Nucleic Acids Research, 2026, **54**, gkaf1430
https://doi.org/10.1093/nar/gkaf1430
Nucleic Acid Enzymes

OXFORD

Modular determinants of cleavage preference in GIY-YIG nucleases revealed by block-based DNA shuffling and directed evolution

Kurt W. Loedige[1], **Mairwyn A. Hall**[1], **Sabin Ghimirey**[1], **Brent E. Stead**[2], **David R. Edgell**[1,*]

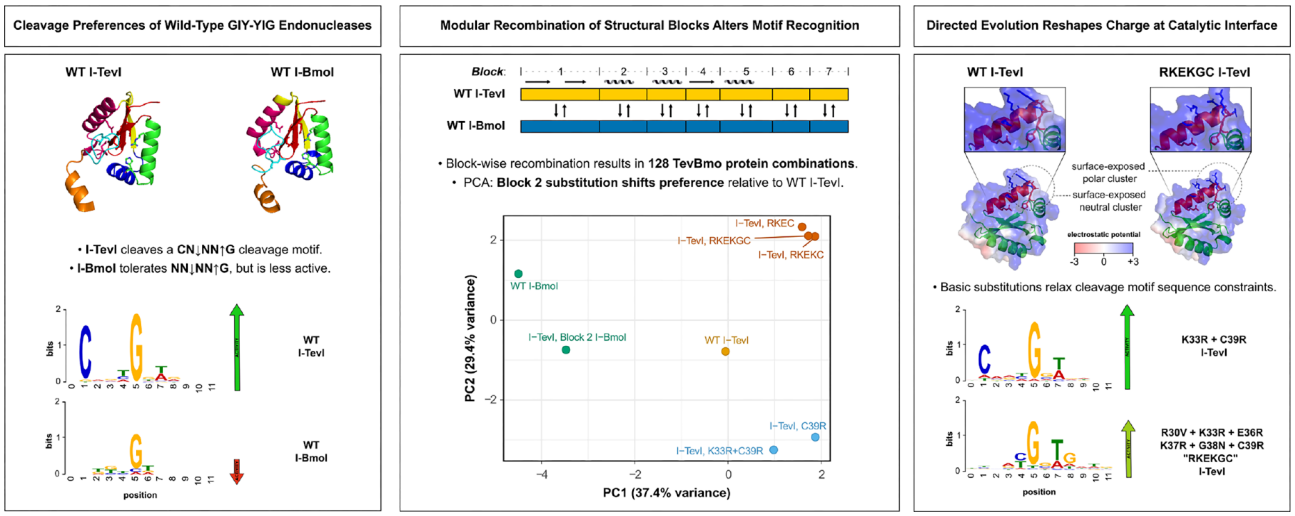
[1]Department of Biochemistry, Schulich School of Medicine and Dentistry, Western University, London, ON N6A 5C1, Canada
[2]Specific Biologics Inc., MaRS Discovery District, Toronto, ON M5G 0B7, Canada

*To whom correspondence should be addressed. Email: dedgell@uwo.ca

Abstract

GIY-YIG homing endonucleases are mobile genetic elements found in phage, bacterial, and organellar genomes. Their modular architecture and sequence-tolerant DNA cleavage properties likely represent evolutionary adaptations to tolerate genetic drift at target sites and enable target-site switching. To investigate modular determinants of GIY-YIG nuclease cleavage preference, we constructed 128 chimeric nucleases by shuffling structural blocks between the prototypical GIY-YIG homing endonuclease I-TevI (CNNNG motif preference) and its isoschizomer I-BmoI (NNNNG preference). Chimeras containing a swapped α-helix1 and adjacent loop exhibited altered motif preferences, highlighting this region as a modular determinant, whereas swaps in other regions disrupted activity without altering cleavage preference. Directed evolution of nonconserved residues within this region identified a cluster (R30, K33, E36, C39) where substitutions enabled cleavage of targets poorly recognized by wild-type I-TevI, including variants with reprogrammed preference toward TNNNG and GNNNG motifs. Our findings define a modular and structural basis for DNA cleavage preference in the GIY-YIG nuclease domain and suggest that recombination of structural subunits could accelerate adaptation to new target sites over evolutionary time scales. These findings further support a strategy for engineering GIY-YIG nuclease domains with expanded cleavage motif selectivity.

Graphical abstract



Introduction

Homing endonucleases (HEases) are compact, site-specific enzymes that generate double-strand breaks or nicks at long (14–40 bp) target sequences [1]. HEase genes are abundant in bacterial, phage, and organellar genomes, and are typically associated with self-splicing group I or II introns, inteins, or exist as free-standing genes that are not embedded within intervening elements [2, 3]. These enzymes promote their own mobility by cleaving target sites in genomes that lack the associated HEase gene, triggering homologous recombination and gene conversion [4]. Some HEases have evolved alternative functions as RNA chaperones [5–7] or transcriptional repressors [8]. In

Received: July 22, 2025. **Revised:** November 10, 2025. **Accepted:** November 22, 2025

general, HEases interact with their target sites in a sequence-tolerant manner [9–12], accommodating nucleotide substitutions throughout the binding site—an evolutionary strategy thought to facilitate mobility by adapting to genetic drift in related genomes [13, 14]. However, to function effectively as selfish elements, HEases must balance the ability to adapt to nucleotide changes in their target sites with sufficient DNA targeting selectivity to avoid toxic off-target cleavage.

Interestingly, insights into the evolvability of HEases have come from engineering studies focused on altering their DNA recognition properties [15, 16]. In single-chain LAGLIDADG endonucleases, the active site is embedded within the DNA-binding interface, making enzyme redesign challenging: substitutions that alter DNA recognition can disrupt catalytic function, creating catalytically compromised enzymes that require further improvement [17, 18]. Similar engineering challenges are seen in other endonuclease families, where specificity changes are possible through minimal mutational changes, but only after significant structural analysis, screening or directed evolution [19, 20]. In contrast, modular endonuclease architectures—where DNA recognition and cleavage are functionally separated into distinct structural units—offer a different evolutionary solution. In such systems, domains with distinct activities can be shuffled, or point substitutions can be introduced within one domain without compromising overall enzymatic function, due to functional buffering [21–23].

Among homing endonuclease (HEase) proteins [16], the GIY-YIG family is distinguished by a modular architecture, featuring a compact N-terminal nuclease domain (∼90 amino acids) with a β-sheet scaffold flanked by α-helices. This domain is linked to a DNA-binding domain via a flexible linker [24–28]. In the case of the prototypical GIY-YIG HEase, I-TevI, the nuclease domain cleaves a 5′-CN↓NN↑G-3′ cleavage motif (CNNNG) positioned upstream from the DNA-binding site and has a macroscopic preference for A/T-rich central NNN triplets [13, 29–31] (Fig. 1C). In contrast, I-BmoI, an isoschizomer of I-TevI, recognizes a 5′-NN↓NN↑G-3′ motif and targets a distinct GC-rich sequence (GCCCG), but is ∼750-fold less active than I-TevI *in vitro* [32, 33] (Fig. 1D). Structural and sequence comparisons suggest that the two enzymes share a conserved fold but diverge in key residues responsible for substrate engagement [32, 34]. While structural and biochemical studies of I-TevI and other GIY-YIG containing enzymes have identified key residues involved in catalysis [24, 27, 28, 35], the mechanism of DNA cleavage site selection by GIY-YIG homing endonucleases is not well understood.

The GIY-YIG nuclease superfamily comprises enzymes involved in DNA repair, restriction, replication, and mobility [26]. The GIY-YIG domain of homing endonucleases are most similar to the domain found in repair nucleases (UvrC, Slx1) [36, 37], restriction-like enzymes (Eco29kI, Hpy188I) [24, 38], and selfish elements [3, 39]. All share a compact catalytic core defined by the GIY and YIG motifs and a conserved glutamate (E75) that coordinates metal ions for phosphodiester hydrolysis. The domain is associated with diverse architectures, from single-domain dimers that cleave both DNA strands cooperatively, to modular fusions, such as I-TevI and I-BmoI, where a GIY-YIG core is linked to an independent DNA-binding domain—highlighting the evolutionary versatility of the GIY-YIG domain [26, 32]. Structural studies of GIY-YIG restriction enzymes revealed DNA recognition by extended loops that connect structurally and evolutionarily conserved segments of the GIY-YIG nuclease domain [24, 35]; these loops are not conserved in GIY-YIG homing endonucleases. These findings suggest that the loops that connect structurally defined units within GIY-YIG nuclease domains may be key for regulating DNA interactions, and that GIY-YIG domains could evolve new DNA specificity by shuffling of these regions over evolutionary timescales.

Here, to dissect how individual structural elements of the GIY-YIG nuclease domain influence substrate preference, we used block-based DNA shuffling and directed evolution with I-TevI and I-BmoI as evolutionary scaffolds [40, 41]. Although chimeric fusions between homologous domains often exhibit intermediate or destabilized phenotypes, we hypothesized that recombining structurally conserved "blocks" would enable a systematic exploration of structure–function relationships while preserving the catalytic core [42, 43]. This strategy builds on prior observations that cleavage activity and specificity is contained within the GIY-YIG nuclease domain [34, 44]. We constructed and screened a 128-member library of chimeric I-TevI/I-BmoI nuclease domains by recombining seven structural blocks. Characterization of substrate preference of the chimeric enzymes using a high-resolution capillary electrophoresis (CE) assay and nontargeted cleavage mapping by Oxford Nanopore sequencing revealed a single block swap consisting of an alpha helix and adjacent loop that influenced cleavage preference. Directed evolution of this block identified 4 residues (R30, K33, E36 and C39) where substitutions shifted preference from CNNNG to TNNNG or GNNNG cleavage motifs. Collectively, our findings establish a structural basis for programming cleavage preference of GIY-YIG nucleases and demonstrate how block-wise recombination and selection can overcome intrinsic specificity constraints.

Materials and methods

Bacterial strains

Escherichia coli NEB5α F′I′ [F′ *proA+B+lacI*^{q} Δ *(lacZ)M15 zzf::Tn10 (TetR) / fhuA2Δ(argF-lacZ)U169 phoA glnV44 ϕ80 Δ(lacZ) M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*], *E. coli* ER2566 [*F*$^-$ *mcrA* (*mrr-hsdRMS-mcrBC*) 80*lacZ M15 lacX74 recA1 ara139 (ara-leu)7697 galU galK rpsL* (*StrR*) *endA1 nupG*] (NEB), and *E. coli* BW25141(λDE3) [*F*$^-$ *λ(DE3) rrnB3 ΔlacZ4787 hsdR514 ΔaraBAD AH33 Δ(rhaBAD)::T7 rhaSR AH33*] were used for cloning, protein expression, and bacterial two-plasmid selection and were grown in Luria Broth (LB) media.

Plasmid construction

Primers used for cloning are listed in Supplementary Table S2 and a complete list of plasmids used in this study is provided in Supplementary Table S3. Control plasmids, including pTox and pEndo constructs [e.g. xTev–xOnu (R27A–E22Q) and E2 I-OnuI], were obtained from prior work [34]. All TevBmo nuclease variants were synthesized as *E. coli* codon-optimized gene fragments by Twist Bioscience (Supplementary Table S1). These fragments were polymerase chain reaction (PCR)-amplified using primers designed to append BsaI-HFv2 recognition sites and were cloned into a modified pACYC expression vector encoding catalytically inactive E2 I-OnuI domain (xOnu) with a C-terminal 6X-histidine tag using BsaI-HFv2 (NEB) Golden Gate assembly, following the manufacturer's

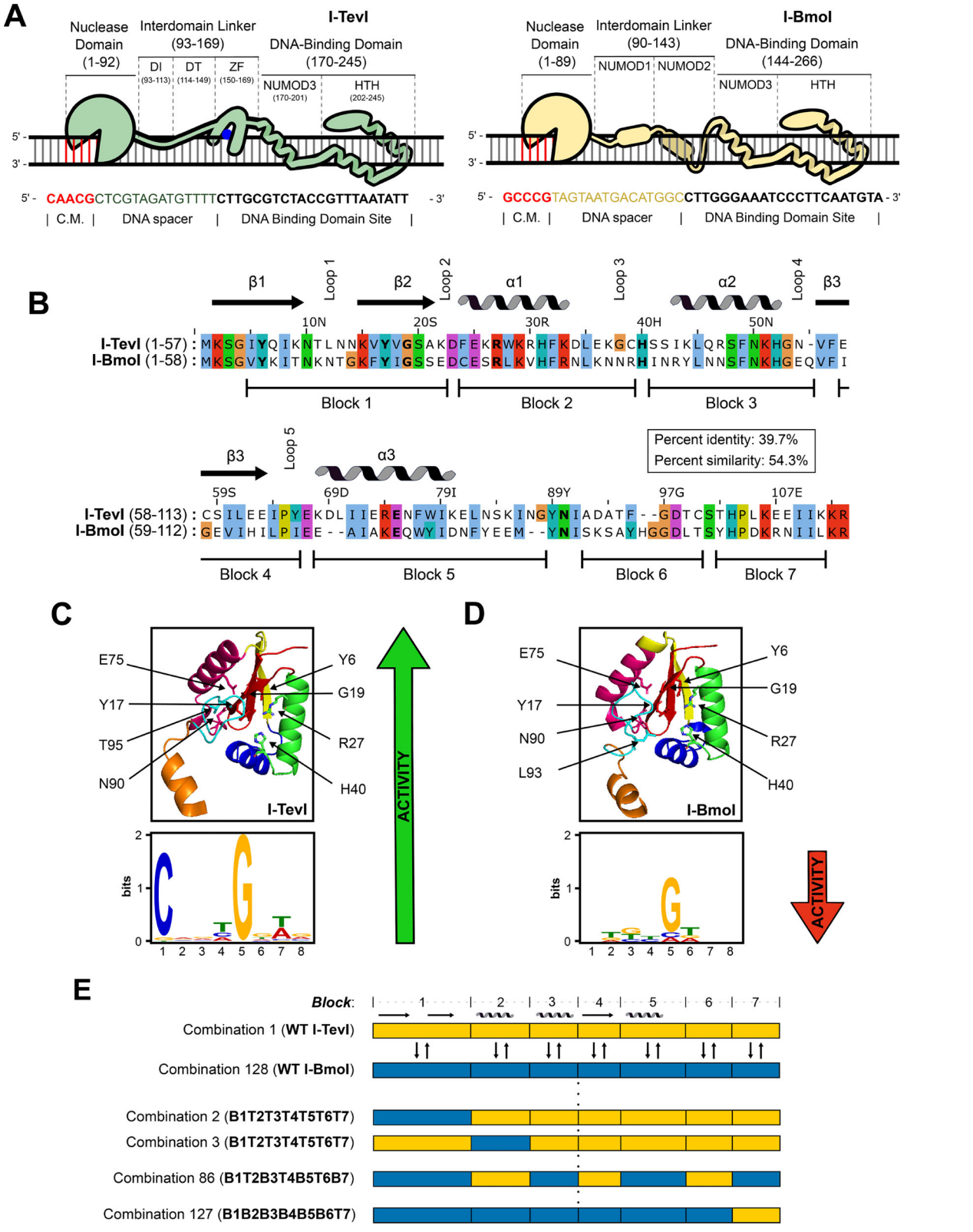


Figure 1. Design of chimeric GIY-YIG nuclease domains via block-wise domain shuffling. (**A**) Schematic diagram of full-length I-TevI and I-BmoI homing endonucleases with annotated structural subdomains and cognate target sites. DI, deletion-intolerant region. DT, deletion-tolerant region. ZF, zinc finger. HTH, helix-turn-helix. C.M., cleavage motif. (**B**) Structure-guided alignment of I-TevI and I-BmoI nuclease domains (residues 1–113) divided into seven conserved blocks used for recombination. Secondary structure elements are shown above. Catalytic residues are bolded. AlphaFold-predicted structures of wild-type I-TevI (**C**) and I-BmoI (**D**) with catalytic residues labeled. Associated sequence logos show preferred cleavage motifs. Green and red arrows illustrate relative I-TevI and I-BmoI nuclease domain cleavage activities. (**E**) Schematic of chimeric domain construction using block-wise recombination between I-TevI (yellow) and I-BmoI (blue).

protocol. To generate TevBmo–SaCas9 fusions, TevBmo variants (amino acids 1–169) were PCR-amplified with a C-terminal GGSGGTGGSG linker and fused in-frame to the N-terminus of an *E. coli* codon-optimized SaCas9 sequence with a C-terminal 6X-histidine tag in pET11a (Amp^R) by Gibson assembly (NEB). All constructs were verified by Sanger sequencing of the insert (London Regional Genomics Facility, London, ON) or by whole plasmid sequencing (Plasmidsaurus).

Construction of the degenerate Block 2 GIY-YIG library

To explore the functional contribution of Block 2 to substrate preference, a degenerate library encoding shuffled I-TevI and I-BmoI residues across positions 24–39 of the GIY-YIG nuclease domain was constructed. The design targeted natural variation between I-TevI and I-BmoI, focusing on the α1 helix and the downstream unstructured loop (loop 3). Conserved positions were retained, while variable residues were encoded using degenerate codons to broadly sample functional diversity. The degenerate region was generated using oligonucleotides DE8011 and DE8012, each containing a BsaI recognition site to enable Golden Gate assembly. The target vector, a preamplified pACYC backbone encoding wild-type I-TevI (residues 1–169, with a deletion spanning residues 24–39), was prepared by PCR using primers DE8013 and DE8014 under conditions described previously. Primer DE8011 (forward) incorporated a randomized sequence spanning residues 24–39, with IUPAC ambiguity codes (e.g. K, W, M, R, Y, S) encoding the consensus variability of I-TevI and I-BmoI. Primer DE8012 (reverse) annealed just downstream of the degenerate segment. Second-strand synthesis was performed using Klenow polymerase, and Golden Gate assembly was conducted as described above. Aliquots (10 µl) of the transformation recovery were plated on LB agar supplemented with chloramphenicol (25 µg/ml) to confirm library complexity, aiming for $\geq 10\times$ coverage of the theoretical 1024 unique protein sequence (8192 nucleotide sequence) combinations.

Bacterial two-plasmid selection

A two-plasmid selection system was used to evaluate the cleavage activity of engineered GIY-YIG endonuclease variants in *E. coli*. The system consisted of a toxic plasmid (pTox) and an endonuclease expression library plasmid (pEndo Library), co-transformed into BW25141(λDE3) *E. coli*. The pTox plasmid contained a pBAD-inducible *ccdB* toxin gene downstream of a non-canonical (AAACG, TAACG, or GAACG) TevOnu target site and carried an ampicillin resistance marker. The pEndo Block 2 plasmid library encoded GIY-YIG variants under the control of a T7 promoter and conferred chloramphenicol resistance.

pTox plasmid stocks were diluted to 20 ng/µl in nuclease-free water. For each transformation, 2 µl of plasmid was added to 28 µl of chemically competent BW25141(λDE3) cells. Cells were incubated on ice for 3 min, heat shocked at 42°C for 35 s, then returned to ice for 5 min. Transformed cells were recovered in 270 µl super optimal broth with catabolite repression (SOC) and shaken at 37°C for 1 h before plating on LB agar containing 100 µg/ml ampicillin and 0.2% glucose to repress *ccdB* expression.

Following transformation with pTox, colonies were made chemically competent. A second transformation was performed by adding 2 µl of 10 ng/µl pEndo Library to 28 µl of competent cells. Transformed cells were recovered in 220 µl SOB for 1 h at 37°C. Cultures were then split into 100 µl aliquots and transferred into 100 µl of either 2× repressive media (2×YT + 50 µg/ml chloramphenicol + 0.4% glucose) or 2× inducing media (2×YT + 50 µg/ml chloramphenicol + 2 mM isopropyl-β-D-thiogalactopyranoside [IPTG]). Cultures were incubated at 37°C in a floor shaker for 4 h.

For growth curve analysis, 20 µl of each culture was transferred into 180 µl of M9 minimal media, either selective (25 µg/ml chloramphenicol +0.2% arabinose) or nonselective (25 µg/ml chloramphenicol +0.2% glucose), and incubated at 37°C with 225 rpm shaking in an Epoch2 BioTek microplate reader. Absorbance at 600 nm was measured at 20-min intervals.

For directed evolution, the remaining 180 µl of inducing outgrowth was transferred to 500 ml of selective M9 minimal media and grown overnight at 37°C with 225 RPM shaking. Plasmids were purified from overnight cultures using the NEB Monarch Miniprep Kit according to the manufacturer's instructions. Purified plasmid DNA was retransformed for a total of five rounds of selection. Following round three, shuffled regions were subcloned into freshly prepared pACYC-TevxOnu backbones using traditional restriction enzyme cloning using BamHI-HF and BstAPI for digest, and T4 DNA ligase for ligation according to manufacturer's protocols (NEB).

Preparation of substrates for capillary electrophoresis and Oxford Nanopore sequencing

Primer sequences used for cloning and amplicon generation are listed in Supplementary Table S2. DNA substrates for CE cleavage assays were cloned into the pSP72 vector (Promega) using a modular, barcode-compatible strategy. Substrate inserts were generated by annealing a forward oligonucleotide containing, in order: a 5′ SphI-HF site, a variable-length DNA barcode, a unique cleavage motif, a synthetic I-TevI DNA spacer (corresponding to the intervening sequence between the cleavage motif and the I-OnuI binding site), and the 5′ portion of an I-OnuI binding site to a universal reverse oligonucleotide encoding the reverse complement of the spacer, the remaining I-OnuI site, and a 3′ EcoRI-HF site. Annealed oligonucleotides were extended using Klenow fragment (3′ → 5′ exo–) in NEBuffer 2 (NEB), purified through a Monarch column (NEB), and digested with SphI-HF and EcoRI-HF. Digested products were ligated into pSP72 that had been predigested with the same enzymes and gel-purified. Nanopore-derived targets were cloned identically, without DNA size barcodes. All enzymatic steps were performed following the manufacturer's protocols (NEB). Constructs were confirmed by Sanger sequencing.

To generate fluorescently labeled DNA substrates for CE, plasmid constructs were PCR-amplified using 5′-fluorescently labeled forward primers [6-FAM (DE7183), TET (DE8010), or HEX (DE8009)] and an unlabeled universal reverse primer (DE7185). Reactions were performed using PrimeSTAR GXL DNA Polymerase (Takara Bio) according to the manufacturer's protocol, using 10 ng of plasmid template and one-quarter of the recommended polymerase amount. PCR was carried out for 30 cycles. Amplified products were separated by agarose gel electrophoresis, and the correct bands were excised and purified using the Monarch DNA Gel Extraction Kit (NEB) to remove lower-molecular-weight contaminants

that interfered with product resolution. DNA concentrations were measured using a Qubit fluorometer (Thermo Fisher Scientific). Fluorophore incorporation and product quality were confirmed by test CE runs and 12% 8M urea-denaturing polyacrylamide gel electrophoresis (PAGE).

Guide RNA synthesis

Guide RNAs were synthesized by one-pot *in vitro* transcription as described previously [45]. For this study, only the SaCas9 λ guide was used. Briefly, overlapping oligonucleotides encoding the T7 promoter, spacer, and scaffold were extended with Klenow Fragment and transcribed using the T7 RiboMAX® Express kit (Promega). Products were purified with the Monarch RNA Cleanup Kit (NEB), and RNA integrity was confirmed by agarose gel electrophoresis.

Protein purification

C-terminal 6× His-tagged TevBmo-xOnu variants and TevBmoSaCas9 (expressed from pACYC and pET11a, respectively) were produced under the control of a T7 promoter in *E. coli* ER2566. Cultures (100 ml) were grown in a 2.5:1 flask-to-volume ratio at 37°C until OD_{600} reached 0.6, induced with 1 mM IPTG, and incubated overnight at 16°C. Cells were harvested by centrifugation and lysed by sonication in binding buffer [20 mM Tris–HCl (pH 8.0), 500 mM NaCl, 1 mM dithiothreitol (DTT), 5% (v/v) glycerol, 10 mM imidazole] supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF). Clarified lysates were applied to 1 ml HisPur™ Ni-NTA Spin Columns (Thermo Fisher Scientific), and proteins were eluted using a stepwise gradient of imidazole (200 and 500 mM). Eluted fractions were flash-frozen in a –80°C ethanol bath and stored at –80°C. Protein purity was assessed by 12% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE), and band intensities were quantified using Bio-Rad ImageLab software. Protein concentrations were calculated based on relative intensity to a known purified protein standard of identical molecular weight.

In Vitro cleavage reactions for capillary electrophoresis

Pooled substrate libraries (2×) were prepared at a final concentration of 33.2 nM, with equimolar representation of each substrate (~1.19 nM each for 28 substrates). Cleavage reactions were performed in triplicate, with each replicate corresponding to a distinct fluorophore for multiplexing. Reactions were assembled at a 50:1 molar ratio of protein to substrate library and contained 16.6 nM DNA substrate, 100 mM NaCl, 20 mM Tris–HCl (pH 8.0), and 1 mM DTT. Reactions were incubated at 37°C for 30 min and quenched with stop solution containing 16.7 mM EDTA and 50 μg/ml proteinase K, followed by a 30-min incubation at 37°C. Replicates were pooled post-quench and purified using the Monarch PCR & DNA Cleanup Kit (NEB), then eluted in 5 μl of nuclease-free water.

For CE, 2 μl of purified product was mixed with 1 μl of LIZ500 size standard (Thermo Fisher Scientific) and 7 μl of formamide. Samples were denatured and run on an ABI 3730xl DNA Analyzer at the London Regional Genomics Centre (London, ON). Trace quality and fragment detection were assessed using PeakScanner software (Applied Biosystems). Fragment sizes and intensities were exported and analyzed using custom Python scripts provided as Supplemetnary Code. Percent cleavage for each substrate was calculated as the ratio of cleaved product peak area to total substrate peak area (cleaved + uncleaved), based on expected fragment lengths and fluorophore–label combinations. Each dataset represents the mean of assays acquired from three independent fluorescent channels (6-FAM, HEX, TET). Samples showing standard deviation of greater than 5% of total signal discarded and repeated. Representative denaturing PAGE analyses illustrating raw substrate cleavage and qualitative agreement with CE-derived profiles are provided in Supplementary Fig. S4.

Capillary electrophoresis data normalization and clustering

To assess global differences in substrate preference among evolved TevBmo variants, cleavage activity was quantified for a panel of barcoded CM substrates using fluorescence-based CE. A matrix of average percent cleavage was generated for each protein variant. To emphasize relative substrate preferences rather than absolute activity, data were *Z*-score normalized across substrates within each protein. Principal component analysis (PCA) was performed using the `prcomp()` function in R. The number of retained components was chosen based on scree plots (Supplementary Fig. S6A). K-means clustering was applied to the PC scores using the `kmeans()` function, with the optimal number of clusters determined by elbow method analysis [`fviz_nbclust()`, `factoextra` package; Supplementary Fig. S6B]. Cluster assignments were overlaid on PCA projections to reveal specificity groupings among single- and multi-residue variants. Figures were generated using `ggplot2`.

In vitro timecourse cleavage assays

Cleavage activity of TevBmo variants fused to xOnu or SaCas9 was assessed by timecourse analysis. Where appropriate, ribonucleoproteins (RNPs) were assembled at a 1:1.5 molar ratio of protein to guide RNA by incubating components at room temperature for 10 min. Cleavage reactions contained 2.5 nM DNA substrate, 50 mM NaCl, 20 mM Tris–HCl (pH 8.0), and 1 mM DTT. For reactions using SaCas9 fusions, 1× rCutSmart Buffer (New England Biolabs; 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 100 μg/ml BSA) was used instead of NaCl-based buffer. Reactions were incubated at 37°C, and aliquots were removed at designated time points and quenched with stop solution (16.7 mM EDTA, 50 μg/ml RNase A, 50 μg/ml proteinase K), followed by incubation at 37°C for 30 min. Cleavage products were resolved by electrophoresis on 1.2% (w/v) agarose gels in Tris-acetate–EDTA (TAE) buffer. Gels were imaged, and band intensities were quantified using ImageLab (Bio-Rad) by manual band calling based on fragment size. Cleavage fractions were calculated from the relative intensity of cleaved versus uncleaved bands.

Oxford Nanopore sequencing of non-targeted cleavage and feature-based motif discovery

Nontargeted cleavage assays were performed as described previously [45]. Briefly, TevBmo variants were fused to catalytically active SaCas9 to provide a programmable DNA-binding scaffold that targets the λ phage tip attachment protein (J) gene, thereby supplying a defined positive-control cleavage

site while preserving the intrinsic cleavage behavior of the GIY-YIG nuclease domain.

TevBmoSaCas9 RNPs were incubated with λ-phage genomic DNA, and cleavage products were prepared for sequencing using the Oxford Nanopore SQK-LSK114 protocol with native barcoding (EXP-NB104). Sequencing was performed on MinION flow cells (R9.4.1), and basecalling, read alignment, and cleavage site parsing were conducted using the same custom scripts and pipeline described previously [45].

To evaluate intrinsic cleavage preferences, we used an unweighted feature set comprising a single instance of each of the 1000 most abundant read termini for each variant extracted from λ-phage genomic alignments to increase information content at positions enriched among the highest-frequency sites (e.g. C1 corresponding to the first cytosine of the CNNNG cleavage motif) [45]. Flanking sequences were retrieved and analyzed using MEME (v5.5.7) in "one occurrence per sequence" (oops) mode with a fixed motif width of 12 nucleotides. Motif logos were visualized from the resulting position weight matrices. Cleavage events occurring within ±50 bp of the Cas9 cut site were excluded from downstream analysis to avoid contributions from the single-site programmed Cas9 cleavage event.

Clustering of nucleases by λ-cleavage profiles

To compare variant-specific cleavage patterns, the nontargeted cleavage product matrix was transposed such that rows represented nucleases and columns represented genomic features. Features cleaved fewer than 35 times across all variants were removed. The resulting matrix was *Z*-score-normalized across features and subjected to PCA. The top five principal components (explaining >90% of variance; Supplementary Fig. S6C) were retained. K-means clustering was performed, and the optimal number of clusters ($k = 4$) was determined by elbow analysis (Supplementary Fig. S6D). Cluster identities were projected onto PCA plots (e.g. PC1 versus PC2, PC1 versus PC3) to reveal grouping patterns among variants.

Clustering of genomic features by cleavage profiles across variants

To determine whether subsets of genomic features were cleaved similarly across variants, we collected the top 250 cleaved sequences per nuclease, removed duplicates, and compiled a high-confidence feature set. Cleavage values were log-transformed using `log1p` and *Z*-score-normalized across features. The top five components (Supplementary Fig. S6E) from PCA analysis were used for k-means clustering. Four clusters were selected based on elbow analysis (Supplementary Fig. S6F). Cluster identities were used to annotate PCA projections and a heatmap of *Z*-scored cleavage activity. MEME (v 5.5.7) was used to extract representative motifs for each cluster.

Structural contact and interface analysis

Inter-block contacts within the I-TevI nuclease domain were analyzed from the crystal structure (PDB: 1MK0) using PyMOL v2.5. Blocks 2, 4, and 5 were defined as residues 24–39, 57–66, and 68–88, in wild-type I-TevI numbering, respectively. Pairwise atomic contacts were identified with the `dist` command using a 4.5 Å cutoff, generating three interface sets (B2–B4, B2–B5, B4–B5). Residues within this distance were displayed as sticks (radius 0.22 Å) and labeled on Cα atoms, with grey dashed lines indicating inter-block contacts.

Results

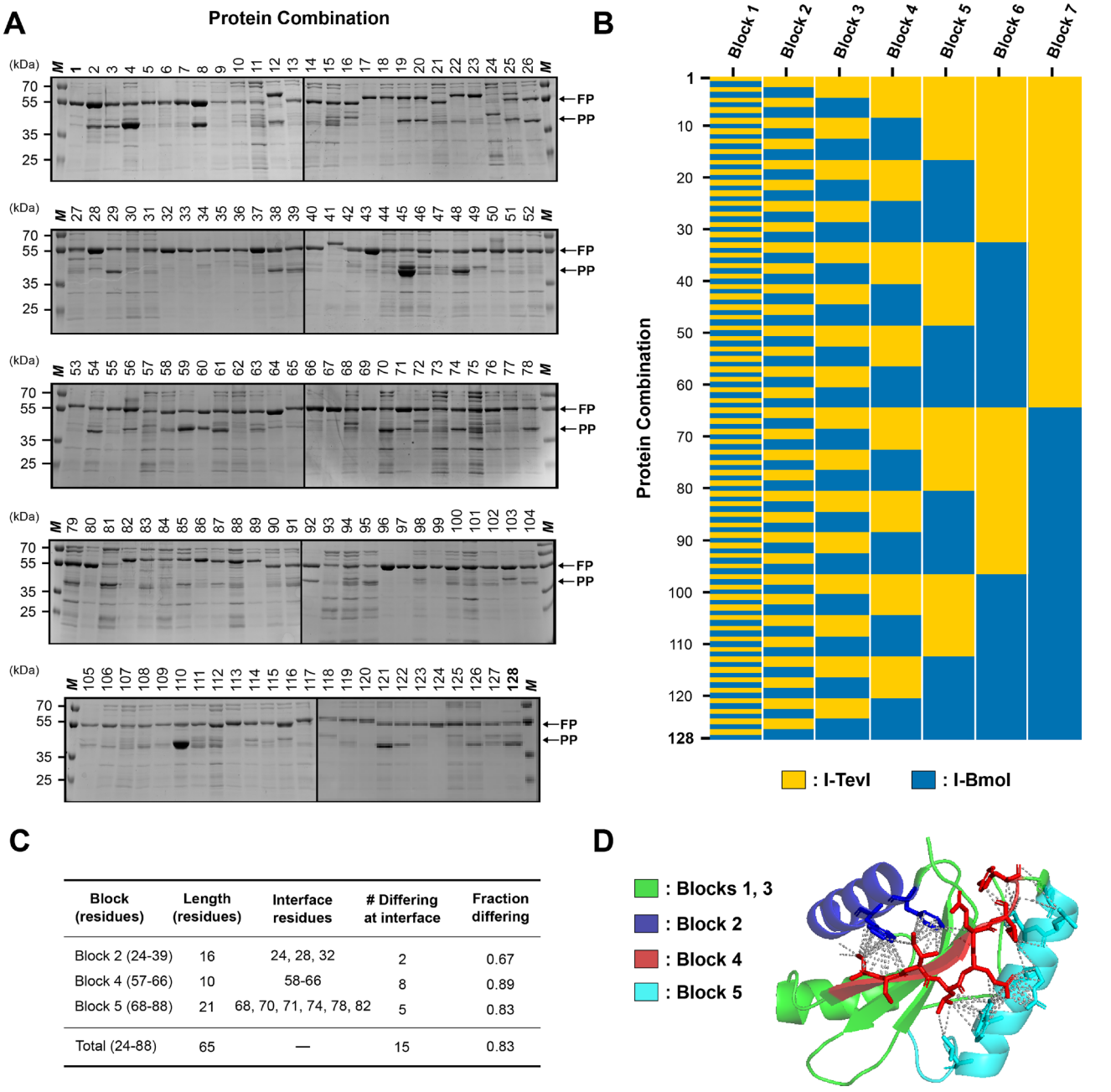
Design of a block-shuffled GIY-YIG nuclease library

As the first step in our design, we reasoned that a block-based library would enable systematic exploration of sequence–function relationships within the GIY-YIG scaffold. Given that I-TevI and I-BmoI evolved in distinct sequence contexts but recognize similar target sequences, we hypothesized that recombination could uncover modular determinants of specificity and activity, while also exposing structural dependencies that limit sub-domain portability. We constructed a chimeric library of 128 members by recombining seven sequence blocks from the I-TevI and I-BmoI nuclease domains based on secondary structure elements (Fig. 1A and B). The last block (Block 7) included the deletion intolerant (DI) region of the I-TevI linker [46] and the corresponding region in I-BmoI (Fig. 1B). Block boundaries were placed between α-helix and β-sheet junctions to preserve folding and designed to contain comparable levels of sequence diversity (Fig. 1B–D), with each block harboring between 5 and 11 amino acid differences. The exception was Block 5, encompassing the third α-helix, that had 18 residue differences (Fig. 1B). In total, 128 combinations were constructed (Figs 1E and, 2A and B) and fused to the I-TevI linker (residues 114–169) in the context of the I-OnuI meganuclease [30] to alleviate the toxicity issues associated with the overexpression of full-length WT I-TevI in *E. coli*. The wild-type I-TevI and I-BmoI nuclease domains were included as references (combinations 1 and 128, respectively) (Fig. 2B).

Expression and purification of chimeric GIY-YIG constructs

We cloned and purified each chimeric nuclease domain to enable downstream biochemical testing. SDS–PAGE analysis of all 128 block-shuffled combinations revealed a high success rate of soluble expression and recovery (Fig. 2A). The majority of variants migrated at the expected molecular weight (∼55 kDa) with minimal degradation. A subset of constructs exhibited decreased expression or partial proteolysis, evidenced by reduced purity or a secondary band migrating near ∼40 kDa. These lower-molecular-weight species likely represent liberated Onu DNA-binding domains resulting from proteolytic cleavage of the N-terminal nuclease domain, inter-domain linker, or synthetic GGSGGS linker (∼11–20 kDa), as the C-terminal 6× His tag remains intact. Such fragments are catalytically inactive and may only reduce apparent cleavage activity by occluding substrate DNA in a cleavage motif sequence-independent manner. Shifts in the apparent molecular weight of some constructs (e.g. lanes 12, 17–20) are consistent with block-composition-specific mobility differences typical of basic, linker-rich proteins.

Several block combinations consistently purified poorly or showed significant proteolysis. These included constructs in which Block 4 from one parental enzyme (e.g. Bmo) was introduced into a background retaining Blocks 2 and 5 from the other (e.g. Tev), or vice versa. Examples include combinations 9 (T1 T2 T3 B4 T5 T6 T7), 45 (T1 T2 B3 B4 T5 B6 T7), and 78 (B1 B2 T3 T4 B5 T6 B7), which yielded low recovery and/or visible degradation products (Fig. 2A). The recurring



Block (residues)	Length (residues)	Interface residues	# Differing at interface	Fraction differing
Block 2 (24-39)	16	24, 28, 32	2	0.67
Block 4 (57-66)	10	58-66	8	0.89
Block 5 (68-88)	21	68, 70, 71, 74, 78, 82	5	0.83
Total (24-88)	65	—	15	0.83

Figure 2. Expression, composition, and structural prediction of block-shuffled TevBmo-xOnu chimeric proteins. (**A**) SDS–PAGE analysis of affinity-purified chimeric constructs. Each lane represents a unique block-shuffled combination ($n = 128$), arranged in order of increasing combination number. Vertical black bars delineate independent gels. FP, full protein. PP, partial protein. M, molecular weight marker (kDa). Uncropped gel images are available in Supplementary Fig. S9. (**B**) Block compositions corresponding to each construct in panel (A). Yellow boxes indicate blocks derived from I-TevI, and blue boxes indicate blocks derived from I-BmoI. (**C**) Summary of inter-block contact analysis based on the I-TevI crystal structure (PDB: 1MK0). Shown are block boundaries, residue counts, interface residues within 4.5 Å, residues differing at those interfaces, and the fraction of differing positions between I-TevI and I-BmoI. Residues numbered according to wild-type I-TevI. (**D**) Structural mapping of Tev–Bmo residue differences on the I-TevI nuclease-domain crystal structure (PDB: 1MK0). Blocks 1 and 3 (green), 2 (blue), 4 (red), and 5 (cyan) are shown. Grey dotted lines indicate atomic contacts (within 4.5 Å). Side chains displayed as sticks correspond to residues that differ between I-TevI and I-BmoI.

instability of these constructs suggests that specific blocks—particularly those spanning Blocks 2, 4, and 5—form structurally interdependent interfaces whose disruption compromises proper folding or proteolytic resistance. Overall, however, the GIY-YIG nuclease scaffold proved broadly tolerant to block-wise recombination, with the majority of variants expressing solubly and retaining integrity following purification.

To investigate the structural basis for this instability, we mapped the positions of these blocks and their sequence differences onto the experimentally determined I-TevI nuclease-domain crystal structure (PDB: 1MK0, Fig. 2C and D). This analysis provided a residue-level view of inter-block contacts and offered a mechanistic explanation for why recombination across certain boundaries resulted in purification difficulties.

Mapping I-TevI/I-BmoI residue differences onto the I-TevI nuclease-domain crystal structure revealed that Block 4 is predominantly interfacial: nine of its ten residues form contacts with Blocks 2 and 5 (Fig. 2C and D). Eight of these nine contact residues differ between I-TevI and I-BmoI and are largely solvent-exposed, indicating that sequence divergence is concentrated within an interface-dense region connecting otherwise conserved sub-domains. Conversely, Block 2, comprising 16 residues, contributes only three interfacial positions, two of which vary between I-TevI and I-BmoI. This pattern suggests that substitutions in Block 4 co-vary with neighboring residues to preserve packing geometry within a structurally coupled sub-fold. The dense network of contacts centered on Block 4 likely represents a co-adapted interface between Blocks 2 and 5, explaining why chimeras that disrupt this boundary frequently exhibited poor solubility or proteolysis.

Single-block substitutions reveal determinants of cleavage preference

To evaluate how specific structural elements influence substrate preference, we performed a high-resolution CE cleavage assay on a panel of 28 synthetic TevCM substrates with each of the 128 variants (Fig. 3 and Supplementary Figs S2, S3, and S12). Each substrate contained a cleavage motif (CNNNG) with variations in the central triplet (NNN) previously shown to perturb Tev activity [30, 34], or substitutions in the first position (AAACG, TAACG, GAACG) that are poor substrates for cleavage by wild-type I-TevI. The cognate CAACG I-TevI cleavage site and the AAACA site that abolishes I-TevI cleavage were used as control substrates. Each cleavage motif was flanked by a constant synthetic spacer (lacking CNNNG sites) and an I-OnuI binding site. Substrates were uniquely size-barcoded with each barcode increasing in length by 2 bp and labeled at the 5′ end with 6-FAM, HEX, or TET fluorophores to enable multiplexed detection in pooled reactions (Fig. 3A and Supplementary Table S2). The CE format allowed us to rapidly and quantitatively assess cleavage activity of the variants on a wide range of substrates in a more cost and time effective workflow than using next generation sequencing or PAGE analysis (representative denaturing PAGE analyses are shown in Supplementary Fig. S4).

Screening of the 128-member library revealed that constructs containing multiple Bmo-derived blocks were inactive below the limit of detection (as determined using the catalytically inactive R27A I-TevI variant), consistent with structural incompatibility or absence of required combinatorial effects (Supplementary Fig. S1). We therefore focused on single-block I-BmoI substitution variants within the I-TevI scaffold. Percent cleavage for each variant was determined for the 28-member substrate panel and visualized as a heatmap (Fig. 3D).

Substitution of I-BmoI Block 1 (combination 2) reinforced preference for the canonical CAACG motif, whereas Block 7 (combination 65) increased cleavage activity without altering substrate rank order. The Block 3 substitution (combination 5) similarly reinforced CAACG preference and diminished activity at cleavage motifs with first-position substitutions, with minimal impact on triplet preference. In contrast, the Block 2 substitution (combination 3) induced a reproducible shift in specificity toward substrates with first-position substitutions, including AAACG, TAACG, and GAACG. This shift was more evident when cleavage activity for each variant was normalized to its own activity on the cognate CAACG substrate, allowing relative substrate preferences to be compared independent of total catalytic output (Fig. 3E). Collectively, these data demonstrate that Block 2 modulates recognition of the CM, while other blocks play secondary roles in tuning activity or maintaining folding.

Directed evolution of Block 2 variants enhances preference toward noncognate motifs

To probe the role of Block 2 in substrate recognition, we constructed a combinatorial library that included the two possible residues at all 10 positions that differed between I-BmoI and I-TevI. This generated a library of 1024 combinations encompassing positions 24–39 that spanned αhelix1 and the adjacent loop (Fig. 4A). This region includes several surface-exposed residues previously implicated in modulating cleavage specificity [34], including C39R, which arose alone and in combination with other substitutions that conferred improved activity on nonpermissive substrates.

The library was subjected to directed evolution against three noncognate targets, AAACG, TAACG, and GAACG, using a two-plasmid bacterial survival assay (Fig. 4B) [30, 47]. In this system, *E. coli* cells were cotransformed with a CcdB toxin-expressing plasmid (pTox) containing the target motif and a pEndo plasmid expressing chimeric TevBmo-xOnu proteins. Cleavage of the toxic plasmid conferred survival under inducing conditions (arabinose + IPTG) whereas repressive conditions (glucose) served as a negative control. Selections were carried out for five rounds, with a subcloning step at round 3 to eliminate plasmid artifacts and reset the genetic background by recloning the selected inserts into freshly prepared pACYC–Tev–xOnu backbones before continuing selection.

Across all three targets (AAACG, TAACG, GAACG), growth curves demonstrated progressive enrichment of active variants relative to wild-type I-TevI (Fig. 4C–E). By round five, enriched libraries showed significantly enhanced growth rates on noncognate targets, confirming successful selection for variants with altered motif preferences. Sanger sequencing of enriched clones revealed a set of recurrent mutations specific to each selection condition (Fig. 4F). For the AAACG substrate, the dominant substitution was C39R. Selections on the TAACG and GAACG substrates yielded multiresidue substitutions, including RKEC, RKEKC, and RKEKGC variants. These residues at positions R30, K33, E36, K37, G38, and C39 clustered at the end of α-helix1 and in the following loop region, suggesting that they contribute to reshaping the local electrostatic environment or DNA backbone interactions near the catalytic interface (Fig. 5A and B). The recurrent selection of C39R across distinct targets highlights this position as a key residue for modulation of substrate preference. Notably, C39R alone accounted for 87% of enriched variants in the AAACG selection, underscoring its sufficiency in enabling cleavage of this noncognate substrate. These data further demonstrate that directed evolution of the Block 2 hotspot residues could be used to access target sites not cleaved efficiently by wild-type I-TevI.

Evolved Block 2 variants exhibit distinct structural features and substrate preferences

To understand how directed evolution of Block 2 variants reprograms substrate preference, we modeled the structure of enriched variants using AlphaFold2. Comparison of wild-type

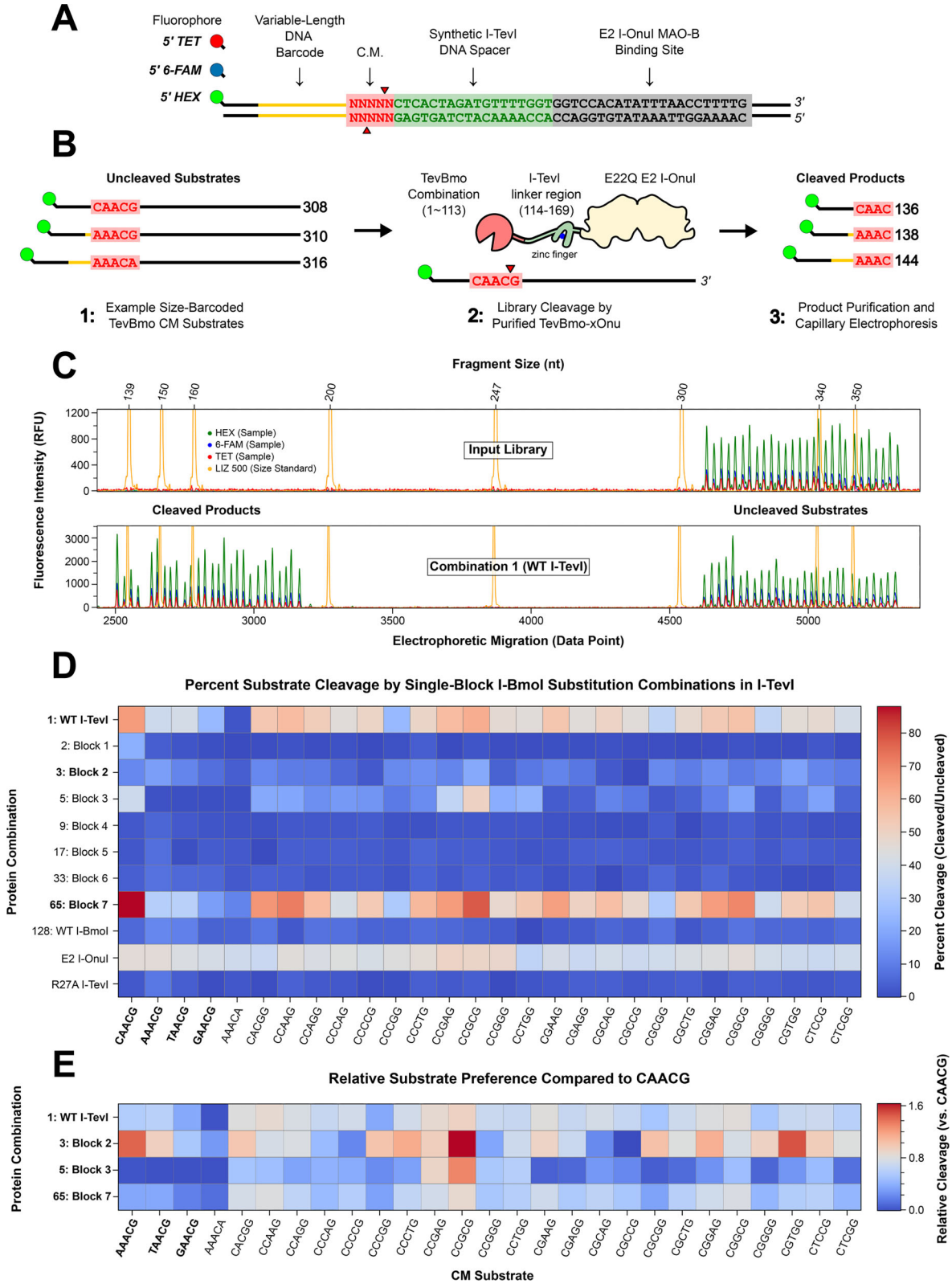
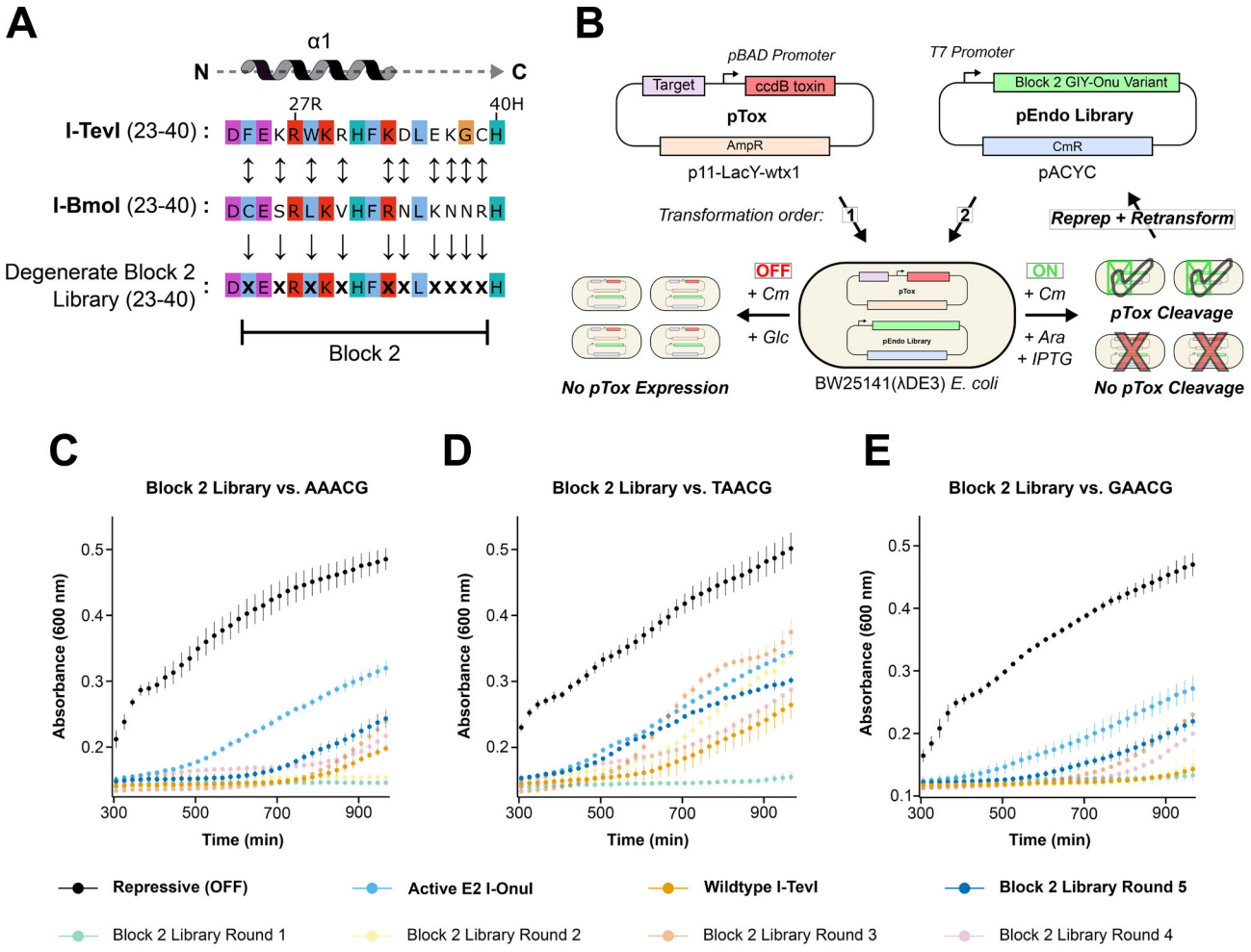


Figure 3. CE-based cleavage profiling of single-block TevBmo-xOnu variants. (**A**) Schematic of fluorescently labeled, size-barcoded CM substrate library. Each substrate includes a 5′ fluorophore (6-FAM, HEX, or TET), a unique length barcode, a central cleavage motif (CM, red), a constant I-TevI spacer (green), and an I-OnuI binding site (gray). (**B**) Overview of the CE workflow. Pooled substrates were incubated with purified TevBmo-xOnu proteins and analyzed by CE. (**C**) Representative electropherograms showing substrate length distributions for the input library (top) and cleavage by wild-type I-TevI (bottom). Cleaved and uncleaved substrates are annotated. Internal LIZ-500 size standard peaks are labeled. Independent electropherograms for each fluorophore channel are shown in Supplementary Fig. S2 (uncleaved) and Supplementary Fig. S3 (cleaved). (**D**) Heatmap showing percent cleavage across all 28 CM substrates for wild-type and single-block substitution variants. (**E**) Substrate preference profiles for selected variants (WT, Blocks 2, 3, 7), normalized to CAACG.



F

Sequence	Target	Variant	Frequency (%)	Substitution(s)
FEKRWKRHFKDLEKGRH	AAACG	"C39R"	87	C39R
FEKRWKRHFRDLEKGRH	AAACG	"K33R + C39R"	13	K33R + C39R
FEKRWKVHFRDLKKGRH	TAACG, GAACG	"RKEC"	67, 27	R30V + K33R + E36K + C39R
FEKRWKVHFRDLKNGRH	TAACG, GAACG	"RKEKC"	33, 60	R30V + K33R + E36K + K37R + C39R
FEKRWKVHFRDLKNNRH	GAACG	"RKEKGC"	13	R30V + K33R + E36K + K37R + G38N + C39R

Figure 4. Directed evolution of Block 2 variants for activity on noncognate CM targets. (**A**) Alignment of I-TevI and I-BmoI amino acid sequences spanning residues 24–39 (Block 2). Positions marked with "X" were randomized in the combinatorial library. (**B**) Dual-plasmid selection strategy schematic. A toxic plasmid (pTox) expressing the CcdB toxin under an arabinose-inducible promoter was cotransformed with a pEndo library expressing chimeric variants. Cleavage of pTox confers survival under inducing conditions (Ara + IPTG); glucose (Glc) serves as a repression control. Growth curves for Block 2 library and control constructs targeting AAACG(**C**), TAACG (**D**), or GAACG (**E**). Absorbance at 600 nm (A600) measured over time. Rounds of enrichment preceding round 5 less opaquely colored. (**F**) Aligned amino acid sequences of top enriched variants after five rounds of selection for each target. Mutations relative to wild-type I-TevI are shown in red. Frequencies indicate relative abundance among sequenced clones.

I-TevI and the RKEKGC variant revealed that enriched substitutions cluster on a surface-exposed loop region downstream of the α1-helix (Fig. 5A and B). These residues (e.g. K33R, E36N, K37R, G38N) form a polar patch adjacent to conserved catalytic residue R27, suggesting that changes in local electrostatics or DNA backbone interactions may underlie altered cleavage motif recognition.

To validate these functional changes, we assessed cleavage activity of evolved variants across the original 28-member CE substrate panel (Fig. 5C). We used PCA to compare variant substrate preferences (Fig. 5D), finding that the first two principal components explained 66.8% of the variance and revealed clear separation between wild-type I-TevI, I-BmoI, and the evolved variants (Fig. 5D). These analyses identified variants that clustered with similar preferences, including the RKEC, RKEKC, and RKEKGC combinations that had increased activity on the TAACG and GAACG substrates as compared to wild-type I-TevI (Fig. 5C). The C39R and K33R, C39R variants showed marginal increases in activity on substrates with C1 (corresponding to the first cytosine of the CNNNG cleavage motif) substitutions, but more significant activity improvements on CNNNG substrates poorly cleaved

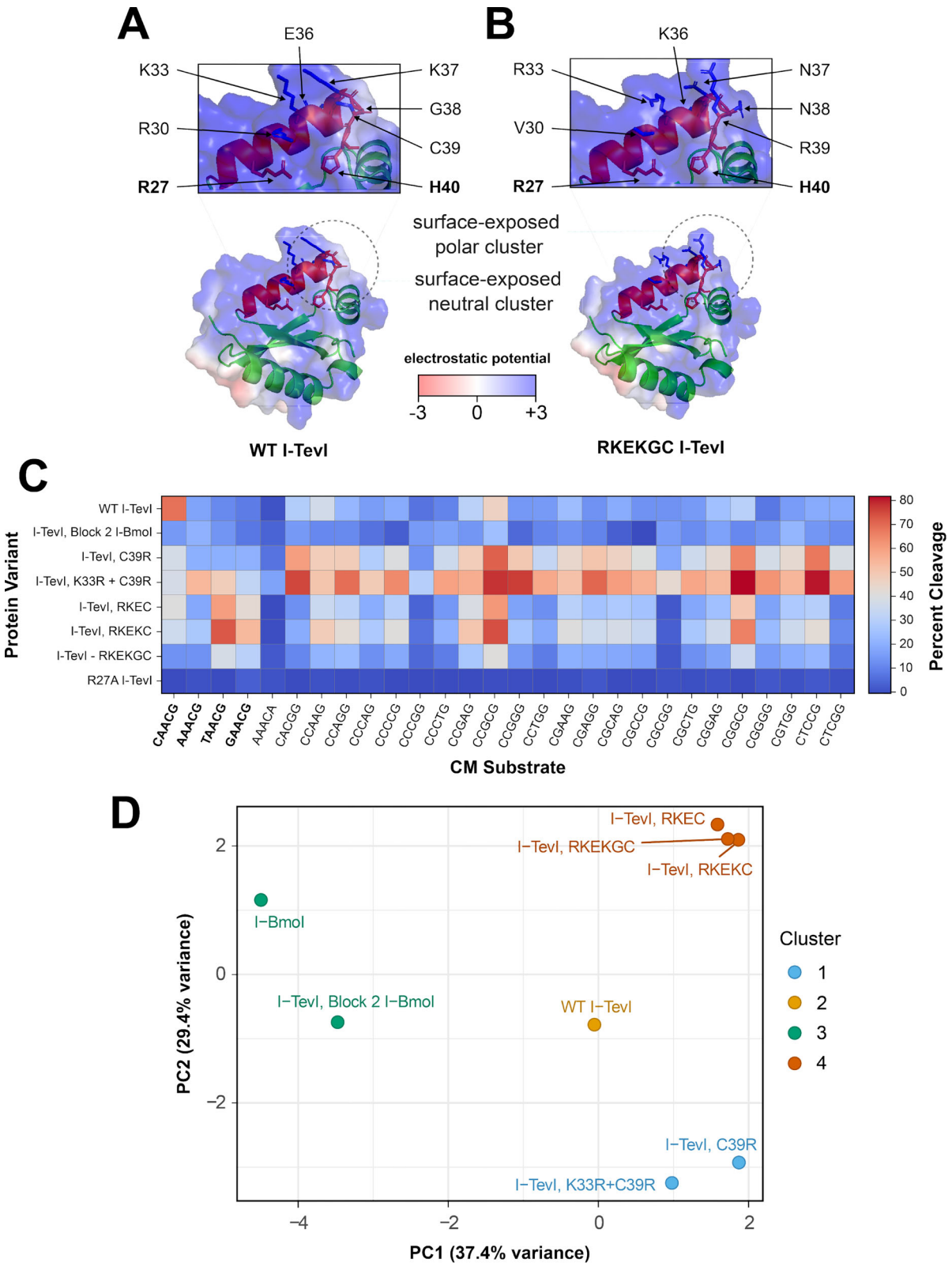


Figure 5. Structural modeling and functional clustering of evolved TevBmo-xOnu variants. AlphaFold-predicted structures of wild-type I-TevI (**A**) and the RKEKGC variant (**B**), with catalytic residues (R27, H40) in red and evolved substitutions (blue sticks) highlighted within Block 2 (red). (**C**) Heatmap of percent cleavage across CM substrates for selected evolved variants. Increased activity on non-cognate motifs is observed for C39R, K33R+C39R, and RKEC-class variants. (**D**) PCA of normalized cleavage preferences. K-means clustering ($k = 4$) separates variants by substrate preference.

by the wild-type scaffold (Fig. 5C). Variants incorporating I-BmoI Block 2 in I-TevI resulted in clustering with WT I-BmoI, and not WT I-TevI, highlighting the importance of this region for the determination of preference. Block 2-swapped I-TevI also clustered separately from those with RKEC-like substitutions.

These results collectively demonstrate that the preferred substrates for variants were consistent with the selection target used during directed evolution, especially among RKEC-class variants, suggesting that target-driven enrichment directly shapes sequence preference. Block 2 acts as a specificity determinant, and selected substitutions enable cleavage of sequences inaccessible to the parental enzymes.

Nontargeted cleavage mapping reveals variant-specific motif preferences

To more directly profile the cleavage preferences of our evolved GIY-YIG variants, we performed cleavage mapping using Oxford Nanopore sequencing of nontargeted cleavage products (NTCPs) as described [45]. We re-cloned the nuclease domain variants as N-terminal fusions to Cas9 from *Staphylococcus aureus* (SaCas9). Purified protein and *in vitro* synthesized sgRNA targeting the *J* gene (cleavage position 15905) were used to form RNP (Supplementary Figs S5A and S13, and Fig. 6A). RNPs for each variant were incubated with λ-phage genomic DNA to identify a timepoint with ∼50% cleavage for sequencing (Fig. 6B–J). Nontargeted cleavage events producing DNA fragments were mapped and ranked by frequency, then filtered to identify top 1000 high-confidence sites for motif analysis.

All tested variants retained activity in a dual-active SaCas9 fusion context, and cleaved λ DNA over a 60-min window; however, this set of nucleases exhibited distinct kinetics and product profiles (Fig. 6B–J). Wild-type Tev, catalytically enhanced variants (e.g. KTQ, VKN) [48], C39R and K33R+C39R variants completely converted λ DNA in to NTCPs by 60 min. The RKEC, RKEKC, and RKEKGC variants showed slower but sustained accumulation of NTCPs over the 60 min timecourse, consistent with a trade-offs of activity for relaxation in cleavage preference, as has been observed while evolving other nucleases [22, 49]. Swapping of I-BmoI 2 Block 2 into I-TevI (Fig. 6E) resulted in reduced accumulation of NTCP, consistent with the attenuated activity profile observed in CE-based assessments (Figs 3 and 5). Notably, a discrete cleavage product of ∼3.1 kb was uniquely observed in the VKN variant (Fig. 6D), indicating a variant-specific shift in cleavage site preference relative to wild-type I-TevI.

To identify the sequence preferences informing these cleavage events, we extracted the most frequently cleaved positions from each dataset and generated sequence logos from aligned 21-bp windows surrounding these positions. These motif profiles revealed divergence in sequence preferences among the tested variants (Fig. 6B–J). Wild-type Tev retained a canonical CNNNG-like motif, while RKEC-class variants exhibited no information content at position 1 of the CM, consistent with evolved preference toward TAACG, and GAACG motifs. Notably, the logos for C39R and K33R+C39R and KTQ variants displayed intermediate sequence features, with slightly diminished information at position 1, reflecting partial reprogramming of substrate preference consistent with the variant-specific separation observed in the PCA (Fig. 7A). In all cases, we continued to observe a preference for thymine at position 7 (DNA spacer position 2), implying that this is largely a linker-mediated contact unchanged across Block 2-derived variants.

These high-throughput cleavage profiling data reinforce findings from the CE assay and structural modeling; namely, that mutations within Block 2 reshape cleavage-motif preference to enable directed evolution of nucleases with access to novel targets.

Unsupervised clustering reveals functionally distinct classes of cleaved targets

To compare how evolved nucleases differed in their nontargeted cleavage profiles, we applied unsupervised learning to normalized cleavage frequency data derived from the Oxford Nanopore sequencing, in two ways. First, PCA revealed a clear separation between the wild-type nuclease domain (WT) and evolved variants (Fig. 7A). Importantly, variant clustering patterns observed from nontargeted cleavage mapping (Fig. 7A) mirrored those obtained from CE-based profiling (Fig. 5D), indicating that both independent approaches capture preference relationships among evolved variants.

Second, unsupervised k-means clustering of normalized cleavage frequency data grouped individual cleavage sites into four distinct sequence clusters (S1–S4) with similar cleavage profiles across variants (Supplementary Fig. S7). The number of clusters ($k = 4$) was determined using the elbow criterion (Supplementary Fig. S6D). Cluster–variant activity relationships, visualized by heat map analysis, revealed that each sequence cluster was preferentially cleaved by a subset of variants (Fig. 7B and C). For example, RKEKC and RKEKGC showed strong activity on Cluster S2, while wild-type I-TevI primarily cleaved Cluster S3 targets. Motif analysis of the top-cleaved sequences revealed a distinct sequence logo for each group (Fig. 7D–G). Cluster S1 featured a relaxed C1 constraint, notably including a C1T substitution and relazed information content in the spacer (positions >6). Cluster S2, strongly cleaved by RKEC-class variants, was enriched for T and G at the first position, similar to substrates TAACG and GAACG that were used for directed evolution. In contrast, cluster S2 was inaccessible to other nucleases, including C39R and K33R+C39R, which were not evolved on the C1T substrate. However, the KTQ variant, with substitutions K26R, T95S, and Q158R that tolerate C1T target sites, outperformed these nucleases here, suggesting increased activity, or a convergent but incomplete transition in preference logic toward that of the RKEC class. Cluster S4 was characterized by an increase in information content in the DNA spacer with the VKN variant showing high activity on this cluster. This increased activity is consistent with previous directed evolution studies that identified the VKN variant (V117F, K135R, and N140S) that modulates preference at position 7 the DNA spacer region [48].

Kinetic confirmation of cleavage activity on a nanopore-derived cluster S2 target

To validate the activity of evolved GIY-YIG variants on a target exclusive of the canonical I-TevI CNNNG motif, we selected a frequently cleavaged sequence identified by nanopore cleavage profiling (Supplementary Fig. S5B and Fig. 8A). Cleavage assays with purified GIY-xOnu chimeras and supercoiled plasmid containing the selected target showed that the engineered variants (RKEC, RKEKC, RKEKGC) con-

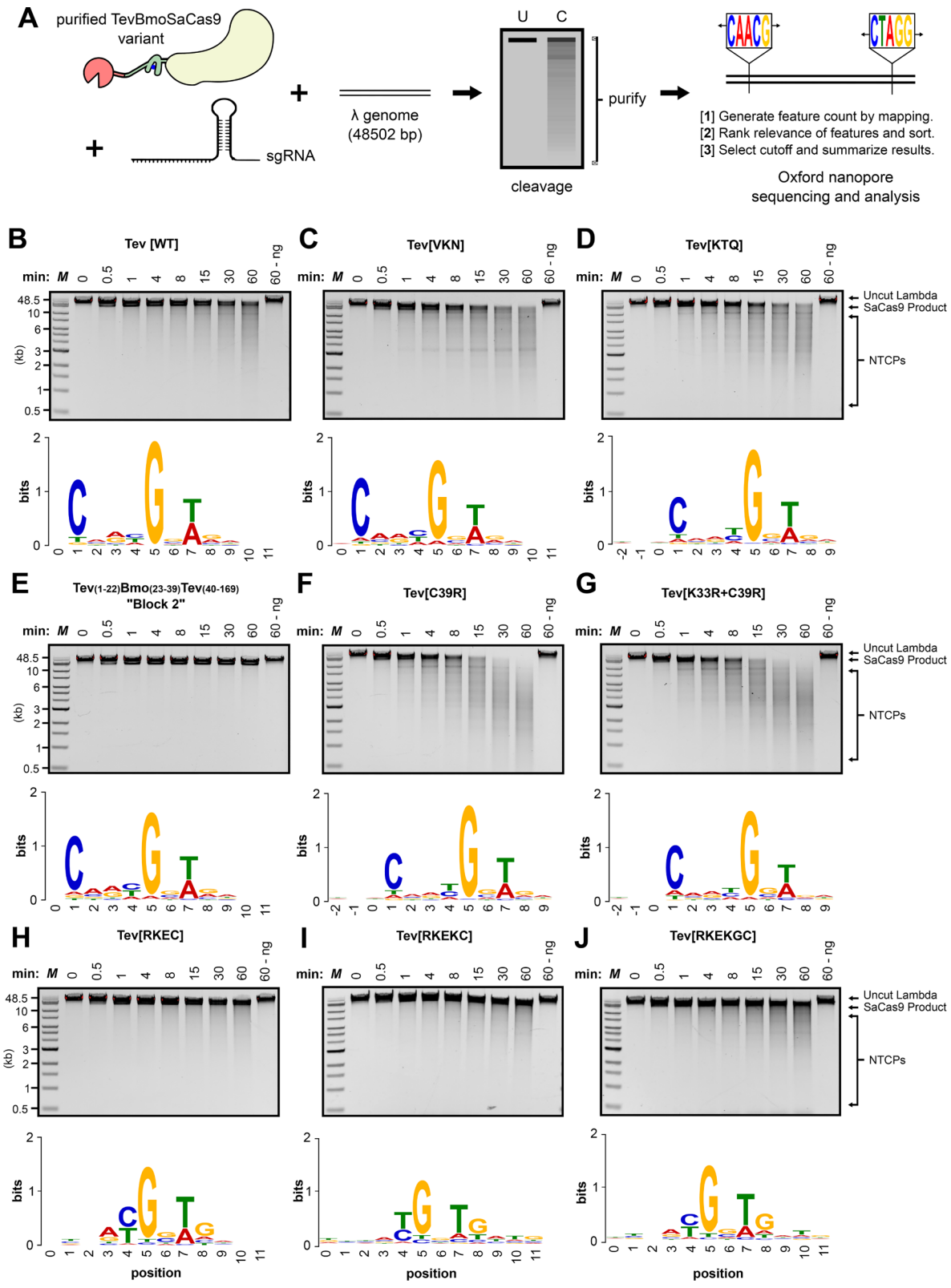


Figure 6. λ Cleavage timecourses and feature extraction. (**A**) Schematic of the Oxford Nanopore-based profiling workflow. (**B**–**J**) Representative agarose gels of timecourse cleavage assays of λ genomic DNA using wild-type and engineered TevBmoSaCas9 variants, including previously described I-TevI nuclease and linker domain variants (KTQ, VKN) [48], block-shuffled (Block 2), and directed-evolution derivatives (C39R, K33R+C39R, RKEC, RKEKC, RKEKGC). Each panel shows the progression of cleavage over time, with reaction times (minutes) indicated above each lane. Uncut λ DNA, targeted SaCas9 cleavage products, and NTCPs are annotated. Below each gel, representative sequence logos summarize the consensus cleavage motifs. Uncropped gel images are available in Supplementary Fig. S10.

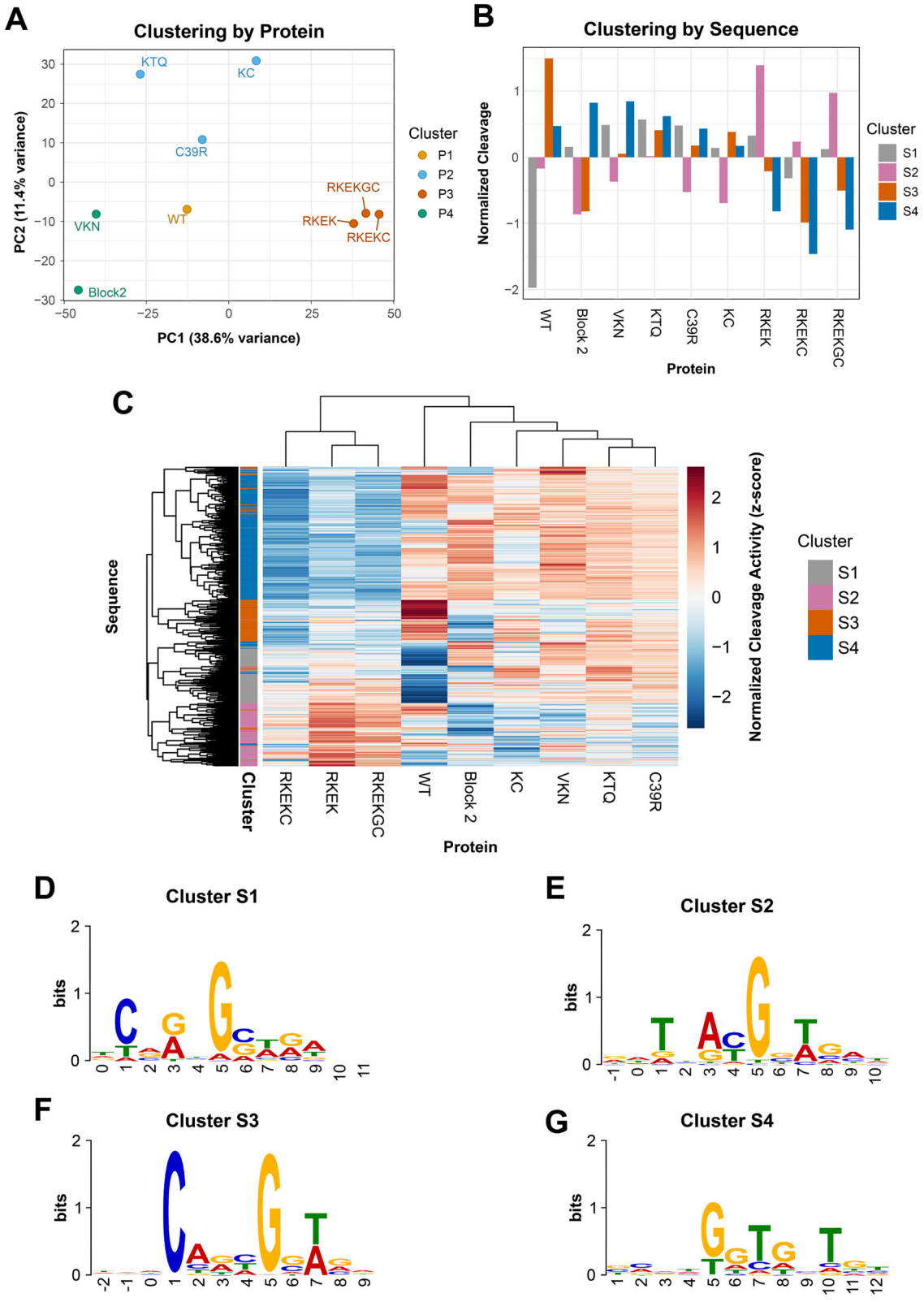


Figure 7. Functional clustering of evolved variants and cleaved targets by Oxford Nanopore sequencing analysis. (**A**) PCA of protein variants based on normalized cleavage across top genomic targets. KC, K33R+C39R. (**B**) Bar plot showing average normalized cleavage activity (*z*-score) for each nuclease across the four k-means sequence clusters identified in Supplementary Fig. S3. (**C**) Heatmap showing *z*-score-normalized cleavage activity across all high-frequency targets (rows) and protein variants (columns). Columns clustered hierarchically; rows grouped by sequence clusters defined in panel (B) and Supplementary Fig. S3. (**D**–**G**) Sequence logos representing motif composition of Clusters 1–4, respectively.

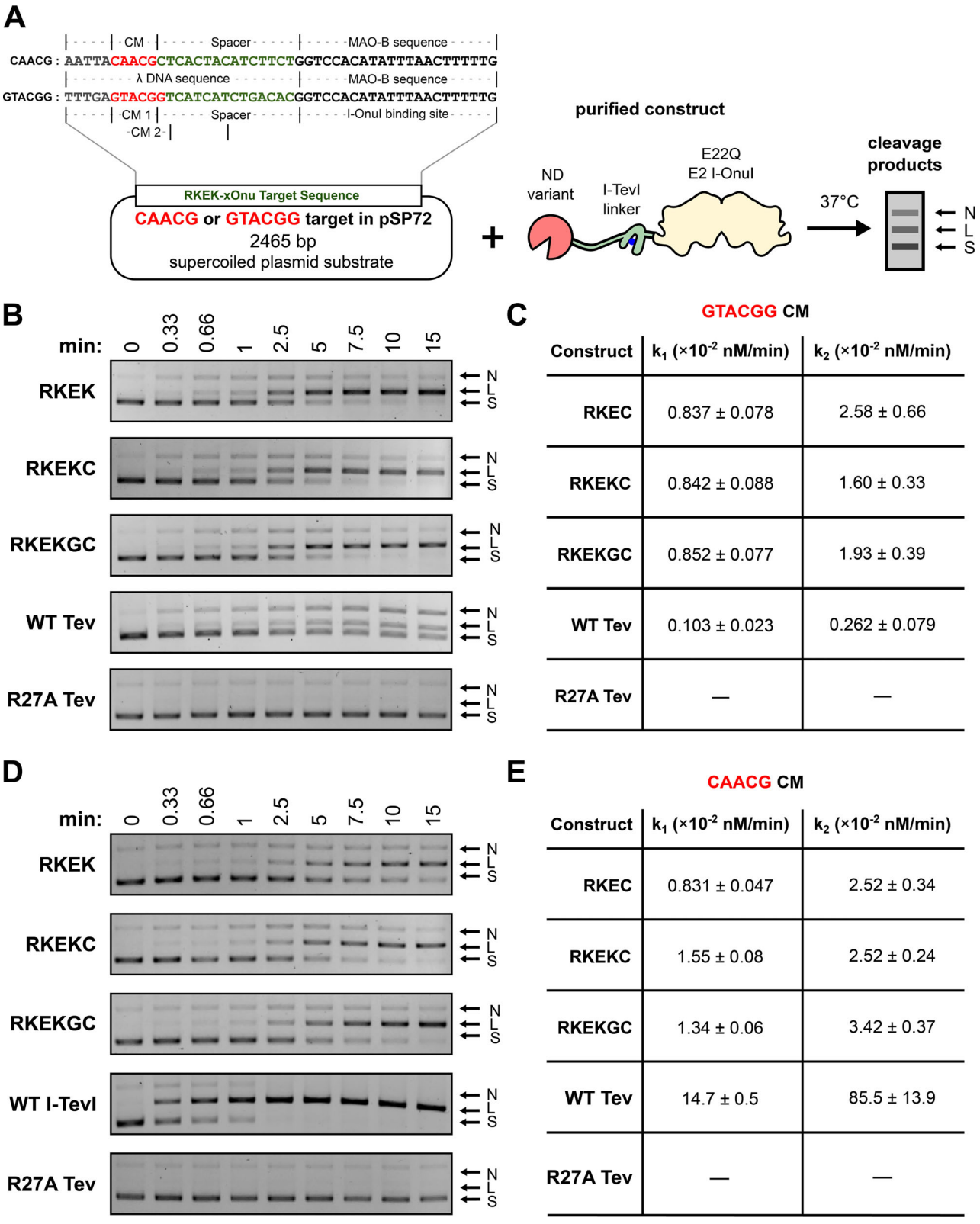


Figure 8. Kinetic analysis of evolved GIY-YIG variants on a nanopore-derived target. (**A**) Sequence schematic of Oxford Nanopore-derived RKEC-xOnu target, or cognate CAACG target embedded in pSP72 plasmid. The Nanopore-derived target contains a set of nested cleavage motifs (CM 1 and 2), with flanking λ sequence fused to the I-OnuI binding site. The cleavage assay measures progressive conversion of supercoiled DNA (S) to nicked (N) and fully linearized (L) products. (**B**) Timecourse cleavage of supercoiled plasmid substrate containing the RKEC-xOnu target resolved by agarose gel electrophoresis. N, L, S are defined as in panel (A) and indicated. (**C**) Rate constants for supercoiled loss (k_1) and linear product accumulation (k_2) are shown as mean ± SE (nM/min, $n = 3$). GTACGG substrate. Values are normalized to 2.56 nM substrate. (**D**) Timecourse cleavage of supercoiled plasmid substrate containing cognate I-TevI CAACG target resolved by agarose gel electrophoresis. Labeled as in panel (B). (**E**) Rate constants for supercoiled loss (k_1) and linear product accumulation (k_2) are shown and normalized as in panel (C). CAACG substrate. Uncropped gel images for panels (B) and (D) are available in Supplementary Fig. S11.

verted substrate to nicked and linear forms efficiently within 15 min, whereas wild-type I-TevI showed reduced activity (Fig. 8B). Kinetic analyses indicated ~10-fold faster rate constants for conversion of supercoiled to nicked intermediate (k_1) and nicked intermediate to linear product (k_2) for all variants relative to wild-type I-TevI (Supplementary Fig. S8). These findings confirm that evolved RKEC-class variants efficiently recognize and cleave a noncanonical target, and that activity improvements extend to sequences without CNNNG motifs.

Evolved variants (RKEC, RKEKC, RKEKGC) exhibited comparable activity on the CAACG substrate (Fig. 8A) relative to GTACGG, whereas the wild-type I-TevI ND cleaved CAACG ~147-fold faster (Fig. 8B–E). These results indicate a trade-off between catalytic efficiency and relaxed motif preference—an activity-specificity trade-off typical of evolved nucleases. This pattern is consistent with the NTCP accumulation profiles observed in Fig. 6 and confirms that the evolved activities are maintained across distinct C-terminal DNA-binding domains.

Discussion

The GIY-YIG nuclease domain is associated with various proteins that function in DNA repair, replication, retrotransposition, restriction systems, and regulation of DNA competence [26, 39, 50, 51], highlighting a remarkable adaptability to engage DNA through different mechanisms. For GIY-YIG homing endonucleases, the nuclease domain is characterized by sequence–tolerant interactions at a short cleavage motif (CNNNG for I-TevI) [9, 13]. Here, we use a layered approach to pinpoint elements within the GIY-YIG nuclease domain that regulate cleavage preference; first, by shuffling structural units between I-TevI and I-BmoI, we identified Block 2 and the adjacent loop as impacting cleavage preference; and second, by screening a library of 1024 possible substitutions at positions that differed between I-TevI and I-BmoI in this region, we isolated variants with altered cleavage profiles relative to the parental enzyme, including high activity on TAACG and GAACG sites.

The Block 2 substitutions (positions 30, 33, 36, 37, 38, 39) we identified clustered as a series of basic residues adjacent to the catalytic R27 and likely engage the negatively charged DNA backbone. This pattern of substitutions suggests a mechanism by which localized electrostatic tuning improves DNA substrate engagement without requiring specific base contacts. Indeed, the sequence logos profiles for the RKEC series of variants show a loss of information content across the cleavage site (apart from the G at position 5) that is consistent with relaxation of preference rather than a change in specificity. Of these residues, C39 was previously identified in a directed evolution study [34] but not tested for changes in substrate preference. Isolation of this residue in two separate directed evolution experiments reinforces the importance of this region in cleavage preference. The Block 2 loop may function as a transferable "preference switch" within GIY-YIG nucleases, enabling modular redefinition of cleavage behavior through rational recombination or mutagenesis. Residues near the Block 2–3 boundary (e.g. C39R) may participate in cooperative stabilization or preference tuning across this junction. Although Block 3 substitutions did not yield intermediate cleavage phenotypes, expanding block-exchange analyses beyond residue 39 may capture such coupling.

Structural studies of the GIY-YIG restriction enzymes Eco29kI and Hpy188I revealed that each enzyme reads out DNA substrate by a small number of base-specific contacts as compared to other restriction enzymes [24, 35]. Interestingly, the residues contacting DNA were localized in loops that connected structurally conserved regions; however, the position of these loops differed from the loop region in I-TevI and I-BmoI where we identified residues impacting cleavage preference. These findings indicate that the GIY-YIG nuclease domain can be divided into two functional units; the well-structured catalytic core [52, 53] and accessory loops that dictate sequence preference. From an evolutionary perspective, localizing residues that regulate substrate preference to variable loop regions provides a mechanism to rapidly evolve sequence specificity without directly impacting catalysis. Moreover, the separation of DNA cleavage from DNA binding function by the modular structure of GIY-YIG homing endonucleases [25] could further accelerate evolution of both cleavage and binding specificity.

For our study, we used two approaches to profile cleavage preference of evolved variants across synthetic and genomic substrates. CE is an improvement over past bar-coding approaches [10, 34] in that it offers a rapid, multiplexed, and scalable alternative to generate quantitative competitive cleavage activity across a defined panel of substrates without variable motif density of randomized libraries, or coverage requirements associated with NGS-based preference assessments. We assayed cleavage activity on a panel of 28 barcoded substrates, but we estimate this panel could be expanded to hundreds of substrates by using compatible fluorophores and substrate design to resolve individual cleavage products. In addition, the nontargeted cleavage assay [45] provides complementary information to the CE method, and profiles cleavage activity and specificity within the context of a complex genomic sequence. This approach provides different types of information from those obtained from NGS-based assays, which often use short, artificial substrates with low-complexity sequence composition. While our results were generated using λ genomic DNA, our methodology is easily scalable to more complex genomes such as *E. coli*, with ~100-times more information than λ.

Our studies were performed in the context of fusions of the chimeric GIY-YIG domains to different DNA-binding platforms (meganucleases and Cas9) to avoid toxicity of full-length I-TevI in *E. coli* and to highlight the portability of the GIY-YIG domain to different scaffolds. As such, our findings have implications for the design of programmable GIY-YIG-based genome editors. The overall specificity of GIY-YIG based editors is composed of interactions from both the nuclease and DNA-binding domains [30, 48, 54, 55]. This separation of cleavage from binding functions is especially important for applications where catalytic selectivity needs to be finely tuned; for example, in therapeutic genome editing scenarios requiring precise cleavage at variant motifs. Importantly, the underlying design principles for modifying GIY-YIG nuclease domain preference only requires accurate secondary structure predictions based on conservation analysis between related parent proteins. This low barrier to entry establishes a general strategy for structure-guided recombination and mutagenesis across diverse GIY-YIG nuclease scaffolds.

While this study provides a modular framework for rewiring nuclease preference within the GIY-YIG family, several limitations should be noted. First, we did not analyze

the contributions of the DNA spacer (located between the CNNNG motif and DNA-binding site and contacted by the I-TevI linker domain) to substrate preference. Previous directed evolution experiments identified residues in the I-TevI linker (positions 93–169) that enhanced activity on suboptimal DNA spacer sequences [48], including the VKN variant tested here. We cannot rule out the possibility that the I-TevI linker also contributes to cleavage site recognition and preference, namely at position T7 that is adjacent to the cleavage site [8, 48, 56]. Second, although recombination between I-TevI and I-BmoI revealed robust modularity, these proteins are isoschizomers and bind and cleave related target sites in thymidylate synthase genes [32]. It is unclear whether the preference-switching behaviour observed here will generalize to more distantly related GIY-YIG homing endonucleases. Third, while we were successful in modulating cleavage preference at the first position C and the central NNN triplets of the CNNNG motif, we have not yet identified variants that can tolerate substitutions in the 5th position G. This may be an inherent limitation of the I-TevI and I-BmoI scaffolds and domain shuffling with other GIY-YIG endonucleases that target non-GC-rich substrates may identify positions that regulate preference at this position. Fourth, the structural and biochemical diversity within the GIY-YIG superfamily may limit the portability of our findings to domains that function in different aspects of nucleotide metabolism, as discussed above for the GIY-YIG restriction enzymes Eco29kI and Hpy188I that function as dimers. Finally, while many recombinants were catalytically active, we did not systematically explore the determinants of inactivity in poorly performing variants, such as folding defects, active-site misalignment, or disruption of co-evolving residue networks previously identified in the GIY-YIG domain [47]. Expanding the parental pool, linker space, domain boundary diversity, and restoring integrity of co-evolving networks will be critical for testing the applicability of our strategy across the GIY-YIG superfamily.

In summary, our study illustrates how block-based DNA shuffling can exploit structural modularity within the GIY-YIG nuclease domain to evolve new substrate preferences. By systematically testing modular blocks and their impact on sequence preference, we identified key residues and structural regions that define a preference switch within the nuclease domain. With a broader repertoire of parental GIY-YIG nuclease domains, it may be possible to construct targeted libraries to yield variants with desired cleavage preferences against diverse substrates.

Acknowledgements

The authors thank Jenn Biltcliffe (London Regional Genomics Centre) for advice with capillary electrophoresis, and members of the Edgell lab and SBI for comments on the manuscript.

Author contributions: Kurt W. Loedige (Conceptualization [lead], Data curation [lead], Formal analysis [lead], Investigation [lead], Methodology [lead], Project administration [supporting], Software [lead], Supervision [supporting], Validation [lead], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal]), Mairwyn A. Hall (Conceptualization [supporting], Data curation [supporting], Formal analysis [supporting], Investigation [supporting], Methodology [supporting], Validation [supporting], Writing—review & editing [supporting]), Sabin Ghimirey (Conceptualization [equal], Formal analysis [supporting], Investigation [supporting], Methodology [supporting], Software [equal], Validation [supporting], Visualization [supporting], Writing—review & editing [supporting]), Brent E. Stead (Conceptualization [supporting], Formal analysis [supporting], Funding acquisition [equal], Project administration [supporting], Visualization [supporting], Writing—original draft [supporting], Writing—review & editing [supporting]), and David R. Edgell (Conceptualization [equal], Formal analysis [supporting], Funding acquisition [equal], Investigation [supporting], Methodology [supporting], Project administration [lead], Resources [supporting], Software [supporting], Supervision [lead], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal])

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

B.E.S. is an employee of Specific Biologics Inc. D.R.E. is a co-founder of Specific Biologics Inc. All other authors (K.W.L., M.A.H., S.B.) declare no conflict of interest.

Funding

This work was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) to D.R.E. (RGPIN-2022-05459) and an Alliance Grant from NSERC and MITACS Canada to D.R.E. and B.E.S. (ALLRP 571374-21). K.W.L. was supported by an Ontario Graduate Scholarship from the Government of Ontario. Funding for open acces charges: NSERC RGPIN-2022-05459.

Data availability

The Oxford Nanopore sequencing datasets have been deposited in the Sequence Read Archive under accession PRJNA1291786. Computer code for analyzing sequencing data is provided in the supplementary information.

References

1. Chevalier BS, Stoddard BL. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res* 2001;29:3757–74. https://doi.org/10.1093/nar/29.18.3757
2. Lambowitz AM, Belfort M. Introns as mobile genetic elements. *Annu Rev Biochem* 1993;62:587–622. https://doi.org/10.1146/annurev.bi.62.070193.003103
3. Sharma M, Ellis RL, Hinton DM. Identification of a family of bacteriophage T4 genes encoding proteins similar to those present in group I introns of fungi and phage. *Proc Natl Acad Sci USA* 1992;89:6658–62. https://doi.org/10.1073/pnas.89.14.6658
4. Belfort M, Derbyshire V, Parker MM *et al.* Mobile introns: pathways and proteins, *Mobile DNA II*, 2007, 759–83.
5. Bolduc JM, Spiegel PC, Chatterjee P *et al.* Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing factor. *Genes Dev* 2003;17:2875–88. https://doi.org/10.1101/gad.1109003
6. Ho Y, Kim SJ, Waring RB. A protein encoded by a group I intron in *Aspergillus nidulans* directly assists RNA splicing and is a DNA endonuclease. *Proc Natl Acad Sci USA* 1997;94:8994–9. https://doi.org/10.1073/pnas.94.17.8994

7. Lambowitz AM, Perlman PS. Involvement of aminoacyl-tRNA synthetases and other proteins in group I and group II intron splicing. *Trends Biochem Sci* 1990;15:440–4. https://doi.org/10.1016/0968-0004(90)90283-H
8. Edgell DR, Derbyshire V, Roey PV *et al.* Intron-encoded homing endonuclease I-TevI also functions as a transcriptional autorepressor. *Nat Struct Mol Biol* 2004;11:936–44. https://doi.org/10.1038/nsmb823
9. Bryk M, Quirk S, Mueller J *et al.* The *td* intron endonuclease I-TevI makes extensive sequence-tolerant contacts across the minor groove of its DNA target. *EMBO J* 1993;12:2141–9. https://doi.org/10.1002/j.1460-2075.1993.tb05862.x
10. Li H, Ulge UY, Hovde BT *et al.* Comprehensive homing endonuclease target site specificity profiling reveals evolutionary constraints and enables genome engineering applications. *Nucleic Acids Res* 2012;40:2587–98. https://doi.org/10.1093/nar/gkr1072
11. Argast GM, Stephens KM, Emond MJ *et al.* I-PpoI and I-CreI homing site sequence degeneracy determined by random mutagenesis and sequential *in vitro* enrichment. *J Mol Biol* 1998;280:345–53. https://doi.org/10.1006/jmbi.1998.1886
12. Chevalier B, Turmel M, Lemieux C *et al.* Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-CreI and I-MsoI. *J Mol Biol* 2003;329:253–69. https://doi.org/10.1016/S0022-2836(03)00447-9
13. Edgell DR, Stanger MJ, Belfort M. Coincidence of cleavage sites of intron endonuclease I-TevI and critical sequences of the host thymidylate synthase gene. *J Mol Biol* 2004;343:1231–41. https://doi.org/10.1016/j.jmb.2004.09.005
14. Scalley-Kim M, McConnell-Smith A, Stoddard BL. Coevolution of a homing endonuclease and its host target sequence. *J Mol Biol* 2007;372:1305–19. https://doi.org/10.1016/j.jmb.2007.07.052
15. Bogdanove AJ, Bohm A, Miller JC *et al.* Engineering altered protein–DNA recognition specificity. *Nucleic Acids Res* 2018;46:4845–71. https://doi.org/10.1093/nar/gky289
16. Stoddard BL. Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. *Structure* 2011;19:7–15. https://doi.org/10.1016/j.str.2010.12.003
17. Takeuchi R, Choi M, Stoddard BL. Redesign of extensive protein–DNA interfaces of meganucleases using iterative cycles of *in vitro* compartmentalization. *Proc Natl Acad Sci* 2014;111:4061–6. https://doi.org/10.1073/pnas.1321030111
18. McMurrough TA, Brown CM, Zhang K *et al.* Active site residue identity regulates cleavage preference of LAGLIDADG homing endonucleases. *Nucleic Acids Res* 2018;46:11990–2007.
19. Saravanan M, Vasu K, Nagaraja V. Evolution of sequence specificity in a restriction endonuclease by a point mutation. *Proc Natl Acad Sci USA* 2008;105:10344–7. https://doi.org/10.1073/pnas.0804974105
20. Zhu Z, Zhou J, Friedman AM *et al.* Isolation of BsoBI restriction endonuclease variants with altered substrate specificity. *J Mol Biol* 2003;330:359–72. https://doi.org/10.1016/s0022-2836(03)00595-3
21. Chan Sh, Bao Y, Ciszak E *et al.* Catalytic domain of restriction endonuclease BmrI as a cleavage module for engineering endonucleases with novel substrate specificities. *Nucleic Acids Res* 2007;35:6238–48. https://doi.org/10.1093/nar/gkm665
22. Chevalier BS, Kortemme T, Chadsey MS *et al.* Design, activity, and structure of a highly specific artificial endonuclease. *Mol Cell* 2002;10:895–905. https://doi.org/10.1016/S1097-2765(02)00690-1
23. Li L, Wu LP, Chandrasegaran S. Functional domains in Fok I restriction endonuclease. *Proc Natl Acad Sci USA* 1992;89:4275–9. https://doi.org/10.1073/pnas.89.10.4275
24. Mak ANS, Lambert AR, Stoddard BL. Folding, DNA recognition, and function of GIY-YIG endonucleases: crystal structures of R. Eco29kI. *Structure* 2010;18:1321–31. https://doi.org/10.1016/j.str.2010.07.006
25. Derbyshire V, Kowalski JC, Dansereau JT *et al.* Two-domain structure of the *td* intron-encoded endonuclease I-TevI correlates with the two-domain configuration of the homing site. *J Mol Biol* 1997;265:494–506.https://doi.org/10.1006/jmbi.1996.0754
26. Dunin-Horkawicz S, Feder M, Bujnicki JM. Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics* 2006;7:1–19. https://doi.org/10.1186/1471-2164-7-98
27. Kowalski JC, Belfort M, Stapleton MA *et al.* Configuration of the catalytic GIY-YIG domain of intron endonuclease I-TevI: coincidence of computational and molecular findings. *Nucleic Acids Res* 1999;27:2115–25. https://doi.org/10.1093/nar/27.10.2115
28. Van Roey P, Meehan L, Kowalski JC *et al.* Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI. *Nat Struct Biol* 2002;9:806–11.
29. Bell-Pedersen D, Quirk S, Clyman J *et al.* Intron mobility in phage T4 is dependent upon a distinctive class of endonucleases and independent of DNA sequences encoding the intron core: mechanistic and evolutionary implications. *Nucleic Acids Res* 1990;18:3763–70. https://doi.org/10.1093/nar/18.13.3763
30. Wolfs JM, DaSilva M, Meister SE *et al.* MegaTevs: single-chain dual nucleases for efficient gene disruption. *Nucleic Acids Res* 2014;42:8816–29. https://doi.org/10.1093/nar/gku573
31. Bell-Pedersen D, Quirk SM, Bryk M *et al.* I-TevI, the endonuclease encoded by the mobile *td* intron, recognizes binding and cleavage domains on its DNA target. *Proc Natl Acad Sci USA* 1991;88:7719–23. https://doi.org/10.1073/pnas.88.17.7719
32. Edgell DR, Shub DA. Related homing endonucleases I-BmoI and I-TevI use different strategies to cleave homologous recognition sites. *Proc Natl Acad Sci* 2001;98:7898–903. https://doi.org/10.1073/pnas.141100098
33. Edgell DR, Stanger MJ, Belfort M. Importance of a single base pair for discrimination between intron-containing and intronless alleles by endonuclease I-BmoI. *Curr Biol* 2003;13:973–8. https://doi.org/10.1016/S0960-9822(03)00340-3
34. Roy AC, Wilson GG, Edgell DR. Perpetuating the homing endonuclease life cycle: identification of mutations that modulate and change I-TevI cleavage preference. *Nucleic Acids Res* 2016;44:7350–9.
35. Sokolowska M, Czapinska H, Bochtler M. Hpy188I–DNA pre-and post-cleavage complexes—snapshots of the GIY-YIG nuclease mediated catalysis. *Nucleic Acids Res* 2011;39:1554–64. https://doi.org/10.1093/nar/gkq821
36. Verhoeven EE, van Kesteren M, Moolenaar GF *et al.* Catalytic sites for 3′ and 5′ incision of *Escherichia coli* nucleotide excision repair are both located in UvrC. *J Biol Chem* 2000;275:5120–3. https://doi.org/10.1074/jbc.275.7.5120
37. Gaur V, Wyatt HD, Komorowska W *et al.* Structural and mechanistic analysis of the Slx1–Slx4 endonuclease. *Cell Rep* 2015;10:1467–76. https://doi.org/10.1016/j.celrep.2015.02.019
38. Ibryashkina EM, Zakharova MV, Baskunov VB *et al.* Type II restriction endonuclease R. Eco29kI is a member of the GIY-YIG nuclease superfamily. *BMC Struct Biol* 2007;7:1–10. https://doi.org/10.1186/1472-6807-7-48
39. Pyatkov KI, Arkhipova IR, Malkova NV *et al.* Reverse transcriptase and endonuclease activities encoded by Penelope-like retroelements. *Proc Natl Acad Sci* 2004;101:14719–24. https://doi.org/10.1073/pnas.0406281101
40. Stemmer WP. Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* 1994;370:389–91. https://doi.org/10.1038/370389a0
41. Crameri A, Raillard SA, Bermudez E *et al.* DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* 1998;391:288–91. https://doi.org/10.1038/34663
42. Barriault D, Plante MM, Sylvestre M. Family shuffling of a targeted bphA region to engineer biphenyl dioxygenase. *J Bacteriol* 2002;184:3794–800. https://doi.org/10.1128/JB.184.14.3794-3800.2002
43. Carbone MN, Arnold FH. Engineering by homologous recombination: exploring sequence and function within a

conserved fold. *Curr Opin Struct Biol* 2007;17:454–9. https://doi.org/10.1016/j.sbi.2007.08.005
44. Liu Q, Derbyshire V, Belfort M *et al.* Distance determination by GIY-YIG intron endonucleases: discrimination between repression and cleavage functions. *Nucleic Acids Res* 2006;34:1755–64. https://doi.org/10.1093/nar/gkl079
45. Loedige KW, White AL, McMurrough TA *et al.* A buffer-tuning strategy to profile domain-specific activity of chimeric I-TevI/CRISPR gene editors *in vitro*. *Sci Rep* 2025;15:42742. https://doi.org/10.1038/s41598-025-26759-7
46. Liu Q, Dansereau JT, Puttamadappa SS *et al.* Role of the interdomain linker in distance determination for remote cleavage by homing endonuclease I-TevI. *J Mol Biol* 2008;379:1094–106. https://doi.org/10.1016/j.jmb.2008.04.047
47. Kleinstiver BP, Fernandes AD, Gloor GB *et al.* A unified genetic, computational and experimental framework identifies functionally relevant residues of the homing endonuclease I-BmoI. *Nucleic Acids Res* 2010;38:2411–27. https://doi.org/10.1093/nar/gkp1223
48. Wolfs JM, Hamilton TA, Lant JT *et al.* Biasing genome-editing events toward precise length deletions with an RNA-guided TevCas9 dual nuclease. *Proc Natl Acad Sci USA* 2016;113:14988–93. https://doi.org/10.1073/pnas.1616343114
49. Lanio T, Jeltsch A, Pingoud A. On the possibilities and limitations of rational protein design to expand the specificity of restriction enzymes: a case study employing EcoRV as the target. *Protein Eng* 2000;13:275–81. https://doi.org/10.1093/protein/13.4.275
50. Chenal M, Rivera-Millot A, Harrison LB *et al.* Discovery of the widespread site-specific single-stranded nuclease family Ssn. *Nat Commun* 2025;16:2388. https://doi.org/10.1038/s41467-025-57514-1
51. Aravind L, Walker DR, Koonin EV. Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res* 1999;27:1223–42. https://doi.org/10.1093/nar/27.5.1223
52. Van Roey P, Waddling CA, Fox KM *et al.* Intertwined structure of the DNA-binding domain of intron endonuclease I-TevI with its substrate. *EMBO J* 2001;20:3631–7. https://doi.org/10.1093/emboj/20.14.3631
53. Truglio JJ, Rhau B, Croteau DL *et al.* Structural insights into the first incision reaction during nucleotide excision repair. *EMBO J* 2005;24:885–94. https://doi.org/10.1038/sj.emboj.7600568
54. Kleinstiver BP, Wolfs JM, Kolaczyk T *et al.* Monomeric site-specific nucleases for genome editing. *Proc Natl Acad Sci USA* 2012;109:8061–6. https://doi.org/10.1073/pnas.1117984109
55. Kleinstiver BP, Wang L, Wolfs JM *et al.* The I-TevI nuclease and linker domains contribute to the specificity of monomeric TALENs. *G3: Genes, Genomes, Genetics* 2014;4:1155–65. https://doi.org/10.1534/g3.114.011445
56. Dean AB, Stanger MJ, Dansereau JT *et al.* Zinc finger as distance determinant in the flexible linker of intron endonuclease I-TevI. *Proc Natl Acad Sci USA* 2002;99:8554–61. https://doi.org/10.1073/pnas.082253699

Received: July 22, 2025. **Revised:** November 10, 2025. **Accepted:** November 22, 2025