Biotechnology Reports 45 (2025) e00877

Contents lists available at ScienceDirect

Biotechnology Reports

journal homepage: www.elsevier.com/locate/btre

Research Article

Identification and purification of a novel bacteriophage T7 endonuclease from the Kogelberg Biosphere Reserve (KBR) biodiversity hotspot

Priyen Pillay [a], Maabo Moralo [a], Sibongile Mtimka [a], Taola Shai [a], Kirsty Botha [a,b], Lusisizwe Kwezi [a], Tsepo L. Tsekoa [a,*]

[a] Chemicals Cluster, Council for Scientific and Industrial Research (CSIR), Pretoria, South Africa
[b] Department of Plant and Soil Sciences, University of Pretoria, Hillcrest, South Africa

ARTICLE INFO

Keywords:
T7 endonuclease I
Holliday junction resolvase/nuclease
DNA-protein interaction
Nucleases
Junction-resolving enzyme
Genome editing detection
Crispr/Cas9

ABSTRACT

The four-way (Holliday) DNA junction is a key intermediate in homologous recombination, a ubiquitous process that is important in DNA repair and generation of genetic diversity. The final stages of recombination require resolution of the junction into nicked-duplex species by the action of a junction-resolving enzyme. The enzymes involved are nucleases that are highly selective for the structure of branched DNA. Here we present the isolation, expression and purification of a novel T7 endonuclease from the Kogelberg Biosphere Reserve (KBR), which possesses junction resolving capabilities. An initial approach was employed where the process was scaled up to 3 L with IPTG concentration of 0.1 mM at 30 °C and purified via immobilised metal affinity chromatography (IMAC). Expression titres of 20 ± 0.003 $\mu g.L^{-1}$ culture were achieved with the amount of KBR-T7 endonuclease required per reaction ranging from as low as 10 to 100 nanograms. The solubility of the enzyme was relatively poor; however, enzyme activity was not affected. A derivative for improved solubility and efficacy was then designed from this original wild-type version, MBP-KBR-T7 and was expressed under similar conditions at 20 °C yielding 1.63 ± 0.154 $mg.L^{-1}$ of formulated enzyme. This novel high value enzyme derivative is a valuable asset within the molecular reagent space as a tool for confirming both *in vivo* and *in vitro* genome editing; therefore, a means to produce it recombinantly in a scalable and technoeconomicaly viable process is highly desirable.

1. Introduction

The four-way (Holliday) DNA junction is a key intermediate in homologous recombination, a ubiquitous process that is important in DNA repair and generation of genetic diversity [1]. The final stages of recombination require resolution of the junction into nicked-duplex species by the action of a junction-resolving enzyme [2]. The enzymes involved are nucleases that are highly selective for the structure of branched DNA [3–6]. DNA endonucleases capable of catalyzing the cleavage of DNA mismatches and non-β DNA structures including Holliday junctions and cruciform leaving 3′-OH and 5′-phosphate are becoming key enzymes for the use in genome editing mutation detection workflows prior to extensive sequencing efforts. This necessitates the need to identify and utilize existing and novel DNA endonucleases as key complements to the genome editing workflow. Using this T7 endonuclease I (T7EI) genomic-cleavage based detection method, one can quickly measure on-target genome editing efficiency generated by non-homologous end joining (NHEJ) activity. There are some popular kits that have been developed by Invitrogen[1] and New England Biolabs (NEB).[2] Briefly, the required materials behind these detection workflows are: a taq polymerase, the T7 Endonuclease I, in some kits EDTA, purified genomic DNA from untargeted and targeted cells, PCR primers which amplify a ~1 kb region around the target site where the target site should ideally be within the center of the amplicon which facilitate resolution of DNA fragments post T7 endonuclease I digestion. The premise behind the assay lies in amplifying non-edited and edited targets separately and then bringing those two elements together in a PCR thermocycler through annealing. Those templates are then restricted with the T7 endonuclease enzyme which is designed to identify Holliday junctions around artificial mismatches created by the annealing process between edited and non-edited amplicons.

There are many methods for termination of the T7 endonuclease

* Corresponding author at: P.O Box 395, Pretoria, 0001, South Africa.
E-mail address: TTsekoa@csir.co.za (T.L. Tsekoa).

[1] https://www.thermofisher.com/order/catalog/product/A24372
[2] https://www.neb.com/en/protocols/2014/08/11/determining-genome-targeting-efficiency-using-t7-endonuclease-i

https://doi.org/10.1016/j.btre.2025.e00877
Received 16 September 2024; Received in revised form 7 January 2025; Accepted 22 January 2025
Available online 23 January 2025

reaction which employ the use of EDTA, Proteinase K and 10 % SDS. In this study, we found that a combination of all three reagents yielded the best results. Furthermore, we also found that using a hybrid purification strategy employing both the conventional ethanol precipitation and a purification kit not only accentuates bands upon resolution, but also preserves them from accidental loss during subsequent ethanol precipitation centrifugation steps and recommend this hybrid strategy or other appropriate clean kits to users. DNA fragments are then analysed on Bioanalyzer instruments or standard agarose/acrylamide gel electrophoresis [7,8]. A key consideration for the success of such genome editing detection workflows lies within careful primer design taking cognisance of the target organism's ploidy level especially homologous genes that result from allopolyploidy which are commonly referred to as 'homoeologs' [9]. Allopolyploidy, which is a type of whole-genome duplication via hybridization followed by genome doubling [10], results in multiple copies of genes bearing considerable similarity to one another. Therefore, in order for the T7 genome editing assay to effectively work for scenarios such as these, primers must be designed to specifically detect a single version of a copy of a gene within the genome of interest, be it humans, plants, bacteria or viruses [11].

The Kogelberg Biosphere Reserve (KBR) in the Cape Floral Kingdom in South Africa is known for its unique plant biodiversity and the potential presence of unique microbial and viral biodiversity associated with this unique plant biodiversity led us to explore the fynbos soil using existing metaviromic resources for novel DNA endonucleases [12]. Segobola *et al.* (2018) found through functional analysis and other metaviromes showed a relatively high frequency of phage-related and structural proteins. The novel phage sequences detected, present an opportunity for future studies aimed at targeting novel genetic resources for applied biotechnology.

The dataset from the KBR was analysed and translated into all six open reading frames. The T7 endonuclease motif, 'TKGLWXXXD,' was used to identify novel T7 endonucleases, with specific annotation of the residues critical for Holliday junction (HJ) cleavage activity (D and K). The full-length novel T7 endonuclease I with an N-terminal 6x histidine affinity (HIS) tag was then expressed in *Escherichia coli* (*E. coli*, Rosetta DE3 pLysS) Here, we present the identification and purification of a novel KBR-T7 endonuclease as well as a second engineered derivative, MBP-KBR-T7 that indicated HJ cleavage activity. The wildtype novel T7 endonuclease was used as a foundation for adding a solubility fusion partner MBP (maltose binding protein) [13], a domain for improving affinity to Holliday junctions as well as an additional Histidine tag to the N-terminal of the protein for better purification. The MBP protein is an approximately 42 kDa, naturally occurring protein in *E. coli*, encoded by the *malE* gene, which is responsible for the uptake, breakdown and transport of maltodextrin, a carbohydrate. It was originally developed as a protein expression tag in the 1980s, but now, it is known to significantly enhance the solubility of a variety of fusion proteins [14]. The Histidine tag is routinely used for protein purification and in detection and the possession of two His tags are usually found to be sufficient for stable binding [15].

2. Materials and methods

2.1. Identification of KBR-T7 endonuclease, secondary and tertiary structure analysis

The dataset 5549_whole from the KBR was analysed and translated into all six frames. Endonuclease (Enterobacteria phage T7.1, Genbank Acc. Number - AAZ32838.1) was used as a comparative model sequence. The T7 endonuclease motif TKGLWXXXD, was used where the residues that significantly affect Holliday junction cleavage activity (D, K) are highlighted in blue to search in CLC bio for potential candidates. A candidate was found, Ext1_S37_L001_R1_001_paired_trimmed_p_8076 (+1) selection. Multiple sequence alignments were generated with PROMALS3D using standard parameters [16]. We submitted both the novel KBR T7 endonuclease (Genbank Acc. Number - PP165528) to NCBI for (https://www.ncbi.nlm.nih.gov/nuccore/PP165528.1/) blast homology analysis [17] and also used BLASTP 2.2.26 [18,19]. We also used Uniprot (https://www.uniprot.org/) and used the blast function (https://www.uniprot.org/blast/). We also confirmed that the native KBR-T7 DNA sequence is flanked with stop codons at the 5′ and 3′ ends. Secondary structure prediction and secondary-structure-based threading were carried out by using the NPS@: Network Protein Sequence Analysis [20]. Swiss-Model was also used to annotate and structurally characterize the protein [21]. I-TASSER modelling was used which starts from the structure templates identified by LOMETS from the PDB library [22]. The LOMETS server is built from compiling predictions from nine different servers that represent a diverse array of state-of-the-art threading algorithms in order to allow the user to construct the best possible 3D models with high accuracy and works in tandem with I-TASSER [23]. Within I-TASSER, the confidence of each model is quantitatively measured by C-score that is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. The C-score is a confidence score metric for estimating the quality of predicted models produced by I-TASSER calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-scores are typically in the range of [−5,2], where C-scores of higher values signify models with a high confidence levels and vice-versa.

2.2. Expression and purification of novel T7 endonuclease

The pET-30b(+)-KBR-T7 and pET-30b(+)-MBP-KBR-T7 vectors were synthesized by GenScript with flanking restriction enzyme sites, *Nde*I (CATATG) and *Xho*I (CTCGAG). The standard transformation protocol was followed, using MAX Efficiency® DH5α™ Competent Cells (Thermo Fisher Scientific, MA, USA) and Rosetta™(DE3) pLysS cells (Merck). Restriction enzyme digests were conducted to confirm the sequence identity of the constructs within each respective host *E. coli* strain. Bacterial synthesis of KBR-T7 was carried out in the T7 RNA polymerase expression system [24]. A volumentric flask supplemented with Kanamycin (Sigma-Aldrich, 30 μg.ml-1) and Chloramphenicol (Sigma-Aldrich, 34 μg.ml-1) with pET-30b(+)-KBR-T7 Rosetta was incubated overnight at 37 °C. A volume of inoculum from the flask was used to further inoculate Luria-Bertani (LB) medium (1 L) supplemented with the above-mentioned antibiotics and grown for ~4 h at 37 °C to an optical density OD of 0.4 at 200 rpm. This was scaled to a final volume of 3 L for pET-30b(+)-KBR-T7 and 1 L for pET-30b(+)-MBP-KBR-T7, respectively. Expression was induced with 0.1 mM IPTG (Sigma-Aldrich) and was performed for 4 h at 30 °C and 20 °C, respectively with shaking at 200 rpm. A volume of culture was sampled and spun down at 12,000 *g* for 2 min, sonicated and resuspended in 200 μL of B-PER buffer (Thermo Fisher Scientific) and spun down once again at 12,000 *g* for 2 min. The supernatant recovered represented the soluble fraction and the pellet, the insoluble fraction, was resuspended in 1x PBS buffer and kept for further downstream analysis. Cells were harvested at 10,000 *g* for 10 min at 4 °C and pellets were stored at −80 °C. Pellets were resuspended in Equilibration buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM Imidazole, pH 7.4) where the pellet weight represented 10 % of the final volume (10% w/v) and rest was Equilibration buffer. B-PER was also added to Equilibration buffer in a 1:10 ratio to facilitate lysis. A French press (Constant Cell Disruption Systems) was used for cell breakage using a single pass. This mixture was sonicated on ice with the following parameters, 2 × 15 s with a 15 s cooling period between each burst. Samples were centrifuged at 10,000 *g* for 30 min at 4 °C until the supernatant was clear. For column equilibration, 1x Equilibration buffer was allowed to pass through Ni-TED 2000 columns (Machery-Nagel). For binding, the clarified lysate was loaded onto column once. A sample of clarified lysate and flow-through was collected for downstream analysis. For washing, the sample was washed twice with 1x Wash buffer

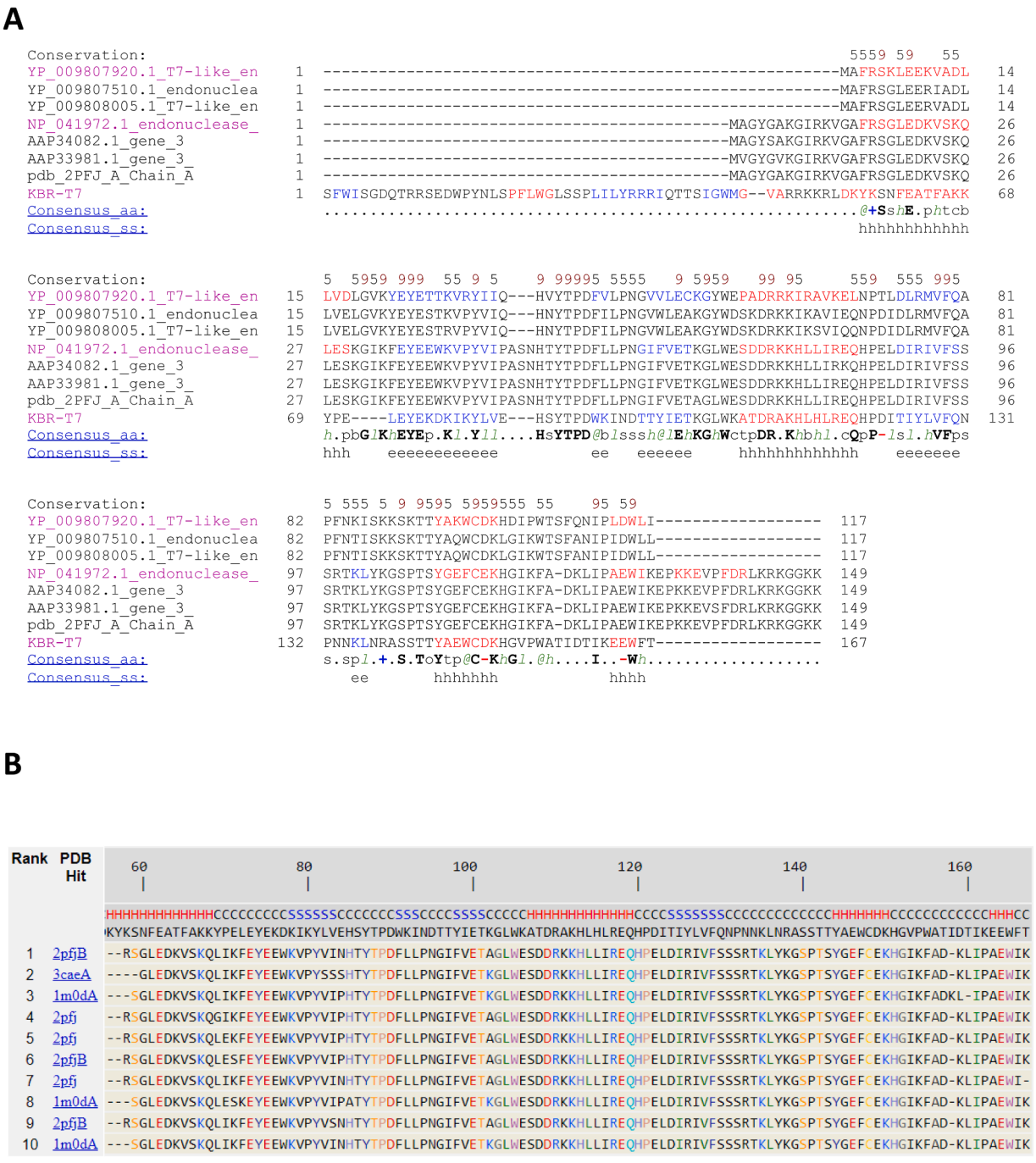

Fig. 1. A. Multiple sequence alignment of the KBR-T7 and T7-like endonucleases. First line in each block shows conservation indices for positions with a conservation index above 5. Each representative sequence has a magenta name and is colored according to PSIPRED secondary structure predictions (red: alpha-helix, blue: beta-strand). A representative sequence and the immediate sequences below it with black names, if there are any, form a closely related group (determined by option "Identity threshold"). The last two lines show consensus amino acid sequence (Consensus_aa) and consensus predicted secondary structures (Consensus_ss). Representative sequences have magenta names and they are colored according to predicted secondary structures (red: alpha-helix, blue: beta-strand). Consensus predicted secondary structure symbols: alpha-helix: h; beta-strand: e. Consensus amino acid symbols are: conserved amino acids are in bold and uppercase letters; aliphatic (I, V, L): l; aromatic (Y, H, W, F): @; hydrophobic (W, F, Y, M, L, I, V, A, C, T, H): h; alcohol (S, T): o; polar residues (D, E, H, K, N, Q, R, S, T): p; tiny (A, G, C, S): t; small (A, G, C, S, V, N, D, T, P): s; bulky residues (E, F, I, K, L, M, Q, R, W, Y): b; positively charged (K, R, H): +; negatively charged (D, E):; charged (D, E, K, R, H): c. B. I-TASSER modeling from the structure templates identified by LOMETS from the PDB library. The templates in this figure are the 10 best templates selected from the LOMETS threading programs.

(50 mM NaH_2PO_4, 300 mM NaCl, 40 mM Imidazole, pH 7.4). After each wash, a sample was also taken. The elution was conducted 3-times with the elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0, 300 mM imidazole).

2.3. Protein analysis, formulation & quantification

Samples were analysed on a Bolt™ 4–12 % Bis-Tris Plus gel system (Thermo Fisher Scientific) using Coomassie Staining Solution. Samples were then transferred onto Immun-Blot® PVDF Membrane (Bio-Rad, Hercules, CA) for western blotting using the Trans-Blot® Turbo™

Table 1
Sequences producing significant alignments from the Uniprot Database with the KBR-T7 protein sequence.

Name	Organism	Description	GO-Molecular Function	Score (bits)	E value
A0A4Q3LT04	Oxalobacteraceae bacterium	[Bacteria] Endonuclease I [Proteobacteria]	*	113	5e-30
A0A2E8A7P9	Opitutae bacterium	[Bacteria] Uncharacterized protein	*	103	9e-26
A0A178HAH5	Agrobacterium rhizogenes	[Bacteria] Uncharacterized protein	*	100	5e-25
A0A1Y0T0 × 9	Pasteurella phage vB_PmuP_PHB02	[Viruses] Endonuclease I [Caudovirales]	*	99	4e-24
A0A2D8HU08	Idiomarina sp.	[Bacteria] Uncharacterized protein	*	99	4e-24
A0A2E0NDY0	Coraliomargarita sp.	[Bacteria] Uncharacterized protein	*	98	6e-24
A0A2D6MFP8	Candidatus Pacearchaeota archaeon	[Archaea] Endonuclease I	*	97	2e-23
A0A2E9IK52	Euryarchaeota archaeon	[Archaea] Endonuclease I [Euryarchaeota]	*	96	3e-23
A0A2R7SDS8	Pseudomonas sp. HMWF010	[Bacteria] Endonuclease I [Proteobacteria]	*	95	2e-22
E3SN95	Cyanophage NATL1A-7	[Viruses] Endonuclease [Caudovirales]	*	94	3e-22
A0A430DBK7	Sphingomonas sp. S-NIH.Pt3_0716	[Bacteria] Endodeoxyribonuclease	*	93	3e-22
A0A1M3DS08	Alphaproteobacteria bacterium 43–37	[Bacteria] Uncharacterized protein	*	94	4e-22
A0A0F7L9U9	Uncultured marine virus	[Viruses] Endodeoxyribonuclease	*	93	8e-22

* GO-Molecular function - deoxyribonuclease IV (phage-T4-induced) activity, DNA Binding.

Blotting System (Bio-Rad, Hercules, CA). Samples were transferred with the following conditions: 1.3 Amps (A), 25 Volts (V) for 20 min. The membrane was blocked in PBST buffer (pH 8) with 5 % fat-free skim milk for 1 hr at 4 °C and then incubated with Anti-6x-His-HRP Tag Mouse Monoclonal Antibody (Sigma) (1:3333 dilution) with 5 % milk O/N at 4 °C and then washed with PBST for 3 times for 10 min each time at room temperature (RT). Detection was performed using the Clarity™ ECL Western Blotting Substrate Kit (Bio-Rad, Hercules, CA). Western blot images were visualized on the ChemiDoc MP System, Biorad, Universal Hood III (Bio-Rad, Hercules, CA). In addition to the Clarity™ ECL Western Blotting Substrate, 3,3′,5,5′-Tetramethylbenzidine (TMB) Substrate Systems (Sigma-Aldrich) was also used for sample detection. The concentrations for MBP-KBR-T7 and KBR-T7 endonuclease I were determined using the Bradford method using appropriate serial dilutions of BSA (20 $mg.mL^{-1}$ stock concentration) as a standard. The eluted proteins were dialyzed and formulated into 20 mM Tris–HCl / 200 mM NaCl / 0.1 mM EDTA / 50 % glycerol / 0.15 % Triton® X-100 / 1 mM DTT, pH 7.5.

2.4. Preparation of annealed DNA substrates and cleavage assays

Plant DNA was extracted from plants using the DNAeasy® Plant Mini Kit (Qiagen, Germany). Primers that span a protease gene region 250 bp on either side of the guide RNA were designed giving rise to a 520 bp amplicon. A PCR was set up using the standard procedure for the Q5 High-Fidelity 2X Master Mix (NEB) for the amplification of the test gene from genomic DNA from negative control cells and edited genomic DNA from targeted cells. The following PCR conditions were used: initial denaturation – 30 s at 98 °C, for 35 cycles Denaturation – 5 s at 98 °C, Annealing – 10 s at 67 °C, Elongation – 72 °C, final extension – 2 min at 72 °C. These amplicons were purified using the GeneJET PCR Purification Kit (Thermo Scientific™) with a final elution volume of 30 µL. To determine genome targeting efficiency of the novel T7 Endonuclease I (KBR-T7), DNA fragments post hybridization and T7 Endonuclease I (NEB) digestion were eluted in 20 µl of nuclease-free water and visualized on either a 5 % TBE acrylamide gel or a 2.5 % agarose gel. The reaction buffer (1x or 10x) was prepared as follows, 50 mM NaCl / 10 mM Tris–HCl / 10 mM $MgCl_2$ / 1 mM DTT, pH 7.9 for 1x and 500 mM NaCl / 100 mM Tris–HCl / 100 mM $MgCl_2$, 10 mM DTT pH 7.9. The reaction setup to test efficacy of KBR-T7 endonuclease efficacy was conducted as follows: 200 ng of each amplicon with appropriate controls, 1x or 10x Reaction buffer, ddH_2O up till 19 µL. The PCR products were then annealed in a thermocycler using the following conditions: Initial Denaturation – 5 min at 95 °C, Annealing: 95 – 85 °C (Ramp rate −2 °C/second), 85 – 25 °C (Ramp rate −0.1 °C/second). Thereafter, T7 endonuclease was added to each reaction in varying amounts. After 45 min at 37 °C, the reactions were terminated by the addition of a stop solution, 1.5 µl of 0.25 M EDTA, 1 µl Proteinase K and 2 µL of 10 % SDS. Thereafter, a hybrid purification method using ethanol precipitation of DNA was conducted overnight at −20 °C using 0.1 vol of 3 M NaOAc solution and 2.5 - 3 vol of ice cold 100 % Ethanol. Thereafter, we used the Thermo Fisher PCR purification kit, as per the manufacturer's guidelines, to purify the digested DNA fragments which were eluted in 20 µl of water and analysed on either a 5 % acrylamide gel or a 2.5 % agarose gel.

3. Results

3.1. Discovery and annotation of novel T7 endonuclease from the KBR

Sequences producing significant alignments with T7-like endonuclease annotation were used to create a multiple sequence alignment with the KBR-T7 endonuclease protein sequence (Fig. 1). Table 1 lists the sequences producing significant alignments from the Uniprot Database. The characteristic motifs of the RuvC family include acidic residues that are usually an aspartate near the end of strand 1, a glutamate near the end of the conserved strand 4, and two aspartates (DXXD) embedded in the C-terminal helix (Fig. 1). The catalytic residues and the adjacent secondary structural elements are perfectly conserved in the KBR-T7 sequence (Fig. 1) which shows that this protein is an active resolvase. Fig. 2a and b illustrate structural elements of the novel KBR-T7 endonuclease modelled by Network Protein Sequence Analysis and Swiss-Model, respectively. Swiss-Model found that the KBR-T7 sequence was closely related to the Bacteriophage T7 endonuclease I [25] having a 43 % sequence similarity with a 67 % sequence coverage. The oligomeric state of the protein was predicted to be a homo-dimer according to Swiss-Model. The full list of templates used by Swiss-Model matching the KBR-T7 target sequence includes the following templates (Table 2). The theoretical pI/Mw: 9.30 / 20,054.71 was determined using Expasy (https://web.expasy.org/compute_pi/) [26]. I-TASSER modelling was also used which starts from the structure templates identified by LOMETS from the PDB library (Table 3). I-TASSER only used the templates of the highest significance in the threading alignments, the significance of which are measured by the Z-score, *i.e.* the difference between the raw and average scores in the unit of standard deviation. LOMETS is a meta-server threading approach containing multiple threading programs, where each threading program can generate tens of thousands of template alignments. The templates in this section were the ten best templates selected from the LOMETS threading programs. For each target, I-TASSER uses the SPICKER program to cluster all the decoys based on the pair-wise structure similarity and reports up to five models which corresponds to the five largest structure clusters. The confidence of each model is quantitatively measured by C-score that is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [−5, 2], where a C-score of a higher value signifies a model with a higher confidence and vice-versa (Fig. 2c and d). I-TASSER modelling starts from the structure templates

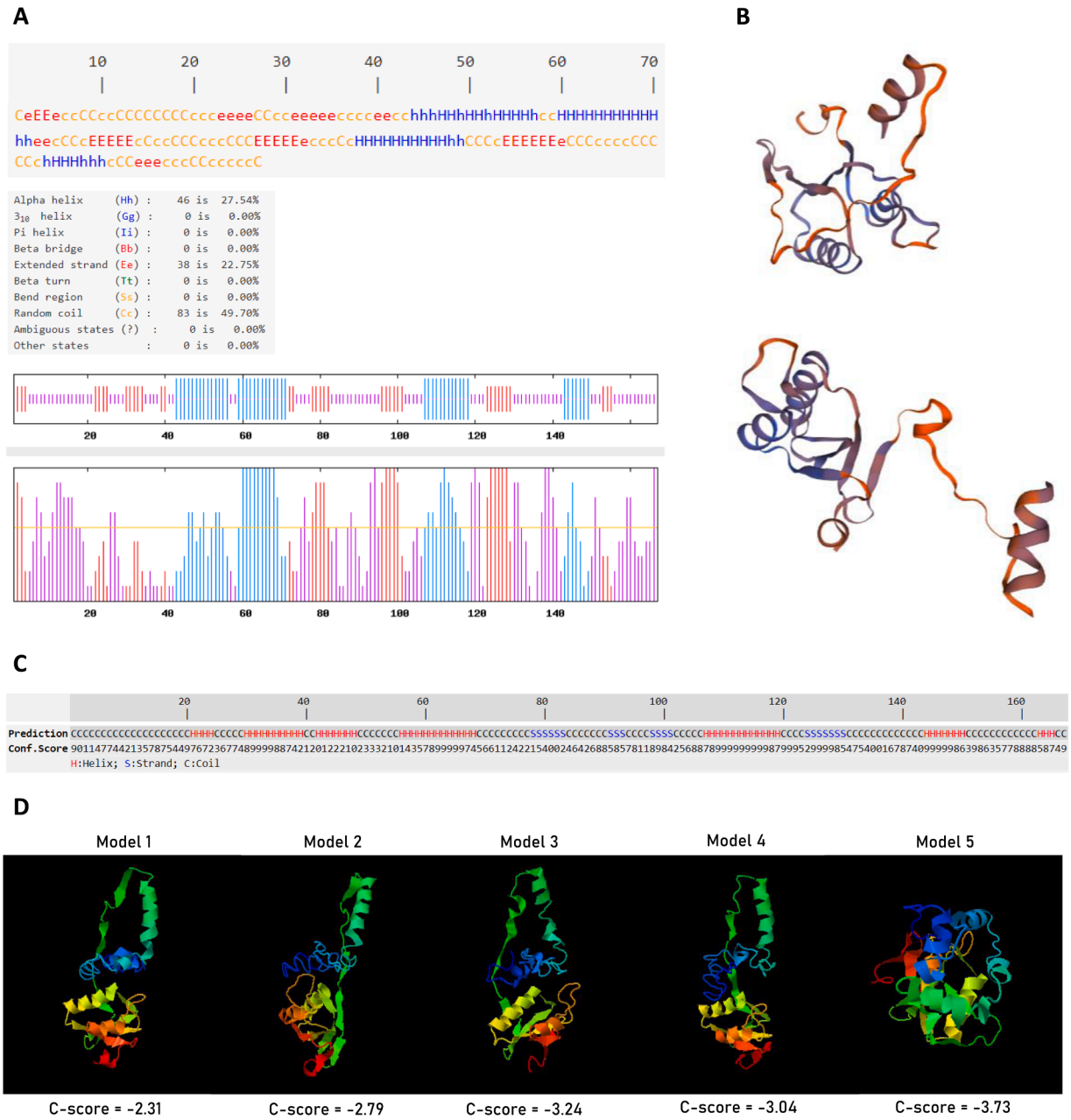

Fig. 2. A. Secondary structure analysis of the KBR-T7 protein sequence using the NPS@: Network Protein Sequence Analysis. B. Tertiary structure analysis of the KBR-T7 protein sequence using the Swiss-Model Sequence Analysis Tool. B. Secondary structure analysis of the KBR-T7 protein sequence using I-TASSER. B. Tertiary structure analysis of the KBR-T7 protein sequence using I-TASSER.

identified by LOMETS from the PDB library. LOMETS is a meta-server threading approach containing multiple threading programs, where each threading program can generate tens of thousands of template alignments. I-TASSER then only used the templates of the highest significance in the threading alignments, the significance of which are measured by the Z-score, *i.e.* the difference between the raw and average scores in the unit of standard deviation. The templates in this section are the ten best templates selected from the LOMETS threading programs. Usually, one template of the highest Z-score is selected from each threading program, where the threading programs are sorted by the average performance in the large-scale benchmark test experiments (Table 3). The Ramachandran plot generated shows the theoretical values and conformations for the ψ and φ angles for the amino-acid residues within the T7 endonuclease protein (Fig. 3). It illustrates the empirical distribution of datapoints observed in a single structure in usage for structure validation outlining theoretically favored regions.

3.2. *Expression and purification of recombinant KBR-T7 endonuclease*

To determine whether the novel KBR-T7 protein encodes for a functional HJ resolvase, we expressed recombinant polyhistidine-tagged KBR-T7 protein (pET-30b(+)-KBR-T7 and its derivative pET-30b (+)-MBP-KBR-T7) in *E. coli.* After affinity purification, the KBR-T7 recombinant proteins of ~19 kDa and ~65 kDa were the major components visualized by SDS-PAGE and immunoblotting (Fig. 4a and b). We then quantified the clarified lysate, eluate and formulated fractions for the KBR-T7 and MBP-KBR-T7 proteins using BSA as a standard. Quantification of the formulated eluates for the KBR-T7 and MBP-KBR-T7 proteins amounted to 20 ± 0.003 $\mu g.L^{-1}$ culture and 1.63 ± 0.154 mg. L^{-1}, respectively. A vast increase in yield was observed between KBR-T7

Table 2
Sequences used as templates from the Swiss-Model Database to model KBR-T7 protein sequence.

Name	Sequence Identity	Sequence Similarity	Coverage	Resolution	Method	GMQE*
2pfj.1.C	39.831	0.423	0.707	3.100	X-ray	0.498
2pfj.1.C	38.462	0.420	0.701	3.100	X-ray	0.479
1m0d.1.A	41.071	0.430	0.671	1.900	X-ray	0.482
1fzr.1.A	40.179	0.428	0.671	2.100	X-ray	0.480
1fzr.2.A	40.179	0.428	0.671	2.100	X-ray	0.478
1m0d.1.A	43.119	0.437	0.653	1.900	X-ray	0.479
1fzr.1.A	42.202	0.434	0.653	2.100	X-ray	0.476
1fzr.2.A	42.202	0.434	0.653	2.100	X-ray	0.477
3cae.1.A	41.667	0.426	0.647	3.000	X-ray	0.475
3cae.1.C	41.667	0.426	0.647	3.000	X-ray	0.459
3cae.1.D	41.667	0.426	0.647	3.000	X-ray	0.459
3cae.1.I	41.667	0.426	0.647	3.000	X-ray	0.468
3cae.1.A	46.914	0.456	0.485	3.000	X-ray	0.355
3cae.1.C	46.914	0.456	0.485	3.000	X-ray	0.344
3cae.1.D	46.914	0.456	0.485	3.000	X-ray	0.346
3cae.1.I	46.914	0.456	0.485	3.000	X-ray	0.356
3jsz.1.A	17.460	0.307	0.377	1.700	X-ray	0.155
3jt1.1.A	17.460	0.306	0.377	2.300	X-ray	0.158
2wzf.1.A	17.460	0.304	0.377	2.100	X-ray	0.157
2xh3.1.A	10.870	0.254	0.275	2.490	X-ray	0.097
2xgr.1.A	10.870	0.246	0.275	1.700	X-ray	0.096
2kng.1.A	23.077	0.315	0.156	NA	NMR	0.054
6qkp.1.A	20.833	0.308	0.144	NA	NMR	0.046
6qkq.1.A	21.739	0.310	0.138	NA	NMR	0.039

* GMQE (Global Model Quality Estimation) is a quality estimation, which combines properties from the target–template alignment and the template structure. The resulting GMQE score is expressed as a number between 0 and 1, reflecting the expected accuracy of a model built with that alignment and template, normalized by the coverage of the target sequence. Higher numbers indicate higher reliability.

Table 3
Sequences identified from the PDB library to model KBR-T7 protein sequence by I-TASSER.

Rank[a,*]	PDB Hit[b]	Identity 1[c]	Identity 2[d]	Coverage[e]	Norm Z-score[f]
1	2pfjB	0.41	0.28	0.65	1.69
2	3caeA	0.42	0.29	0.64	1.94
3	1m0dA	0.43	0.28	0.65	2.32
4	2pfj	0.41	0.28	0.65	6.70
5	2pfj	0.41	0.28	0.65	5.25
6	2pfjB	0.41	0.28	0.65	2.35
7	2pfj	0.42	0.28	0.65	7.39
8	1m0dA	0.42	0.28	0.65	1.97
9	2pfjB	0.41	0.28	0.65	2.10
10	1m0dA	0.43	0.28	0.65	1.51

[a] Rank of templates represents the top ten threading templates used by I-TASSER.
[b] Identity 1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence.
[c] Identity 2 is the percentage sequence identity of the whole template chains with query sequence.
[d] Coverage represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein.
[e] Norm. Z-score is the normalized Z-score of the threading alignments.
[f] Alignment with a Normalized Z-score >1 mean a good alignment and *vice versa*.
[*] The top 10 alignments reported above (in order of their ranking) are from the following threading programs: a) MUSTER b) FFAS-3D c) SPARKS-X d) HHSEARCH2 e) HHSEARCH I f) Neff-PPAS g) HHSEARCH h) pGenTHREADER i) wdPPAS j) PROSPECT2.

and MBP-KBR-T7.

3.3. Cleavage of HJs

The pattern of cleavage in each strand of the HJ was examined (Fig. 5). Four identical HJs were incubated with the commercial T7 endonuclease and the novel KBR-T7 endonuclease purified from three independent batches and the products were separated in a 5 % TBE polyacrylamide gel (Fig. 5a). In each reaction, full-length and faster migrating strands were detected. There were two predominant nicks occurring around the 250 bp position resulting in a nicked strand appearing around the 250 bp mark. DNA fragments post hybridization and MBP-KBR-T7 Endonuclease I digestion were eluted in 20 µl of nuclease-free water and visualized on a 2.5 % agarose gel (Fig. 5b). At 0.5 µL (~0.11 µg), the MBP-KBR-T7 endonuclease begins to work at a temperature of 37 °C in two different buffer backgrounds as compared to the competitor endonuclease. The stoichiometry of the reaction needs to be precise for the enzyme to function effectively, however, with the MBP-KBR-T7 derivative, it still functions at both higher amounts and different concentrations of buffer background introducing a great degree of flexibility for optimization by the end-user.

4. Discussion

The present finding of a novel HJ endonuclease from the Kogelberg Biosphere Reserve discovered from the metavirome [12,27] culminates a long search for such an enzyme, which could have multiple roles in the rapidly advancing genome editing field. DNA mismatches arise as part of the genome editing mutation detection workflow, and a HJ endonuclease is required for recognition of DNA mismatches and subsequent cleavage. Orthologs producing significant alignments from the Uniprot Database compared with the KBR-T7 protein sequence show statistically significant similarity to each other (probability of occurrence by chance, E value $<10^{22}$) but not to any other proteins, which strongly suggests a monophyletic origin [28]. TM-score and RMSD are estimated based on C-score and protein length following the correlation observed between these qualities. Since the top 5 models are ranked by the cluster size, it is possible that the lower-rank models have a higher C-score in rare cases. Although the first model has a better quality in most cases, it is also possible that the lower-rank models have a better quality than the higher-rank models as seen in our benchmark tests. If the I-TASSER simulations converge, it is possible to have <5 clusters generated; this is usually an indication that the models have a good quality because of the converged simulations.

Mapping of these conserved sequence motifs on the of RuvC, suggested that the entire family preserves the core structural elements of RuvC separated by poorly conserved spacers [29]. These residues are

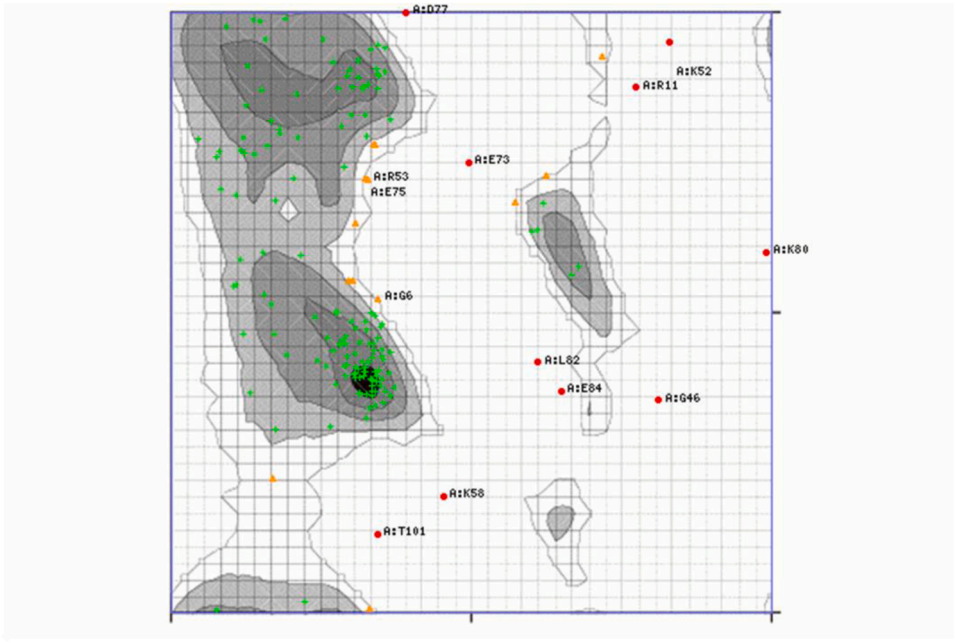

Fig. 3. Ramachandran plot assessing the sterically allowed and disallowed conformations of KBR-T7 endonuclease backbone.

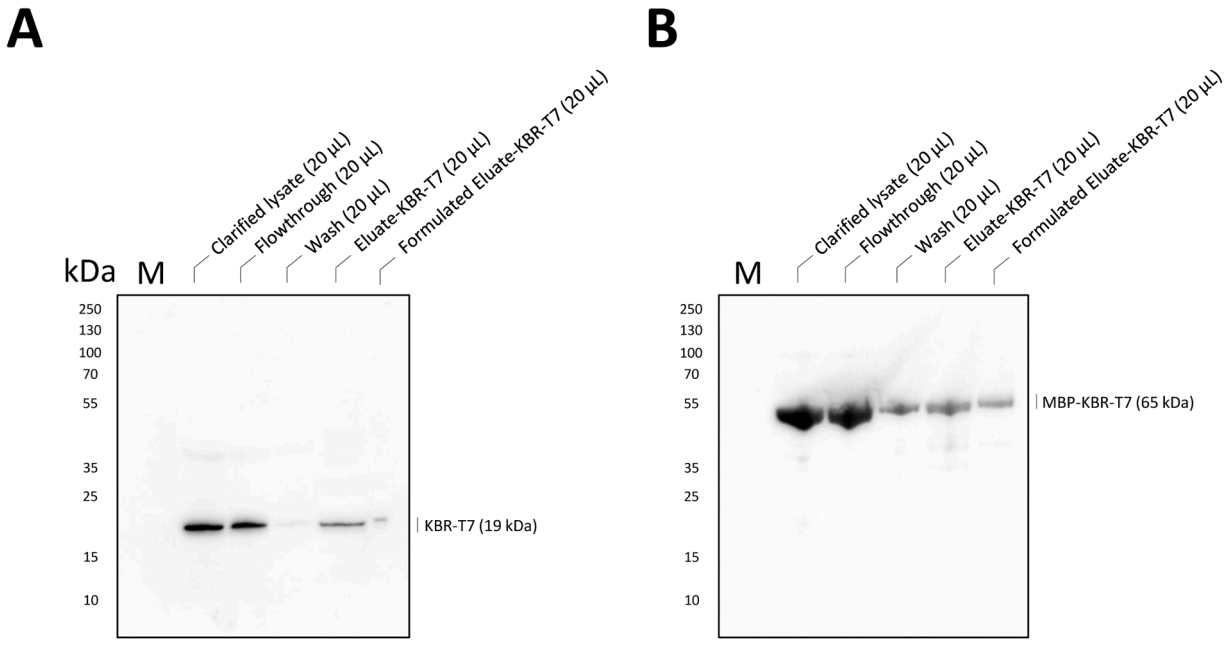

Fig. 4. Recombinant polyhistidine-tagged KBR-T7 expressed in *E. coli* Rosetta™(DE3) pLysS and purified by chromatography on a metal-affinity resin, analyzed on a PVDF immunoblot. A. Purification of KBR-T7. B. Purification of MBP-KBR-T7.

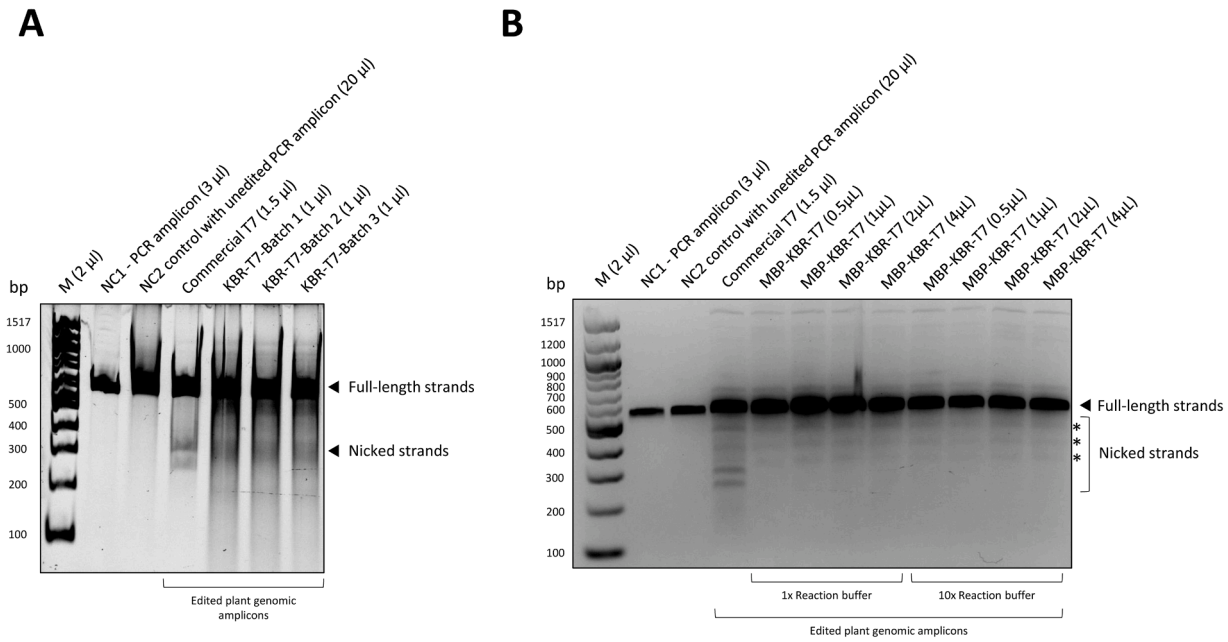

Fig. 5. DNA strands of HJ are nicked symmetrically by KBR-T7. A. Annealed DNA substrates incubated with KBR-T7 and Commercial T7 endonuclease for 45 min and analyzed by electrophoresis on a 5 % TBE polyacrylamide gel. B. Annealed DNA substrates were incubated with MBP-KBR-T7 in two different buffer backgrounds, 1x and 10x buffer and analysed on a 2.5 % agarose gel. *asterisks indicate nicked strands generated by the activity of MBP-KBR-T7.

critical for the resolvase activity of RuvC forming a spatially juxtaposed acidic triad that coordinates a metal ion [30]. Interestingly, there are slight differences in the β-bridge portion and not all potential catalytic centres are conserved (one is different) within the KBR-T7 molecule. Leucine, Lysine & Aspartic Acid are new amino acids within the β-bridge, along with the omission of 1 or 2 amino acids that also make up the bridge could account for observed activity. Mutagenesis experiments have identified five acidic residues that could potentially be involved in catalysis in endonuclease I; Glu 20, Glu 35, Asp 55, Glu 65 and Asp 74 [25]. Furthermore, Pro54, Asp55 - Glu65, Lys67 and Glu20, shows that all five of the conserved amino acid residues are important to the catalytic activity and that mutation of any of these positions leads to reduced cleavage rates by a factor of at least 100 [31]. Studies have shown that the removal of residues 141–149 significantly reduces the affinity of this nuclease for DNA junctions [32]. A third T7 endonuclease derivative was designed with the addition of the C-terminus domain containing the following residues RLKRKGGKK purported to increase affinity of nucleases to DNA junctions; however, no improvements in efficacy were observed (data not shown).

Bacterial expression and metal-affinity purification of recombinant KBR-T7 endonuclease was carried out in a similar manner as previously described with the modification of using the Rosetta™ (DE3) pLysS *E. coli* strain [28]. We observed a similar level of toxicity to the host in our study as previously reported [25] which also necessitated the adjustment of expression conditions to optimize yields. We found that by scaling to a final volume of 3 L produced significant amounts of the novel T7 endonuclease, enough for subsequent DNA recognition and cleavage analysis. The use of agar to facilitate and sustain growth as previously described was circumvented in this regard [33]. Furthermore, we opted to attach a solubility fusion partner, MBP onto the KBR-T7 endonuclease which significantly increased the solubility and yields, concomitantly.

5. Conclusions

In conclusion, both the KBR-T7 and MBP-KBR-T7 variants show efficacy in being able to detect genome editing through Holliday junction identification and cleavage potential. With great interest in Crispr/Cas9 technology, there is an opportunity to add value by providing the user accompanying reagents such as novel T7 endonucleases to facilitate confirmation workflows. Furthermore, our results pave the way for the inclusion of these variants in such genome editing detection pipelines which are readily used prior to whole genome sequencing. The molecular biology reagents market relies on the development of these complimentary tools for the user. In this regard, this manufacturing process for a novel MBP-KBR-T7 endonuclease with precise specificity has been developed to fulfil this complimentary need.

CRediT authorship contribution statement

Priyen Pillay: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Maabo Moralo:** Writing – review & editing, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Sibongile Mtimka:** Writing – review & editing, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Taola Shai:** Writing – review & editing, Validation, Methodology, Formal analysis, Data curation. **Kirsty Botha:** Writing – review & editing, Visualization, Validation, Methodology. **Lusisizwe Kwezi:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Methodology, Investigation, Formal analysis, Data curation. **Tsepo L. Tsekoa:** Writing – review & editing, Supervision, Resources, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Tsepo Tsekoa reports financial support was provided by Technology Innovation Agency. Tsepo Tsekoa reports financial support was provided by South Africa Department of Science and Innovation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the Technology Innovation Agency (TIA) and the Department of Science and Innovation (DSI) for funding the study.

Data availability

Data will be made available on request.

References

[1] R. Holliday, A mechanism for gene conversion in fungi, Genet. Res. 5 (2) (1964) 282–304.
[2] D.M. Lilley, M.F. White, The junction-resolving enzymes, Nat. Rev. Mol. Cell Biol. 2 (6) (2001) 433–443.
[3] J.M. Hadden, A.-C. Déclais, S.B. Carr, D.M. Lilley, S.E. Phillips, The structural basis of Holliday junction resolution by T7 endonuclease I, Nature 449 (7162) (2007) 621–624.
[4] D.R. Duckett, M.J.E.G. Panis, D.M. Lilley, Binding of the junction-resolving enzyme bacteriophage T7 endonuclease I to DNA: separation of binding and catalysis by mutation, J. Mol. Biol. 246 (1) (1995) 95–107.
[5] R.G. Pöhler, M.J.E. Giraud-Panis, D.M. Lilley, T4 endonuclease VII selects and alters the structure of the four-way DNA junction; binding of a resolution-defective mutant enzyme, J. Mol. Biol. 260 (5) (1996) 678–696.
[6] M.F. White, D.M. Lilley, The structure-selectivity and sequence-preference of the junction-resolving enzyme CCE1 of Saccharomyces cerevisiae, J. Mol. Biol. 257 (2) (1996) 330–341.
[7] M.B. Izydorczyk, E. Kalef-Ezra, D. Horner, X. Zheng, N. Holmes, M. Toffoli, Z. G. Sahin, Y. Han, H.H. Mehta, D.M. Muzny, Single cell long read whole genome sequencing reveals somatic transposon activity in human brain, medRxiv (2024), 2024.11. 11.24317113.
[8] T. Sakuma, A. Nishikawa, S. Kume, K. Chayama, T. Yamamoto, Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system, Sci. Rep. 4 (1) (2014) 5400.
[9] N.M. Glover, H. Redestig, C. Dessimoz, Homoeologs: what are they and how do we infer them? Trends Plant. Sci. 21 (7) (2016) 609–621.
[10] P.S. Soltis, D.E. Soltis, The role of hybridization in plant speciation, Annu. Rev. Plant. Biol. 60 (1) (2009) 561–588.
[11] M.R. Jamee, Z. Ansari, M.I. Qureshi, From design to validation of CRISPR/gRNA primers towards genome editing, Bioinformation 18 (5) (2022) 471.
[12] J. Segobola, E. Adriaenssens, T. Tsekoa, K. Rashamuse, D. Cowan, Exploring viral diversity in a unique South African soil habitat, Sci. Rep. 8 (1) (2018) 111, https://doi.org/10.1038/2Fs41598-017-18461-0.
[13] K.D. Pryor, B. Leiting, High-level expression of soluble protein inescherichia coliusing a his6-tag and maltose-binding-protein double-affinity fusion system, Protein Expr. Purif. 10 (3) (1997) 309–319.
[14] R.B. Kapust, D.S. Waugh, Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused, Prot. Sci. 8 (8) (1999) 1668–1674.
[15] L. Nieba, S.E. Nieba-Axmann, A. Persson, M. Hämäläinen, F. Edebratt, A. Hansson, J. Lidholm, K. Magnusson, Å.F. Karlsson, A. Plückthun, BIACORE analysis of histidine-tagged proteins using a chelating NTA sensor chip, Anal. Biochem. 252 (2) (1997) 217–228.
[16] J. Pei, B.-H. Kim, N.V. Grishin, PROMALS3D: a tool for multiple protein sequence and structure alignments, Nucleic Acids Res. 36 (7) (2008) 2295–2300.
[17] E.W. Sayers, J. Beck, E.E. Bolton, D. Bourexis, J.R. Brister, K. Canese, D.C. Comeau, K. Funk, S. Kim, W. Klimke, Database resources of the national center for biotechnology information, Nucleic Acids Res. 49 (D1) (2021) D10.
[18] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (17) (1997) 3389–3402.
[19] S.F. Altschul, J.C. Wootton, E.M. Gertz, R. Agarwala, A. Morgulis, A.A. Schäffer, Y. K. Yu, Protein database searches using compositionally adjusted substitution matrices, FEBS J. 272 (20) (2005) 5101–5109.
[20] C. Combet, C. Blanchet, C. Geourjon, G. Deleage, NPS@: network protein sequence analysis, Trends Biochem. Sci. 25 (3) (2000) 147–150.
[21] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling, Electrophoresis 18 (15) (1997) 2714–2723.
[22] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, Y. Zhang, The I-TASSER Suite: protein structure and function prediction, Nat. Methods 12 (1) (2015) 7–8.
[23] S. Wu, Y. Zhang, LOMETS: a local meta-threading-server for protein structure prediction, Nucleic Acids Res. 35 (10) (2007) 3375–3382.
[24] A.H. Rosenberg, B.N. Lade, C. Dao-shan, S.-W. Lin, J.J. Dunn, F.W. Studier, Vectors for selective expression of cloned DNAs by T7 RNA polymerase, Gene 56 (1) (1987) 125–135.
[25] J.M. Hadden, M.A. Convery, A.-C. Déclais, D.M. Lilley, S.E. Phillips, Crystal structure of the Holliday junction resolving enzyme T7 endonuclease I, Nat. Struct. Biol. 8 (1) (2001) 62–67.
[26] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel, A. Bairoch, ExPASy: the proteomics server for in-depth protein knowledge and analysis, Nucleic Acids Res. 31 (13) (2003) 3784–3788.
[27] Segobola, M.P.J., Identification and Characterisation of Nucleic Acid Manipulating Enzymes from Metaviromic DNA. 2019, University of Pretoria.
[28] A.D. Garcia, L. Aravind, E.V. Koonin, B. Moss, Bacterial-type DNA Holliday junction resolvases in eukaryotic viruses, Proc. Natl. Acad. Sci. 97 (16) (2000) 8926–8931.
[29] M. Ariyoshi, D.G. Vassylyev, H. Iwasaki, H. Nakamura, H. Shinagawa, K. Morikawa, Atomic structure of the RuvC resolvase: a Holliday junction-specific endonuclease from E. coli, Cell 78 (6) (1994) 1063–1072.
[30] A. Saito, H. Iwasaki, M. Ariyoshi, K. Morikawa, H. Shinagawa, Identification of four acidic amino acids that constitute the catalytic center of the RuvC Holliday junction resolvase, Proc. Natl. Acad. Sci. 92 (16) (1995) 7470–7474.
[31] A.-C. Déclais, J. Hadden, S.E. Phillips, D.M. Lilley, The active site of the junction-resolving enzyme T7 endonuclease I, J. Mol. Biol. 307 (4) (2001) 1145–1158.
[32] M.J. Parkinson, J.R.G. Pöhler, D.M. Lilley, Catalytic and binding mutants of the junction-resolving enzyme endonuclease I of bacteriophage T7: role of acidic residues, Nucleic Acids Res. 27 (2) (1999) 682–689.
[33] J.M. Hadden, A.C. Déclais, S.E. Phillips, D.M. Lilley, Metal ions bound at the active site of the junction-resolving enzyme T7 endonuclease I, EMBO J. 21 (13) (2002) 3505–3515.