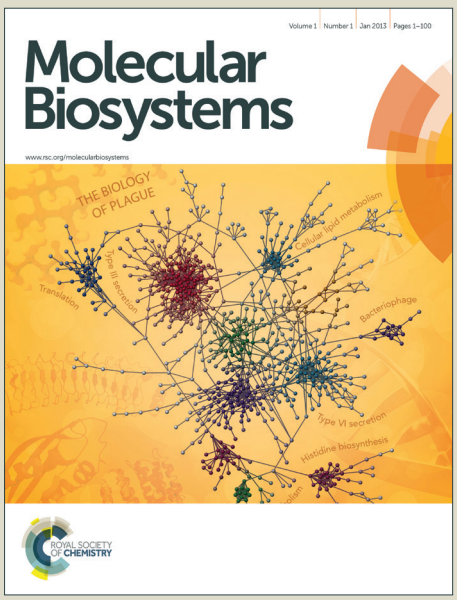
Molecular BioSystems

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Molecular Biosystems

PAPER

The effect of sequence context on the activity of cytosine DNA glycosylases

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Scott T. Kimber[a] Tom Brown[b] and Keith R Fox[a]

We have prepared single (N204D) and double (N204D:L272A) mutants of human uracil DNA glycosylase (hUDG), generating two cytosine DNA glycosylases (hCDG and hCYDG). Both these enzymes are able to excise cytosine (but not methylcytosine), when this base is part of a mismatched base pair. hCDG is more active than the equivalent *E. coli* enzyme (eCYDG) and also has some activity when the cytosine is paired with guanine, unlike eCYDG. hCDG also has some activity against single stranded DNA, while having poor activity towards an unnatural base pair that forces the cytosine into an extrahelical conformation (in contrast to eCYDG for which a bulky base enhances the enzyme's activity). We also examined how sequence context affects the activity of these enzymes, determining the effect of flanking base pairs on cleavage efficiency. An abasic site or a hexaethylene glycol linker placed opposite the target cytosine, also causes an increase in activity compared with an AC mismatch. Flanking an AC mismatch with GC base pairs resulted in a 100-fold decrease in excision activity relative to flanking AT base pairs and the 5′-flanking base pair had a greater effect on the rate of cleavage. However, this effect is not simply due to the stability of the flanking base pairs as adjacent GT mismatches also produce low cleavage efficiency.

Introduction

Uracil DNA glycosylase (UDG) removes uracil from DNA. This base arises from deamination of cytosine, generating G.U mispairs[1] which, if not repaired, would result in a cytosine to thymine transition mutations. UDG is also able to excise uracil from A.U base pairs that results from misincorporation during DNA replication,[2] though this is not mutagenic. The enzyme is highly specific for uracil and shows no activity towards any other base. Thymine is excluded from the enzyme's active site as a result of steric clash between its C5 methyl group and a tyrosine.[3] UDG's specificity for uracil comes from shape complementarity, which excludes purines and thymine,[3,4] and specific hydrogen bonding that recognises U and excludes C. A critical asparagine (N204 in the human enzyme, N123 in *E. coli*) forms specific hydrogen bonds from its amine and carbonyl oxygen to the O4 and N3 of uracil respectively (Fig. 1).[4-6] Mutation of this asparagine to aspartate changes the hydrogen bond donor-acceptor pattern and allows for recognition of cytosine (*3*), thereby generating a cytosine DNA glycosylase (CDG).[3,7] Previous studies have shown that it is not possible to express the *E. coli* N123D mutant in *E. coli*, presumably because eCDG would be toxic,

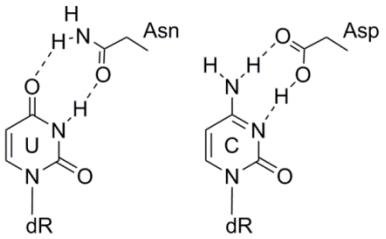


Fig. 1. Interaction of asparagine (N123 in UDG or N204 in hUDG) with uracil and the proposed recognition of cytosine by aspartate.

degrading the host genome. However the enzyme's activity can be attenuated by mutating L191[8, 9] since this amino acid is responsible for flipping the base into the enzyme's active site.[9-11] This allows expression of the double mutant (N123D;L191A; designated eCYDG).[7] Although this enzyme has greatly reduced activity compared with eUDG it is able to excise cytosine when it is in a mismatched base pair (such as AC), but shows no activity when paired with guanine.[12] eCYDG does not cleave 5-methylcytosine in any base pair context, as there is a steric clash with the 5-methyl group, in the same way that thymine is excluded from UDG.

The human CDG (hCDG) variant (N204D) can be expressed in *E. coli*[3] and we have determined its activity against various cytosine-containing substrates, as well as that of the (N204D:L272A) double mutant. We have previously suggested that cytosine DNA glycosylase could be used in a technique for the detecting 5-methylcytosine.[12] We have therefore also investigated the effect of sequence context on the activity of these CDGs. We find that the

[a.] *Centre for Biological Sciences, Life Science Building 85, University of Southampton, Highfield, Southampton SO17 1BJ, UK*
[b.] *Department of Chemistry, University of Oxford, Chemistry Research Laboratory, 12 Mansfield Road, Oxford, OX1 3TA, UK*
† Footnotes relating to the title and/or authors should appear here.
Electronic Supplementary Information (ESI) available: [Sequences of oligonucleotides that were used to construct hUDG]. See DOI: 10.1039/x0xx00000x

Molecular BioSystems Accepted Manuscript

Please do not adjust margins

Substrate	Sequence
X.Y	5′-CCGAATCAGTGCGCAXAGTCGGTATTTAGCC-3′ 3′-GGCTTAGTCACGCGTYTCAGCCATAAATCGG-5′
A.C(G)	5′-CCGAATCAGTGCGCGCGGTCGGTATTTAGCC-3′ 3′-GGCTTAGTCACGCGCACCAGCCATAAATCGG-5′
A.C(AG)	5′-CCGAATCAGTGCGCACGGTCGGTATTTAGCC-3′ 3′-GGCTTAGTCACGCGTACCAGCCATAAATCGG-5′
A.C(GA)	5′-CCGAATCAGTGCGCGCAGTCGGTATTTAGCC-3′ 3′-GGCTTAGTCACGCGCATCAGCCATAAATCGG-5′
A.C(GT)	5′-CCGAATCAGTGCGCGCGGTCGGTATTTAGCC-3′ 3′-GGCTTAGTCACGCGTATCAGCCATAAATCGG-5′
HEG.C(G)	5′-CCGAATCAGTGCGCGCGGTCGGTATTTAGCC-3′ 3′-GGCTTAGTCACGCGCHCCAGCCATAAATCGG-5′
HEG.C(AP)	5′-CCGAATCAGTGCGCGCGGTCGGTATTTAGCC-3′ 3′-GGCTTAGTCACGCGΦHΦCAGCCATAAATCGG-5′
HEG.C(-G-)	5′-CCGAATCAGTGCGCGCGGTCGGTATTTAGCC-3′ 3′-GGCTTAGTCACGCG-H-CAGCCATAAATCGG-5′
I.C(G)	5′-CCGAATCAGTGCGCGCGGTCGGTATTTAGCC-3′ 3′-GGCTTAGTCACGCGCICCAGCCATAAATCGG-5′
G.C(AT)	5′-CGAATAATTATATAACATATATATATTTAGC-3′ 3′-GCTTATTAATATATTGTATATATATAAATCG-3′
gap.C	5′-CCGAATCAGTGCGCACAGTCGGTATTTAGCC-3′ 3′-GGCTTAGTCACGCGT TCAGCCATAAATCGG-5′
Long gap.C	5′-CCGTACTGAATCAGTGCGCACAGTCGGTATTTACGATAGCC-3′ 3′-GGCATGACTTAGTCACGCGT TCAGCCATAAATGCTATCGG-5′
ssC(polyA)	5′-AAAAAAAAAAAAAAAACAAAAAAAAAAAAAAA-3′
ssC(GAT)	5′-GGATAAATAGGGAGTCTGAGAAGTGATTAGG-3′

Table 1. DNA oligonucleotide substrates used in this study to characterise the cleavage rates of CDGs. The target bases are shown in bold and underlined. X = U, C, ^{Me}C. Y = G, A, AP (abasic site, Φ), Z (anthraquinone pyrollidine), HEG (hexaethylene glycol, H), I (inosine).

ability of these enzymes to excise cytosine is affected by the surrounding sequences.

Results

Excision Properties of hCDG and hCYDG

We expected that hCDG would be more active than eCYDG, as it only contains a single mutation relative to the wild type enzyme and retains the leucine (L272, equivalent to L191 in the *E.coli* enzyme) that is required to facilitate base flipping. We therefore investigated the activity of hCDG against the same substrates (shown in Table 1) that were previously used with eCYDG.[12] Representative results of the cleavage assays are shown in Figs. 2 and 3, and k_{cat} values determined from these are summarised in Table 2. It can be seen that hCDG is about 30-fold more active than eCYDG at cleaving cytosine from an A.C mismatch (Table 2). hCDG still excises uracil, as previously seen with eCYDG,[12] and it is 10- and 60-fold more active than eCYDG with G.U and A.U respectively. Although hCDG showed no activity against the substrate with a central GC base pair, we also examined its effect on a GC pair that is embedded within a block of AT residues (sequence G.C(AT)). hCDG was able to excise cytosine from this GC base pair, albeit at less than 1% of the rate produced at AC, in contrast to eCYDG, which showed no activity against this substrate. This enzyme also showed some activity against an AC mismatch that

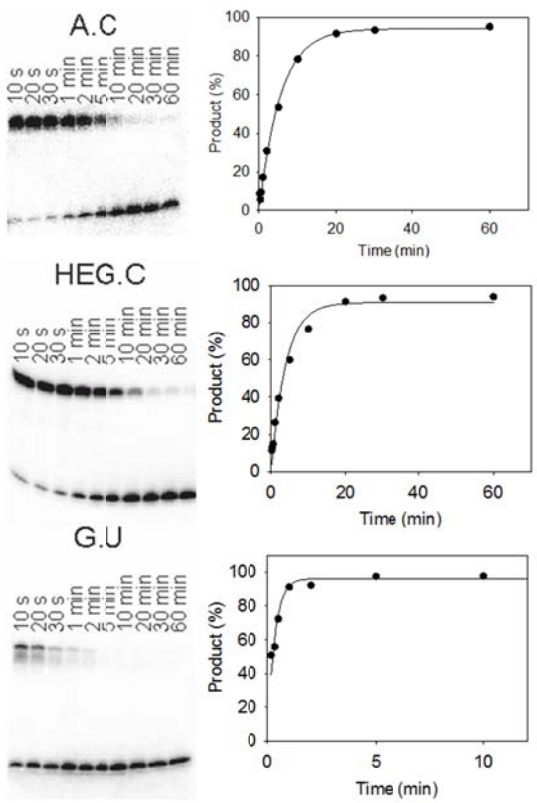


Fig. 2. hCDG cleavage of fragments A.C, HEG.C and G.U. In each gel the 32P labelled duplex substrates (~50 nM) were incubated with ~1.25 µM hCDG for up to 24 hours and cleaved by heating at 95°C in 10% piperidine for 20 mins. The products were resolved on 12.5% denaturing polyacrylamide gels. The graphs are derived from phosphorimage analysis of the gels and show the rate of formation of the cleavage product. These are fitted with single exponential curves.

Substrate	Rate of excision (min^{-1}) hCDG	hCYDG	*e*CYDG
A.C	0.17 ± 0.02	0.013 ± 0.001	0.006 ± 0.001
AP.C	0.04 ± 0.004	0.0001[1]	0.014 ± 0.003
Z.C	0.005 ± 0.001[4]	0.02 ± 0.005	0.10 ± 0.02
HEG.C	0.26 ± 0.02	0.013 ± 0.002	0.13 ± 0.001
G.C	ND	ND	ND
G.C(AT)	0.0006	ND	ND
gap.C	0.02 ± 0.003	NI	0.016 ± 0.002
long gap.C	0.03 ± 0.007	NI	0.0072 ± 0.0007
G.U	3.8 ± 0.7	0.51 ± 0.01	0.36 ± 0.04
A.U	1.27 ± 0.08	NI	0.020 ± 0.004
ssC(polyA)	0.03 ± 0.005	NI	0.0003 ± 0.0001[5]
ssC(GAT)	0.02 ± 0.002	ND	0.0001[3]

Table 2. Excision rates for hCDG and hCYDG cleavage of the different DNA substrates compared to *e*CYDG (*12*). The sequences of the oligonucleotides are shown in Table 1. No cleavage was observed for any substrate containing methylcytosine. ND - no cleavage detected after 24 hours; NI – not investigated. Values represent the average of three independent determinations. k_{cat} rate values was estimated from a single time point at 24 hrs[1] 8 hrs[2], 4 hrs[3], 2 hrs[4] and 1 hr[5] assuming a simple exponential. The data for eCYDG are taken from.[12]

Molecular BioSystems Accepted Manuscript

This journal is © The Royal Society of Chemistry 20xx

is flanked by GC base pairs (A.C(G)).

hCDG also showed more than 100-fold greater activity than eCYDG against the single-stranded substrates that contained cytosine in the centre of an oligodA tract (ssC(polyA)) or in a mixed sequence of GAT residues (ssC(GAT)). This greater cleavage efficiency most likely arises because the presence of L272 allows the enzyme to retain the cytosine within its active site, acting as a plug allowing excision to occur.[10,11]

However, hCDG did not display greater activity than eCYDG against all the substrates; Z.C (where Z is anthraquinone pyrollidine; a bulky synthetic nucleoside analogue), was cleaved at a 20-fold lower rate than with eCYDG. We also examined the effect of leaving the target cytosine unpaired, placing this opposite the non-nucleosidic linker hexaethylene glycol. This combination produced one of the fastest rates of cleavage for both eCYDG and hCDG. Other arrangements with unpaired cytosine (AP.C, gap.C and long gap.C) were also cleaved, though at slower rates and were only two- to four-fold better substrates for hCDG than eCYDG. As with the single stranded substrates, this is likely to be due to the ability of hCDG to correctly intercalate L272 into the DNA duplex.

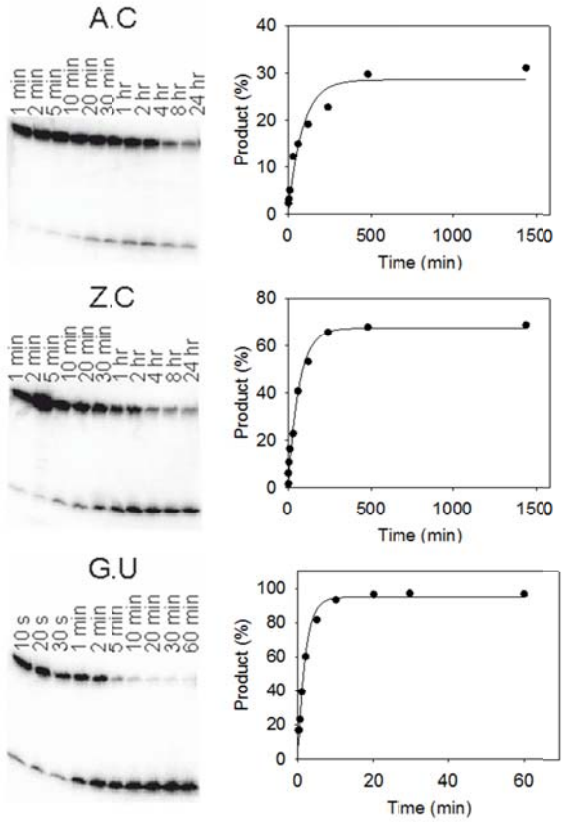


Fig. 3. hCYDG cleavage of fragments A.C, Z.C and G.U. In each gel the 32P labelled duplex substrates (~50 nM) were incubated with ~1.25 µM hCYDG for up to 24 hours and cleaved by heating at 95°C in 10% piperidine for 20 mins. The products were resolved on 12.5% denaturing polyacrylamide gels. The graphs are derived from phosphorimage analysis of the gels and show the rate of formation of the cleavage product. These are fitted with single exponential curves.

The results of similar experiments with the double mutant hCYDG are shown in Fig. 4 and Table 2. Since this is a double mutant, it is not surprising that it is approximately 10-fold less active than hCDG. Although hCYDG's activity at GU was comparable to that of eCYDG, its activity at other mismatches was more variable. It showed a 2-fold greater activity at A.C than eCYDG, but was 5-fold less active with Z.C, though it is more active than hCDG against this bulky substrate. hCYDG is also about 10-fold less active than eCYDG against AP.C and about 100-fold less active at HEG.C. It is worth noting, that as with eCYDG [7] and hCDG, hCYDG did not cleave 5-methylcytosine in any base pair combination.

Effects of sequence context on eCYDG activity

Although hCDG was generally more active than eCYDG against cytosine-containing mismatches, we considered that it would not be suitable for use in any detection assay for 5-methylcytosine as the DNA target would also be degraded by its (weak) activity at GC base pairs and its ability to degrade single-stranded DNA. We therefore decided to explore further the effect of sequence context on the rate of cytosine excision by eCYDG.

Previous results ([12] and Table 2) have suggested that base pair stability affects the rate of base excision. An example of this is the rate of cleavage of G.U (a wobble base pair), which is faster than that of the more stable A.U base pair. This is also highlighted by the observation that the various CDGs cleave the mismatched, unpaired or unnatural pairs A.C, AP.C, Z.C and HEG.C, but show no activity against G.C in the same sequence context. To investigate this further we examined the effect of pairing the target cytosine with inosine, as the I.C base pair has only two hydrogen bonds compared to three for G.C. However, no cytosine excision could be detected with eCYDG, though hCDG produced a very low rate of cleavage (Table 3).

We therefore focussed on how the flanking base pairs affect cleavage efficiency by eCYDG, looking first at the A.C combination. Changing the base pairs flanking the A.C mismatch from AT to GC (sequence A.C(G)) caused a 60-fold decrease in the rate of cleavage by eCYDG and over 100-fold decrease for hCDG. A similar though less pronounced effect was seen on changing the bases flanking an AU pair to GC (sequence A.U(G)), for which the rate was decreased 5-fold compared to flanking AT base pairs. Changing the base pair on the 5'-side of the AC mismatch to GC (sequence A.C(GA)) had a greater effect than changing the 3'-base pair (sequence A.C(AG)).

Since the flanking base pairs appear to affect the enzyme's activity, with more stable base pairs reducing the cleavage rate, we examined the effect of placing GT mismatches on either side of the AC pair (sequence A.C(GT)). This produced a small increase in rate compared to flanking GC base pairs for both eCYDG and hCDG, but this was still much less than seen with flanking AT pairs.

We performed similar experiments with the HEG.C (Fig. 4) combination, since this produced the fastest cleavage rate when flanked by AT pairs. In this case changing the flanking

Molecular BioSystems Accepted Manuscript

Please do not adjust margins

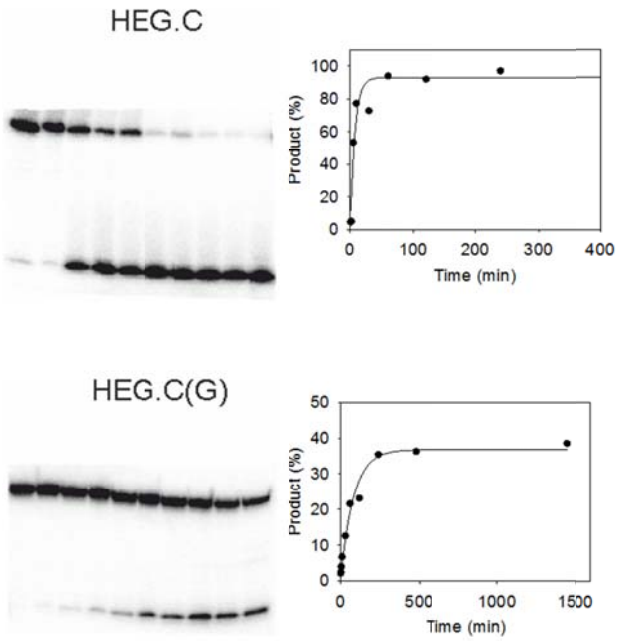


Fig 4. eCYDG cleavage of fragments HEG.C and HEG.C(G). In each gel the 32P labelled duplex substrates (~50 nM) were incubated with ~1.25 µM hCDG for up to 24 hours and cleaved by heating at 95°C in 10% piperidine for 20 mins. The products were resolved on 12.5% denaturing polyacrylamide gels. The graphs are derived from phosphorimage analysis of the gels and show the rate of formation of the cleavage product. These are fitted with single exponential curves.

Substrate	eCYDG	hCDG
I.C	ND	0.0004
A.C*	0.006 ± 0.001	0.17 ± 0.02
A.C(G)*	0.0001[1]	0.0012
A.C(AG)	0.005 ± 0.001	
A.C(GA)	0.0003[1]	
A.U*	0.020 ± 0.004	
A.U(G)	0.004	
A.C(GT)	0.0003[2]	0.006 ± 0.001
HEG.C*	0.13 ± 0.001	
HEG.C(G)	0.01 ± 0.001[2]	
HEG.C(-G-)	ND	
HEG.C(AP)	ND	

Table 3. Rates of excision for CDG cleavage of different substrates assessing sequence context. The substrates are based on the X.Y oligonucleotides shown in Table 1; base(s) in brackets designate the context of the flanking base pair. No cleavage was observed for any substrate containing methylcytosine. ND - no cleavage detected after 24 hours. Values represent the average of three independent determinations. k_{cat} rate values were estimated from a single time point at 24 hrs[1] and 5 mins[2] assuming a simple exponential. *Taken from Table 2.

base pairs from AT to GC caused a 13-fold decrease in the rate of cleavage. Surprisingly, replacing the flanking GC pairs with G.AP (thereby removing any base pairing) greatly reduced the rate of cleavage. Similarly, since the hexaethylene glycol linker has the same number of bonds as three nucleotides we omitted the nucleotides on either side, bridging the ACA target with a single HEG (sequence HEG.C(-G-) (Fig. 4)). This too abolished the cleavage. The cleavage efficiency is therefore not solely dictated by the stability of the flanking base pairs.

Discussion

These data show that hCDG has 30-fold greater CDG activity than eCYDG against the AC substrate, though this increased activity was not seen with all cytosine-containing substrates as it was 20-fold less active with Z.C. This is most likely due to clashes with the leucine that is responsible for base flipping, which is retained in hCDG but missing in eCYDG. We assume that the anthraquinone pyrollidine (Z) fills the space that is normally occupied by the base pair (as proposed for pyrene; (*7*)) and so forces the opposing cytosine into an extrahelical conformation to aid excision. Z.C may be a poor substrate for hCDG as L272 is unable to intercalate into the space left by the extrahelical cytosine as this will be occupied by the anthraquinone pyrollidine nucleoside. This is consistent with the observation that hCYDG (containing the L272A mutation) cleaves at Z.C faster than hCDG. It also explains hCDG's ability to excise cytosine from single-stranded DNA substrates, as L272 will aid in keeping cytosine in the active site for a longer period of time, allowing it to adopt the correct conformation for cleavage of the N-glycosidic bond. With eCYDG, the base will more easily enter and exit the active site, increased by the flexibility of ssDNA, reducing the enzyme's ability to form the correct enzyme-substrate complex for base excision. These data therefore suggest that the major role of the leucine is to act as a "plug" to retain the base in the active site, rather than actively "pushing" out the base.[9-11] These results suggest that eCDG would be the most active cytosine DNA glycosylase against AC (and probably HEG.C), but could be inferior to eCYDG at a substrate containing a bulky base analogue such as Z.C.

These data show that these CDG enzymes are most efficient when the target cytosine is part of an unstable base pair, with very little activity against G.C and efficient cleavage at A.C, AP.C and HEG.C, in which the cytosine will be more readily flipped out from the DNA helix. However the rate of excision is also affected by the sequence context in which the target cytosine is located and a similar effect is seen with A.C and HEG.C. Changing the flanking base pairs from AT to GC greatly reduced the efficiency of cleavage (by 100-fold for A.C and 10-fold for HEG.C). This effect was also seen with uracil when comparing A.U to A.U(G), showing that it is context and not target base dependent. The greater change with A.C is most likely because this base pair contains one hydrogen bond; while there will be no direct contacts in the HEG.C pair. While the stability of the flanking base pairs does affect the rate of excision (flanking AT is a much better substrate that flanking GC) this cannot be the main factor influencing the reaction, as unstable combinations (such as G.T, G.AP and G-) produce poor rates of cleavage. Base stacking between the target cytosine and its flanking bases may therefore be an important factor influencing the rate of excision.

Molecular BioSystems Accepted Manuscript

This journal is © The Royal Society of Chemistry 20xx

Please do not adjust margins

These data suggest that the 5′-flanking base has a greater effect as changing the base pair on the 5′-side of the AC mismatch to GC (sequence A.C(GA)) had a 10-fold greater effect than changing the 3′-base pair (sequence A.C(AG)). This is consistent with previous observations with UDG, which show that the 5′ flanking base has an important role in the rate of excision.[18-20]

Experimental

Complete Gene Synthesis

The clone expressing hUDG was generated by total gene synthesis[13,14] based on the cDNA sequence of human placental UDG.[15] Residues 3-84 were omitted as this N-terminal signal sequence is responsible for translocation to the nucleus/mitochondria.[16,17] NdeI and EcoRI restriction sites were added to the 5′ and 3′ termini respectively. The sequence was prepared as 18 oligonucleotides, each of approximately 60 bp that overlapped each other by 20 bp (Supplementary material). A mixture containing all the oligonucleotides (100 nM each) was amplified by PCR as follows: one cycle at 98°C for 2 min, then 55 cycles of 98°C for 30 sec, 50°C for 30 sec and 72°C for 2 min, and finally one cycle at 72°C for 10 min. A small amount of this mixture was then further amplified using only the first and last oligonucleotides: one cycle at 98°C for 2 min, 23 cycles 98°C for 30 sec, 50°C for 30 sec and 72°C for 2 min, and finally one cycle at 72°C for 10 min and held at 4°C. The final product underwent PCR clean-up (QIAGEN) and stored at -20°C. This was then cloned between the NdeI and EcoRI sites of pET28a. Site-directed mutagenesis was then used to generate the N204D (hCDG) and N204D:L272A (hCYDG) mutants.

Preparation of enzymes

The enzymes were expressed in BL21(DE3)pLysS cells, which were induced with 1 mM IPTG for three hours. The cells were lysed by sonication, purified using an Ni-NTA (His Trap FF Crude; GE Healthcare) and eluted in 250 mM imidazole. The enzymes were concentrated and further purified using a 20 mL 10000 MW Vivaspin column (Fisher Scientific). This produced enzymes of suitable purity, as analysed by SDS polyacrylamide gel electrophoresis, with a yield of 1.5 mg per 500 mL culture.

Radiolabelled oligonucleotides

Oligonucleotides were synthesized on an Applied Biosystems ABI 394 automated DNA/RNA synthesiser on the 0.2 or 1 µmol scale using standard methods. Phosphoramidite monomers and other reagents were purchased from Applied Biosystems or Link Technologies. The pyrrolidine anthraquinone phosphoramidite was purchased from Berry & Associates. Each 31 mer oligonucleotide was radiolabelled at its 5′-end using γ-^{32}P[ATP] and T4 polynucleotide kinase (New England Biolabs), purified by denaturing PAGE, and resuspended in 10 mM MES pH 6.3 containing 25 mM NaCl and 2.5 mM $MgCl_2$. These were mixed with an excess of the unlabelled complementary oligonucleotides and annealed by slowly cooling from 95°C to 4°C.

Enzyme cleavage

Radiolabelled DNA (approximately 50 nM in 10 mM MES pH 6.3 containing 25 mM NaCl and 2.5 mM $MgCl_2$) was digested with hCDG or hCYDG (typically 1.25 µM) for up to 24 h, removing samples from the reaction mixture at various time intervals. The reactions were stopped using 10% piperidine (v/v) and heated at 95°C for 20 min to cleave the phosphodiester backbone. The samples were lyophilised, resuspended in 5 µL loading buffer (80% (v/v) formamide, 10 mM EDTA, 10 mM NaOH and 0.1% (w/v) bromophenol blue) and run on a 12.5% denaturing polyacrylamide gel containing 8 M urea. The gel was then fixed, dried, subjected to phosphorimaging and analysed using ImageQuantTL. Experiments were performed in triplicate and k_{cat} values were determined using SigmaPlot by fitting to a single exponential rise to maximum to plots of percentage cleaved against time. The rate of cleavage of some substrates was very low (less than 10% cleaved after 24 hours incubation). In these instances an estimate of the rate constant was obtained from the fraction cleaved at a given time, assuming a simple exponential process.

Acknowledgements

STK was supported by a research studentship from the BBSRC.

References

1 T. Lindahl, Instability and decay of the primary structure of DNA, *Nature,* 1993, **362**, 709-715.
2 L. H. Pearl, Structure and function in the uracil-DNA glycosylase superfamily, *Mutat. Res.,* 2000, **460**, 165-181.
3 B. Kavli, G. Slupphaug, C. D. Mol, A. S. Arvai, S. B. Peterson, J. A. Tainer and H. E. Krokan, Excision of cytosine and thymine from DNA by mutants of human uracil-DNA glycosylase, *EMBO J,* 1996, **15**, 3442-3447.
4 C. D. Mol, A. S. Arvai, G. Slupphaug, B. Kavli, I. Alseth, H. E. Krokan and J. A. Tainer, Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis, *Cell,* 1995, ***80***, 869-878.
5 A. C. Drohat, J. Jagadeesh, E. Ferguson and J. T. Stivers, Role of electrophilic and general base catalysis in the mechanism of Escherichia coli uracil DNA glycosylase, *Biochemistry,* 1999, **38**, 11866-11875.
6 R. Savva, K. McAuley-Hecht, T. Brown and L. H. Pearl, The structural basis of specific base-excision repair by uracil-DNA glycosylase, *Nature*, 1995, **373**, 487-493.
7 K. Kwon, Y. L. Jiang, and J. T. Stivers, Rational engineering of a DNA glycosylase specific for an unnatural cytosine:pyrene base pair, *Chem. Biol.,* 2003, **10**, 351-359.
8 Y. L. Jiang, J. T. Stivers and F. H. Song, Base-flipping mutations of uracil DNA glycosylase: substrate rescue using a pyrene nucleotide wedge, *Biochemistry* 2002, **41**, 11248-11254.
9 Y. L. Jiang and J. T. Stivers, Mutational analysis of the base-flipping mechanism of uracil DNA glycosylase, *Biochemistry* 2002, **41**, 11236-11247.
10 S. S. Parikh, C.D. Mol, G. Slupphaug, S. Bharati, H. E. Krokan and J. A. Tainer, Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J* 1998, **17**, 5214-5226.

Molecular BioSystems Accepted Manuscript

Please do not adjust margins

11 H. E. Krokan, R. Standal and G. Slupphaug G, DNA glycosylases in the base excision repair of DNA. *Biochem. J.* 1997, **325**, 1-16.
12 S. T. Kimber, T. Brown and K. R. Fox, A mutant of uracil DNA glycosylase that distinguishes between cytosine and 5-methylcytosine, *PloS one* 2014, **9**, e95394.
13 D. G. Gibson, H. O. Smith, C. A. Hutchison, J. C. Venter and c. Merryman, Chemical synthesis of the mouse mitochondrial genome, *Nature methods,* 2010, **7**, 901-903.
14 C. Prodromou and L. H. Pearl, Recursive PCR: a novel technique for total gene synthesis, *Protein Eng.,* 1992, **5**, 827-829.
15 L. C. Olsen, R. Aasland, C. U. Wittwer, H. E. Krokan and D. E. Helland, Molecular cloning of human uracil-DNA glycosylase, a highly conserved DNA repair enzyme, *EMBO J* 1989, **8,** 3121-3125.
16 H. E. Krokan, M. Otterlei, H. Nilsen, B. Kavli, F. Skorpen, S. Andersen, C. Skjelbred, M. Akbari, P. A. Aas and G. Slupphaug, Properties and functions of human uracil-DNA glycosylase from the UNG gene, *Prog. Nucleic Acid Res. Mol. Biol.,* 2001, **68**, 365-386.
17 H. Nilsen, M. Otterlei, T. Haug, K. Solum, T. A. Nagelhus, F. Skorpen and H. E. Krokan, Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the UNG gene, *Nucleic Acids Res.,* 1997, **25,** 750-755.
18 Y. L. Jiang and J. T. Stivers, Reconstructing the substrate for uracil DNA glycosylase: tracking the transmission of binding energy in catalysis, *Biochemistry* 2001, **40**, 7710-7719.
19 R. M. Werner, Y. L. Jiang, R. G. Gordley, G. J. Jagadeesh, J. E. Ladner, G. Xiao, M. Tordova, G. L. Gilliland and J. T. Stivers, Stressing-out DNA? The contribution of serine-phosphodiester interactions in catalysis by uracil DNA glycosylase, *Biochemistry* **39**, 2000, 12585-12594.
20 G. Slupphaug, I. Eftedal, B. Kavli, S. Bharati, N. M. Helle, T. Haug, D. W. Levine and H. E. Krokan, Properties of a recombinant human uracil-DNA glycosylase from the UNG gene and evidence that UNG encodes the major uracil-DNA glycosylase. *Biochemistry* 1995, **34**, 128-138..

Molecular BioSystems Accepted Manuscript

This journal is © The Royal Society of Chemistry 20xx