the HNH catalytic domain (protein ID: YP_535176), which may confer additional base recognition (40). Supplementary Figure S3A shows that the two additional blocks of aa sequences (5-aa and 11-aa residues) at the N-terminus and one extra block of 32-aa residues located at the C-terminal end compared with phage N.\(\phi\)Gamma. Phage Sal2 HNHE (N.\(\phi\)Sal2 or N.\(\phi\)Sal2) was purified and used to digest pBR322. N.\(\phi\)Sal2 shows robust nicking activity in \(Mn^{++}\) buffer, which efficiently nicks DNA sites TG\(\rightarrow\)CTC (GAG\(\rightarrow\)GT) (Supplementary Figure S3B). The five nicking sites, which spans 14-bp DNA with a core recognition site of GAGCA DNW (SS\(\rightarrow\)GAGC\(\rightarrow\)DNW) (S, C; G, D, A/G/T, not C; W, A/T). The sixth base of the core is intolerant of a C because no GAGCA-related sites were nicked (data not shown). In summary, N.\(\phi\)Sal2 shows optimal nicking activity in \(Mn^{++}\) buffer, nicking at TG\(\rightarrow\)CTC (GAG\(\rightarrow\)CA) or variant sites with 1- to 2-bp mismatches.

### HNH Bth193 from of Bacillus thuringensis strain T13001

We identified another N.\(\phi\)Gamma homolog encoded on a prophage in the genome of B. thuringensis strain T13001 (protein ID: ZP_04118662). This enzyme, named as N.BthT13001I (a.k.a. Bth193 HNHE), has an additional N-terminal region relative to N.\(\phi\)Gamma containing four ZR motifs (Supplementary Figure S4A). N.BthT13001I was expressed in E. coli and partially purified. It nicks DNA sequences GG\(\rightarrow\)GT or CG\(\rightarrow\)GT in \(Mn^{++}\) buffer as determined by run-off sequencing (Supplementary Figure S4B and D, data not shown). The nicking sites in pBR322 were compiled by WebLogo as SG\(\rightarrow\)GT (Supplementary Figure S4B). In \(Mg^{++}\) buffer, however, N.BthT13001I recognizes and nicks DNA sites with a core sequence of CSG\(\rightarrow\)GT (AC\(\rightarrow\)CSG) (Supplementary Figure S4C). Based on the high sequence similarity with N.\(\phi\)Gamma, we tentatively concluded that the HNH catalytic domain recognizes the SGGT sequence, and the additional ZRs at the N-terminus might contribute to the extra base C recognition in CSGGT.

We also tested whether the HNH catalytic domain is interchangeable between N.BthT13001I and N.\(\phi\)Gamma. The N-terminus ZR of N.BthT13001I was fused with the 76-aa N.\(\phi\)Gamma catalytic domain (F5) to form a chimeric protein (the arrow in Supplementary Figure S4A indicates the fusion junction). The nicking sites of the fusion NEase are nearly identical to that of N.BthT13001I, but with preference for CCG\(\rightarrow\)GT (AC\(\rightarrow\)CGG) sequence, further confirming the \(\text{C}\) base recognition conferred by the extra ZRs (data not shown). We concluded that the HNH catalytic domain is readily exchangeable between Bth193 and N.\(\phi\)Gamma and the chimeric enzyme N-N.BthT13001I::C-N.\(\phi\)Gamma (F5) prefers CCG\(\rightarrow\)GT sites for nicking.

### HNH nicking endonuclease gp54 from Lactobacillus phage Lrm1 and other HNHEs

In the sequenced genome of Lactobacillus phage Lrm1 (phage genome accession number: EU246945) (41), the last gene (gp54, protein ID: ABY84355) was annotated as a terminase \(S\) subunit. However, two other genes encoding terminase small and large subunits are located at the left arm of the phage genome (Supplementary Figure S5A). This predicted protein (gp54, 264-aa long) contains a HNH catalytic domain near the N-terminus (H\(\rightarrow\)77-N\(\rightarrow\)93-H\(\rightarrow\)102) and a AAA-NTPase
domain (also termed Walker A motif GxxxxGK[S/T]), but lacking the Walker B motif: h4[D/E] (h4 = four hydrophobic aa residues) at the C-terminus. In addition, the C-terminus contains a putative zinc-binding motif (2xZR [HxxxH], [RxxxH] or [HxxxD]; and seven other Lactobacillus phage encode close homologs carrying the 2xZR motif [HxxxH] (data not shown). Phage Lrm1 HNHE (N.φLrm1 or N.phILrm1) was expressed in E. coli, partially purified by affinity chromatography through a chitin column, and used to digest pBR322 DNA; It displays DNA nicking activity in Mn++ buffer and low activity in Mg++ buffer (Supplementary Figure S5B and C). Addition of NTP (ATP, GTP, CTP or UTP) failed to stimulate N.φLrm1 nicking activity (data not shown). The partially nicked DNA was gel-purified and subjected to run-off sequencing. Supplementary Figure S5D summarizes the nicking sites as HSSG↓GT (AC↑CSSD), which resembled the nicking sites of Bth193 CSG↓GT (AC↑CSG). We do not know the function of the AAA-superfamily domain (Walker A motif) at the C-terminus of N.φLrm1 as neither NTP nor dNTP have any stimulatory effect on DNA nicking activity. This domain could serve as an accessory domain interacting with DNA or other proteins. We have not studied the interaction of this HNHE with the phage Lrm1 terminase.

Figure 2. Characterization of phage Sal2 HNHE (N.φSal2). (A) Run-off sequencing of two nicked sites TG↓CTC (GAG↑CA) in pBR322 which was digested by N.φSal2 in Mn++ buffer. (B) Nicking sites in pBR322 (Mg++ buffer) compiled by WebLogo.
We also purified the HNHEs from *Bacillus* phage phi105 (protein ID: ADF59184), *Geobacillus* virus E2 HNHE (GBVE2 gp070, protein ID: YP_001522898) (42), *Clostridium* phage phi3626, gp50 (protein ID: NP_612879), *Staphylococcus aureus* strain Y74T prophage, ORF Sap040a_009 (protein ID: ACZ58991) (data not shown). The nicking sites of these HNHEs are summarized in Table 1.

### Construction of Zif268 and N. φGamma (F5, 76 aa) fusion nicking endonuclease

A large number of natural ZFPs have been identified in the past 25 years consisting primarily of eukaryotic transcription factors (see ZF database) (43), and additional ZFPs with high DNA binding affinity have been engineered by genetic selection or phage display (44–47). Thus, ZFPs provide a collection of DNA binding modules (arrays) ranging from 6 to 12 bp with two to four individual ZFs linked together. ZFNs have been engineered to cleave large recognition sequences 12–24 bp (48–52). Individual ZFP forms a DNA recognition module with two antiparallel β-sheets and one α-helix (ββzZn), where the second β-sheet contacts the sugar phosphate backbone and the α-helix contacts the base pairs in the DNA major groove as revealed by the structure of the Zif268–DNA complex (53).

We next attempted the fusion of the minimal N. φGamma nicking domain (F5, 76 aa) to Zif268 (3 ZFs, 117 aa) (partial ZFP sequence of North American songbird *Vireo cassinii*, GenBank accession: protein ID: AA084760, 210 aa) to generate the chimeric enzyme Zif268III::N. φGamma F5 (designated as ZifIII for short). Two fusion proteins were constructed, R1 fusion with 13-aa linker and R2 fusion with 3-aa linker between the two functional domains (Figure 3A). The two fusion proteins were purified to near homogeneity by chromatography (Figure 3B) and used to digest pUC19 (Figure 3C) and pUC19-Zif3 (containing a Zif268 recognition sequence GCGGGGCGG). Run-off sequencing of the nicked products indicated no apparent nicking taking place within or near the GCGGGGCGG sequence (data not shown). However, a clear nicking site was detected at ATCG site is indicated by a blue line in Figure 3F, data not shown). Nicking reactions performed in high salt buffer (B3) could minimize the promiscuous nicking activity as some smearing was detected in prolonged incubation in B1, B2 and B4 (data not shown). Double-stranded breaks could be minimized by shorter digestion (~30 min) and in 150 mM NaCl (~80% supercoiled DNA converted to nicked circular in 2 h in 150 mM NaCl with minimal dsDNA breaks, data not shown). These results indicate that a ZF nickase can be constructed from a ZFP (three fingers) and the N. φGamma minimal nicking domain to create a chimera capable of nicking a 12- to 13-bp target site. Addition of two or more ZFs may further increase the nicking specificity, which is a prerequisite for applications in genome editing.

### Off-target nicking sites by Zif268::N. φGamma F5

To investigate potential ‘star’ sites of ZifIII R1, λ DNA was nicked by the chimeric NEase and the digested DNA was subjected to run-off sequencing at the suspected ‘star’ sites (9- to 11-bp matches). One strong off-target nicking site was found with the sequence AC†TG-N6-GCCGTTGAG (CA†GT, 2-bp matches in each ZF binding site, λ coordinate 3981, Supplementary Figure S6A). Two ‘star’ sites with the sequence AC†CC-N6-GCCCGGTTT (GG†GT; 3-bp, 2-bp, 1-bp matches to ZF3, ZF2 and ZF1 sites, respectively; λ coordinate 4486) and ACTG-N6-GCGGGGGTA (CA†GT; 3-bp, 3-bp, 1-bp matches in ZF3, ZF2 and ZF1 sites, respectively; λ coordinate 11111) were partially nicked by the ZifIII R1 NEase (Supplementary Figure S6B, data not shown). No apparent nicking was found in the suspected ‘star’ sites listed in Supplementary Figure S6C, which contains 5- to 7-bp matches to the Zif268 binding site (data not shown). These off-target sites indicate that 2-bp matches in each ZF are likely strong ‘star’ sites. The 5- to 6-bp matches in ZF3 and ZF2 in conjunction with 1-bp match in ZF1 are probable ‘star’ sites.

### DISCUSSION

The minimal HNH catalytic domain with its own sequence specificity

The HNH motif is highly conserved among the HNHEs, although in some cases the second His residue can be substituted by a Lys or Asn residue (H-N-K/N). The first conserved His residue acts as a general base to activate a water molecule for nucleophilic attack of the DNA phosphodiester bond, and the Asn orients the catalytic His residue and scissile phosphate in the correct position for DNA hydrolysis (54,55). The second His (or Lys/Asn) stabilizes the leaving group. HNH domains, which generally fold into a secondary structure described as ‘ββz–metal’, bind the DNA backbone in the minor
Figure 3. Characterization of Zif268::N.\textalpha Gamma F5 chimeric nickase (ZifIII). (A) Amino acid sequence of the fusion HNHE (Zif268 sequence shown in blue). The sequence was derived from partial ZFP of \textit{Vireo cassini} (GenBank: gb AAO84760). Two fusion constructs were made, R1 with a 13-aa linker (7 aa: DKKAEKA from the native songbird Zif268) and R2 with 3-aa linker between the two functional domains. (B) SDS–PAGE analysis of purified ZifIII R1 and R2 HNHEs. Lanes 2–5, 6 and 7, purified R1 and R2 fusion proteins from chitin columns. Lanes 8–11, purified R1 fusion protein from a heparin column (HiTrap Heparin HP, 5 ml). (C) Nicked pUC19 DNA in a time course (lanes 2 and 6, 20 min to 2 h). Lanes 7 and 8, Nt.BspQI nicked and SphI-digested DNA, respectively. (D) Digestion of pUC-ATCG-N6-Zif3(GCGTGGGCG) and pUC-ATCG-N6-Zif3star (GCGCGGGGA) in a time course (20 min to 2 h) in Buffer 3 (100 mM NaCl). Small amount of linear DNA (<5% of total DNA) appeared after 20 min digestion. Note: there is a pre-existing ‘star’ site in pUC19 in addition to the cognate site or ‘star’ site inserted at the multiple cloning sites. (E) Digestion of pUC-ATCG-N6-Zif3(GCGTGGGCG) by varying concentration of ZifIII R1 in NEB Buffer 3 for 30 min. The Nt.BspQI-nicked DNA was used as a marker. The slow migrating DNA in (E) and (F) is probably nicked dimer. (F) Run-off sequencing of nicked DNA (B3, 1 h, 37°C) with forward and reverse primers. Red line, Zif268 cognate site; green line, Zif268 ‘star’ site; blue line, N.\textalpha Gamma site. Arrows indicate the nicking sites where double peaks appeared. SC, supercoiled; L, linear; NC, nicked circular DNA.
groove, whereas additional residues mediate sequence-specific interactions with the DNA (56). The zinc metal ion is coordinated by a tetrahedral coordinating geometry by the two ZRs [CxxC]$_2$ or [CxxH, CxxC] (55,57,58).

Secondary structure prediction by the Phyre server (59) revealed that the N.omega Gamma catalytic domain (F6, 66 aa) contains a $\beta$-$\beta$-$\alpha$-$\alpha$ structure, as predicted from known structures of a HNH from Geobacter metallireducens GS-15 (a HNH without the extra ZR domain, pdb accession: 2qgpB, Northeast Structural Genomics Consortium target GmR87) (A. P. Kuzin et al., unpublished results) and Pacl (a HNH family REase) (57). The active deletion variant F5 (76 aa) contains an extra $\alpha$-helical structure that may be involved in increased DNA-binding affinity or dimerization. It is somewhat unexpected that the deletion of the N-terminal 51-aa ZR domain does not totally abolish the specificity of N.omega Gamma. The deletion only weakened the activity in Mg$^{++}$ buffer and relaxed the sequence recognition. The 76-aa minimal nicking domain has been used in construction of fusion protein with ZFP. The ‘star’ sites of N.omega Gamma in Mn$^{++}$ are usually one base different from the cognate site AC$^+$CG, e.g. AT$^+$CG, CC$^+$CG, AC$^+$TG or GC$^+$CG. The off-target sites of Zif268::N.omega Gamma in lambda DNA, however, appear to be AY$^+$YS (e.g. AC$^+$CC, AC$^+$TG, AT$^+$CG). This limited spectrum of ‘star’ sites could be result of few ‘star’ sites interrogated or lower ‘star’ activity in the high salt buffer.

We constructed two versions of Zif268::N.omega Gamma fusion, R1 with 13-aa linker and R2 with 3-aa linker between the ZFs and N.omega Gamma catalytic domains. R1 is more active than R2 in DNA nicking and both require a N$_6$ DNA spacer between the ZFP binding site and the N.omega Gamma target sequence. More Zif268 and N.omega Gamma fusions with variable linker length ranging from 4 to 20 aa are probably necessary to pinpoint the optimal spacing between the two functional domains. The optimal linker between ZF arrays and FokI cleavage domain appears to be 4-8 aa residues for efficient cleavage of target sites with a 6-bp spacer by ZFNs (60). To better understand how N.omega Gamma is likely positioned in the chimeric structure, we built three-dimensional homology-based models of DNA-bound and the apo Zif268::N.omega Gamma (ZifIII R1). As shown in Supplementary Figure S7, this model (monomer) adopts an elongated shape and covers roughly two turns of the DNA helix.

The ZF nickase by fusion of a ZF arrays and a FokI cleavage domain requires dimerization in order to nick the target sites, i.e. formation of a heterodimer by FokI catalytic-proficient (R$^+$) and catalytic-deficient (R$^-$) monomers. Similarly, TALE-FokI nickase (to be constructed and functionally tested) may also require dimer formation (TALE-FokIR$^+$ and TALE-FokIR$^-$) in order to nick target sites. The advantages of using a HNH nicking domain are: (i) a single molecule construction, and thus the cost of cloning, expression and purification may be cut in half, (ii) the wide selection of natural HNH nicking domains existing among the phage and prophage-encoded RNases and DNA/RNA metabolic enzymes and transcription factors (62–65). The ZF found in McrA endonuclease, a member of the HNH superfamily, had been studied previously by computer modeling and structure prediction (34).

Metal ion requirement of HNHEs

A few HNHEs are only active in Mn$^{++}$ or Co$^{++}$ buffers, including N.BceSVIII, Lactobacillus phage Lrm1 gp54 (N.omega Lrm1), S. aureus prophage ORF Sap040a_009 HNH. N.BceSVIII and phage N.omega Gamma are active in a range of 1–10 mM Mn$^{++}$ tested. Gp54 of Lactobacillus phage Lrm1 (N.omega Lrm1), however, prefers a low concentration of Mn$^{++}$ (1 mM) in nicking. Thus, the optimal Mn$^{++}$ concentration supporting maximum endonuclease activity depends on the individual enzyme. The non-specific HNHE encoded by phage HK97 contains the zinc-binding motif CxxC and CxxH in the catalytic domain. No additional zinc-binding sites (ZRs) were found at the N-terminus. HK97 gp74 HNH prefers Ni$^{++}$ as a co-factor, although lower activity was also detected in other divalent cations (33,66). HK97 gp74 HNH also nicks plasmid DNA, but the sequence specificity of nicking sites (if any) has not been yet determined. Previously, we found HpyAV (CCTTC N$_8$/N$_5$), a member of the HNH family REase also prefers Ni$^{++}$ for optimal endonuclease activity (67). Interestingly, the modification-dependent endonuclease Sco McrA, an HNH family enzyme, prefers Mn$^{++}$ or Co$^{++}$ in cleaving M.Dcm-modified DNA (68).

In N.omega Gamma, the conserved ‘$\beta$-$\beta$-$\alpha$-metal’ fold is expected to display a tetrahedral coordination of a Zn$^{++}$ ion for structural folding and an Mg$^{++}$/Mn$^{++}$ divalent cation for catalytic function. The magnesium concentration is estimated to be ~100 mM in actively growing E. coli cells, although most of the Mg$^{++}$ is in a bound state to nucleic acids. The Mn$^{++}$ and Zn$^{++}$ ions in E. coli are estimated at much lower concentration, at ~0.2–0.4 mM for Mn$^{++}$ and 0.1 mM for Zn$^{++}$ ions (cells grown in LB broth) (Bionumbers database at the web site: http://bionumbers.hms.harvard.edu). Based on this estimation of metal ion concentration, it is speculated that Mg$^{++}$ is more readily available for the N.omega Gamma
catalytic activity than Mn$^{2+}$, which may explain the observation that most of the phage/prophage-encoded HNH described in this work are not toxic or lethal to E. coli, even though some of the HNHs have frequent nicking sites, and in the absence of companion methylase protection (or prophage gene expression may be tightly repressed). Previously, we expressed Nt.CviPII (CCD, a frequent nickase encoded by a chlorella virus NYs-1) in E. coli (Nt.CviPII requires Mg$^{2+}$ for catalytic activity and belongs to the PDExK family endonucleases). The over-expression of Nt.CviPII requires the presence of a cognate methylase for host DNA protection (7).

**Off-target sites of Zif268::N.$\varphi$Gamma F5 fusion NEase**

The N.$\varphi$Gamma catalytic domain F5 has its own specificity (ACCC, ACCC, ACTG or ATCG) and the Zif268 has 9-bp specificity (GCGTGGGCG or GCGGGGGCG). The fusion enzyme has the combined sequence specificities with a 6-bp spacer. Although we have not analyzed many ‘star’ sites of Zif268::N.$\varphi$Gamma F5 chimeras, the following observation was made in the limited number of off-target sites in λ and pUC19 DNA. For a strong ‘star’ site, 2-bp matches in each finger (6-bp matches total) are better than 6-bp matches in only two fingers and 0-bp match in one finger (6-bp matches total). The 3-bp matches inZF3, 2-bp matches inZF2 and 1-bp match inZF1 (6-bp matches total) appear to be a ‘star’ site (e.g. AC$^\uparrow$CG-N$_6$-GCGSGGGNN, ACCG can be substituted by ACCC, ATCG or ACTG). It is possible that ZF2 and ZF3 are more important than ZF1 in determination of the nicking specificity due to the proximity to the N.$\varphi$Gamma nicking domain or higher binding affinity to DNA. Sequencing more ZifIII R1 ‘star’ sites is necessary to confirm this observation.

**Fusion of N.$\varphi$Gamma F5 nicking domain to other DNA binding elements**

In theory, the N.$\varphi$Gamma minimal nicking domain can be fused to other DNA binding proteins such as a cleavage-deficient NotI REase. Because NotI forms a dimer, the NotI (D160N)::N.$\varphi$Gamma F5 fusion may nick the symmetric site AC$^\uparrow$CG-N$_6$-GCGSGGGCG-N$_6$ CG$^\downarrow$GT (estimated $x$ = 6–12 bp). The N.$\varphi$Gamma nicking domain may also be fused to a TALE effector protein, which recognizes 14-to 24-bp recognition sequences. The resulting fusion protein is expected to nick DNA at rare sites. A number of labs have developed methods for the modular construction of custom-designed TALE effector proteins and TALE nucleases (69–76). The optimal linker between TALE effector protein and the N.$\varphi$Gamma nicking domain needs to be tested experimentally in future work. The homeodeomain fold commonly found in eukaryotic transcription factors consists of a 60-aa helix–turn–helix structure that binds to DNA or RNA. A large number of new specificities of homeodeomain have been selected that can serve as the target-recognizing domain (TRD) of the fusion nicking enzyme (77). In addition, the N.$\varphi$Gamma nicking domain can be conjugated to LNA (locked nucleic acid) which binds to dsDNA by triple-helix formation. Similarly, the N.$\varphi$Gamma nicking domain can be coupled to 5mC-recognition domain such as MBD (methyl-binding domain protein) (78) or the 5mC specificity domain of MspII-family REases, McRb, McRA, SauUSI or SRA protein, generating 5mC-specific nickases.

**Possible biological function of phage and prophage-encoded HNH**

Table 2 shows the gene organization of the phage/prophage-encoded HNH, adjacent ZR protein, terminase small and large subunits and phage portal protein. The biological function of these phage or prophage-encoded HNH NEases is unknown. Because these NEases are located near the DNA packaging enzyme terminase, we speculate that these NEases may play a role in DNA packaging, e.g. by relieving the supercoiled tension built up during DNA translocation into the phage prohead. The nicks may stimulate

---

**Table 2. Genes (ORFs) in proximity to the phage- or prophage-encoded HNHs**

| Phage or prophage | ZR protein$^a$ | HNH | ZR protein$^a$ | Terminase S | Terminase L | Portal protein |
|-------------------|----------------|-----|----------------|-------------|-------------|---------------|
| *B. cereus* ATCC 10978 | γLSU_0048 | Bce_0389 | γLSU_0050 | Bce_0390 | _0397 | _0398 |
| *Bacillus* phage γ | phi105_00255 | phi105_00260 | phi105_00055 | phi105_00060 | _00010 | _00020 |
| *Bacillus* phage 105 | phi105_00255 | phi105_00260 | phi105_00055 | phi105_00060 | _00010 | _00020 |
| *Bacillus subtilis* subspecies spizizenii W23 | phi105_00255 | phi105_00260 | phi105_00055 | phi105_00060 | _00010 | _00020 |
| *B. thuringensis* str T13001 (Bth193) | unknown | Bth0005_53510 | known | Bth0005_53510 | _53540 | unknown |
| *Clostridium* phage phi3626 | phi3626_p49 | phi3626_p50 | phi3626_p60 | phi3626_p70 | _09630 | _09640 |
| *Geobacillus* virus E2 | phi105_00255 | phi105_00260 | phi105_00055 | phi105_00060 | _00010 | _00020 |
| *Lactobacillus* phage Sal2 | phi105_00255 | phi105_00260 | phi105_00055 | phi105_00060 | _00010 | _00020 |
| *Lactobacillus* phage Lmr1 | phi105_00255 | phi105_00260 | phi105_00055 | phi105_00060 | _00010 | _00020 |
| *Staphylococcus aureus* Y74T | phi105_00255 | phi105_00260 | phi105_00055 | phi105_00060 | _00010 | _00020 |
| Prophage Sap040a_009 | phi105_00255 | phi105_00260 | phi105_00055 | phi105_00060 | _00010 | _00020 |
| Phage HK97 (AF069529) | phi105_00255 | phi105_00260 | phi105_00055 | phi105_00060 | _00010 | _00020 |

$^a$The ZR motifs of the ZR proteins located either upstream or downstream of the HNHs contain the amino acid sequence: CxxC, CxxxC, HxxC, HxxxC, or CXXXCH.

$^b$There are two HNHs in the shot-gun genome sequences of *B. thuringensis* str T13001. The N.Bth13001H (Bth193) activity has been verified. The nicking activity of the shorter HNH (67 aa) and the 110-aa ZR protein is unknown.

$^c$The gp74 HNH encoded by phage HK97 contains the zinc-binding motif CxxC and CxxH. Gp74 has nicking activity on plasmids (in Ni$^{2+}$ buffer).
homologous recombination and enhance gene conversion of HNHE-minus incoming phage to HNHE-plus phage, which may also occur during mixed infection. HEases make dsDNA breaks in alleles that are homologous to the endonuclease gene but lack the intron or intein element, which encodes the gene for the HEase. One example of HNH NEase-stimulated trans-homing was reported previously: an HNH nicking enzyme (mobE) encoded by T4 phage introduces a strand-specific nick in the non-coding strand of the nrdB gene of a related phage T2. MobE promotes the mobility of the neighboring non-functional homing endonuclease I-TevIII, which was encoded within a Group I intron interrupting the gene of phage T4 (79). In this case, the nicking enzyme functions as a ‘helper’ HEase. The HNH nicking enzymes may play a role in the lysogenic life cycle of phage. It was proposed that phage HK97 HNHE, gp74 generated dsDNA breaks (or ssDNA nicks in certain metal ions), thereby initiating bacterial SOS repair response that allows for homologous recombination at the cleavage site to occur, resulting in integration of the phage genome (66). Another possibility yet to be ruled out is introduction of dsDNA breaks by HNH nicking endonuclease in conjunction with another nickase such as ParB homologs or Vrr_nuc domain endonucleases, which together function as restriction enzymes to restrict invading DNA.

We have not studied the function of the ZRP located next to the HNHE except for the Bth193 HNHE (N.BthT13001I), which is a fusion of ZRP to HNHE with four putative zinc-binding sites (8xZRs). The adjacent ZRP could serve as the transcription regulator of the HNHE gene or as a mini-specificity subunit that could interact with the HNHE and further extend the sequence specificity. Experiments are in progress to express and purify these small ZRPs to elucidate their function and to construct chimeric fusion NEases from existing HNHEs by domain swapping.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online: Supplementary Methods and Supplementary Figures 1–7.

ACKNOWLEDGEMENTS
We thank Richard Roberts, Bill Jack, Siu-Hong Chan, Jim Samuelson, Ted Davis and Brian Anton for critical reading of the manuscript; Nancy Badger and Barton Slatko in DNA sequencing core lab for sequencing; Siu-Hong Chan for running the molecular size exclusion column; Rebecca Nugent and Geoff Wilson for discussion; Don Comb for general interest in engineering genome-editing reagents; Jim Ellard for support.

FUNDING
Funding for open access charge: New England Biolabs, Inc.

Conflict of interest statement. Some nicking enzymes described in this paper may become commercial products.

REFERENCES
1. Smith, H.O. (1974) Restriction endonuclease from Hemophilus influenzae Rd. Methods Mol. Biol., 7, 71–85.
2. Wilson, G.G. and Murray, N.E. (1991) Restriction and modification systems. Ann. Rev. Genet., 25, 585–627.
3. Van Etten, J.L., Xia, Y., Burbank, D.E. and Narva, K.E. (1988) Chloraure viruses code for restriction and modification enzymes. Gene, 74, 113–115.
4. Pingoud, A., Alves, J. and Geiger, R. (1993) Restriction enzymes. Methods Mol. Biol., 16, 107–200.
5. Roberts, R.J., Vincze, T., Posfai, J. and Maceil, D. (2010) REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res., 38, D234–D236.
6. Xu, S.Y., Zhu, Z., Zhang, P., Chan, S.H., Samuelson, J.C., Xiao, J., Ingalls, D. and Wilson, G.G. (2007) Discovery of natural nicking endonucleases Nb.BsrDI and Nb.BtsI and engineering of top-strand nicking variants from BsrDI and BtsI. Nucleic Acids Res., 35, 4608–4618.
7. Chan, S.H., Zhu, Z., Van Etten, J.L. and Xu, S.Y. (2004) Cloning of CviP11 nicking and modification system from cholera virus NYS1 and application of Nt.CviP11 in random DNA amplification. Nucleic Acids Res., 32, 6187–6199.
8. Kachalova, G.S., Rogulina, E.A., Yunusova, A.K., Artyukh, R.I., Perevyazova, T.A., Matvienko, N.I., Zhelemaya, L.A. and Bartunik, H.D. (2008) Structural analysis of the heterodimeric type III restriction endonuclease R.BspD6I acting as a complex between a monomeric site-specific nickase and a catalytic subunit. J. Mol. Biol., 384, 498–502.
9. Xu, Y., Lunnen, K.D. and Kong, H. (2001) Engineering a nicking endonuclease N.Awil by domain swapping. Proc. Natl Acad. Sci. USA, 98, 12980–12985.
10. Zhang, P., Too, P.H., Samuelson, J.C., Chan, S.H., Vincze, T., Doucette, S., Backstrom, S., Potamouis, K.D., Schramm, T.M., Forrest, D. et al. (2010) Engineering BspQI nicking enzymes and application of N.BspQI in DNA labeling and production of single-strand DNA. Protein expression and purification, 69, 226–234.
11. Heiter, D.F., Lunnen, K.D. and Wilson, G.G. (2005) Site-specific DNA-nicking mutants of the heterodimeric restriction endonuclease R.BbvCI. J. Mol. Biol., 348, 631–640.
12. Kostiuk, G., Sasnauskas, G., Tamulaitiene, G. and Siksnys, V. (2011) Degenerate sequence recognition by the monomeric restriction enzyme: single mutation converts BcnI into a strand-specific nicking endonuclease. Nucleic Acids Res., 39, 3744–3753.
13. Sanders, K.L., Catto, L.E., Bellamy, S.R. and Halford, S.E. (2009) Targeting individual subunits of the FokI restriction endonuclease to specific DNA strands. Nucleic Acids Res., 37, 2105–2115.
14. Greenstein, D. and Horiiuchi, K. (1989) Double-strand cleavage and strand joining by the replication initiator protein of filamentous phage F1. J. Biol. Chem., 264, 12627–12632.
15. Nash, R.P., Niblock, F.C. and Redinbo, M.R. (2011) Tyrosine partners coordinate DNA nicking by the Salmonella typhimurium plasmid pCU1 relaxase enzyme. FEMS Lett., 585, 1216–1222.
16. Chan, S.H., Stoddard, B.L. and Xu, S.Y. (2011) Natural and engineered nicking endonucleases—from cleavage mechanism to engineering of strand-specificity. Nucleic Acids Res., 39, 1–18.
17. Taylor, G.K. and Stoddard, B.L. (2012) Structural, functional and evolutionary relationships between homing endonucleases and proteins from their host organisms. Nucleic Acids Res., 40, 5189–5200.
18. Landthaler, M., Lau, N.C. and Shub, D.A. (2004) Group I intron homing in Bacillus phages SPOI and SP2: a gene conversion event initiated by a nicking homing endonuclease. J. Bacteriol., 186, 4307–4314.
19. Wang, J., Friedman, G., Doyon, Y., Wang, N.S., Li, C.J., Miller, J.C., Hua, K.L., Yan, J.J., Babiarz, J.E., Gregory, P.D. et al. (2012) Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. Genome Res., 22, 1316–1326.
20. Ramirez, C.L., Certo, M.T., Mussolinio, C., Goodwin, M.J., Crudick, T.J., McCaffrey, A.P., Cathomen, T., Scharenberg, A.M. and Jonug, J.K. (2012) Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Res.*, 40, 5560–5568.

21. Kim, E., Kim, S., Kim, D.H., Choi, B.S., Choi, I.Y. and Kim, J.S. (2010) Precision genome engineering with programmable DNA-nicking domains. *Gene*, 459, 1327–1333.

22. Metzger, M.J., McConnell-Smith, A., Stoddard, B.L. and Miller, A.D. (2011) Single-strand nicks induce homologous recombination with less toxicity than double-strand breaks using an AAV vector template. *Nucleic Acids Res.*, 39, 926–935.

23. Xu, S.Y., Nugent, R.L., Kasamkattil, J., Fomenkov, A., Gupta, Y., Maizels, N., Monnat, R.J. Jr and Stoddard, B.L. (2009) Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniL LAGLIDADG homing endonuclease. *Proc. Natl Acad. Sci. USA*, 106, 5099–5104.

24. Roy, A., Kucukural, A. and Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.*, 5, 839–866.

25. Gersbach, C.A., Gaj, T., Gordley, R.M., Mercer, A.C. and Barbas, C.F. 3rd (2011) Targeted plasmid integration into the human genome by an engineered zinc-finger recombinase. *Nucleic Acids Res.*, 39, 7868–7878.

26. Prorocie, M.M., Wenlong, D., Olorunniyi, F.J., Akopian, A., Schleotl, J.G., Hennigan, A., McPherson, A.L. and Stark, W.M. (2011) Zinc-finger recombinase activities in vitro. *Nucleic Acids Res.*, 39, 9316–9328.

27. Xu, S.Y., Nugent, R.L., Kasamkattil, J., Fomenkov, A., Gupta, Y., Aggarwal, A., Wang, X., Li, Z., Zheng, Y. and Morgan, R. (2012) Characterization of type II and III restriction-modification systems from Bacillus cereus strains ATCC 10987 and ATCC 14579. *J. Bacteriol.*, 194, 49–60.

28. Roy, A., Kucukural, A. and Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.*, 5, 725–738.

29. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, 66, 486–501.

30. Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., Blumenthal, R.M., Degtarev, S., Dryden, D.T., Dyvib, K. et al. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.*, 31, 1805–1812.

31. Moodley, S., Maxwell, K.L. and Kanelis, V. (2012) The protein gp74 from the bacteriophage HK97 functions as a HNH endonuclease. *Protein Sci.*, 21, 809–821.

32. Bujnicki, J.M., Radlinska, M. and Rychlewski, L. (2000) Atomic model of the 5-methylcytosine-specific restriction enzyme McrA reveals an atypical zinc finger and structural similarity to betabetaalphaMe endonucleases. *Mol. Microbiol.*, 37, 1280–1281.

33. Fouts, D.E., Rasko, D.A., Cer, R.Z., Jung, L., Fedorova, N.B., Shvartsbeyn, A., Vamathevan, J.J., Tallon, L., Althoff, R., Porteus, M.H. Jr, Dobbs, D., Voytas, D.F. et al. (2006) Standardized reagents and protocols for engineering zinc-finger nucleases by modular assembly. *Nat. Protoc.*, 1, 1637–1652.

34. Carroll, D. (2011) Genome engineering with zinc-finger nucleases. *Genetics*, 188, 773–782.

35. Pavletich, N.P. and Babo, C.O. (1991) Zinc-finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science*, 252, 809–817.

36. Huang, H. and Yuan, H.S. (2007) The conserved asparagine in the HNH motif serves an important structural role in metal finger endonucleases. *J. Mol. Biol.*, 368, 821–827.

37. Kleanthous, C., Kuehlmann, U.C., Pomer, A.J., Ferguson, N., Radford, S.E., Moor, G.R. and Hemmings, A.M. (1999) Structural and mechanistic basis of immunity toward endonuclease colicins. *Nat. Struct. Biol.*, 6, 243–252.

38. Kuhn, M., Moor, G.R., James, R., Kleanthous, C. and Hemmings, A.M. (1999) Structural parsimony in endonuclease active sites: should the number of homing endonuclease families be redefined? *FEBS Lett.*, 463, 1–2.

39. Hegde, S., Heiter, D.F., Chan, S.H., Wang, H., Xu, S.Y., Morgan, R.D., Wilson, G.G. and Stoddard, B.L. (2010) Unusual target site disruption by the rare-cutting HHN restriction endonuclease Psn. *Structure*, 18, 734–745.

40. Sokolowska, M., Czapinska, H. and Bochtler, M. (2009) Crystal structure of the beta beta-alpha-Me type II restriction endonuclease Hpy99I with target DNA. *Nucleic Acids Res.*, 37, 3799–3810.
59. Kelley, L.A. and Sternberg, M.J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc., 4, 363–371.

60. Handel, E.M., Alwin, S. and Cathomen, T. (2009) Expanding or restricting the target site repertoire of zinc-finger nucleases: the inter-domain linker as a major determinant of target site selectivity. Mol. Ther., 17, 104–111.

61. Veith, T., Martin, R., Wurm, J.P., Weis, B.L., Duchardt-Ferner, E., Safferthal, C., Hennig, R., Mirus, O., Bohnsack, M.T., Wohnert, J. et al. (2012) Structural and functional analysis of the archaeal endonuclease Nob1. Nucleic Acids Res., 40, 3259–3274.

62. Elsby, L.M. and Roberts, S.G. (2008) Interaction of the TFIIB zinc ribbon with RNA polymerase II. Biochem. Soc. Trans., 36, 595–598.

63. Vanamee, E.S., Hsieh, P., Zhu, Z., Yates, D., Garman, E., Xu, S. and Aggarwal, A.K. (2003) Glucocorticoid receptor-like Zn(Cys)4 motifs in BslI restriction endonuclease. J. Mol. Biol., 334, 595–603.

64. Saravanan, M., Vasu, K., Ghosh, S. and Nagaraja, V. (2007) Dual role for Zn2+ in maintaining structural integrity and inducing DNA sequence specificity in a promiscuous endonuclease. J. Biol. Chem., 282, 32320–32326.

65. Grishin, N.V. (2000) C-terminal domains of Escherichia coli topoisomerase I belong to the zinc-ribbon superfamily. J. Molecular Biol., 299, 1165–1177.

66. Moodley, S. (2010) Biochemical investigation of the bacteriophage protein HK97 gp74. M.S. Thesis. University of Toronto, 1-102.

67. Chan, S.H., Opitz, L., Higgins, L., O’Loane, D. and Xu, S.Y. Cofactor requirement of HpyAV restriction endonuclease. PLoS One, 5, e9071.

68. Liu, G., Ou, H.Y., Wang, T., Li, L., Tan, H., Zhou, X., Rajakumar, K., Deng, Z. and He, X. (2010) Cleavage of phosphorothioated DNA and methylated DNA by the type IV restriction endonuclease SccmcrA. PLoS Genet., 6, e1001253.

69. Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J. and Voytas, D.F. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. Genetics, 186, 757–761.

70. Morbitzer, R., Elsaesser, J., Hausner, J. and Lahaye, T. (2011) Assembly of custom TALE-type DNA binding domains by modular cloning. Nucleic Acids Res., 39, 5790–5799.

71. Li, T., Huang, S., Jiang, W.Z., Wright, D., Spalding, M.H., Weeks, D.P. and Yang, B. (2011) TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Res., 39, 359–372.

72. Sanjana, N.E., Cong, L., Zhou, Y., Cunniff, M.M., Feng, G. and Zhang, F. (2012) A transcription activator-like effector toolbox for genome engineering. Nat. Protoc., 7, 171–192.

73. Mahfouz, M.M., Li, L., Shamimuzzaman, M., Wibowo, A., Fang, X. and Zhu, J.K. (2011) De novo-engineered transcription activator-like effector (TALE) hybrid nucleases with novel DNA binding specificity creates double-strand breaks. Proc. Natl Acad. Sci. USA, 108, 2623–2628.

74. Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J. et al. (2011) A TALE nuclease architecture for efficient genome editing. Nat. Biotechnol., 29, 143–148.

75. Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D. and Joung, J.K. (2012) FLASH assembly of TALENs for high-throughput genome editing. Nat. Biotechnol., 30, 460–465.

76. Briggs, A.W., Rios, X., Chari, R., Yang, L., Zhang, F., Mali, P. and Church, G.M. (2012) Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. Nucleic Acids Res., 40, e117.

77. Chu, S.W., Noyes, M.B., Christensen, R.G., Pierce, B.G., Zhu, L.J., Weng, Z., Stormo, G.D. and Wolfe, S.A. (2012) Exploring the DNA-recognition potential of homeodomains. Genome Res., 22, 1889–1898.

78. Fomenkov, A., Too, P.H., Chan, S.H., Vaisvila, R., Cantin, B.A., Mazzola, L., Tam, V. and Xu, S.Y. (2008) Targeting DNA 5mCpG sites with chimeric endonucleases. Anal. Biochem., 381, 135–141.

79. Wilson, G.W. and Edgell, D.R. (2009) Phage T4 mobE promotes trans homing of the defunct homing endonuclease I-TevIII. Nucleic Acids Res., 37, 7110–7123.