Crystal structure and initial characterization of a novel archaeal-like Holliday junction-resolving enzyme from *Thermus thermophilus* phage Tth15-6

Josefin Ahlqvist, Javier A. Linares-Pasteñ, Maria Håkansson, Andrius Jasilionis, Karolina Kwiatkowska-Semrau, Ölafrur H. Friðjónsson, Anna-Karina Kaczorowska, Slawomir Dabrowski, Arnþór Ævarsson, Guðmundur Ó. Hreggviðsson, Salam Al-Karadaghi, Tadeusz Kaczorowski and Eva Nordberg Karlsson

This study describes the production, characterization and structure determination of a novel Holliday junction-resolving enzyme. The enzyme, termed Hjc_15-6, is encoded in the genome of phage Tth15-6, which infects *Thermus thermophilus*. Hjc_15-6 was heterologously produced in *Escherichia coli* and high yields of soluble and biologically active recombinant enzyme were obtained in both complex and defined media. Amino-acid sequence and structure comparison suggested that the enzyme belongs to a group of enzymes classified as archaeal Holliday junction-resolving enzymes, which are typically divalent metal ion-binding dimers that are able to cleave X-shaped dsDNA–Holliday junctions (Hjs). The crystal structure of Hjc_15-6 was determined to 2.5 Å resolution using the selenomethionine single-wavelength anomalous dispersion method. To our knowledge, this is the first crystal structure of an Hj-resolving enzyme originating from a bacteriophage that can be classified as an archaeal type of Hj-resolving enzyme. As such, it represents a new fold for Hj-resolving enzymes from phages. Characterization of the structure of Hjc_15-6 suggests that it may form a dimer, or even a homodimer of dimers, and activity studies show endonuclease activity towards Hjs. Furthermore, based on sequence analysis it is proposed that Hjc_15-6 has a three-part catalytic motif corresponding to E–SD–EVK, and this motif may be common among other Hj-resolving enzymes originating from thermophilic bacteriophages.

1. Introduction

Holliday junction-resolving enzymes are nucleases that cleave four-way DNA–Holliday junctions (Hjs) into two unconnected DNA duplexes. Hjs are common intermediates during meiotic and mitotic genetic recombination, and Hj-resolving enzymes have been isolated from all types of eukaryotic and prokaryotic cells and their viruses (Wyatt & West, 2014). Hjs were first presented by Robin Holliday in 1964, when he suggested a model for gene conversion during meiosis in fungi (Holliday, 1964), claiming that two homologous chromosomes paired between complementary sequences lead to the formation of a cross-stranded structure that physically links the two component helices. Since then, other models of how and why two dsDNA helices may cross-link have been reported. This type of four-way DNA structure at the point of strand
exchange has become known as an Hj, and the enzymes that resolve them are referred to as Hj-resolving enzymes.

Apart from playing a major role in gene conversion during meiosis, it has been suggested that Hjs and Hj-resolving enzymes are involved in processes such as DNA repair and the introduction of plasmid and phage DNA into host genomes (Wyatt & West, 2014), and Hj-resolving enzymes are now considered to be key enzymes in DNA recombination (Aravind et al., 2000). For instance, the 157-residue endo VII encoded by gene 49 of bacteriophage T4 resolves branched multimeric T4 DNA before packaging it into phage heads (Biertümpfel et al., 2007; Kemper & Brown, 1976; Kemper & Janz, 1976; Mizuuchi et al., 1982). In 2001, Birkenbihl and coworkers presented the characterization, including the isolation and purification, of two similar Hj-cleaving enzymes from two different viruses, SIRV1 and SIRV2, that infect the archaeon Sulfolobus islandicus (Birkenbihl et al., 2001). At the time, the phage enzymes endonuclease VII from phage T4 and endonuclease I from phage T7 (T7 endo I), the bacterial proteins RuvC and RusA from Escherichia coli and the yeast enzymes CCE1 from Saccharomyces cerevisiae and YPD2 from Schizosaccharomyces pombe 2.3 were the best-studied members of the growing group of structure-specific endonucleases. While the latter two enzymes are homologous, no significant sequence similarity exists between the other proteins.

Hj-resolving enzymes are diverse both in sequence and in structural organization (Lilley & White, 2000). However, this may not be surprising considering the different evolutionary origins, purposes and outcomes of the creation and resolution of Hjs. Aravind and coworkers analysed the structural and evolutionary relationships of various Hj-resolving enzymes and related nucleases, suggesting that an Hj-resolving enzyme function has evolved independently from at least four distinct structural folds (RNase H, endonuclease, endonuclease VII–colicin E and RusA; Aravind et al., 2000). From this work, the endonuclease fold (the structural prototype of which is the phage λ exonuclease) was shown to encompass a far greater diversity of nucleases than previously suspected, including archaeal Hj-resolving enzymes, repair nucleases such as the RecB and Vsr enzymes, and a variety of predicted nucleases. The authors state that the structural prototype for archaeal Hj-resolving enzymes originates from Pyrococcus furiosus, and was isolated, cloned and characterized by Komori et al. (1999).

Even though the origin, amino-acid sequence, fold and overall biological functionality of Hj-resolving enzymes vary, these nucleases often share a number of characteristics (Lilley, 2017). For instance, they most often have a high proportion of positively charged amino acids, enabling them to bind DNA with high affinity. Analyses of the three-dimensional structures of Hj-resolving enzymes have shown that their active sites generally contain three or four acidic residues that are required for metal binding and catalysis. In addition, a divalent metal ion, usually Mg$^{2+}$ or Mn$^{2+}$, that is essential for DNA cleavage but not for DNA binding is present. Furthermore, Hj-resolving enzymes have been reported to be dimeric, allowing the use of twin active sites to catalyse two coordinated incisions (Wyatt & West, 2014).

Several non-archaeal-like Hj-resolving enzymes from viruses have had their structures determined and/or their activities confirmed following endonuclease VII from phage T4 (Kemper & Brown, 1976; Kemper & Janz, 1976; Raaijmakers et al., 1999; Ariyoshi et al., 1994; Mizuuchi et al., 1982), which was the first enzyme known to resolve an Hj. In recent years, structures of fowlpox Hj-resolving enzyme (Culyba et al., 2009; Li et al., 2020) and canarypox Hj-resolving enzyme (Li et al., 2016), for example, have been determined.

In this study, carried out within the frame of the VIRUS-X project (Aevarsson et al., 2021), a novel archaeal-like Hj-resolving enzyme, Hjc_15-6, originating from phage Tth15-6 is described. Phage Tth15-6, which infects the thermophilic bacterium Thermus thermophilus, was originally isolated from a coastal hot spring at Reykjanes in Isafjardardjúp, Iceland (G. Ó. Hreggviðsson, personal communication). Tth15-6 belongs to the long-tailed phage morphotype of the Siphoviridae family in the Caudovirales order (Yu et al., 2006). T. thermophilus is a Gram-negative, thermophilic heterotrophic bacterium that is found in coastal hot springs all around the world. The genomes of other viruses infecting T. thermophilus strains have been sequenced, for example Thermus phage φYS40 (Naryshkina et al., 2006), Thermus virus IN93 (Matsushita & Yanase, 2008), Thermus phage φTMA (Tamakoshi et al., 2011), Thermus virus P23-77 (Jalasvuori et al., 2009), Thermus phage TSP4 (Lin et al., 2010), Thermus phage G20c (Xu et al., 2017), Thermus phage P23-45 and Thermus phage P74-26 (Minakhin et al., 2008), but only a few enzymes have so far been characterized from these bacteriophages.

The study of Hjc_15-6 reported here encompasses structural analysis, including the crystal structure, and some studies of the biological function of this putative Hj-resolving enzyme. Furthermore, the sequence relationship between this enzyme and related Hj-resolving enzymes of viral and bacterial origin, including the putative enzymes encoded in the deposited genomes of Thermus phage P23-45 and Thermus phage P74-26, is investigated. This paper describes, to our knowledge, the first crystallized Hj-resolving enzyme from a thermophilic phage that may be defined as an archaeal type both from a sequence and structural perspective. Therefore, it also presents a new fold among Hj-resolving enzymes originating from phages.

2. Materials and methods

2.1. Isolation of phage Tth15-6

An overnight culture of T. thermophilus MAT15 was mixed with 4 ml of water from a coastal hot spring at Reykjanes in Isafjardardjúp, Iceland. The sampled water had been filtered through a 0.22 μm pore-size filter prior to mixing with the overnight culture. The obtained suspension was incubated at 65°C for 30 min. 5 ml soft agar with 10 mM MgCl₂ was then added and poured onto Medium 166 solidified with agar and
the plates were incubated at 65°C overnight (Hjorleifsdottir et al., 2001). Subsequently, plaques were aseptically transferred with a Pasteur pipette into microcentrifuge tubes with 100 μl 10 mM MgCl₂ and stored at 4°C. The plaque solution was then diluted 10⁻³–10⁻⁴ with 10 mM MgCl₂, and 100 μl of the diluted phage solution was added to 900 μl of an overnight culture of *T. thermophillus* MAT15 and incubated at 65°C for 30 min. Soft agar was then added and the mixture was poured onto plates with Medium 166. After overnight growth, the plaques were picked and the whole procedure was repeated four times to purify the phage particles.

2.2. DNA isolation, sequencing and identification of the putative DNA-resolving enzyme gene

Phage-amplification enrichment for DNA isolation was achieved using a 10⁻³–10⁻⁴ phage dilution, which gave confluent lysis on soft agar plates. The soft agar was scraped from 20 plates and suspended in 100 ml 10 mM MgCl₂. The mixture was incubated at room temperature and 600 rev min⁻¹ for 2 h and was then centrifuged at 11 000g for 20 min. The supernatant was filtered through a sterile 0.22 μm pore-size filter and the titre was analysed. Phage DNA was isolated from the supernatant with a titre of >10⁹ pfu ml⁻¹ using the NucleoSpin Plasmid kit (Macherey-Nagel), performing binding in NT2 buffer. A sequencing library was constructed using the MiSeq system (Illumina) sequencing platform using the MiSeq Reagent Kit v3 and 2 × 300 bp sequencing cycles. Sequences were trimmed for quality using *Trimmomatic* version 0.36 and assembled using the *SPAdes* version 3.12.0 assembly algorithm to produce a linear consensus sequence of 76 134 bp.

Analysis of the predicted ORF in the sequenced genome of phage *Tth15-6* was performed using the *RAST* server (https://rast.nmpdr.org/; Aziz et al., 2008). Genome annotation suggested the presence of a gene encoding a putative DNA-resolving enzyme/helicase-like enzyme. Gene-sequence annotation allowed the design of primers for PCR amplification of the gene from the isolated genomic DNA.

2.3. Gene cloning

The gene encoding Hjc_15-6 was amplified from the phage *Tth15-6* genomic DNA by PCR with forward primer 5'-TCT CATATGCTAAAGATAAAGGA-3' (the NdeI restriction site is underlined) and reverse primer 5'-TAGCTCCCTCGAG ACCTGTGAAGTC-3' (the *XhoI* restriction site is underlined). The amplified gene was inserted into the pET-21b(+) vector (Novagen), giving the construct pET-21b:Hjc_15-6. A PCR reaction mixture with a final volume of 20 μl was prepared with 10 μl Master Mix iProof (2 ×) (Bio-Rad), 1 ng DNA template, 0.5 μM of each primer, 3% (w/v) DMSO and nuclease-free water. The reaction was preheated at 98°C for 120 s and 29 cycles of amplification were then applied with denaturation at 98°C for 20 s, annealing at 65°C for 30 s and elongation at 72°C for 15 s, with a final elongation at 72°C for 420 s. The PCR products were purified with the QiAquick PCR purification kit (Qiagen). Both the ampiclon and the pET-21b(+) vector were digested with the NdeI and *XhoI* restriction endonucleases. The digestion products were purified from the agarose gel with a QIAquick Gel Extraction kit (Qiagen) and ligated with T4 ligase according to the manufacturer’s instructions. *E. coli* DH5α (Novagen) cells were transformed with the ligation products. Colonies harbouring the construct pET-21b::Hjc_15-6 were determined by colony PCR. The expression-construct sequence was verified by sequencing, confirming that the gene was cloned in frame with the C-terminal hexahistidine tag encoded by the expression vector. The gene was deposited in the NCBI GenBank with accession number MW788030.

2.4. Protein production

The gene encoding Hjc_15-6 with an attached C-terminal hexahistidine tag was initially expressed in *E. coli* BL21(DE3) cells (Novagen) using lysogony broth (LB) for cultivation in shake flasks and was subsequently upscaled to 2.5 l in a 31 fermenter using mAT medium.

The cultivation in shake flasks was performed in baffled Erlenmeyer flasks filled to not more than a quarter of the flask volume with LB (10 g l⁻¹ tryptone, 5 g l⁻¹ yeast extract, 5 g l⁻¹ NaCl) supplemented with 100 μg ml⁻¹ ampicillin. Expression cultures were inoculated to 0.5%(v/v) with overnight culture and were cultivated at 37°C and 200 rev min⁻¹. Heterologous overexpression of Hjc_15-6 was induced with 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) when the culture reached an OD₆₂₀ nm of 0.7–0.9. Induction was performed for 4 h at 37°C and 200 rev min⁻¹.

Target enzyme production was upscaled to 2.5 l in a 31 fermenter (Belach Biotechnik) using defined mAT medium [2 g l⁻¹ ammonium sulfate, 14.6 g l⁻¹ K₂HPO₄, 3.2 g l⁻¹ Na₂HPO₄·H₂O and 0.5 g l⁻¹ ammonium hydroxide] supplemented with 2 ml l⁻¹ 1 M MgSO₄ and 2 ml l⁻¹ trace-elements solution (Holme et al., 1970) as well as 20 ml l⁻¹ 50%(w/v) n-glucose (de Maré et al., 2005) supplemented with 100 μg ml⁻¹ ampicillin. Cultivation in the fermenter was performed at a constant pH of 7.0 to 37°C with at least 20% dissolved oxygen tension (DOT). The aeration was set to a constant 1 v.v.m. (volume of air per volume of liquid per minute). The pH was maintained by titration with 12.5%(w/v) ammonia solution, while the DOT was controlled by the stirrer speed (initial speed 300 rev min⁻¹). The fermenter was inoculated to 4%(v/v) with overnight culture cultivated using mAT medium at 30°C and 200 rev min⁻¹. Heterologous overexpression of Hjc_15-6 was induced with 1 mM IPTG when the culture in the fermenter reached an OD₆₂₀ nm of 3. Induction was performed for 1 h at 37°C.

After termination of cultivation, harvesting was performed by centrifugation of the chilled cultures at 3800g for 10 min at 4°C. The collected cell pellets were stored frozen before purification of Hjc_15-6.

2.5. Selenium derivatization of Hjc_15-6

Hjc_15-6 was derivatized with selenium by seleno-1-methionine (SeMet) incorporation, producing recombinant
protein in the methionine-auxotrophic E. coli strain B834(DE3) (Novagen) cells using mAT medium supplemented with 50 mg l⁻¹ SeMet. Cultivation in shake flasks was performed for derivatization (Turner et al., 2007). Expression cultures were inoculated to 1%(v/v) with cell suspension prepared by washing fresh overnight culture from LB agar plates with mAT medium twice. Induction, harvesting and cell-pellet storage were performed as described previously.

2.6. Protein purification

The collected cell pellets were thawed on ice and resuspended in binding buffer (100 mM Tris–HCl pH 7.4, 500 mM NaCl). The cells were disrupted by sonication (20 cycles of 30 s on, 30 s off with 60% amplitude) using an UP400s sonicator (Hielscher Ultrasound Technology) in an ice bath. The soluble protein fraction was separated from the cell debris by centrifugation at 17 000g for 20 min at 4°C. Hjc_15-6 was purified by nickel-affinity chromatography using a 5 ml 16 × 25 mm HisTrap HP column (GE Healthcare Life Sciences). Elution was performed with a linear 8 ml gradient to 100% elution buffer (100 mM Tris–HCl pH 7.4, 500 mM NaCl, 500 mM imidazole). Fractions containing Hjc_15-6 were combined and dialyzed (1:5000) for 12–16 h at 4°C. The dialyzed protein sample was subjected to crystallization trials. The purity of Hjc_15-6 was polished by combining nickel affinity with Heparin Sepharose chromatography for use in activity studies. Hjc_15-6 was purified by Heparin Sepharose chromatography using a 1 ml 7 × 25 mm HiTrap Heparin HP column (GE Healthcare Life Sciences). Elution was performed with a linear salt gradient. Desalting was performed by dialysis. The purity and integrity of Hjc_15-6 were assessed by glycine SDS–PAGE using 4–20% gradient gels. The protein concentration was estimated by measuring A₂₈₀nm with a NanoDrop 1000 (Thermo Fisher Scientific).

2.7. Mass spectrometry

Purified samples of Hjc_15-6, produced using complex (LB) medium or defined (mAT) medium as described previously, were analysed by mass spectrometry (MS). The protein sample concentration was adjusted to 5 mg ml⁻¹ and it was frozen in liquid nitrogen and stored frozen at −80°C prior to MS analysis. As a precaution, the samples were spun down after thawing on ice.

MS spectra were acquired using an Autoflex Speed MALDI–TOF/TOF mass spectrometer (Bruker Daltonics) in positive linear mode. 0.5 µl matrix solution consisting of 5 mg ml⁻¹ α-cyano-4-hydroxycinnamic acid, 80%(v/v) acetonitrile, 0.1%(v/v) trifluoroacetic acid (TFA) was added to 1 µl Hjc_15-6 sample on a MALDI stainless-steel plate. A total of 5000 laser shots were collected per spectrum and were calibrated using the Protein I calibration standard (Bruker Daltonics) containing six internal standard proteins (insulin, m/z 5734.52; cytochrome c, m/z 6181.05; myoglobin, m/z 8476.66; ubiquitin I, m/z 8565.76; cytochrome c, m/z 12 360.97; myoglobin, m/z 16 952.31).

2.8. Dynamic light scattering

Purified samples of Hjc_15-6 produced in complex (LB) medium or defined (mAT) medium as described previously were analysed by dynamic light scattering (DLS) at a 173° backscattering angle at 25°C using a Zetasizer Nano ZS instrument (Malvern Panalytical). The concentration of the protein sample after filtration through a 0.22 µm pore-size filter was adjusted to 0.25 mg ml⁻¹ with SPG (succinic acid–NaH₂PO₄–glycine) buffer pH 7.0. The purified protein from defined (mAT) and complex (LB) medium had an initial concentration of 5.4 and 11.4 mg ml⁻¹, respectively (from the A₂₈₀nm measured using a NanoDrop). The purified protein produced in defined medium had been stored for six months at −20°C, while the protein produced in complex medium was freshly made. After concentration adjustment, the protein samples were centrifuged at 14 000g for 5 min. DLS measurements were performed twice in triplicate at 20°C using ZEN0040 micro-cuvettes (Malvern Panalytical). In addition, the 50 µl samples were incubated at 25°C for 1 min before measurement. The size of the protein particles was calculated from the average size distribution obtained from six measurements: two measurements in triplicate.

2.9. Screening for resolving activity

The resolving activity of Hjc_15-6 was studied using fluorescent X-shaped DNA as a Holliday junction substrate. The substrate was assembled from DNA oligomers 1–4, with fluorescent-tagged CGA triplets at the 5’ position, as detailed in Table 1.

100 mM oligo DNA was mixed to give a concentration of 10 mM DNA. The solution was heated for 3 min at 94°C and then cooled for hybridization. 1 µl of 10 mM hybridized DNA containing approximately 10 pmol X-shaped DNA was used in a reaction mixture with 1 µl 1.3 mg ml⁻¹ purified Hjc_15-6 (~35 pmol) that had been produced in complex medium and purified once or twice as described above. The final buffer concentration in the reaction mixture was 10 mM Tris–HCl pH 8.5, 100 mM KCl, 10 mM MgCl₂ supplemented, when necessary, with 5 mM ATP. The reaction mixtures were incubated at 50°C for 30 min. The result of the reactions was visualized on an 8% TBE polyacrylamide gel (Sambrook et al., 1989). As a control for the digestion of nonbranched dsDNA, 1 µg blunt-end double-stranded phage λ DNA (a polymerase-treated phage λ DNA from A&A Biotechnology, catalogue No. 3500-500, DNA marker λ/HindIII) was used as a substrate DNA (instead of X-shaped DNA) with Hjc_15-6 that had been purified twice. The reaction mixture and incubation time were as before, although the incubation temperature was kept at

### Table 1

| DNA oligomers used to build X-shaped DNA. |
|------------------------------------------|
| Oligo 1 5’-CGAGTGTGTGTCAGATACGTTATGCTCCTGCG-3’ |
| Oligo 2 5’-GCAGACATGCTACATACCTGATCGTCCG-3’ |
| Oligo 3 5’-CGACGCGAATGATACCAGTGTTACGCTG-3’ |
| Oligo 4 5’-TGCGTCGAGGTACGCTGGTATTTCGCAAGAATCG-3’ |

Acta Cryst. (2022). D78, 212–227

Josefin Ahlqvist et al. • Holliday junction-resolving enzyme 215
37°C. The result was visualized on a 1% agarose gel (Sambrook et al., 1989).

2.10. Crystallization and X-ray diffraction data collection

Native crystals were obtained using Hjc_15-6 produced in defined (mAT) medium and purified as described previously. The crystals were grown using protein solution consisting of 5.4 mg ml⁻¹ Hjc_15-6 in 20 mM SPG buffer system pH 7.0 (Molecular Dimensions). The crystallization drops were set up using 200 nl protein solution, 50 nl seed solution and 150 nl reservoir solution (1.8 M ammonium sulfate, 0.1 M sodium acetate pH 4.2). The seed solution (crushed native Hjc_15-6 crystals in 100 mM sodium acetate pH 4.8, 2 M ammonium sulfate, 0.5 mg ml⁻¹ native Hjc_15-6) was obtained using seed beads (Hampton Research) from initial native crystals grown under the same conditions. The crystals were grown at 20°C in MRC 3-well plates over 40 µl reservoir solution. Since attempts to use the molecular-replacement method to determine the structure did not succeed, selenium-derivatized (SeMet) Hjc_15-6 was produced. One SeMet Hjc_15-6 crystal was grown in the same way as described for the native Hjc_15-6 crystals but using protein at 3 mg ml⁻¹. Another SeMet Hjc_15-6 crystal was grown without seeding in a 200 nl protein drop with 200 nl reservoir solution (1.8 M ammonium sulfate, 0.1 M sodium acetate pH 4.2) added. All crystals appeared in 1–2 weeks. Prior to data collection and just before flash-cooling in liquid nitrogen, the crystals were briefly transferred to a cryosolution [100 mM sodium acetate pH 4.2, 1.8 M ammonium sulfate, 25% (v/v) glycerol].

Ten data sets for native Hjc_15-6 were collected to 2.55 Å resolution at a wavelength of 1.7701 Å on beamline I04 at Diamond Light Source (DLS), United Kingdom. The data were processed in XDS (Kabsch, 2010) and all data sets were scaled using XSICALE (Kabsch, 2010). These data were collected in an attempt to obtain sulfur SAD (single anomalous determination) phases, but turned out to give highly redundant and good native data, although the sulfur signal was too weak for phasing using the CRANK2 pipeline (Skubák et al., 2018). Instead, data sets for SeMet Hjc_15-6 were collected on beamline I04 at DLS at the peak wavelength for selenium (0.9795 Å). The data sets were processed in XDS (Kabsch, 2010) and two of these data sets were combined using XSICALE (Kabsch, 2010) to 2.50 Å resolution. Significant anomalous signal extended to 3.2 Å resolution. The structure was determined using the CRANK2 pipeline (Skubák et al., 2018) included in the CCP4 suite using the combined peak data sets and the Hjc_15-6 sequence. Subsequent model building was performed in Coot (Emsley et al., 2010) with refinement in REFMAC5 (Murshudov et al., 2011) followed by refinement in BUSTER (Bricogne et al., 2011).

3. Results and discussion

3.1. Gene identification, expression and purification

Genome sequencing of phage Tth15-6 resulted in one contig of about 76 kb. Analysis of the nucleotide sequence revealed a close similarity (95.2% identity with 70% query coverage) to the Thermus bacteriophage TSP4 isolated from hot springs in Tengchong, People’s Republic of China (Lin et al., 2010). In addition, the phages G20c and P23-45 from the Geyser Valley in Iceland and P74-26 from the Uzon Valley in Kamchatka, Russia that infect T. thermophilus HB8 (Mina-khin et al., 2008) are related, although these homologous proteins have diverged significantly in amino-acid sequence from their counterparts from phages Tth15-6 and TSP4. ORFs within the phage Tth15-6 genome were predicted by the RAST server and suggested the presence of a gene encoding a putative DNA-resolving enzyme/helicase-like enzyme of 155 amino-acid residues extending from coordinate nucleotide 18521 to nucleotide 18988 in the phage Tth15-6 genome. A sequence-similarity search with BlastP confirmed that the encoded protein, termed Hjc_15-6, showed sequence similarity to other deposited sequences of putative Hj-resolving enzymes.

et al., 2005). A theoretical molecular-mass calculation was also performed using ProtParam. A sequence-similarity search was performed using BlastP (https://blast.ncbi.nlm.nih.gov/Blast.cgi; NCBI Resource Coordinators, 2016) against the nonredundant protein-sequence database (in September 2021), excluding models (XM/XP), nonredundant RefSeq proteins (WP) and uncultured/environmental sample sequences, but otherwise using the default settings. Evolutionary analysis was performed and a phylogenetic tree was obtained using MEGA version X (Kumar et al., 2018) using the data from the BlastP search. Alignments were performed using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) through Jalview version 2.11.14 (Waterhouse et al., 2009) using the ClustalO web service with the default settings, unless stated otherwise. Protein attribution to (super)families was performed with InterPro version 83.0 (http://www.ebi.ac.uk/interpro/; Blum et al., 2021). Conservative sequence motifs were investigated in the NLM Conserved Domain Database (CDD) version 3.18 (Lu et al., 2020). Illustrations of protein structures were prepared with CCP4MG (McNicholas et al., 2011) and UCSF Chimera version 1.15 (Petterson et al., 2004). Protein structure comparisons were made with the PDB coordinates for chain A of Hjc_15-6 using the DALI server (http://ekhidna2.biocenter.helsinki.fi/dali/; Holm, 2020). All protein illustrations and protein structure comparisons were made using the SeMet-derivatized model of Hjc_15-6 (PDB entry 7bgs), unless stated otherwise. The predicted oligomeric state of Hjc_15-6 was analysed with the coordinates of the native and SeMet-derivatized structures using the Protein Interfaces, Surfaces and Assemblies service (PIASA) at the European Bioinformatics Institute (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html; Krissinel & Henrick, 2007).
The gene encoding the putative Hj-resolving enzyme from phage Tth15-6 was subsequently PCR-amplified and cloned in frame with a C-terminal hexahistidine tag in a pET-series expression vector; after heat-shock transformation it was expressed in E. coli BL21(DE3) cells using either 1 l complex (LB) medium in shake flasks or 2.5 l defined (mAT) medium in a fermenter. The yield was comparatively high using both cultivation strategies, and the crude cell extract constituted almost 50% of the soluble target protein according to qualitative estimation after visualization with SDS–PAGE (Fig. 1a). Typically, a total of approximately 75 mg of Hjc_15-6 was obtained after protein purification by nickel-affinity chromatography using both cultivation strategies. The production of selenium-derivatized Hjc_15-6 using 0.5 l defined (mAT) medium supplemented with SeMet resulted in a total of approximately 0.8 mg protein after nickel-affinity purification.

3.2. Molecular mass, pI and oligomeric state

The deduced amino-acid sequence of full-length Hjc_15-6 (155 amino acids) corresponds to a theoretical average molecular mass of 17 580 Da (without the hexahistidine tag), which increases to 18 646 Da when the C-terminal hexahistidine tag (LEHHHHHHH) is included. The molecular mass of purified recombinant Hjc_15-6 in solution was successfully determined by MS. The purified Hjc_15-6 included the hexahistidine tag, although the first amino acid (Met) of the full-length protein was lost, most likely due to N-terminal methionine excision (NME) after overproduction (Bonissone et al., 2013). With these modifications, the recombinant protein has an expected theoretical average molecular mass of 18 514 Da and a theoretical monoisotopic mass of 18 502 Da. The calibrated MS spectrum confirmed that the actual molecular mass of recombinant Hjc_15-6 is between these values. Furthermore, molecular masses of 18 507 Da for recombinant protein produced in complex (LB) medium and 18 510 Da for protein produced in defined (mAT) medium were determined.

Figure 1
(a) Lanes 1 and 2: SDS–PAGE of crude extract supernatant and purified Hjc_15-6 protein produced in shake flasks using complex (LB) medium. Lanes 3 and 4 correspond to lanes 1 and 2 where the crude extract supernatant and purified Hjc_15-6 protein were instead produced in a fermenter using defined (mAT) medium. Lanes L contain molecular-mass markers (Bio-Rad). (b) Internally calibrated mass spectrogram (z = +1) of recombinant Hjc_15-6 protein overproduced in shake flasks using complex (LB) medium (upper black spectrogram) and Hjc_15-6 protein overproduced in a fermenter using defined (mAT) medium (lower pink spectrogram). (c) Externally calibrated mass spectrogram (z = +2) of overproduced Hjc_15-6 protein. The lower blue spectrogram (graph A) corresponds to Hjc_15-6 that was overproduced in shake flasks using complex (LB) medium and the upper green spectrogram (graph B) corresponds to Hjc_15-6 protein that was overproduced in a fermenter using defined (mAT) medium. Peaks A2 and B2 have an additional mass of 96.4 and 97.3 Da compared with peaks A1 and B1, respectively, and peaks B3 and B4 have an additional mass of 97.6 and 93.2 Da compared with peaks B2 and B3, respectively.
by MS, showing that no misincorporation of amino acids occurred in either medium (Fig. 1b).

Interestingly, MS analysis revealed a difference in the additional groups interacting with Hjc_15-6 from the two production strategies. The MS spectra of Hjc_15-6 expressed in defined (mAT) medium had three additional peaks with higher molecular masses, suggesting that some of the protein had one or two additional molecular groups with estimated molecular masses of 97.3 and 97.6 Da and also a third group with a calculated molecular mass of 93.2 Da (peaks B2, B3 and B4 in Fig. 1c). The corresponding MS spectra of Hjc_15-6 expressed in complex (LB) medium had only one additional peak, indicating a molecular group with an estimated molecular mass of 96.4 Da (peak A2 in Fig. 1c). These groups may be either sulfate (molecular mass 96 Da) or phosphate (molecular mass 95 Da), as discussed further below. The pl value of Hjc_15-6 was estimated as 9.7.

Dynamic light-scattering (DLS) measurements demonstrated that the polydispersity was only 9 ± 1%, confirming the monodispersity of the pure Hjc_15-6 sample regardless of the production medium and/or the cultivation mode [complex (LB) medium and shake flasks or defined (mAT) medium and fermenter]. Furthermore, the radius of 3.5 ± 0.5 nm in both cases indicates that the oligomeric state of the resolvase is either a dimer or a dimer of dimers. Analysis of the polydispersity of the coordinates in the Semet-derivatized structure (PDB entry 7bgs) and the native structure (PDB entry 7bnx) of Hjc_15-6 derived using PISA suggests that Hjc_15-6 is more likely to occur in a tetrameric form, with \( \Delta G^{2u} \) values of 6.1 and 8 kcal mol\(^{-1}\), respectively, for a tetrameric assembly and −1.7 and −1.3 kcal mol\(^{-1}\), respectively, for a dimeric assembly. However, since Hjc-resolving enzymes are reported to use only two active sites in the dimeric state (Wyatt & West, 2014), it may be that Hjc_15-6 adopts a tetrameric form for purposes other than its Hj-resolving activity. Due to this uncertainty, Hjc_15-6 is discussed and displayed below both as a homodimer and a dimer of homodimers.

3.3. Sequence analysis

A search with BlastP shows that the amino-acid sequence of Hjc_15-6 has the highest sequence similarity to putative Hj-resolving enzymes from other thermophilic phages that infect *Thermus* species, including *Thermus* phage TSP4 (GenBank QAY181291; 96.1% identity, 100% query coverage). *Thermus* phage P74-26 (GenBank ABU96992; Minakhin et al., 2008; 51.9% identity, 98% query coverage), *Thermus* virus P23-45 (GenBank ABU96877; Minakhin et al., 2008; 52.6% identity, 98% query coverage) and *Thermus* phage G20c (GenBank API81850; Xu et al., 2017; 51.3% identity, 98% query coverage). Their total alignment score is 99 ± 1% and the E-values are 3.0 \( \times 10^{-35} \) or less. The putative Hj-resolving enzyme from the thermophilic phage *Thermus* phage phiFa (GenBank ATJ74718.1) has a relatively low sequence similarity to Hjc_15-6, despite having quite a high query coverage (29.4% identity, 87% query coverage). In the phylogenetic tree (Fig. 2a) all proteins with a sequence query cover of 45% or more are displayed, along with four structurally related enzymes of archaeal origin. In Fig. 2(b) the amino-acid sequences of the proteins in the phylogenetic tree are displayed along with three additional relevant conservative domains found in the NML Conserved Domain Database (CDD). Hjc_15-6, which is displayed in the top row in Fig. 2(b), is aligned with the sequences in rows 2–9, which include the conservative domains and Hj-resolving enzymes from *P. furiosus* and all phage proteins found in BlastP except for *Thermus* phage phiFa.

Sequence analysis of Hjc_15-6 indicates that approximately the first 60 amino acids (the N-terminus) of the 155-residue sequence represent the more conserved part of the enzyme. In this part, a search in InterPro shows homology with two superfamilies: the tRNA endonuclease-like domain superfamily (IPR011856) and the restriction endonuclease type II-like superfamily (IPR011335). A search in the NLM CDD (version 3.18, 55570 PSSMs, default settings and full mode) generated a match with the restriction endonuclease-like superfamily (cd01037) and the three conserved domains COG1591, Archeal_HJR (cd00523) and Hjc (pfam1870) (mentioned above and displayed in Fig. 2b), which are all described as Hj-resolving enzymes of archaeal type. The Hjc domain is known to be the archaeal equivalent of RuvC. However, it has quite a different amino-acid sequence (Komori et al., 2000), and the Archeal_HJR domain has been described to show structural similarity to type II restriction endonucleases.

According to InterPro, the tRNA_endonuclease-like domain superfamily (IPR011856) has overlapping entries with the three domain types tRNA intron endonuclease-catalytic domain-like (IPR006677), VRR-NUC domain (IPR014883) and PD(D/E)/K domain endonuclease (IPR021671). It has also been described to represent a structural domain found in three groups of endonucleases: TsnA endonucleases (Ronning et al., 2004; Xu et al., 2017), tRNA-intron endonucleases (Li et al., 1998) and Hjc-type resolving enzymes (classified as archaeal Holliday junction endonucleases; Nishino et al., 2001; Xu et al., 2017). The Restrct_endonuc-II-like superfamily (IPR011335) has been described to represent the core structure found in most type II restriction endonucleases.

Type II restriction endonucleases are involved in protecting bacteria and archaea against invading foreign DNA. Most of them are homodimeric or tetrameric enzymes with a similar structural core. They require Mg\(^{2+}\) ions for catalysis and cleave DNA independently of ATP at defined sites of 4–8 bp in length. Even if they differ in the details of the recognition process, it has been suggested that they evolved from a common ancestor (Xu et al., 2017; Pingoud & Jeltsch, 2001; Pingoud et al., 2005; Nakonieczna et al., 2009).

A comparison of type II restriction endonucleases and other restriction endonucleases belonging to the PD-(D/E)/K superfamily illustrates that these enzymes have a similar core which harbours the active site (one per subunit) and comprise a five-stranded mixed \( \beta \)-sheet flanked by \( \alpha \)-helices (Pingoud et al., 2005; Venclovas et al., 1994; Pingoud & Jeltsch, 2001; Nishino et al., 2001; Winkler, 1992), which matches Hjc_15-6,
Figure 2

(a) Phylogenetic tree made in MEGA X based on proteins found in BlastP with a query coverage of 45% or more (obtained in September 2021) along with four structurally related enzymes found by the DALI server. (b) All protein sequences in the phylogenetic tree are displayed along with three relevant conservative domains found in the NML Conserved Domain Database (CDD). Rows 1–9 correspond to Hjc_15-6 aligned with the Hj-resolving enzyme from P. furiosus (found by the DALI server) and all phage proteins found in BlastP, except for Thermus phage phiFA. The alignment was performed in Jalview (Waterhouse et al., 2009) using the ClustalO web service with default settings and aligned output order, with colouring by ClustalX.

In this study, the conserved residue Glu10 is suggested to complement the PD-(D/E)X catalytic motif and the respective SD-EVK motif (found in the phage sequences except for that of Thermus phage phiFA) to build a proposed signature. The proposed signature is marked with black boxes and consists of residues E$_{10}$-SD$_{40}$-EVK$_{55}$ in Hjc_15-6 and other aligned Hj-resolving enzymes originating from phages (rows 1–2 and rows 6–8) and residues E$_{10}$-PD$_{40}$-EVK$_{55}$ in the conservative domains (rows 3–5).
as discussed below. Proteins belonging to the PD-(D/E)XK superfamily have various functions such as replication, restriction, DNA repair and tRNA–intron splicing (Steczkie-wicz et al., 2012). The PD-(D/E)XK motif is known to be a catalytic sequence motif that is involved in binding Mg²⁺ and in cleavage of the phosphodiester bond in the substrate DNA backbone (Dupureur & Dominguez, 2001; Pingoud & Jeltsch, 2001; Pingoud et al., 2014). Based on extensive analysis, Pingoud and coworkers pointed out that a majority of the experimentally characterized type II restriction endonucleases for which full-length sequences are available belong to the PD-(D/E)XK phosphodiesterase superfamily, which also includes other nucleases; for instance Sulfolobus solfataricus Hj-resolving enzyme (Pingoud et al., 2014).

To summarize, amino-acid sequence analysis with InterPro and CDD suggests that Hjc_15-6 may be classified as an archaeal type of Hj-resolving enzyme that shares common features with many other endonucleases, not least with the type II restriction endonucleases that belong to the PD-(D/E)XK superfamily.

In the alignment between Hjc_15-6 and the conservative domain sequences (Fig. 2b), it is obvious that all conservative domains align with the PD-(D/E)XK motif. Hjc_15-6 residues 39–40 and 53–55 align rather well with this motif; however, the hydrophobic proline is now replaced by the polar residue serine (Fig. 2b, row 1).

Several attempts have been made to improve the classification of the PD-(D/E)XK superfamily (Aravind et al., 2000; Kosinski et al., 2005; Laganekas et al., 2011; Steczkiewicz et al., 2012). Daiyasu and coworkers investigated whether Asp40, Glu53 and Lys55 (in the alignment above) were critical for activity in the archaean Hj-resolving enzyme from P. furiosus (the residues are marked with black boxes in row 9 of the alignment in Fig. 2b) by replacing them with alanines. The study showed that all of these residues were crucial for catalytic activity (Daiyasu et al., 2000). Hence, the archaean Hj-resolving enzyme from P. furiosus is related to the type II restriction endonucleases. However, as seen in the alignment, its catalytic motif is VD-(D/E)XK rather than PD-(D/E)XK (Fig. 2b).

It has been suggested that archaean Hj-resolving enzymes, together with Mrr (methylated adenine-recognition and restriction)-like endonucleases, should define a new family (Aravind et al., 2000). This family has the endonuclease fold and the notable conservation of three motifs [which are part of the PD-(D/E)XK catalytic motif identified above] centred around a constellation of charged residues that could form the active site. These three motifs are described as follows: (G/P)X₁₀,X₉₋₁₁,G(F/Y) (motif I), hDhhXₚ (motif II) and hh(EE/D/O)KK (motif III), where X is any amino acid, p is any polar amino acid and h is a hydrophobic amino acid. Glu in motif I is strictly conserved (Aravind et al., 2000) and may therefore be crucial for catalytic activity. The N-terminal motif (motif I) is predicted to form a helix, and the other two motifs form β-sheets. By comparing these results with other closely related families, it has been suggested that archaean Hj-resolving enzymes coordinate a divalent cation (Mg²⁺) via Asp in motif II, Glu or Gln in motif III and one of the O atoms of the scissile phosphodiester group. Lys in motif III is likely to contact the phosphate of the DNA backbone (Aravind et al., 2000).

The conclusions mentioned above align rather well with Hjc_15-6, giving the following three motifs: (G)X₁₀,X₁₀,G(F/Y) (motif I), pDhpXₚ (motif II) and pphEHK (motif III). The exceptions in Hjc_15-6 are that some surrounding residues are classified as polar in motifs II and III rather than hydrophobic. Despite this difference in hydrophobicity, it may be suggested that a putative catalytic site for Hjc_15-6 is Glu₁₀–Asp₄₀–Glu₅₃–Lys₅₅, where Asp₄₀ takes part in binding an Mg²⁺ ion together with Glu₅₃ and one of the O atoms of the scissile phosphodiester group, and Lys₅₅ makes contact with the DNA backbone.

The close relationship between the sequences of Hjc_15-6 and putative Hj-resolving enzymes from thermophilic phages can be seen in rows 1–2 and 6–8 in the alignment in Fig. 2b. Interestingly, except for the putative Hj-resolving enzyme from Thermus phage phiFa, all of the phage proteins display the three-part catalytic motif E-SD-EVK. Yet, at the same time, the corresponding motifs found in Hj-resolving enzymes from other species are E-PD-EVK or even E-VD-EVK, as in the case of the archaean Hj-resolving enzyme from P. furiosus. Based on this observation and the discussion above, we propose that Hjc_15-6 and several other Hj-resolving enzymes originating from thermophilic phages will have a three-part signature motif corresponding to E-SD-EVK. Furthermore, based on previous studies, as discussed above, it is suggested that Glu from motif I, Asp from motif II and Glu and Lys from motif III are the crucial catalytic site residues responsible for phosphodiester bond cleavage in conjunction with Mg²⁺ on the substrate DNA.

3.4. Crystal structure

The crystal structure of Hjc_15-6 was determined to 2.54 Å resolution using SAD with selenium-derivatized protein (Table 2). Eight Se atoms were found in the asymmetric unit with an occupancy of >25%, and two Hjc_15-6 molecules. After solvent flattening the mean figure of merit (FOM) was 0.59, and after automatic model building Rmerge and Rfree were 0.33 and 0.34, respectively. After further model building and refinement, the final model included two Hjc_15-6 molecules, polypeptide chains A and B, with residues 5–28, 39–85, 98 and 100–149 of chain A and residues 5–29, 39–90, 98 and 100–150 of chain B visible in the electron density as well as eight sulfate or phosphate ions (modelled as sulfates) and 37 water molecules (PDB entry 7bgs). This model was then used to determine the native structure of the Hjc_15-6 protein. The native structure is similar to the SeMet-derivatized structure and includes polypeptide chains A and B, with residues 5–28, 39–85, 98 and 100–149 of chain A and residues 5–28, 39–80, 98 and 100–150 of chain B visible in the electron density as well as eight sulfate or phosphate ions (modelled as sulfates) and 32 water molecules (PDB entry 7bmx).
3.5. Structural analysis

The secondary structure of the Hjc_15-6 monomer is composed of seven β-strands and six α-helices (Figs. 3a and 3b). Hj-resolution enzymes have so far been reported to function as dimers (Wyatt & West, 2014); however, both the DLS and PISA results in this study suggest that Hjc_15-6 may occur as a dimer of homodimers. Hjc_15-6 is presented as a homodimer in Fig. 3(c) and as a dimer of homodimers in Fig. 3(d), both coloured according to secondary structure. It can be seen in Fig. 3(d) that the Hjc_15-6 tetramer has a X-shaped appearance, in which the two dimers seem to cross each other at the centre of the tetramer with an angle of 90°. However, in Fig. 4(a) it can be seen that dimer I (dark and light blue) is in the front of dimer II (dark red and pink).

As discussed previously, the putative active-site residues in the proposed motif Eββ-SDββ-DEVK53-55, except for residue 39, agree with archaeal Hj-resolution enzymes, as predicted by others (Aravind et al., 2000). Moreover, this signature is situated within the most conservative part of Hjc_15-6, where the putative active-site residue Glu10 (motif I) is located in α-helix 1, Asp40 (motif II) in β-sheet 2 and Glu53 and Lys55 (motif III) in β-sheet 3 (Fig. 3b). However, Ser39 prior to Asp40 (motif II) is not shown to be part of β-sheet 2. This may be due to the gap between residues 30 and 37, which are not seen in any chain in the two models. It can be seen in Figs. 3(a) and 3(b) that the first 60 residues (N-terminus) of Hjc_15-6, which represent the conservative part of the protein, and the core around the signature motif have the typical endonuclease fold built by an αββαβαβ topology (Kinch et al., 2005).

The electron-density map of Hjc_15-6 suggested that each monomer has two bound sulfate or phosphate groups with full occupancy. In addition, there are two further sulfate- or phosphate-binding sites in both monomers that have an occupancy of 0.5 and/or have high B values. Nishino and coworkers suggested that sulfate groups in the P. furiosus Hjc enzyme act as scaffolds for the N-terminal conformation. They also suggested that the sulfate ions are essential for successful crystallization (Nishino et al., 2001). The latter observation was also experienced in this study when crystallizing Hjc_15-6. The fully occupied sites are displayed in Fig. 4(b), along with the proposed active-site residues Glu10, Asp40, Glu53 and Lys55 in the suggested catalytic motif Eββ-SDββ-DEVK53-55 of the Hjc_15-6 protein, viewed as a tetramer (a dimer of homodimers). The image is drawn at an angle where the two dimers are situated alongside each other. It is interesting to note that at this angle it can be seen that the tetramer has a tunnel through its centre. Fig. 4(c) shows the surface electron density of the tetramer displayed at the same angle as in Fig. 4(b). In the close-up of the tunnel (Fig. 4d), several of the active-site residues from chain A in both dimer I and dimer II are oriented close to the surface of the tunnel. The overall orientation of the active-site residues is displayed in Fig. 4(e), where the electrostatic surface potential is shown along with the active-site residues represented as blooids. The distances between the closest residues across the tunnel are estimated to be around 20 Å, as calculated using CCP4MG and displayed in Fig. 4(d). It might be that DNA binds in this cavity. However, this is uncertain for several reasons. Firstly, as discussed above, we are not convinced that Hjc_15-6 is a tetramer. Secondly, the tunnel may be smaller than modelled since some residues are missing. Finally, the Hj-resolving activity should be covered by the first third of the conservative part of the Hjc_15-6 protein, viewed as a tetramer in dimer I (Fig. 4d). If the tetramer is rotated one more time by 90° to the left, it can also be seen that there is another cleft between chain A in dimer II and chain B in dimer I. However, it cannot be excluded that DNA may bind in the tunnel, and if this is the case it might be that Hjc_15-6 has an additional function beyond acting as an Hj-resolving enzyme.

Table 2

| Protein     | SeMet Hjc_15-6 | Native Hjc_15-6 |
|-------------|----------------|-----------------|
| PDB code    | 7bgs           | 7bnx            |
| Data collection† |                |                 |
| Space group  | C222,          | C222,           |
| a, b, c (Å)  | 97.58, 103.71, 84.85 | 96.71, 104.17, 84.86 |
| α, β, γ (%)  | 90, 90, 90     | 90, 90, 90      |
| Resolution  | 30.2-50 (2.60-2.50) | 29-2.55 (2.66-2.55) |
| Rmerge      | 0.111 (1.735)  | 0.067 (0.698)   |
| Rfree‡      | 0.027 (0.560)  | 0.009 (0.186)   |
| CC1/2       | 0.999 (0.631)  | 1.000 (0.953)   |
| (I/i(I))    | 18.2 (1.1)     | 47.2 (2.9)      |
| Completeness (%) | 99.9 (100.0) | 98.8 (98.7) |
| Multiplicity | 16.7 (10.3)    | 50.9 (12.7)     |
| No. of Se sites | 8               |                 |
| FOM (CRANK2) | 0.59            |                 |
| Refinement  |                |                 |
| No. of unique reflections | 15236           | 14269          |
| Rmerge/Rfree‡ | 0.226/0.258    | 0.220/0.254    |
| No. of atoms |                |                 |
| Protein chains A/B | 961/1025       | 952/1006       |
| Sulphates/water | 40/37           | 40/32          |
| B factors (Å²) | 91.9/95.2      | 90.9/95.9      |
| Protein chains A/B | 100.0/67.9     | 98.8/66.2      |
| Sulphates/water |                |                 |
| R.m.s. deviations |                |                 |
| Bond lengths (Å) | 0.009           | 0.009          |
| Bond angles (°) | 2.0             | 0.97           |
| Ramachandran statistics (%) |                |                 |
| Favourable | 97.0            | 96.6           |
| Allowed | 3.0             | 2.6            |
| Disallowed | 0.0             | 0.4            |
| MolProbity score | 1.68 [99th percentile] | 1.80 [99th percentile] |
| Clashscore | 5.29 [99th percentile] | 5.15 [99th percentile] |

† Two crystals were used for the SeMet Hjc_15-6 data and one crystal was used for the native data. ‡ 5% of the data were used for the refinement set.
to four unique UniProt entries that are all thermophilic archaeal Hj-resolving enzymes. Two of the Hj-resolving enzymes, Q97YX6 (HJE_SACS2) and Q7LXU0 (HJC_SACS2), have a significantly different substrate specificity from each other, cleaving Holliday junctions either in a sequence-independent (HJE_SACS2) or a sequence-dependent manner, even though both proteins originate from genes in *Saccharolobus solfataricus* (Kvaratskhelia & White, 2000). The other two Hj-resolving enzymes originate from *Archaeoglobus fulgidus* and *P. furiosus* (see Table 3 for an overview of the PDB entries and the corresponding proteins).

All archaeal Hj-resolving enzymes from the best structurally matching molecules (as obtained in DALI) were superposed on the Hjc_15-6 monomer (chain A), and are displayed in Figure 3.

![Figure 3](image-url)

**Figure 3**

(a) The secondary structure of the Hjc_15-6 monomer is displayed based on data from both chains A and B as described in UCSF Chimera and CCP4MG. (b) Hjc_15-6 protein monomer coloured by secondary structure with α-helices (red helices) and β-sheets (blue arrows) numbered in order from the N-terminus to the C-terminus. The image also includes the putative active site with the catalytic residues Glu10, Asp40, Glu53 and Lys55 labelled and presented in ball-and-stick representation (Fig. 3b was produced in UCSF Chimera with the native coordinates of PDB entry 7bnx chain B). (c) Hjc_15-6 protein displayed as a homodimer. Chains A and B are slightly differently modelled. (d) Hjc_15-6 protein displayed as a dimer of homodimers. In (c) and (d), Hjc_15-6 is also coloured by secondary structure, with α-helices as red tubes and β-sheets as blue arrows; both images were produced in CCP4MG.

Table 3

All PDB entries that match Hjc_15-6 (chain A) with a Z-score of >4, a %id PDB of ≥20 and a Lali (length of alignment) of ≥73 using the DALI server.

| PDB entry | Protein name | UniProtKB entry | Protein length | R.m.s.d. (Å) | Species |
|-----------|--------------|-----------------|----------------|-------------|---------|
| 2wiz (chain B) | Holliday-junction resolving enzyme Hjc | Q28314 (HJC_ARCFU) | 139 | 2.8–3.1 | *Archaeoglobus fulgidus* |
| 1ob8 (chain B) | Holliday-junction resolving enzyme Hjc | Q97YX6 (HJE_SACS2) | 135 | 3.0–3.1 | *Saccharolobus solfataricus* |
| 1ipi (chain B) | Holliday-junction resolving enzyme Hjc | E7FHX4 (HJC_PYRFU) | 123 | 3.0–3.1 | *Pyrococcus furiosus* |
| 1ob9 (chain A) | Holliday-junction resolving enzyme Hjc | Q7LXU0 (HJC_SACS2) | 143 | 3.0–4.1 | *Saccharolobus solfataricus* |
Figure 4
Images of Hjc_15-6 as a tetramer. Chains A and B in dimer I are shown as dark and light blue ribbons and chains A and B in dimer II are shown as dark red and pink ribbons in (a), (b), (d), (f) and (h). The active-site residues Glu10, Asp40, Glu53 and Lys55 are drawn as cylinders and coloured according to atom type in (b), (d) and (g). Phosphate/sulfate groups are drawn as sulfate groups with cylinders coloured according to atom type in (b), (e), (f) and (g). In (e) the electrostatic surface potential is transparent, with bloboids of active-site residues coloured in green (dimer I) and pink (dimer II). In (a) it can be seen that dimer I is in front of dimer II. In (b)–(e) all images are displayed at the same angle, revealing a tunnel in the centre of the Hjc_15-6 tetramer that is viewed as a close-up in (d), where two distances have been measured. (f) and (g) present views rotated from the bottom to the top by 90° from those in (b)–(e), revealing a cleft between chain A in dimer I and chain B in dimer II. In (h) and (i) the tetramer is further rotated 90° to the left compared with (f) and (g), revealing another cleft between chain A in dimer II and chain B in dimer I.
up to residue 60 in Fig. 5(a). It can be seen that all enzymes correspond very well to each other in this conservative part. When superposing only the 60-residue conservative part in CCP4MG, it was found that Hjc from *P. furiosus* matches Hjc_15-6 best. The structural orientation of Glu10, Glu53, Val54 and Lys55 from the catalytic motif of Hjc_15-6 is strikingly similar to the corresponding residues from Hjc of *P. furiosus* as seen in Fig. 5(b). However, all archaeal Hj-resolving enzymes give an r.m.s.d. (root-mean-square deviation) value between 1.74 and 1.88 Å. Even if the conservative parts of Hjc_15-6 and the archaeal Hj-resolving enzymes are very similar to each other, it can be seen that their dimeric organizations are somewhat divergent (Figs. 5c–5g), except for HJC_ARCFU (*A. fulgidus*) and HJC_SACS2 (*S. solfataricus*), which are quite similar in their dimeric organizations (Figs. 5d and 5g). Based on the observations above, we may conclude that Hjc_15-6 has a typical archaeal endonuclease fold and a dimer organization that resembles those of other archaeal Hj-resolving enzymes.

To date, there are (to the best of our knowledge) three available structures of Hj-resolving enzymes originating from phages. One of them is a mutant variant from phage bIL67 that resembles the *E. coli* RuvC fold (Green *et al.*, 2013) and the other two are (and have the folds of) T4 endonuclease VII and T7 endo I, respectively (Bond *et al.*, 2001). Hjc_15-6 should be most closely related to T7 endo I, since the active site of T7 endo I is much more similar to archaeal Hjc-resolving enzymes than to T4 endonuclease VII (which is almost totally α-helical) and the RuvC fold (which has a different topology; Hadden *et al.*, 2001). Yet, T7 endo I did not come up as a result in the DALI search. Furthermore, their dimeric organizations are quite different from each other, as can be seen in Figs. 5(c) and 5(h). The difference between Hjc_15-6 and T7 endo I shows that this study presents a new fourth fold for Hj-resolving enzymes originating from phages. This aligns well with the previous report that both T7 endo I and archaeal Hj-resolving enzymes belong to the endonuclease fold, although T7 endo I is quite divergent within this class (Aravind *et al.*, 2000).

### 3.6. Activity screening for resolving enzyme activity

Activity screening demonstrated that purified recombinant Hjc_15-6 acts as a resolvase, cleaving Holliday-junction oligomers. Two affinity-chromatography techniques were exploited for Hjc_15-6 purification as a precaution against contamination when recombinant Hjc_15-6 was prepared for...
activity characterization. The reactions were performed with Hjc_15-6 in the presence of Mg$^{2+}$, using marked fluorescent DNA shaped as Holliday junctions, with or without ATP. After 30 min of incubation at 50°C, the Hjc_15-6 digestion of DNA was visualized by TBE–PAGE as fragmented DNA that has migrated down the gel (Fig. 6a). The cleavage reaction seems to be independent of the presence of ATP, which is a feature that Hjc_15-6 shares with type II restriction endonucleases (Pingoud & Jeltsch, 2001). Since no endonuclease activity was observed on polymerase-treated blunt-end dsDNA (Fig. 6b) the activity screening demonstrates that the Hjc_15-6 sample is not contaminated with other endonucleases and that Hjc_15-6 has specificity towards Holliday junctions in contrast to nonbranched dsDNA. However, further studies are necessary for detailed characterization of the resolving activity and specificity of Hjc_15-6.

4. Conclusion

This study presents the first structure of an Hj-resolving enzyme, that of Hjc_15-6, originating from a T. thermophilus phage. Based on both sequence conservation and structural architecture, it can be assigned to the archaeal Holliday junction-resolving enzymes. The structure was initially solved using selenium-derivatized Hjc_15-6 (PDB entry 7bgs) and subsequently also with native Hjc_15-6 (PDB entry 7bnx). The enzyme demonstrates typical features of type II restriction endonucleases, in particular the endonuclease fold. Activity screening confirmed the Hj-resolving activity of Hjc_15-6; it does not cleave blunt-end λ/HindIII DNA fragments.

Sequence analysis and comparison with other publically available sequences of putative Hj-resolving enzymes from thermophilic phages and related studies have led us to propose that most Hj-resolving enzymes originating from thermophilic phages will have a three-part signature motif corresponding to E–SD–EVK and that the Glu (motif I), Asp (motif II) and Gln and Lys (motif III) residues are crucial residues responsible for phosphodiester bond cleavage in conjunction with Mg$^{2+}$ on the substrate DNA. Furthermore, since Hjc_15-6 has an archaeal fold, this study demonstrates a new fourth fold for Hj-resolving enzymes originating from phages.

Acknowledgements

We are grateful for discussions with members of the VIRUS-X project consortium and for the technical assistance of Katja Bernfur, Lund University in the MS analysis. The authors declare that they have no conflicts of interest with the contents of this article.

Funding information

The following funding is acknowledged: European Union Horizon 2020 Research and Innovation Programme Virus-X: Viral Metagenomics for Innovation Value (grant no. 685778).

References

Aevarsson, A., Kaczorowska, A., Adalsteinsson, B. T., Ahlqvist, J., Al-Karadaghi, S., Altenbuchner, J., Arsin, H., Atlason, U., Brandt, D., Cichowicz-Cieslak, M., Cornish, K. A. S., Courtin, J., Dabrowski, S., Dahle, H., Djefiane, S., Dorawa, S., Dusaucy, J., Enault, F., Fedty, A., Freitag-Pohl, S., Fridjonsson, O. H., Galiez, C., Glomsaker, E., Guérin, M., Gundeson, S. E., Gudmundsdóttir, E. E., Gudmundsson, H., Hákansson, M., Henke, C., Helleux, A., Henriksen, J. R., Hjörleifsdóttir, S., Hreggvidsson, G. O., Jasilionis, A., Jocheheim, A., Jónsdóttir, I., Jónsdóttir, L. B., Jureczak-Kurek, A., Kaczorowski, T., Kalinowski, J., Kozlowski, L. P., Krupovic, M., Kwiatkowska-Semrau, K., Laine, O., Lange, J., Lebrat, J., Linares-Pastén, J., Liu, Y., Lorentsen, S. A., Lutterman, T., Mas, T., Merré, W., Mirdita, M., Morzywolek, A., Ndela, E. O., Karlsson, E. N., Olgudóttir, E., Pedersen, C., Perler, F., Petursdóttir, S. K., Plotka, M., Pohl, E., Prangishvili, D., Ray, J. L., Reynisson, B.,
Röbertsdöttrir, T., Sandaa, R., Szycba, A., Skirnisdóttir, S., Ödering, J., Solstad, T., Steen, J. H., Stafnebø, S. K., Steinegger, M., Overá, K. S., Stribley, B., Svensson, A., Szadkowski, M., Tarrant, E. J., Terzian, P., Tourigny, M., Bergh, T., Vanhalst, J., Vincent, J., Vrolijk, B., Walse, B., Wang, L., Watzlawick, H., Welin, M., Werbowy, O., Wons, E. & Zhang, R. (2021). FEBS Microbiol. Lett. 568, fnab067.

Aralvin, L., Makarova, K. S. & Koonin, E. V. (2000). Nucleic Acids Res. 28, 3417–3432.

Ariyoshi, M., Vassylyev, D. G., Iwasaki, H., Nakamura, H., Shinagawa, H. & Morikawa, K. (1999). Cell, 78, 1063–1072.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Dism, T., Edwards, R. A., Formma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, E. K., Paarmann, D., Paczian, T., Parchello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, W., Wilke, A. & Zagnitko, O. (2008). BMC Genomics, 9, 75.

Biertümpfel, C., Yang, W. & Suck, D. (2007). Nature, 449, 616–620.

Birkenbihl, R. P., Neef, K., Prangishvili, D. & Kemper, B. (2001). J. Mol. Biol. 309, 1067–1076.

Blum, M., Chang, H.-Y., Chuangarusk, S., Grego, T., Kandasamy, S., Mitchell, A., Nuka, G., Pasyan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G. A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D. H., Letunic, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Necci, M., Orenso, C. A., Pandurangan, A. P., Rivoire, C., Sigrist, C. J. A., Stillioe, I., Thanki, N., Thomas, P. D., Tosatto, S. C. E., Wu, C. H., Bateman, A. & Finn, R. D. (2021). Nucleic Acids Res. 49, D344–D354.

Bond, C. S., Kvaratskhelia, M., Richard, D., White, M. F. & Hunter, W. N. (2001). Proc. Natl Acad. Sci. USA, 98, 5509–5514.

Bonnissone, S., Gupta, N., Romine, M., Bradshaw, R. A. & Pevzner, P. A. (2013). Mol. Cell Proteomics. 12, 14–28.

Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O. S. & vonHrneh, C. (2011). BUSTER. Global Phasing Ltd, Cambridge, UK.

Culby, M. J., Hwang, Y., Minkah, N. & Bushman, F. D. (2009). J. Biol. Chem., 284, 1190–1201.

Daiyasu, H., Komori, K., Sakae, S., Ishino, Y. & Toh, H. (2000). Nucleic Acids Res. 28, 4540–4543.

Duprere, C. M. & Dominguez, M. A. (2001). Biochemistry, 40, 387–394.

Emsley, P., Lohakamp, B., Scott, W. G. & Cowtan, K. (2010). Acta Cryst. D66, 486–501.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). The Proteomes Protocols Handbook, edited by J. M. Walker, pp. 571–607, Totowa: Humana Press.

Green, V., Curtis, F. A., Sedelnikova, S., Rafferty, J. B. & Sharples, G. J. (2013). Mol. Microbiol. 89, 1240–1258.

Hadden, J. M., Convery, M. A., Décals, A.-C., Lilley, D. M. J. & Phillips, E. S. V. (2001). Nat. Struct. Biol. 8, 62–67.

Hjorleifsdottir, S., Skirnisdottir, S., Hreggvidsson, G. O., Holst, O. & Kristjansson, J. K. (2001). FEBS Lett. 284, 1190–1201.

Holliday, R. (1964). Genet. Res. 5, 282–304.

Holm, L. (2000). Protein Sci. 9, 128–140.

Holme, T., Arvidson, S., Lindholm, B. & Pavlul, B. (1970). Process Biochem. 5, 62–66.

Jalasvuori, M., Jaatinen, S. T., Laurinavicius, S., Ahola-Iivarinen, E., Kalkkinen, N., Bamford, D. H. & Bamford, J. K. H. (2009). J. Virol. 83, 9388–9397.

Kabsch, W. (2010). Acta Cryst. D66, 125–132.

Kemper, B. & Brown, D. T. (1976). J. Virol. 18, 1000–1015.

Kemper, B. & Janz, E. (1976). J. Virol. 18, 992–999.

Khromov, I., Ackermann, H.-W., Washburn, M., Mushegian, A. & Severinov, K. (2006). J. Mol. Biol. 364, 667–677.

NCBI Resource Coordinators (2016). Nucleic Acids Res. 44, D7–D19.

Nishino, T., Komori, K., Tsuchiya, D., Ishino, Y. & Morikawa, K. (2001). Structure, 9, 197–204.

Pavlova, N. V., Inman, R., Minakhin, L., Kozyavkin, S. A., Khromov, I., Ackermann, H.-W., Washburn, M., Mushegian, A. & Severinov, K. (2008). J. Mol. Biol. 378, 468–480.

Pingoud, A., Jeltsch, A. (2001). Nucleic Acids Res. 29, 3705–3727.

Pingoud, A., Wilson, G. G. & Wende, W. (2014). Nucleic Acids Res. 42, 7489–7527.

Pingoud, A., Vix, O., Törö, L., Goltz, S., Kemper, B. & Suck, D. (1999). EMBO J. 18, 1447–1458.

Raoikjjarmer, H., Vix, O., Törö, L., Golz, S., Kemper, B. & Suck, D. (1999). EMBO J. 18, 1447–1458.

Ronning, D. R., Li, Y., Perez, Z. N., Ross, P. D., Hickman, A. B., Craig, N. L. & Dyda, F. (2004). EMBO J. 23, 2972–2981.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual. New York: Cold Spring Harbor Laboratory Press.

research papers
Skubík, P., Araç, D., Bowler, M. W., Correia, A. R., Hoelz, A., Larsen, S., Leonard, G. A., McCarthy, A. A., McSweeney, S., Mueller-Dieckmann, C., Otten, H., Salzman, G. & Pannu, N. S. (2018). *IUCrJ*, 5, 166–171.

Steczkiewicz, K., Muszewska, A., Knizewski, L., Rychlewski, L. & Ginalska, K. (2012). *Nucleic Acids Res.* **40**, 7016–7045.

Tamakoshi, M., Murakami, A., Sugisawa, M., Tsunecizumi, K., Takeda, S., Saheki, T., Izumi, T., Akiba, T., Mitsuoka, K., Toh, H., Yamashita, A., Arisaka, F., Hattori, M., Oshima, T. & Yamagishi, A. (2011). *Bacteriophage*, 1, 152–164.

Turner, P., Pramhed, A., Kanders, E., Hedström, M., Karlsson, E. N. & Logan, D. T. (2007). *Acta Cryst.* **F63**, 802–806.

Venclovas, Č., Timinskas, A. & Siksnys, V. (1994). *Proteins*, **20**, 279–282.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. (2009). *Bioinformatics*, **25**, 1189–1191.

Winkler, F. K. (1992). *Curr. Opin. Struct. Biol.* **2**, 93–99.

Wyatt, H. D. M. & West, S. C. (2014). *Cold Spring Harb. Perspect. Biol.* **6**, a023192.

Xu, R. G., Jenkins, H. T., Chechik, M., Blagova, E. V., Lopatina, A., Klimuk, E., Minakhin, L., Severinov, K., Greive, S. J. & Antson, A. A. (2017). *Nucleic Acids Res.* **45**, 3580–3590.

Yu, M. X., Slater, M. R. & Ackermann, H. W. (2006). *Arch. Virol.* **151**, 663–679.