

# Restriction endonuclease TseI cleaves A:A and T:T mismatches in CAG and CTG repeats

Long Ma<sup>1</sup>, Kai Chen<sup>1</sup>, David J. Clarke<sup>1</sup>, Christopher P. Nortcliffe<sup>1</sup>, Geoffrey G. Wilson<sup>2</sup>, J. Michael Edwardson<sup>3</sup>, A. Jennifer Morton<sup>3</sup>, Anita C. Jones<sup>1</sup> and David T. F. Dryden<sup>1,\*</sup>

<sup>1</sup>EaStChem School of Chemistry, University of Edinburgh, The King's Buildings, Edinburgh EH9 3JJ, UK, <sup>2</sup>New England BioLabs, Inc., 240 County Road, Ipswich, MA 01938, USA and <sup>3</sup>Department of Pharmacology, University of Cambridge, Cambridge CB2 1PD, UK

Received April 13, 2012; Revised and Accepted February 25, 2013

## ABSTRACT

The type II restriction endonuclease TseI recognizes the DNA target sequence 5'-G<sup>^</sup>CWGC-3' (where W = A or T) and cleaves after the first G to produce fragments with three-base 5'-overhangs. We have determined that it is a dimeric protein capable of cleaving not only its target sequence but also one containing A:A or T:T mismatches at the central base pair in the target sequence. The cleavage of targets containing these mismatches is as efficient as cleavage of the correct target sequence containing a central A:T base pair. The cleavage mechanism does not apparently use a base flipping mechanism as found for some other type II restriction endonuclease recognizing similarly degenerate target sequences. The ability of TseI to cleave targets with mismatches means that it can cleave the unusual DNA hairpin structures containing A:A or T:T mismatches formed by the repetitive DNA sequences associated with Huntington's disease (CAG repeats) and myotonic dystrophy type 1 (CTG repeats).

## INTRODUCTION

Type II restriction endonucleases have found many uses in molecular biology because of their ability to cleave DNA molecules with extraordinary precision at specific sequences of base pairs. Thousands of restriction enzymes have been discovered, and many are available commercially (1). The restriction enzyme TseI, isolated originally from a thermophilic bacterium of the genus *Thermus*, displays optimal activity at ~65°C. TseI recognizes the symmetric but ambiguous five-base pair sequence 5'-G<sup>^</sup>CWGC-3' in double-strand (ds) DNA (where W = A or T) and cleaves after the first G to produce

fragments with three-base 5'-overhangs (1). ApeKI, a similar enzyme of identical specificity from the archaeon *Aeropyrum*, displays optimum activity at even higher temperatures (2).

X-ray crystallography has revealed two distinct structural classes of restriction enzymes that recognize quasi-symmetric DNA sequences of this kind. Restriction enzymes MvaI (specificity: CC<sup>^</sup>WGG) and BcnI (CC<sup>^</sup>SGG where S = G or C) bind to DNA and are active as small monomeric proteins (3–5). The proteins contain only one catalytic site and accomplish ds cleavage by sequential nicking and hydrolysing first one DNA strand and then the other in separate binding events of opposite orientation (6,7). Each monomer interacts with all five base pairs of the recognition sequence. The interactions leading to recognition of the four defined base pairs are straightforward (3,4), but those leading to recognition of the central ambiguous base pair are less well understood, and thought perhaps to involve reversals in hydrogen-bond configurations (6).

Restriction enzymes PspGI (<sup>^</sup>CCWGG) and Ecl18kI (<sup>^</sup>CCNNG where N = any base), in contrast, bind to DNA as homodimeric proteins (8,9). As each subunit has a catalytic site, the homodimers have two and can accomplish ds cleavage simultaneously, in a single binding event, albeit that for Ecl18kI requires additional interactions between neighbouring molecules at flanking sites (10). Each subunit of PspGI and Ecl18kI interact in a conventional way with only half of the base pairs that make up the recognition sequence, namely, the two outer C:Gs that form each half-site. In both enzymes, remarkably, the central bases are flipped from the helix into pockets, and the gap left behind is compressed, effectively reducing the recognition sequences to just CCGG (8,9). Recognition of the central base pair is thought to take place mainly within the pockets; these accommodate any base in Ecl18kI, but in some way discriminate against G or C in PspGI (11). When 2-aminopurine (2-AP) is

\*To whom correspondence should be addressed. Tel: +0131 650 4735; Fax: +0131 650 6453; Email: david.dryden@ed.ac.uk  
Present address:

Long Ma, School of Chemistry, University of St. Andrews, North Haugh, St. Andrews, Fife KY16 9ST, UK

substituted for adenine at the central position in the recognition sequence, binding by PspGI and Ecl18kI produces a marked increase in fluorescence because of base unstacking, a signature of base flipping (12,13).

PspGI and Ecl18kI are similar to each other in both structure and amino acid (aa) sequence. They are distinct from BcnI and MvaI, which themselves are also similar in sequence and structure. The structure of TseI has not been solved, but its aa sequence, and that of the closely similar ApeKI, shows little similarity to those of the other four enzymes, suggesting that TseI might belong to a third structural class, distinct from either of the other two. It is unclear, then, whether TseI binds to its target sequence as a monomer or a homodimer, and whether the central base pair remains intra-helical during recognition or becomes flipped.

Given its recognition sequence, TseI will cleave CAG and CTG trinucleotide repeats, which are involved in the aetiology of a number of neurodegenerative diseases, such as Huntington's disease (HD; CAG repeats) and myotonic dystrophy type 1 [CTG repeats; reviewed in (14,15)]. There is evidence that triplet repeats produce unusual DNA structures, such as triplexes, hairpins, slipped-strand DNA and G-quadruplexes (16). These structures affect the mobility of DNA in agarose gels (e.g. 17), and they have also been directly observed by electron microscopy (18,19). Recently, we used atomic force microscopy (AFM) to analyse DNA samples of various CAG repeat lengths (20). We found that the structural profile of the DNA changed significantly as repeat length increased. DNA from wild-type mice appeared as short linear molecules, whereas when the CAG repeat length increased, various DNA structures, including convolutions, folds and protrusions, became apparent. Over 50% of DNA molecules containing 408 CAG repeats contained one of these unusual structures. We showed that the convoluted DNA was sensitive to mung bean nuclease, indicating that it contained hairpin mismatches. We then used TseI, to further characterize the structures observed in the super-long CAG repeats. We found that at room temperature, TseI preferentially cleaved linear regions of 858-repeat DNA, leaving behind the contorted regions. In contrast, at 80°C, TseI completely digested the DNA. These observations suggested that TseI preferentially cleaves CAG repeats within normal B-form DNA, but that at higher temperatures, it can also cleave DNA containing central A:A or T:T mismatches; mismatches that will be present in the hairpins. We speculated that like PspGI, TseI might also flip out the central bases, and in doing so, it might cleave such mismatched sequences by accommodating adenine in both pockets, in the case of A:A mismatches, or thymine in both pockets, in the case of T:T mismatches. In the current study, we set out to test this idea.

## MATERIALS AND METHODS

### TseI enzyme

The amino acid and DNA sequences for TseI are available in GenBank under accession numbers JN035228 and

AEN19713. We thank David Hough of New England Biolabs for the kind gift of purified protein. The protein (378 aa; 5000 U/ml) had a concentration of 10.8  $\mu\text{M}$  in terms of TseI monomers based on a molecular weight of 41 780 Da and an extinction coefficient at 280 nm of 37930  $\text{M}^{-1} \text{cm}^{-1}$ . ProtParam (21) at <http://web.expasy.org/cgi-bin/protparam/protparam> was used to calculate these values, assuming all cysteine residues were reduced and the N-terminal methionine was unprocessed.

### Preparation of 2-aminopurine labelled oligonucleotides

Apart from oligonucleotides containing 2-aminopurine (2-AP), all oligonucleotides were obtained from ATDbio (Southampton). Solid-phase synthesis of 2-AP-containing oligonucleotides was performed using phosphoramidite chemistry on a MerMade DNA/RNA oligonucleotide synthesiser (BioAutomation, USA) in a 5'-trityl (4,4'-dimethoxytrityl, DMT) group-on manner. Purification of the synthesised oligonucleotides was done using standard two-stage DMT-on/DMT-off reverse phase high-performance liquid chromatography (HPLC). The DMT-on full-length products possessed a prolonged retention time during reverse phase HPLC purification and were easily separated from failed sequences. The reaction was detritylated with 40% acetic acid/water, and a further DMT-off reverse phase HPLC purification step was applied for higher purity. A NAP-10 column (GE Healthcare) was used for desalting, and a Speedvac was used to concentrate the synthesized products. All synthesized oligonucleotides were examined by ESI-FTICR mass spectrometry to confirm an accurate molecular mass. All the synthesized products were quantified using UV-Vis absorbance at 260 nm with the extinction coefficient determined using Integrated DNA biophysics online software (<http://biophysics.idtdna.com/UVSpectrum.html>).

### Size-exclusion chromatography analysis for investigating molecular mass of TseI in solution at different protein concentrations

An analytical HPLC gel filtration column calibrated with protein standards (apoferritin 443 kDa,  $\beta$ -amylase 200 kDa, alcohol dehydrogenase 150 kDa, bovine serum albumin 66 kDa and carbonic anhydrase 29 kDa) was used to determine the molecular weight of TseI in a buffer composed of 20 mM Tris-HCl, 20 mM 2-(*N*-morpholino)ethanesulfonic acid (MES), 10 mM magnesium chloride, 7 mM  $\beta$ -mercaptoethanol, 200 mM sodium chloride and 0.1 mM ethylenediaminetetraacetic acid (EDTA), pH 6.5, at room temperature as previously described (22). The low pH is required for stability of the silica-based chromatography material. The TseI was tested at concentrations from 4000 to 100 nM.

### Measurement of DNA binding to TseI by fluorescent anisotropy

The anisotropy duplex labelled at the 5'-end on one strand with hexachlororfluorescein (HEX) was used, Table 1. Fluorescence anisotropy measurements were performed using an Edinburgh Instrument FS900 photon counting spectrofluorometer using a T-format measurement and

analysed as previously described (23). The excitation wavelength was 535 nm, emission wavelength was 555 nm and bandwidths were 5 nm. Eight hundred microlitres of 10 nM of the anisotropy duplex was placed in a 10-mm path length quartz cuvette at 25°C. Small amounts of TseI were added to the duplex solution using a micro-litre syringe to achieve enzyme concentrations from 5 to 320 nM and gently mixed by magnetic stirring. The buffer was 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM calcium acetate and 1 mM dithiothreitol, pH 7.9.

### Steady-state measurements of 2-AP-substituted DNA and DNA-TseI complexes

The 2-AP-labelled oligonucleotides were synthesized, annealed to fully or partially complementary strands (Table 1) and used for investigating the interaction of TseI with DNA. All duplexes were buffered with 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM calcium

acetate and 1 mM dithiothreitol, pH 7.9. Two hundred microlitres of samples of 250 nM duplex plus 1500 nM TseI was incubated for 20 min at 25°C. Steady-state fluorescence spectra were measured using a FluoroMax (Horiba Jobin Yvon) photon counting spectrofluorometer. Spectra were recorded with a bandpass of 10 nm for both excitation (315 nm) and emission.

### Fluorescence-based TseI activity assay

The substrate duplexes consisted of a 5'-HEX-labelled oligonucleotide annealed to a complementary strand labelled at the 3'-end with a black hole quencher 1 group, Table 1. The buffer was 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate and 1 mM dithiothreitol, pH 7.9. The duplexes are essentially non-fluorescent in the ds-state and highly fluorescent in the single-stranded state achieved by cleavage with TseI at a temperature greater than the melting temperature of

**Table 1.** Oligonucleotides and duplexes used in this study

Duplex name	Sequence
2-AP:T duplex 1	AGGAGTGAAGT <b>CGC2GC</b> CCCGTGCTCAAG (5'-3') TCCTCACTTCAG <b>CGT</b> CGGGCACGAGTTC (3'-5')
A:2-AP duplex 2	AGGAGTGAAGT <b>CGCAGC</b> CCCGTGCTCAAG (5'-3') TCCTCACTTCAG <b>CG2C</b> GGGCACGAGTTC (3'-5')
T:2-AP duplex 3	AGGAGTGAAGT <b>CGCAGC</b> CCCGTGCTCAAG (5'-3') TCCTCACTTC <b>2GC</b> T <b>CG</b> GGGCACGAGTTC (3'-5')
2-AP:A duplex 4	AGGAGTGAAGT <b>CGC2GC</b> CCCGTGCTCAAG (5'-3') TCCTCACTTCAG <b>CGAC</b> GGGCACGAGTTC (3'-5')
2-AP:2-AP duplex5	AGGAGTGAAGT <b>CGC2GC</b> CCCGTGCTCAAG (5'-3') TCCTCACTTCAG <b>CG2C</b> GGGCACGAGTTC (3'-5')
Anisotropy duplex A:T	HEX -AGGAGTGAAGT <b>CGCAGC</b> CCCGTGCTCAAG (5'-3') TCCTCACTTCAG <b>CGT</b> CGGGCACGAGTTC (3'-5')
Anisotropy duplex G:C	HEX -AGGAGTGAAGT <b>CGCGC</b> CCCGTGCTCAAG (5'-3') TCCTCACTTCAG <b>CGC</b> CGGGCACGAGTTC (3'-5')
Anisotropy duplex A:A	HEX -AGGAGTGAAGT <b>CGCAGC</b> CCCGTGCTCAAG (5'-3') TCCTCACTTCAG <b>CGAC</b> GGGCACGAGTTC (3'-5')
Anisotropy duplex T:T	HEX -AGGAGTGAAGT <b>CGTGC</b> CCCGTGCTCAAG (5'-3') TCCTCACTTCAG <b>CGT</b> CGGGCACGAGTTC (3'-5')
Fluorescence assay A:T duplex	HEX -AGGAGTGAAGT <b>CGCAGC</b> CCCGTGCTCAAG (5'-3') BHQ1-TCCTCACTTCAG <b>CGT</b> CGGGCACGAGTTC (3'-5')
Fluorescence assay A:A mismatch duplex	HEX -AGGAGTGAAGT <b>CGCAGC</b> CCCGTGCTCAAG (5'-3') BHQ1-TCCTCACTTCAG <b>CGAC</b> GGGCACGAGTTC (3'-5')
Fluorescence assay T:T mismatch duplex	HEX -AGGAGTGAAGT <b>CGTGC</b> CCCGTGCTCAAG (5'-3') BHQ1-TCCTCACTTCAG <b>CGT</b> CGGGCACGAGTTC (3'-5')
Fluorescence assay product duplex	HEX -AGGAGTGAAGT <b>CG</b> (5'-3') BHQ1-TCCTCACTTCAG <b>CGT</b> (3'-5')
A:T duplex	AGGAGTGAAGT <b>CGCAGC</b> CCCGTGCTCAAG (5'-3') TCCTCACTTCAG <b>CGT</b> CGGGCACGAGTTC (3'-5')
A:A duplex	AGGAGTGAAGT <b>CGCAGC</b> CCCGTGCTCAAG (5'-3') TCCTCACTTCAG <b>CGAC</b> GGGCACGAGTTC (3'-5')
T:T duplex	AGGAGTGAAGT <b>CGTGC</b> CCCGTGCTCAAG (5'-3') TCCTCACTTCAG <b>CGT</b> CGGGCACGAGTTC (3'-5')
G:C duplex	AGGAGTGAAGT <b>CGCGC</b> CCCGTGCTCAAG (5'-3') TCCTCACTTCAG <b>CGC</b> CGGGCACGAGTTC (3'-5')
G:G duplex	AGGAGTGAAGT <b>CGCGC</b> CCCGTGCTCAAG (5'-3') TCCTCACTTCAG <b>CGCG</b> GGGCACGAGTTC (3'-5')
13mer product sequence	AGGAGTGAAGT <b>CG</b> (5'-3')
15mer product sequence	Phos- <b>CAGC</b> CCCGTGCTCAAG (5'-3')
12mer product sequence	<b>GGGCACGAGTTC</b> (3'-5')
16mer product sequence	TCCTCACTTCAG <b>CGT</b> -Phos (3'-5')

Recognition sites of TseI are in bold.

HEX, hexachlorofluorescein group; BHQ1, black hole quencher 1; Phos, phosphate group; 2-AP, 2-aminopurine analogue base.

the cleavage products. The increase of the signal can easily be quantified and initial reaction rates determined. All samples were analysed with an Edinburgh Instrument spectrofluorometer with excitation at 530 nm, emission at 555 nm and 5 nm bandwidths. Each DNA sample (100  $\mu$ l) was placed in a quartz microcuvette with 10-mm path length, put in the fluorometer and allowed to reach the assay temperature of 60°C, whereupon a stable fluorescence background signal was achieved. The fluorescence intensity as a function of time was immediately recorded after addition of TseI to a constant monomer concentration of 32.4 nM. Photon counts were converted to the amount of 5'-HEX-labelled oligonucleotide product released using a calibration curve. The substrate concentration versus initial cutting rate was plotted and fitted using the Michaelis-Menten equation  $\{v = V_{max}(S)/[K_M + (S)]\}$  to determine the kinetic parameters  $K_M$ ,  $V_{max}$  and  $k_{cat}$ . To determine the rate of melting of the products, a small aliquot of the fluorescence assay product duplex (1.12  $\mu$ M), Table 1, was diluted with buffer at 60°C in the microcuvette and the fluorescence signal recorded.

### Mass spectrometry

For mass spectrometry analysis of oligonucleotides, LC-MS was used, with detection performed in the negative mode. Briefly, an Ultimate 3000 HPLC system was used (Dionex Corporation, Sunnyvale, CA, USA), equipped with an Aeris Widepore C18 reverse phase analytical column 50  $\times$  2.1 mm (Phenomenex). Six-microlitre samples containing  $\sim$ 5  $\mu$ M of DNA duplex or TseI-treated duplex were injected onto the column. For chromatography, mobile phase A and B were prepared comprising 98:2 water:ammonia and 49:49:2 water:MeOH:ammonia, respectively. Samples were injected onto the analytical column, washed with mobile phase A for 10 min, followed by a 20-min linear gradient elution (200  $\mu$ l/min) into 100% mobile phase B. MS data was acquired on a Bruker 12 Tesla SolarixQe Fourier Transform Ion Cyclotron Resonance (FT-ICR) instrument (Bruker Daltonics, Billerica, MA, USA) equipped with an electrospray ionization (ESI) source. Desolvated ions were transmitted to a 6-cm Infinity cell Penning trap. Trapped ions were excited (frequency chirp 48–500 kHz at 100 steps of 25  $\mu$ s) and detected between  $m/z$  600 and 2000 for 0.5 s to yield a broadband 512 Kword time-domain data set. Fast Fourier Transforms and subsequent analyses were performed using DataAnalysis (Bruker Daltonics) software. Isotope distributions of specific charge states were predicted from theoretical empirical formulas. These were overlaid on the recorded experimental data.

### Polyacrylamide gel showing the cleavage of matched and mismatched duplexes by TseI

Unlabelled duplexes, Table 1, were used in this experiment at a concentration of 10  $\mu$ M. In all, 2.5  $\mu$ l of TseI stock solution (10.8  $\mu$ M) was added to give a final TseI concentration of 0.26  $\mu$ M. The buffer was 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate

and 1 mM dithiothreitol, pH 7.9. All samples were incubated at 65°C for  $\sim$ 6 h in the reaction buffer and then 2.5  $\mu$ l run on a 15% polyacrylamide gel in 1  $\times$  Tris/Borate/EDTA buffer at 150 V, stained with SYBR Gold and viewed under UV light. Single strands corresponding to the reactants and products, Table 1, were run as markers. The DNA marker ladder was O'RangeRuler five base pairs DNA Ladder (Thermo Scientific).

### Denaturing HPLC analysis for TseI-matched and -mismatched cutting

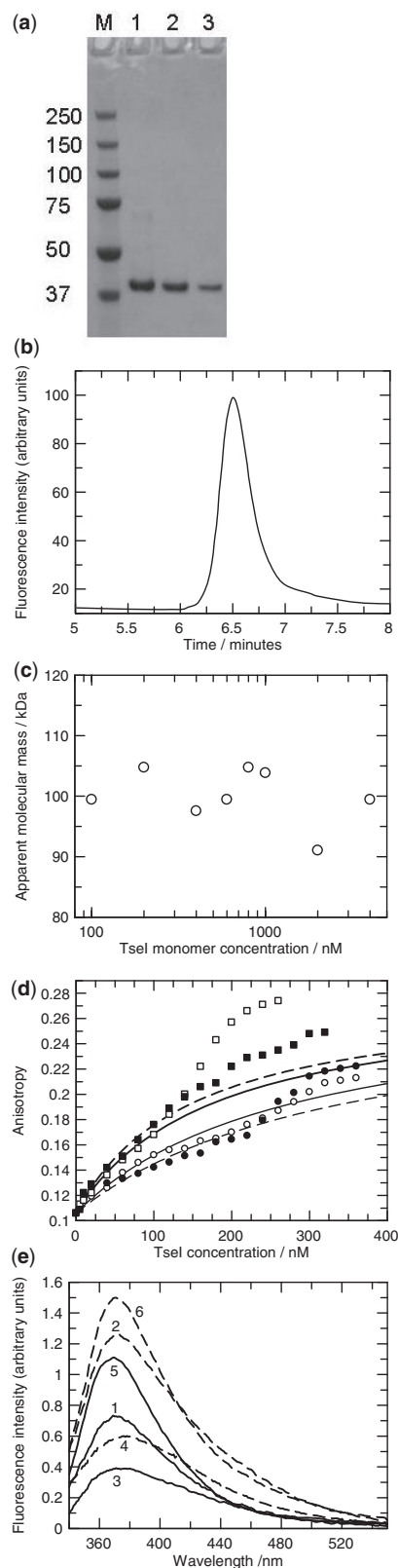
Unlabelled duplexes, Table 1, were used in this experiment at a concentration of 10  $\mu$ M. When cleavage was to be tested, 2.5  $\mu$ l of TseI stock solution (10.8  $\mu$ M) was added to 100  $\mu$ l of the duplex solution to give a final TseI concentration of 0.26  $\mu$ M. All samples were incubated at 65°C for  $\sim$ 12 h. The buffer was 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate and 1 mM dithiothreitol, pH 7.9. Twenty-microlitre samples were injected onto the HPLC column. The injection solely of TseI in the absence of the duplex gave no HPLC signal above the baseline. A Gilson HPLC equipped with absorption detection at 254 nm was used for analysis of DNA cleavage. The reverse phase column (C18 Jones Chromatography) was thermostatted at 65°C. A linear acetonitrile gradient from 5 to 65% acetonitrile was generated by mixing 0.1 M acetic acid, 5% acetonitrile and 0.1 M acetic acid and 65% acetonitrile aqueous solution. The pH of the two solvents was adjusted to 6.5 with triethylamine.

## RESULTS

### Characterization of TseI

A preparation of TseI was analysed in SDS-PAGE (Figure 1a) and by size-exclusion chromatography (Figure 1b and c). The SDS-PAGE gel showed a single band after Coomassie blue staining at a molecular weight of  $\sim$ 42 kDa close to that expected from the amino acid sequence. Further analysis with a calibrated analytical size-exclusion chromatography column showed a single elution peak at a constant elution time irrespective of protein concentration in a range from 100 to 4000 nM. The calibration indicated a molecular weight of  $100 \pm 18$  kDa, suggesting that TseI is a homodimer in the buffer used (Figure 1c).

Binding of TseI to HEX-labelled duplexes containing the target sequence or A:A, T:T or G:C base pairs at the middle of the sequence was examined using the increase in fluorescence anisotropy of the HEX label caused by the slowing of molecular rotation when the mass of the duplex is increased by protein binding (23), Figure 1d. Binding to the target sequence or to the G:C duplex occurred following a normal one-site ligand-binding equation until a protein concentration of  $\sim$ 120 nM was reached. Above this concentration, an additional binding event became visible. Binding to the A:A and T:T mismatched duplexes was weaker than to the properly base paired duplexes, but it again followed a one-site binding equation until a protein concentration of  $\sim$ 220 nM was



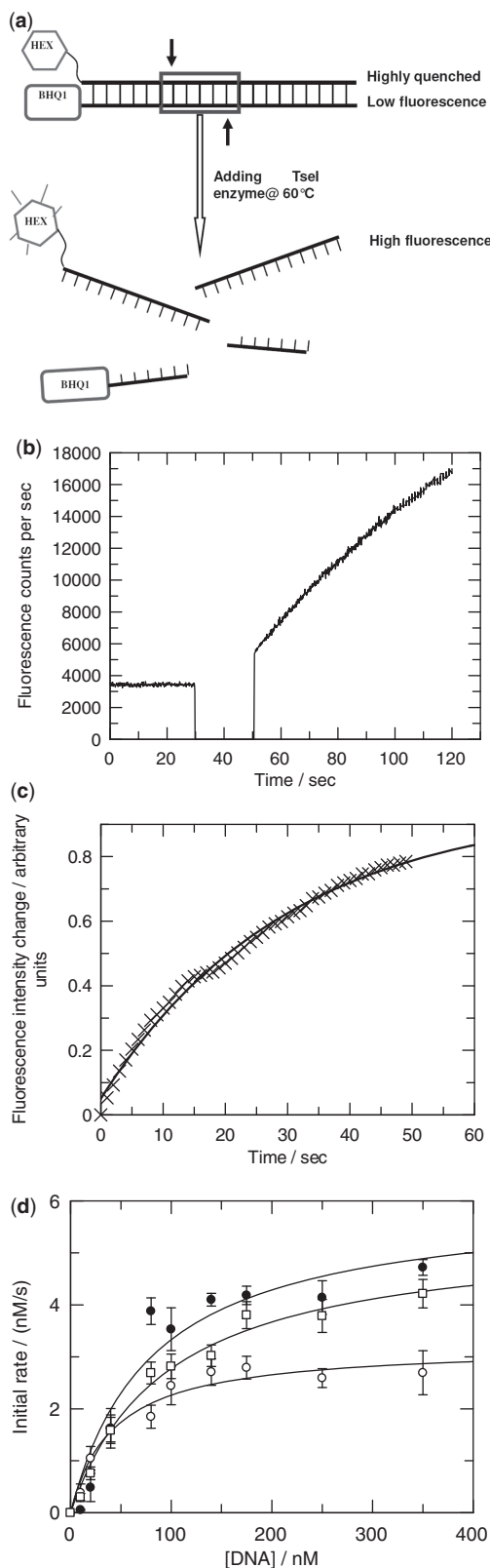
**Figure 1.** Analysis of TseI enzyme. (a) 4–12% gradient SDS–PAGE gel stained with Coomassie blue. Lane 1 shows the molecular mass markers (kDa); lanes 2–4 show TseI samples at 5.4, 2.7 and 1.35  $\mu$ M, respectively. (b) Size-exclusion chromatography, as monitored by fluorescence emission at 350 nm with excitation at 295 nm, to investigate the solution molecular mass. The elution profile (1000 nM shown) showed a single peak at  $\sim$ 6.5 min corresponding to a molecular mass of 100 kDa. (c)

reached. Above this concentration, an additional binding event became visible as the anisotropy increased further. These additional binding events can only be due to an increase in the amount of a species with even greater molecular mass than the complex of one DNA duplex with one enzyme molecule. We attribute this additional binding event to non-specific binding of an extra copy of the protein to the protein–DNA complex. In support of this, we note that the final values of the anisotropy were similar to those previously observed for HEX-labelled DNA binding to M.EcoKI, an enzyme of molecular weight 170 kDa (23). The dissociation constants obtained by fitting the one-site binding equation to data up to 120 nM protein concentration gave values of  $176 \pm 7$ ,  $149 \pm 6$ ,  $279 \pm 9$  and  $347 \pm 30$  nM for the duplexes containing A:T, G:C, A:A and T:T base pairs, respectively, at the central position of the target. Thus, TseI binds less well to the distorted duplexes containing the mismatched base pair. It is of interest that the binding of TseI to the G:C duplex is almost identical to its binding to the A:T duplex, even though the former sequence lacks the recognition sequence for the enzyme. Thus, these data also suggest that TseI must acquire its sequence specificity after the binding event but before or concomitant with the cleavage event.

Several restriction enzymes recognizing DNA target sequences similar to that recognized by TseI have been shown by crystallography and 2-AP fluorescence studies to flip out the central base pair in the target sequence and to collapse the DNA duplex, thus converting a five-base pair recognition sequence into a four-base pair recognition sequence. The intensity increases in 2-AP emission because of enzyme binding, and base flipping ranged from 10- to 1000-fold for these enzymes. No structure is available for TseI; therefore, the possibility that TseI uses base flipping was investigated by replacing the central bases with 2-AP either paired with T (2-AP:T) or mismatched with A (2-AP:A) or with another 2-AP (2-AP:2-AP). As a control, 2-AP was also placed outside of the target sequence and paired with T. The fluorescence of the duplexes containing 2-AP showed a typical 2AP emission spectrum when excited at 315 nm with emission maximum at 370 nm, Figure 1e. The addition of excess TseI to ensure near complete binding of the duplex only

#### Figure 1. Continued

Dependence of molecular mass as a function of protein concentration injected onto the column. (d) Fluorescence anisotropy increase for 10 nM hexachlorofluorescein-labelled 28-bp DNA duplex as a function of TseI monomer concentration. Duplex with the target sequence (open squares, bold solid line), duplex with an A:A mismatch (open circle, solid line), duplex with G:C at centre of the target (solid squares, bold dashed line) and duplex with T:T mismatch (solid circles, thin dashed line). Lines are determined using a one-site binding equation for data up to a TseI concentration of 120 nM. (e) Steady-state fluorescence emission spectra of 250 nM 2-aminopurine-substituted DNA in the absence/presence of 1.5  $\mu$ M TseI enzyme. The solid lines represent the fluorescence intensity of DNA in the absence of TseI for duplexes 2-AP:T duplex 1 (spectrum 1), A:2-AP duplex 2 (spectrum 3) and T:2-AP duplex 3 (spectrum 5). The dotted lines represent the fluorescence intensity of the same duplexes in the presence of excess TseI (spectra 2, 4 and 6, respectively).



**Figure 2.** Fluorescence-based assay of TseI activity. (a) The 5'-HEX-labelled top strand was annealed with 3'-Black Hole Quencher (BHQ) 1-labelled strand and became highly quenched. Adding TseI at elevated temperature (60°C) results in cleavage of the duplex, separation of the fluorophore-quencher pair and the appearance of fluorescence. (b) Fluorescence intensity as a function of time. Initially, a low signal was observed. After 30s, the sample chamber was opened, and an

caused a small increase (up to ~50%) in the emission intensity without changing the emission maximum wavelength. This change in intensity was observed for all locations and base pairings of the 2-AP probe. We suspect that the modest intensity changes observed are due to minor distortion of the DNA by the enzyme and the placing of the 2-AP in a more hydrophilic environment rather than base flipping *per se*. However, it is also possible that the presence of an aromatic aa, such as tryptophan, in proximity to base-flipped 2-AP could quench the 2-AP fluorescence and disguise base flipping. In this case, its fluorescence would be similar in magnitude to that observed for the duplex with 2-AP located outside the recognition site. However, in the absence of further structural information, we take these results to suggest that base flipping does not occur when TseI binds to its DNA target.

#### A continuous fluorescence-based assay demonstrates that TseI cleaves its target sequence even if it contains an A:A or T:T mismatch

The endonuclease activity of TseI on short DNA duplexes containing the target sequence and variations thereof was investigated using a continuous fluorescence assay. The assay uses the difference in thermal stability of the 28 bp substrate and the shorter products, which will be single stranded at the assay temperature, to give a spectroscopic signal, Figure 2a. This assay was initially proposed by Waters *et al.* (23), who used the increase in absorption because of the melting of the short products, and was subsequently converted to fluorescence measurements by, for example, Li *et al.* (24), who used the melting of the short products to remove a fluorescence quencher from contact with a fluorescence HEX reporter. Provided that the assay temperature lies between the melting temperature of the substrate and the products, the products melt on cleavage by the enzyme, and the fluorescence of the fluorophore is greatly enhanced. This assay works well for TseI, Figure 2b, showing a substantial increase in fluorescence from a low-background level as a function of time after addition of the enzyme. The fluorescence assay product duplex melts faster than the cleavage of the substrate by the enzyme with melting complete after ~60 s, Figure 2c. Thus the melting rate of the product does not limit the assay.

The initial rate of fluorescence increase was determined as a function of substrate concentration for both a duplex containing the target sequence and for one containing an A:A or T:T mismatch in the middle of the sequence. All three duplexes were cleaved by the enzyme, and in fact, the

**Figure 2.** Continued  
aliquot of TseI was added to 100 nM duplex (fluorescence assay A:T duplex, Table 1). After closing the sample chamber at 50 s, TseI caused a rapid increase in signal. (c) Time dependence of melting of the fluorescence assay product duplex at 60°C in the absence of any TseI. (d) Michaelis-Menten plots of TseI cleavage for both matched (open circles), A:A-mismatched DNA substrate (black circles) and T:T-mismatched DNA substrate (open square). Error bars are standard deviations for experiments performed in triplicate.

mismatched duplexes were cleaved more quickly, as shown in the Michaelis–Menten analysis, Figure 2d. The maximum velocities for cleavage were  $3.2 \pm 0.2$ ,  $6.0 \pm 0.8$  and  $5.4 \pm 0.3 \text{ nM s}^{-1}$  for the normal duplex, for the A:A mismatched duplex and the T:T mismatched duplex, respectively, giving  $k_{cat}$  values of  $0.20 \pm 0.01$ ,  $0.37 \pm 0.05$  and  $0.33 \pm 0.02 \text{ s}^{-1}$ , respectively, assuming 100% of the enzyme molecules were active as dimers. The Michaelis constants,  $K_M$ , were  $42.2 \pm 9.3$ ,  $80.7 \pm 30.6$  and  $95.2 \pm 14.8 \text{ nM}$  for the normal duplex, the A:A mismatched duplex and the T:T mismatched duplex, respectively. The values of  $k_{cat}/K_M$  were, therefore,  $4.0 \pm 0.95 \times 10^{-3}$ ,  $4.6 \pm 3.4 \times 10^{-3}$  and  $3.5 \pm 0.6 \times 10^{-3} \text{ nM}^{-1} \text{ s}^{-1}$ ; hence, the enzyme has no preference for one substrate over the other, and the weaker dissociation constant observed in the anisotropy assay for the mismatched sequences is mirrored in the poorer  $K_M$  values.

### Mass spectrometry confirms that TseI cleaves mismatched DNA duplexes at the same location as on normal DNA

FT-ICR mass spectrometry was used to identify the nature of the cleaved DNA products. The uncleaved duplexes after denaturation and separation by reverse phase HPLC gave the expected molecular masses, Table 2 and Supplementary Figures S1 and S2. Treatment of the A:T, A:A and T:T duplexes with TseI produced four shorter DNA strands as products. Each of these products corresponded exactly to the mass expected if cleavage occurred at the normal location for TseI, Table 3 and Supplementary Figures S3 and S4.

### TseI cleaves target sequences containing A:A and T:T mismatches but not sequences containing G:C or G:G at the central position of the target sequence

The ability of TseI to cleave target sequences containing a mismatch was unexpected and would indicate a new use for TseI in investigating A:A and T:T mismatches generated by the formation of hairpins in repetitive DNA sequences such  $(\text{CAG})_n$  found in many genetic diseases. For this reason, the action of TseI on further mismatched duplexes was explored.

Rather than perform full enzyme activity studies on each possible mismatch using the fluorescence assay, we performed gel assays, Figure 3, and HPLC assays, Supplementary Figure S5, for cleavage on duplexes containing various base pairs or mismatches in the central position of the target sequence. Figure 3 shows that 28-bp duplexes containing mismatches of A or T were all cleaved into shorter duplexes as effectively as the normal cognate sequence, but that those duplexes containing G or C at the central position were not cleaved as expected, as they lacked the target sequence. Cleavage was complete for the A:T, T:T and A:A duplexes as would be predicted from the kinetic parameters determined in the continuous fluorescence assay.

HPLC analysis of duplex substrates and products using reverse phase chromatography (run at high temperature to denature the DNA) was performed, Supplementary Figure S5. The substrate duplexes did not denature on

**Table 2.** Mass Spectrometry analysis of DNA duplexes

Intact duplex	m/z observed	Molecular mass (Da)	Assignment	Sequence	Elemental formula	Predicted molecular mass (Da)
A:T duplex	1081.435 (8-)	8659.48	Sequence 1	5'-AGG AGT GAA GTC GCA GCC CGT GCT CAA G-3'	$\text{C}_{273}\text{H}_{341}\text{N}_{114}\text{O}_{163}\text{P}_{27}$	8659.48
	1063.053 (8-)	8512.424	Sequence 2	3'-TCC TCA CTT CAG CGT CGG GCA CGA GTT C-5'	$\text{C}_{270}\text{H}_{344}\text{N}_{99}\text{O}_{169}\text{P}_{27}$	8512.43
A:A duplex	1081.442 (8-)	8659.536	Sequence 1	5'-AGG AGT GAA GTC GCA GCC CGT GCT CAA G-3'	$\text{C}_{273}\text{H}_{341}\text{N}_{114}\text{O}_{163}\text{P}_{27}$	8659.48
	1064.183 (8-)	8521.464	Sequence 3	3'-TCC TCA CTT CAG CGA CGG GCA CGA GTT C-5'	$\text{C}_{270}\text{H}_{343}\text{N}_{102}\text{O}_{167}\text{P}_{27}$	8521.44
T:T duplex	1080.31 (8-)	8650.48	Sequence 4	5'-AGG AGT GAA GTC GCT GCC CGT GCT CAA G-3'	$\text{C}_{273}\text{H}_{342}\text{N}_{111}\text{O}_{163}\text{P}_{27}$	8650.47
	1063.063 (8-)	8512.504	Sequence 2	3'-TCC TCA CTT CAG CGT CGG GCA CGA GTT C-5'	$\text{C}_{270}\text{H}_{344}\text{N}_{99}\text{O}_{169}\text{P}_{27}$	8512.43

For each duplex, two species were observed, which displayed molecular masses corresponding to the DNA sequences given. All m/z values and molecular masses quoted are monoisotopic; the charge state for the observed m/z is indicated in the brackets.

**Table 3.** Mass spectrometry analysis of TseI-treated DNA duplexes

Original duplex	m/z observed	Molecular mass (Da)	Assignment	Sequence	Elemental formula	Predicted molecular mass (Da)
A:T duplex	1152.95 (4-)	4615.80	15mer product sequence	5'-Phos-CAG CCC GTG CTC AAG-3'	C <sub>144</sub> H <sub>184</sub> N <sub>57</sub> O <sub>90</sub> P <sub>15</sub>	4615.76
	1210.45 (4-)	4845.80	16mer product sequence	3'-TCC TCA CTT CAG CGT C-Phos-5'	C <sub>153</sub> H <sub>199</sub> N <sub>51</sub> O <sub>100</sub> P <sub>16</sub>	4845.78
	1227.22 (3-)	3684.66	12mer product sequence	3'-GGG CAC GAG TTC-5'	C <sub>117</sub> H <sub>147</sub> N <sub>48</sub> O <sub>70</sub> P <sub>11</sub>	3684.65
A:A duplex	1352.92 (3-)	4061.76	13mer product sequence	5'-AGG AGT GAA GTC G-3'	C <sub>129</sub> H <sub>159</sub> N <sub>57</sub> O <sub>74</sub> P <sub>12</sub>	4061.73
	1152.94 (4-)	4615.76	15mer product sequence	5'-Phos-CAG CCC GTG CTC AAG-3'	C <sub>144</sub> H <sub>184</sub> N <sub>57</sub> O <sub>90</sub> P <sub>15</sub>	4615.76
	1212.71 (4-)	4854.84	16mer product sequence	3'-TCC TCA CTT CAG CGA C-Phos-5'	C <sub>153</sub> H <sub>198</sub> N <sub>54</sub> O <sub>98</sub> P <sub>16</sub>	4854.8
T:T duplex	1227.22 (3-)	3684.66	12mer product sequence	3'-GGG CAC GAG TTC-5'	C <sub>117</sub> H <sub>147</sub> N <sub>48</sub> O <sub>70</sub> P <sub>11</sub>	3684.65
	1352.92 (3-)	4061.76	13mer product sequence	5'-AGG AGT GAA GTC G-3'	C <sub>129</sub> H <sub>159</sub> N <sub>57</sub> O <sub>74</sub> P <sub>12</sub>	4061.73
	1150.69 (4-)	4606.76	15mer product sequence	5'-Phos-CTG CCC GTG CTC AAG-3'	C <sub>144</sub> H <sub>185</sub> N <sub>54</sub> O <sub>92</sub> P <sub>15</sub>	4606.75
	1210.45 (4-)	4845.80	16mer product sequence	3'-TCC TCA CTT CAG CGT C-Phos-5'	C <sub>153</sub> H <sub>199</sub> N <sub>51</sub> O <sub>100</sub> P <sub>16</sub>	4845.79
	1227.22 (3-)	3684.66	12mer product sequence	3'-GGG CAC GAG TTC-5'	C <sub>117</sub> H <sub>147</sub> N <sub>48</sub> O <sub>70</sub> P <sub>11</sub>	3684.65
	1352.92 (3-)	4061.76	13mer product sequence	5'-AGG AGT GAA GTC G-3'	C <sub>129</sub> H <sub>159</sub> N <sub>57</sub> O <sub>74</sub> P <sub>12</sub>	4061.73

For each sample, four species were observed, which displayed molecular masses corresponding to the DNA sequences given. All m/z values and molecular masses quoted are monoisotopic; the charge state for the observed m/z is indicated in the brackets.

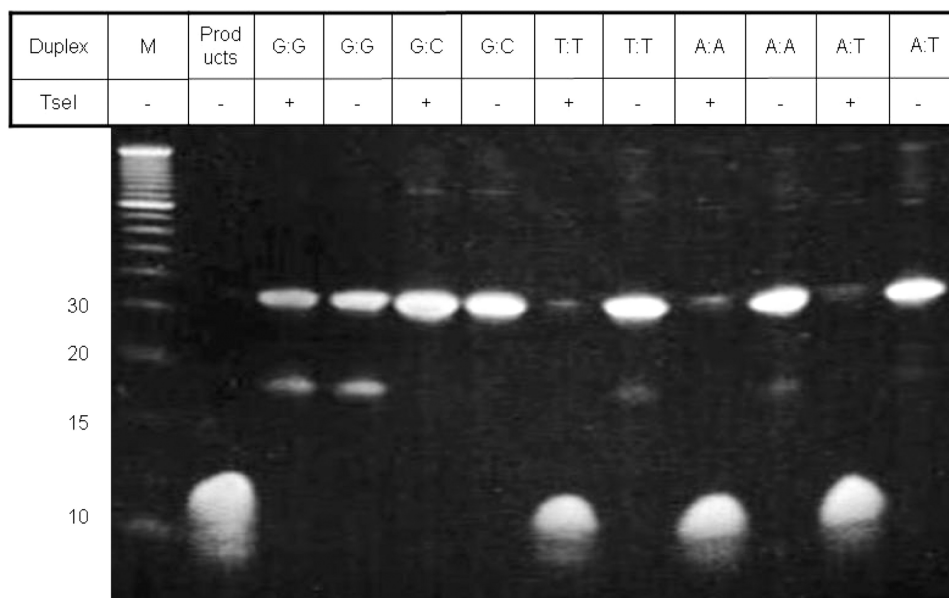
the column and eluted as a single peak for all substrates investigated, namely, A:T, A:A, T:T, G:C and G:G. The shorter product strands eluted as poorly resolved peaks at an elution time substantially lower than that of the uncleaved DNA strands; consequently, cleaved DNA could be clearly distinguished from uncleaved DNA. The 28-base single strands eluted at the same time as the full-length duplexes or slightly later. The absorption elution profiles, Supplementary Figure S5, clearly demonstrated that duplexes containing the normal target sequence or A:A or T:T mismatches were cleaved. The cleavage seemed to be incomplete for these three duplexes in contrast to the assay shown in Figure 3, where the A:T, A:A and T:T duplexes were completely cleaved. This may be due to the enzyme remaining bound to a fraction of the cleavage products and slowing their elution. Duplexes lacking the target site or containing G:G or C:C mismatches present at the central base pair of the target sequence were not cleaved. Duplexes containing 2-AP instead of A were also cleaved when paired with T or mismatched with A or 2-AP (2-AP:T duplex 1, A:2-AP duplex 2, 2-AP:A duplex 4 and 2-AP:2-AP duplex 5 from Table 1) (data not shown).

## DISCUSSION

Data presented here show that TseI cleaves not only its cognate sequence GCWGC but also sequences in which the central base pair is an A:A or T:T mismatch. In the apparent absence of base flipping of the central base pair, as observed for other restriction enzymes recognizing similar sequences (8,9,11–13), the mode of recognition of the central base pair must be rather unusual for discrimination of a base pair containing A and T from one containing G and C. Normally, discrimination of A:T/T:A from G:C/C:G would use minor groove interactions and not rely on DNA distortions, such as base flipping. However, an A:A or G:G mismatch will distort the DNA helix at the centre of the TseI target sequence to approximately similar degrees; therefore, it is difficult to rationalize how this enzyme can recognize the central base pair in such a way as to disfavour G and C. Perhaps the duplex undergoes a degree of bending at the central base pair to provide the necessary discrimination. However, solving this problem will require X-ray crystallography and would reveal whether TseI and the related ApeKI had a third type of fold distinct from the other restriction enzymes recognizing similarly degenerate targets (3–13).

The cleavage properties of TseI allow it to digest not only CAG repeat sequences in normal duplex DNA but also the mismatched sequences caused by hairpin extrusion in long tracts of such repetitive DNA. In fact, we observed this dual action of TseI recently using AFM imaging of the degradation of the highly convoluted DNA structures formed by annealing DNA strands containing long CAG repeats (20). The motivation for this previous study was to understand the change in DNA structure as CAG repeat length increased, a change that might be relevant to the aetiology of HD. In HD, the presence of an expanded CAG repeat in the coding





**Figure 3.** Polyacrylamide gel analysis of matched and mismatched DNA duplexes being cut by TseI. Lane M was the molecular mass marker with 10, 15, 20 and 30 bp DNA indicated. DNA cleavage of A:T, T:T and A:A duplexes by TseI yields four 12–16mer single-strand oligonucleotides matching those shown in the products lane.

region of *HTT*, the gene encoding Huntingtin, produces an expanded polyglutamine sequence in the expressed protein, thereby rendering it neurotoxic (14). The threshold CAG repeat length that causes HD is ~36, and there is an approximate correlation between the length of the CAG tract and the age of onset of the disease (25–27). Confounding this relationship is instability of repeat length, which occurs both in humans (28–32) and in mouse models of HD (33–37). In fact, super-long CAG expansions up to 1000 have been found in neurons from the brains of patients with adult-onset HD (31,35). Interestingly, it was shown recently that mice carrying the first exon of the human HD gene (R6/2 strain) with a super-long (~500) CAG repeat were found to have a delayed onset of the disease phenotype and prolonged survival compared with mice with shorter repeats (38,39). There is a pronounced (>60%) reduction in both mRNA levels and expression in mice with more than ~335 CAG repeats as compared with mice with 150 repeats (39), which might contribute to the observed phenotypic amelioration. We speculated that super-long DNA adopts unusual structures that progressively reduce the efficiency of transcription of the *HTT* gene, and we went on to observe the presence of these structures by AFM imaging (20). Significantly, TseI preferentially cleaved normal duplex DNA and digested the mismatched DNA only at higher temperatures (80°C). In the present work, although TseI showed no preference for digestion of base paired over mismatched substrates at 60°C, it did cleave mismatched substrates faster; hence, the temperature dependence of TseI action is different on the two substrates. By judicious choice of several incubation temperatures, therefore, it should be possible to use the enzyme to discriminate between the two forms of DNA. For this reason, we suggest that TseI should be a useful

tool, complementary to the use of the EcoP15I restriction enzyme (40), for exploring the role of DNA structure in triplet repeat expansion diseases, such as HD and myotonic dystrophy type 1.

#### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–5.

#### ACKNOWLEDGEMENTS

The authors thank Dr John H. White (School of Chemistry Molecular Biology Service, Edinburgh) for his assistance with oligonucleotide synthesis and Dr Pat Langridge-Smith for access to the SIRCAMS mass spectrometry service (School of Chemistry, Edinburgh). They also thank David Hough of New England Biolabs for the kind gift of purified protein.

#### FUNDING

Chinese Scholarship Council and the MTEM International Studentship Scheme (to L.M.); Engineering and Physical Sciences Research Council (EPSRC), Ph.D. studentship (to K.C.); Doctoral Training Centre in Cell and Proteomic Technologies [Biotechnology and Biological Sciences Research Council (BBSRC) and EPSRC (to C.P.N.)]; Wellcome Trust [GR080463MA and 090288/Z/09/ZA to D.T.F.D.]; CHDI Foundation, Inc (to A.J.M.). Funding for open access charge: Wellcome Trust [GR080463MA and 090288/Z/09/ZA to D.T.F.D.].

*Conflict of interest statement.* None declared.

## REFERENCES

- Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2010) REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.*, **38**, D234–D236.
- Kawarabayasi,Y., Hino,Y., Horikawa,H., Yamazaki,S., Haikawa,Y., Jin-no,K., Takahashi,M., Sekine,M., Baba,S., Ankai,A. *et al.* (1999) Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1. *DNA Res.*, **6**, 83–101.
- Kaus-Drobek,M., Czapinska,H., Sokolowska,M., Tamulaitis,G., Szczepanowski,R.H., Urbanke,C., Siksnys,V. and Bochtler,M. (2007) Restriction endonuclease *MvaI* is a monomer that recognizes its target sequence asymmetrically. *Nucleic Acids Res.*, **35**, 2035–2046.
- Sokolowska,M., Kaus-Drobek,M., Czapinska,H., Tamulaitis,G., Szczepanowski,R.H., Urbanke,C., Siksnys,V. and Bochtler,M. (2007) Monomeric restriction endonuclease *BenI* in the Apo form and in an asymmetric complex with target DNA. *J. Mol. Biol.*, **369**, 722–734.
- Sokolowska,M., Kaus-Drobek,M., Czapinska,H., Tamulaitis,G., Siksnys,V. and Bochtler,M. (2007) Restriction endonucleases that resemble a component of the bacterial DNA repair machinery. *Cell. Mol. Life Sci.*, **64**, 2351–2357.
- Kostiuk,G., Sasnauskas,G., Tamulaitiene,G. and Siksnys,V. (2011) Degenerate sequence recognition by the monomeric restriction enzyme: single mutation converts *BcnI* into a strand-specific nicking endonuclease. *Nucleic Acids Res.*, **39**, 3744–3753.
- Sasnauskas,G., Kostiuk,G., Tamulaitis,G. and Siksnys,V. (2011) Target site cleavage by the monomeric restriction enzyme *BcnI* requires translocation to a random DNA sequence and a switch in enzyme orientation. *Nucleic Acids Res.*, **39**, 8844–8856.
- Bochtler,M., Szczepanowski,R.H., Tamulaitis,G., Grazulis,S., Czapinska,H., Manakova,E. and Siksnys,V. (2006) Nucleotide flips determine the specificity of the *Ecl18kI* restriction endonuclease. *EMBO J.*, **25**, 2219–2229.
- Szczepanowski,R.H., Carpenter,M.A., Czapinska,H., Zaremba,M., Tamulaitis,G., Siksnys,V., Bhagwat,A.S. and Bochtler,M. (2008) Central base pair flipping and discrimination by *PspGI*. *Nucleic Acids Res.*, **36**, 6109–6117.
- Zaremba,M., Owsicka,A., Tamulaitis,G., Sasnauskas,G., Shlyakhtenko,L.S., Lushnikov,A.Y., Lyubchenko,Y.L., Laurens,N., van den Broek,B., Wuite,G.J. *et al.* (2010) DNA synapsis through transient tetramerization triggers cleavage by *Ecl18kI* restriction enzyme. *Nucleic Acids Res.*, **38**, 7142–7154.
- Tamulaitis,G., Zaremba,M., Szczepanowski,R.H., Bochtler,M. and Siksnys,V. (2008) How *PspGI*, catalytic domain of *EcoRII* and *Ecl18kI* acquire specificities for different DNA targets. *Nucleic Acids Res.*, **36**, 6101–6108.
- Tamulaitis,G., Zaremba,M., Szczepanowski,R.H., Bochtler,M. and Siksnys,V. (2007) Nucleotide flipping by restriction enzymes analyzed by 2-aminopurine steady-state fluorescence. *Nucleic Acids Res.*, **35**, 4792–4799.
- Neely,R.K., Tamulaitis,G., Chen,K., Kubala,M., Siksnys,V. and Jones,A.C. (2009) Time-resolved fluorescence studies of nucleotide flipping by restriction enzymes. *Nucleic Acids Res.*, **37**, 6859–6870.
- Orr,H.T. and Zoghbi,H.Y. (2007) Trinucleotide repeat disorders. *Annu. Rev. Neurosci.*, **30**, 575–621.
- López Castel,A., Cleary,J.D. and Pearson,C.E. (2010) Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat. Rev. Mol. Cell Biol.*, **11**, 165–170.
- Sinden,R., Potaman,V.N., Oussatcheva,E.A., Pearson,C.E., Lyubchenko,Y.L. and Shlyakhtenko,L.S. (2002) Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA. *J. Biosci.*, **27**, 53–65.
- Pearson,C.E. and Sinden,R.R. (1996) Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry*, **35**, 5041–5053.
- Pearson,C.E., Wang,Y.-H., Griffith,J.D. and Sinden,R.R. (1998) Structural analysis of slipped-strand DNA (S-DNA) formed in (CTG)<sub>n</sub>•(CAG)<sub>n</sub> repeats from the myotonic dystrophy locus. *Nucleic Acids Res.*, **26**, 816–823.
- Pearson,C.E., Tam,M., Wang,Y.-H., Montgomery,S.E., Dar,A.C., Cleary,J.D. and Nichol,K. (2002) Slipped-strand DNAs formed by long (CAG)<sub>n</sub>•(CAG)<sub>n</sub> repeats: slipped-out repeats and slip-out junctions. *Nucleic Acids Res.*, **30**, 4534–4547.
- Duzdevich,D., Li,J., Whang,J., Takahashi,H., Takeyasu,K., Dryden,D.T.F., Morton,A.J. and Edwardson,J.M. (2011) Unusual structures are present in DNA fragments containing super-long Huntingtin CAG repeats. *PLoS ONE*, **6**, e17119.
- Gasteiger,E., Hoogland,C., Gattiker,A., Duvaud,S., Wilkins,M.R., Appel,R.D. and Bairoch,A. (2005) Protein identification and analysis tools on the ExpASY server. In: Walker,J.M. (ed.), *The Proteomics Protocols Handbook*. Humana Press, New York, pp. 571–607.
- Su,T.-J., Tock,M.R., Egelhaaf,S.U., Poon,W.C.K. and Dryden,D.T.F. (2005) DNA bending by M.EcoKI methyltransferase is coupled to nucleotide flipping. *Nucleic Acids Res.*, **33**, 3235–3244.
- Waters,T.R. and Connolly,B.A. (1992) Continuous spectrophotometric assay for restriction endonucleases using synthetic oligodeoxynucleotides and based on the hyperchromic effect. *Anal. Biochem.*, **204**, 204–209.
- Li,J.J., Geyer,R. and Tan,W. (2000) Using molecular beacons as a sensitive fluorescence assay for enzymatic cleavage of single-stranded DNA. *Nucleic Acids Res.*, **28**, E52.
- Wexler,N.S., Lorimer,J., Porter,J., Gomez,F., Moskowitz,C., Shackell,E., Marder,K., Penchaszadeh,G., Roberts,S.A., Gayán,J. *et al.* (2004) Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. *Proc. Natl Acad. Sci. USA*, **101**, 3498–3503.
- Li,J.L., Hayden,M.R., Warby,S.C., Durr,A., Morrison,P.J., Nance,M., Ross,C.A., Margolis,R.L., Rosenblatt,A., Squitieri,F. *et al.* (2006) Genome-wide significance for a modifier of age at neurological onset in Huntington's disease at 6q23-24: the HD MAPS study. *BMC Med. Genet.*, **7**, 71.
- Andresen,J.M., Gayán,J., Djuosse,L., Roberts,S., Brocklebank,D., Cherny,S.S., Cardon,L.R., Gusella,J.F., MacDonald,M.E., Myers,R.H. *et al.* (2007) The relationship between CAG repeat length and age of onset differs for Huntington's disease patients with juvenile onset or adult onset. *Ann. Hum. Genet.*, **71**, 295–301.
- Telenius,H., Kremer,B., Goldberg,Y.P., Theilmann,J., Andrew,S.E., Zeisler,J., Adam,S., Greenberg,C., Ives,E.J., Clarke,L.A. *et al.* (1994) Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. *Nat. Genet.*, **6**, 409–414.
- Aronin,N., Chase,K., Young,C., Sapp,E., Schwarz,C., Matta,N., Kornreich,R., Landwehrmeyer,B., Bird,E., Beal,M.F. *et al.* (1995) CAG expansion affects the expression of mutant Huntingtin in Huntington's disease brain. *Neuron*, **15**, 1193–1201.
- Kono,Y., Agawa,Y., Watanabe,Y., Ohama,E., Nanba,E. and Nakashima,K. (1999) Analysis of the CAG repeat number in a patient with Huntington's disease. *Intern. Med.*, **38**, 407–411.
- Kennedy,L., Evans,E., Chen,C.M., Craven,L., Detloff,P.J., Ennis,M. and Shelbourne,P.F. (2003) Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum. Mol. Genet.*, **12**, 3359–3367.
- Shelbourne,P.F., Keller-McGandy,C., Bi,W.L., Yoon,S.R., Dubeau,L., Veitch,N.J., Vonsattel,J.P., Wexler,N.S., Arnhem,N., Augood,S.J. *et al.* (2007) Triplet repeat mutation length gains correlate with cell-type specific vulnerability in Huntington disease brain. *Hum. Mol. Genet.*, **16**, 1133–1142.
- Mangiarini,L., Sathasivam,K., Mahal,A., Mott,R., Seller,M. and Bates,G.P. (1997) Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. *Nat. Genet.*, **15**, 197–200.
- Wheeler,V.C., Auerbach,W., White,J.K., Srinidhi,J., Auerbach,A., Ryan,A., Duyao,M.P., Vrbancac,V., Weaver,M., Gusella,J.F. *et al.* (1999) Length-dependent gametic CAG repeat instability in the Huntington's disease knock-in mouse. *Hum. Mol. Genet.*, **8**, 115–122.
- Kennedy,L. and Shelbourne,P.F. (2000) Dramatic mutation instability in HD mouse striatum: does polyglutamine load contribute to cell-specific vulnerability in Huntington's disease? *Hum. Mol. Genet.*, **9**, 2539–2544.

36. Wheeler, V.C., Lebel, L.A., Vrbanac, V., Teed, A., te Riele, H. and MacDonald, M.E. (2003) Mismatch repair gene *Msh2* modifies the timing of early disease in Hdh(Q111) striatum. *Hum. Mol. Genet.*, **12**, 272–281.
37. Gonitel, R., Moffitt, H., Sathasivam, K., Woodman, B., Detloff, P.J., Faull, R.L. and Bates, G.P. (2008) DNA instability in postmitotic neurons. *Proc. Natl Acad. Sci. USA*, **105**, 3467–3472.
38. Morton, A.J., Glynn, D., Leavens, W., Zheng, Z., Faull, R.L., Skepper, J.N. and Wight, J.M. (2009) Paradoxical delay in the onset of disease caused by super-long CAG repeat expansions in R6/2 mice. *Neurobiol. Dis.*, **33**, 331–341.
39. Dragatsis, I., Goldowitz, D., Del Mar, N., Deng, Y.P., Meade, C.A., Liu, L., Sun, Z., Dietrich, P., Yue, J. and Reiner, A. (2009) CAG repeat lengths  $\geq 335$  attenuate the phenotype in the R6/2 Huntington's disease transgenic mouse. *Neurobiol. Dis.*, **33**, 315–330.
40. Moencke-Buchner, E., Reich, S., Muecke, M., Reuter, M., Messer, W., Wanker, E.E. and Krueger, D.H. (2002) Counting CAG repeats in the Huntington's disease gene by restriction endonuclease EcoP15I cleavage. *Nucleic Acids Res.*, **30**, e83/1–e83/7.