



**CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS
DEL INSTITUTO POLITECNICO NACIONAL**

UNIDAD IRAPUATO

**Análisis Evolutivo y Bioquímico-Estructural de la Ruta de Reparación del
ADN por Escisión de Bases de *Entamoeba histolytica*: “El Rompecabezas a
través de las Piezas Faltantes”.**

Tesis que presenta

Carlos H. Trasviña-Arenas

Para obtener el grado de

Doctor en ciencias

En la especialidad de

Biología de Plantas.

Director de Tesis: Dr. Luis G. Brieba de Castro.

Irapuato, Gto.

Diciembre de 2017

El presente trabajo de investigación se desarrolló en el laboratorio de Bioquímica Estructural del Laboratorio Nacional de Genómica para la Biodiversidad del CINVESTAV-IPN bajo la dirección del Dr. Luis G. Brieba de Castro, Profesor Titular del Laboratorio Nacional de Genómica para la Biodiversidad y con la asesoría de los Doctores Gabriela Olmedo Álvarez, Luis Delaye Arredondo y Agustino Martínez Antonio, Profesores Titulares del Departamento de Ingeniería Genética de la Unidad Irapuato del CINVESTAV-IPN, El Dr. Rafael Montiel Duarte del Laboratorio Nacional de Genómica para la Biodiversidad del CINVESTAV-IPN y del Dr. Mario Pedraza Reyes de la Universidad de Guanajuato, en el período comprendido de Agosto del 2013 a Diciembre del 2017.

AGRADECIMIENTOS.

Se agradece al CONACYT por la beca para estudios de Doctorado No. 243008 durante el período comprendido de septiembre del 2013 a agosto del 2017.

La presente tesis estuvo financiada por CONACYT con el apoyo a *Problemas Nacionales* # 213915 dado al Dr. Luis Gabriel Brieba de Castro.

AGRADECIMIENTOS ESPECIALES

A mi director de tesis el **Dr. Luis Brieba de Castro**, por todo el apoyo en el transcurso de este trabajo, por los retos que me hicieron apreciar la ciencia de una forma diferente. Gracias por alimentar mis ganas de experimentar.

A mis asesores **Dra. Gabriela Olmedo Álvarez, Dr. Luis Delaye Arredondo, Dr. Agustino Martínez Antonio, Dr. Rafael Montiel Duarte** y al **Dr. Mario Pedraza Reyes**, por su interés en el presente trabajo y sus siempre acertadas observaciones.

A **Corina Díaz**, por su apoyo y paciencia en todos estos años, eres grande Cori, te quiero.

A los integrantes del laboratorio de Bioquímica Estructural del LANGEBIO: **Victor, Yuni, Eunice, Marcel, Noe, Pedro, Antolin, Paola, Alma, Lalo, Francisco, Misraim, Margarita, Maru, Priscila, Gil, Itzel y Ania**, Por toda la paciencia que me han tenido aguantándome todo este tiempo. Aunque no lo crean, siempre fue y será un gusto trabajar al lado de ustedes.

A **Enrico Baruffini** y **Tiziana Lodi** (investigación de la Universidad de Parma), al **Dr. Aldo Arvizu** (Universidad de Sonora), al **Dr. Rogerio Sotelo** (CIAD) y a la **Dra. Sheila David** (Universidad de California) por abrirme las puertas de sus laboratorios y compartir tan cálidamente su grupo de investigación.

A la **Dra. Gloria Yepiz Plascencia** y a la **M. en C. Alma Peregrino Uriarte**, porque a través del tiempo siguen contribuyendo en mi formación académica y personal.

A mis compas que me han acompañado en el transcurso de este tiempo, Al **Internacional Coca Jr.** (¡Siempre coca!), A **Anie Espinal, Yuni, Sergio Campos, Lis Arce, Myriam Rodríguez, Anibal, Humberto Herrera, Thelma Rico y Elohim Bello.** Muchas gracias por las pláticas y las cervezas consumidas en el proceso.

A **mi familia** (tíos, primos y abuelos), por siempre estar al pendiente de lo que me sucede, festejar conmigo o sin mí los aciertos y por escucharme hasta cuando no deben. Los quiero un chingo.

A **Nallely Hollos** por convertirte tanto en mi vida en tan poco tiempo. ¡Gracias por dejarte encontrar!

DEDICATORIA

“Porque si la dicha se midiese y la gracia fuese su hermana,

*no hubiese dicha más pura, ni gracia más humana,
que el tibio amor de la familia que nos ampara”*

Por eso y por todo, dedico este trabajo a **mi madre, padre y mi hermano**. Porque después de todo el único lugar donde quiero estar es a su lado.

*Woke up starin' at this, starin' at this empty room
Looked at thousand different pictures that your mother took of you
You see I had this crazy dream last night, this man he talked to me
He told me everything that's good and bad about my history*

*But he said that you are, you are the future
He said that you are, you are the future
And the future looks good...*

The future looks good (One Republic)

A mi hijo Noha, porque mi mayor logro en la vida será que sepas, con toda la seguridad que tu corazón pueda albergar, lo mucho que te amo y quiero estar contigo.

ÍNDICE.

RESUMEN	xii
ABSTRACT	xv
ABREVIATURAS	xviii
1. INTRODUCCIÓN	1
1.1. <i>Entamoeba histolytica</i> , LGT y sus implicaciones en la salud humana.	1
1.2. Particularidades del metabolismo de ácidos nucleicos de <i>E. histolytica</i> .	5
1.3. Reparación del DNA en <i>E. histolytica</i> , con especial enfoque a BER.	6
1.4. Mecanismos de reparación de 8oxoG.	11
1.5. MutY: Bioquímica y estructura	13
1.6. Clúster [4Fe-4S]	17
1.7. Clúster [4Fe-4S] y transporte de carga del ADN.	20
1.8. Localización puntual del sustrato por MutY y el clúster [4Fe-4S].	23
1.9. Antecedentes Específicos	26
2. OBJETIVOS GENERALES	30
2.1. Objetivos específicos.	30
3. MATERIALES Y MÉTODOS.	31
3.1. Identificación de genes de la Ruta de Escisión de bases de <i>E. histolytica</i> .	31
3.2. Análisis filogenéticos de la Ruta de Escisión de Bases.	31
3.3. Modelado estructural por homología.	32
3.4. Identificación del Gen de MutY de <i>E. histolytica</i> y clonación.	33
3.5. Mutagénesis puntuales y ensamble de quimeras.	34
3.6. Sobre expresión y Purificación de MutYs.	37
3.7. Identificación química de cluster [4Fe-4S]	39
3.8. Preparación de sustratos para los ensayos de glucosidasas.	39
3.9. Ensayos de actividad de glucosilasa.	40
3.10. Determinación de las MutY como Monofuncionales o bifuncionales.	40
3.11. Identificación funcional del clúster [4Fe-4S]	41
3.12. Conceptualización de cinéticas de enfoque mínimo (no Michaelleano) para la determinación de constantes bioquímicas.	42

3.13. Determinación de constantes k_2 y k_3 por cinéticas de intercambio múltiple y simple.	44
3.14. Determinación de constantes de afinidad ($[P_{1/2}]$).	44
3.15. Ensayos <i>in vivo</i> .	46
4. RESULTADOS Y DISCUSIÓN	48
4.1. EVOLUCIÓN DE LA VÍA DE ESCISIÓN DE BASES EN <i>Entamoeba histolytica</i>.	48
4.1.1. Análisis Evolutivo de BER de <i>E. histolytica</i> .	48
4.1.2. El sistema GO de <i>E. histolytica</i> ; una pieza perdida y dos extrañas.	56
4.1.3. Glicosilasa AlkD: dos adquisiciones y una duplicación.	67
4.1.4. Superfamilia UDG; La familia I y VI están presentes en <i>E. histolytica</i> .	76
4.1.5. EndoIII; una historia de duplicación de genes.	84
4.1.6. Observaciones finales y perspectivas del panorama evolutivo de BER de <i>Entamoeba histolytica</i> .	90
4.2. ESTUDIO BIOQUÍMICO-ESTRUCTURAL DE MUTY CON CARENCIA DE CLÚSTER [4Fe-4S].	92
4.2.1. Evolución de las novedosas MutYs carecientes del clúster [4Fe-4S].	92
4.2.2. Análisis de secuencia y del modelado estructural de MutY.	96
4.2.3. Corroboración de la ausencia del clúster [4Fe-4S].	100
4.2.4. Actividad <i>in vitro</i> y mutaciones del sitio activo.	103
4.2.5. Cinética de intercambio simple y múltiple	109
4.2.6. Ensayo de Frecuencia de mutación: ensayo <i>in vivo</i> .	115
4.2.7. La ausencia del clúster [4Fe-4S] puede ser mediada estructuralmente por efecto estérico.	117
4.2.8. Observaciones finales y perspectivas los estudios bioquímicos-estructurales de MutYs carentes de clúster [4Fe-4S] en <i>E. histolytica</i> y <i>L. brevis</i> .	124
5. CONCLUSIONES	125
REFERENCIAS BIBLIOGRÁFICAS.	126
APENDICES.	139
1er producto de investigación	
2do producto de investigación	

ÍNDICES DE FIGURAS

Figura 1. Ciclo de vida de <i>Entamoeba histolytica</i> .	4
Figura 2. Bases canonicas del DNA y lesiones más frecuentes asociadas a la Ruta de Escisión de Bases	9
Figura 3. Representación esquemática de la vía BER.	10
Figura 4. Representación esquemática del sistema GO.	12
Figura 5. Estructura cristalográfica de MutY de <i>E. coli</i> y <i>G. stearothermophilus</i>	15
Figura 6. Mecanismo catalítico de MutY.	15
Figura 7. Aminoácidos involucrados en la estabilización del 8oxoG y estabilización del ADN a través del motivo HhH en MutY.	17
Figura 8. Relación entre el motivo del clúster [4Fe-4S] y la cavidad catalítica de MutY.	21
Figura 9. Estructura cristalográfica de ADN con lesión 8oxoG:C, con apareamientos canónicos y con la lesión 8oxoG:A.	22
Figura 10. Potenciales medios de MutY con ADN y sin ADN (vs NHE) y esquema del mecanismo de localización puntual de la lesión a través del transporte de carga del ADN y del clúster [4Fe-4S].	25
Figura 11. Estructura cristalográfica de GsMutY y EcMutY y modelados estructurales de MutY <i>E. histolytica</i> y <i>L. brevis</i> .	28
Figura 12. Esquema del diseño de las quimeras de MutYs.	36
Figura 13. Esquema del PCR por SOEing utilizado para el ensamble de las quimeras.	37
Figura 14. Representación esquemática del sistema GO de <i>E. histolytica</i> .	57
Figura 15. Análisis estructural de MutT.	65
Figura 16. Análisis filogenético MutT.	66
Figura 17. Análisis filogenético de AlkD.	73
Figura 18. Análisis estructural de AlkD.	74
Figura 19. Análisis estructural y de secuencia del sitio activo.	75
Figura 20. Análisis filogenético de la superfamilia UDG.	81
Figura 21. Análisis estructural de la familia UDG I.	82
Figura 22. Análisis estructural de la glicosilasa de hipoxantina.	83
Figura 23. Alineamiento Representativo de glicosilasas de hipoxantina.	84
Figura 24. Análisis filogenético EndoIII.	88
Figura 25. Alineamiento representativo y análisis estructural de Endonucleasa III.	89
Figura 26. Alineamiento de secuencia y estructural de la región del motivo del clúster [4Fe-4S].	95
Figura 27. Análisis filogenético MutY.	96
Figura 28. Comparación estructural y de secuencia de MutYs canónicos y novedosos.	97
Figura 29. SDS-PAGE de MutYs purificadas y ensayo de quimioluminiscencia para la detección de clúster [4Fe-4S].	101
Figura 30. MutYs sin clúster [4Fe-4S] son glicosilasa activas <i>in vitro</i> .	106

Figura 31. Experimento de reconstrucción <i>in vitro</i> del clúster [4Fe-4S].	107
Figura 32. Cinética de intercambio múltiple y simple sobre 8oxoG:A e isothermas afinidad.	111
Figura 33. Comparación estructural y de secuencia de logo de la región correspondiente al motivo del clúster [4Fe-4S] entre el clado de Lactobacillales, anaeróbico y MutYs canónicos.	119
Figura 34. Representación esquemática del diseño de quimeras y su correspondiente actividad de glicosilasa <i>in vitro</i> sobre 8oG:A.	121

ÍNDICES DE TABLAS

Tabla 1. Identificación de genes de BER en <i>E. histolytica</i> .	54
Tabla 2. Elementos faltantes de BER en <i>E. histolytica</i> .	55
Tabla 3. Relación entre la presencia de OGG1/MutM/MutY y en el contenido de GC en los genomas de diferentes organismos.	60
Table 4. Constantes cinéticas para MutY con 8oxoG:A y constante de afinidad por 8oxoG:THF	114
Tabla 5. Ensayo de resistencia a rifampicina en la cepa CC104 <i>muty mutm</i> .	117
Tabla 5. Ensayo de resistencia a rifampicina en la cepa CC104 <i>muty mutm</i> de las MutYs quiméricas.	123

RESUMEN

Entamoeba histolytica es una amiba patógena, agente causal de la amibiasis que afecta alrededor del 10% de la población humana. Una de las complicaciones en el tratamiento de la amibiasis es la gran plasticidad genómica a través del intercambio de genes por medio de transferencia lateral de genes (LGT, por sus siglas en inglés). De esta forma, *E. histolytica* ha adquirido genes que le han permitido desarrollar resistencia a drogas como el metronidazol. En general, *E. histolytica* es un organismo singular en relación al metabolismo de ácidos nucleicos. Además de alto flujo de intercambio genético con bacterias y arqueas, la arquitectura y organización de su genoma también muestra singularidades, así como la evolución de las enzimas involucradas en la replicación, transcripción y reparación del ADN.

El mantenimiento de la integridad del ADN es una tarea primordial en *E. histolytica* dado que enfrenta respuestas inmunológicas y exposición a drogas que tienen como modo de acción dañar la información genética. Para esto *E. histolytica* tiene todas las vías de reparación del ADN como; ruta de Escisión de Bases (BER, por sus siglas en inglés), por Escisión de Nucleótidos (NER), por recombinación homóloga (HRR) y no homóloga (NHEJ) así como reparación de malos apareamientos (MMR). El estudio de la ruta BER de este organismo ha mostrado ser interesante dado la carencia de genes codificantes para enzimas como la ADN glicosilasa OGG1 y Ap-endonucleasas. Estas enzimas son centrales para la funcionalidad de BER, por lo que en caso de corroborarse la carencia de dichos genes indicaría mecanismos pocos comunes y/o repercusiones genómicas

preponderantes para este patógeno. Abundado en estas particularidades, en este trabajo se reporta la identificación y caracterización funcional de una ADN glicosilasa de adenina MutY de *E. histolytica* (EhMutY) con una posible incapacidad de coordinar el cofactor clúster [4Fe-4S]. Para esta glicosilasa importante en la reparación de Adeninas mal apareadas con la lesión 8-oxo-guanosina (8oxoG) por la vía de BER, se ha demostrado que este cofactor es vital para la catálisis y reconocimiento del sustrato. Por lo que la carencia del clúster [4Fe-4S] en la EhMutY es un aspecto estructural-bioquímico llamativo por sí solo. Mencionado lo anterior, este trabajo se enfoca a estudiar la composición de BER en *E. histolytica* así como su evolución y características bioquímico-estructurales de sus enzimas, con un especial enfoque en glicosilasas de ADN.

En nuestro análisis de los elementos que integran la reparación del ADN por la vía de BER se logró identificar un gen codificante para AP-endonucleasa central en la vía, la cual genera el producto 5'-dRP. En trabajos previos, no se había podido identificar dicho gen. En lo que respecta a la OGG1/MutM, ADN glicosilasa encargada de iniciar la reparación de la lesión 8oxoG:C, no se pudo identificar este gen en el genoma del patógeno. Nuestro análisis indica que la carencia de este gen puede estar contribuyendo fuertemente al bajo contenido de GC en el genoma de *E. histolytica*. Tampoco se identificó genes como XRCC1 y codificantes para ADN polimerasas con capacidad de liasa sobre 5'-dRP, elementos importantes para la reparación del ADN por las subvía del parche corto de BER. Sin embargo, el genoma de este patógeno contiene genes para polimerasas de ADN con capacidad de desplazamiento de hebra y PCNA característicos de la subvía del parche largo.

El análisis filogenético indica que las glicosilasas de ADN; MutY, AlkD's, UDG de la familia VI y la 8oxodGTPasa MutT fueron adquiridas por LGT de bacterias. No obstante, la EndoIII y la UNG de la familia I muestran una evolución vertical. En lo que respecta a la EhMutY se demostró la carencia del clúster [4Fe-4S] por métodos químicos y funcionales. Bioquímicamente, esta enzima es capaz de procesar malos apareamientos de A:8oG y A:G con la misma eficiencia que la MutY de *E. coli* que contiene el cofactor. Análisis estructurales por modelado por homología indican que la carencia del clúster [4Fe-4S] puede ser conducida por un efecto estérico.

En conclusión, la ruta de reparación de BER de *E. histolytica* ha sido moldeada por la pérdida, duplicación y adquisición de genes por LGT. Además, que las enzimas de esta vía han evolucionado a través de rearrreglos estructurales que permiten prescindir de motivos importantes.

ABSTRACT

Entamoeba histolytica is a human pathogenic amoeba that affects about 10% world population. One treatment problem for amebiasis is its genome plasticity especially due to exchange of genetic information through Lateral Gene Transfer (LGT). Thus, the acquisition of foreign DNA has shaped drug resistance like metronidazole and other metabolic processes. In general, regarding to nucleic acid metabolism this pathogen is singular. Besides a high exchange of genetic information with bacteria and archaea, the genome architecture and organization show some singularities, as well as the evolution of enzymes involved in DNA replication, transcription and repair.

To safeguard the DNA integrity is a pivotal task for *E. histolytica* since faces immunological responses and drug exposure whose action mechanisms are to threaten its genomic information. Thus, this pathogen harbors all different mechanisms for DNA repair like Base excision repair (BER), Nucleotide excision repair (NER), Homologous recombination, Non-Homologous End-Joining (NHEJ) and Mismatch repair (MMR) pathways. Analyses of BER of *E. histolytica* (EhBER) pathways have shown interesting absences. Since genes encoding glycosylase OGG1/MutM and Ap-endonuclease. These enzymes are central for BER functionality; therefore, their lack would suggest non-common repairing mechanisms or huge genomic repercussions for this pathogen. Moreover, Bioinformatic analysis has indicated that Adenine glycosylase MutY of *E. histolytica* (EhMutY) lacks the capability of coordinating a [4Fe-4S] cluster. This cofactor is determining for repairing 8oxoG-

related lesions via BER, being important for catalysis and lesion recognition. Therefore, the lack of this cofactor in EhMutY is another striking feature in EhBER.

Taking into account these considerations, this work focuses on studying EhBER elements as well as their evolution, structure and biochemical features with a special focus on DNA glycosylases.

In our analyses of EhBER elements we were able to identify a gene encoding Ap-endonuclease central for repairing apuric/apyrimidinic sites in this pathway yielding 5'-dRP products. Previous works were not able to identify this determining gene. However, we were not able to identify a gene encoding OGG1/MutM glycosylase in charge of repairing 8oxoG:C. Our analysis suggests that this lack could be involved in the low GC-content in *E. histolytica* genome. In addition, genes encoding XRCC1 and DNA polymerases with 5'-dRP lyase activity were not identified. These proteins are important for BER through short-patch sub-pathway. Nonetheless, this pathogen genome harbors genes encoding DNA polymerases with strand displacement activity and PCNA involved in long-patch sub-pathway. Phylogenetic analyses indicate that MutY, AlkDs, UDG glycosylases (Family-VI) and 8oxodGTPase MutT were acquired by LGT from bacteria. Nevertheless, EndoIII and UNG (Family I) have a vertical evolution. Regarding to EhMutY, we demonstrated the lack of [4Fe-4S] cluster through chemical and functional approaches. Biochemically, this enzyme is able to process A:8oxoG and A:G mismatches with similar efficiency to *E. coli* MutY a cofactor-containing glycosylase. Structural homology modeling suggests that the absence of [4Fe-4S] cluster can be driven by steric effect.

In conclusion, EhBER has been shaped for gene loss, duplication and acquisition by LGT. Moreover, these enzymes have evolved through structural rearrangements that allow the dispensability of determining motifs.

ABREVIATURAS.

ADN; Ácido desoxirribonucleico.

SOD-Fe; Superóxido dismutasa dependiente de hierro.

Ab; Sitio abásico.

BER. Reparación por escisión de base.

NER; Reparación por escisión de nucleótido.

NHEJ; Recombinación no homóloga.

MMR; Reparación de malos apareamientos.

A/C/G/T; deoxiadenina/deoxicitosina/deoxiguanina/deoxitimina.

Kd; Constante de disociación.

[P_{1/2}]; Concentración de proteína para alcanzar la mitad de saturación.

HhH; Hélice-horquilla-Hélice.

LGT; Transferencia lateral de genes.

UDG; Glicosilasa de Uracilo.

HxDG; Glicosilasas de hipoxantina.

GO; Guanina oxidada.

8oxoG; 8-oxo-7,8-dihidroguanina.

Clúster [4Fe-4S]; Clúster hierro-azufre.

P³²; Isotopo-32 radioactivo del fósforo.

1. INTRODUCCIÓN

1.1 *Entamoeba histolytica*, LGT y sus implicaciones en la salud humana.

Entamoeba histolytica es un protozoo patógeno anaeróbico, cuya afectación a la salud humana alcanza el 10% de la población mundial, causando alrededor de 50 millones de casos clínicos de disentería y abscesos hepáticos. Además, se ha estimado que es el agente causal de la muerte de alrededor de 100,000 personas por año en países en vías de desarrollo (Stanley 2003). Uno de los mayores problemas en el tratamiento de la amibiasis es la alta resistencia a drogas que presenta *Entamoeba histolytica*. Un gran número de estudios se han hecho para entender los mecanismos de resistencia de este organismo. Se han caracterizado mecanismos de resistencia contra diversos compuestos, como la emita (Perez, Gómez et al. 1998), derivados de 5-nitroimidazol (como metronidazol, y nitaxozamida) (Samarawickrema, Brown et al. 1997, Wassmann, Hellberg et al. 1999, Upcroft and Upcroft 2001, Cedillo-Rivera, Chávez et al. 2002), cloricina, deunomicina, actinomicina (Ayala, Samuelson et al. 1990) y derivados de bifosfonatos (Ghosh, Chan et al. 2004).

Algunos mecanismos de resistencia a drogas de este organismo están ampliamente caracterizados. En estos involucran el uso de transportadores dependientes de ATP que permiten la expulsión de las drogas fuera de célula, regulación transcripcional de genes a partir de mutantes en promotores, y

disminución de la actividad de enzimas modificadoras de drogas para su activación (Edwards 1993). Otro aspecto importante en la resistencia a drogas, es la plasticidad genómica de *E. histolytica*. La plasticidad genómica es la capacidad de reorganizar la información genética mediante inserciones, transposiciones y rearreglos cromosómicos. Otro mecanismo implicado en la plasticidad genómica es la transferencia lateral (u horizontal) de genes (LGT por sus siglas en inglés) que consiste en el intercambio de información genética entre dos organismos del mismo o diferente orden taxonómico. Esta propiedad ha jugado un rol amplio en la patogenicidad y resistencia a drogas de los microorganismos (Dobrindt, Zdziarski et al. 2010).

En *E. histolytica* los genes que codifican para proteínas relacionadas en la resistencia a drogas como; la piruvato:ferredoxina oxidoreductasa, la ferredoxina de 6 kDa (Rosenthal, Mai et al. 1997), nitrorreductasa (Nixon, Wang et al. 2002) y una superóxido dismutasa dependiente de Fe (SOD-Fe) (Tannich, Bruchhaus et al. 1991, Smith, Feng et al. 1992) tienen un origen bacteriano. Si bien, La LTG ha sido un mecanismo que ha contribuido a la patogenicidad de este protozooario, este mecanismo de adquisición de genes ha contribuido más extensamente en su fisiología. También se han identificado que los genes *iscU* e *iscl* que participan en el ensamblaje del clúster [4Fe-4S], un cofactor con una extensa promiscuidad funcional (Van Der Giezen, Cox et al. 2004), la alcohol deshidrogenasa, NADH oxidasa y la acetil-CoA sintetasa importantes para el mantenimiento del metabolismo anaerobio han sido adquiridas por LGT (Field, Rosenthal et al. 2000, Nixon, Wang et al. 2002). Estudios iniciales estimaron que la contribución genómica

de la LGT a *E. histolytica* era de aproximadamente 1% de los genes (Loftus, Anderson et al. 2005). No obstante, estimaciones más recientes sugieren que alrededor del 4% de los genes han sido adquiridos por este mecanismo (Grant and Katz 2014). El impacto funcional que ha tenido la plasticidad genómica en los procesos biológicos de este patógeno es muy amplio, involucrándose en procesos como el metabolismo energético y de ácidos nucleicos (como la replicación), la transducción de señales y virulencia (Alsmark, Sicheritz-Ponten et al. 2009, Grant and Katz 2014).

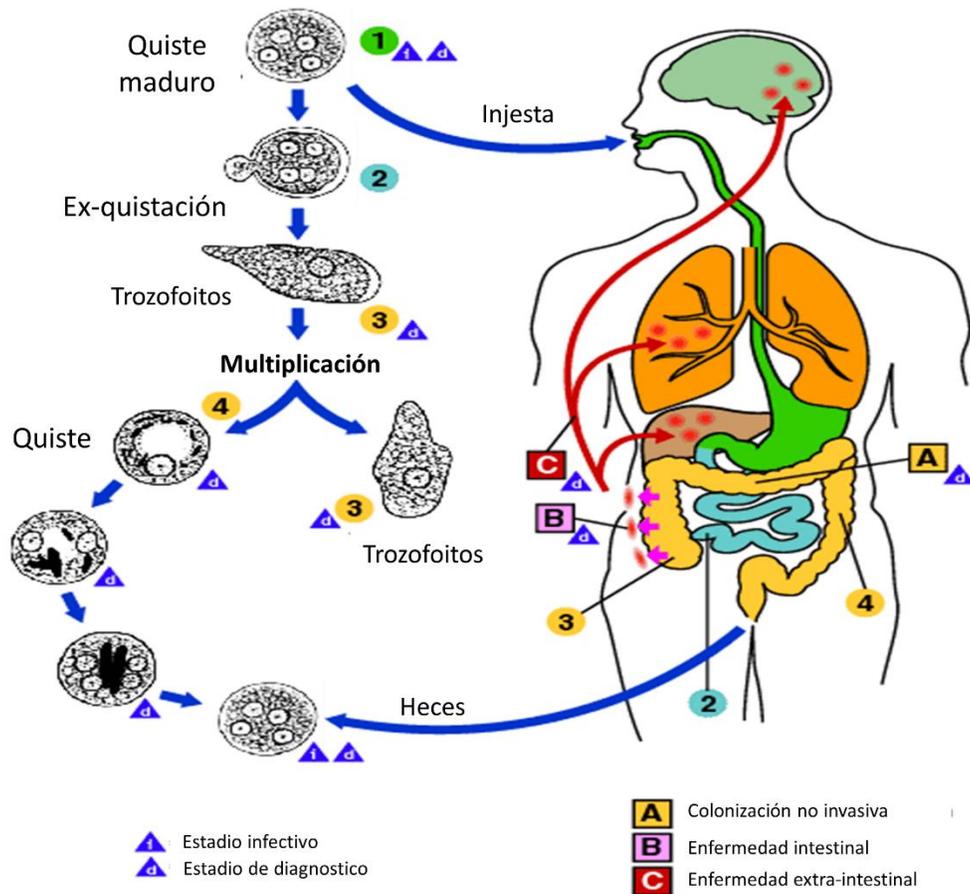


Figura 1. Ciclo de vida de *Entamoeba histolytica*. La infección por *E. histolytica* ocurre por ingestión de quistes maduros por el consumo de alimentos o agua contaminada. Los quistes de este patógeno presentan 4 núcleos y son resistentes al ambiente gastrointestinal. Una vez que los quistes llegan al *caecum* (unión entre el intestino corto y largo) ocurre la ex-quistación generando que la ameba de 4 núcleos se divida por fisión binaria en 4 trozofitos. Los trozofitos migran al intestino largo para alojarse en el tejido submucoso ocasionando una enfermedad intestinal. En este tejido los trozofitos se multiplican. Algunos trozofitos migran al lumen y cambian al estadio de quiste de nuevo para ser desechados a través de las heces. Así pueden iniciar su ciclo infeccioso de nuevo. En caso de Colonización no invasiva los trozofitos permanecen confinados en el lumen intestinal propiciando un cuadro conocido como acarreadores asintomáticos. En casos extremos los trozofitos pueden pasar a otros órganos (hígado, cerebro, corazón y pulmones) a través de la sangre. Modificado de CDC (<https://www.cdc.gov/dpdx/amebiasis/index.html>)

1.2 Particularidades del metabolismo de ácidos nucleicos de *E.*

histolytica.

El estudio del metabolismo de ácidos nucleicos de *E. histolytica* ha mostrado contar con características muy particulares. Como se mencionaba anteriormente, se estima que el genoma de este patógeno tiene aproximadamente un 4% de origen bacteriano. Concerniente a la arquitectura de su genoma, su cromatina está altamente condensada, de tal forma que hasta la fecha no se sabe el número de cromosomas que tiene este organismo (Torres-Guerrero, Peattie et al. 1991). *E. histolytica* contiene ADN circular codificantes, configuración característica de los cromosomas plasmídicos de bacterias (Dhar, Choudhury et al. 1995, Lioutas, Schmetz et al. 1995). Replicativamente, su genoma es mantenido por pocas polimerasas de ADN arquetípicas, capaces de replicar el ADN en presencia de lesiones voluminosas, como glicos de timina y sitios abásicos (Ab) (Pastor-Palacios, Azuara-Liceaga et al. 2010, Pastor-Palacios, López-Ramírez et al. 2012, Pastor-Palacios, López-Ramírez et al. 2015). También presenta una economía genómica en la replicación y simplificación funcional, ya que en eucariontes, hay cuatro ligasas de ADN que están involucradas en procesos específicos como replicación, reparación del ADN y recombinación (incluyendo recombinación en procesos inmunológicos). No obstante, *E. histolytica* solo contiene un gen que codifica para una ligasa de ADN del tipo I que ha demostrado tener una promiscuidad funcional, cumpliendo está sola con todos los procesos involucrados en las transacciones del ADN (Cardona and Brieba 2009). Así mismo molecularmente, cuenta con la capacidad de prescindir de motivos que se han

descrito como fundamentales para la interacción proteína-proteína como la caja de unión al dominio PIP de la ligasa I (Trasviña-Arenas, Cardona-Felix et al. 2017) y de cofactores vitales enzimáticamente hablando como en el caso de la enzima de reparación del ADN MutY y el clúster [4Fe-4S] (Trasviña-Arenas, Lopez-Castillo et al. 2016). Relativo a la transcripción, *E. histolytica*, muestra un alto número de eventos de duplicación de genes que codifican para factores de transcripción de la familia Myb que ha conllevado a una amplia diversificación funcional (Meneses, Cárdenas et al. 2010).

1.3 Reparación del ADN en *E. histolytica*, con especial enfoque a BER.

Como se ha mencionado anteriormente, los aspectos moleculares de los procesos replicativos y de reparación del ADN han resultado ser particulares en *E. histolytica*, mostrando una simplificación y diversificación funcional. Esto sugiere que los aparatos involucrados en el metabolismo de ácidos nucleicos han estado bajo una fuerte presión evolutiva que los ha llevado a adaptarse molecularmente para mantener su función. *E. histolytica* como patógeno humano, se encuentra expuesto a diferentes estreses, como exposición a especies reactivas de oxígeno (ROS, por sus siglas en inglés) por condiciones inmunológicas del huésped, además de exposición a agentes alquilantes y deaminaciones. Esto hace que el ADN de este patógeno este constantemente expuesto a compuestos mutagénicos, por lo que es de vital importancia tener los mecanismos moleculares necesarios para la reparación de la información genética. Es importante mencionar que *E. histolytica*, a pesar de ser un organismo anaerobio, cuenta con mecanismos enzimáticos y no

enzimáticos para protegerse de ROS (Tekwani and Mehlotra 1999), resaltando así la importancia de depurar estos compuestos.

López-Camarillo y colaboradores (2009), así como Marchat y colaboradores (2011) han analizado los diferentes componentes de los mecanismos de reparación del ADN presentes en *E. histolytica*, encontrando que contiene los elementos necesarios para los procesos de reparación del ADN por las vías de Reparación por Escisión de Bases (BER por sus siglas en inglés), reparación por Escisión de Nucleótidos (NER), por recombinación homóloga (HRR) y no homóloga (NHEJ), y la ruta de reparación por malos apareamientos (MMR) (López-Camarillo, Lopez-Casamichana et al. 2009, Marchat, López-Casamichana et al. 2011).

La ruta de reparación de BER es la más versátil siendo capaz de reparar una gran gama de daños sutiles como oxidaciones, alquilaciones y deaminaciones, con enormes repercusiones como transversiones y transiciones (**Figura 2** se muestran algunas de las lesiones más comunes en el ADN asociadas a BER). Esta ruta de reparación esta concertada por un conjunto de enzimas que participan en los diferentes pasos de reparación. Primeramente, una glicosilasa de ADN resuelve la complicada tarea de identificar la sutil lesión en un mar de ADN no dañado. Una vez identificándola, cataliza la hidrólisis del enlace N-glucosídico entre la deoxiribosa y la base nitrogenada dañada (Krokan, Standal et al. 1997). Posteriormente, una endonucleasa apúrica/apirimidimica (AP-endonucleasa) está encargada de procesar un rompimiento del sitio abásico (Ab) mediante la hidrólisis ó lisis del enlace fosfodiester (dependiendo del tipo de AP-endonucleasa) en el extremo 5' del sitio Ab. Este sitió parcial se conoce como producto 5'-dRP. El producto 5'-dRP

puede ser procesado ulteriormente a través de dos vías alternativas de BER, la ruta de parche corto o parche largo (Frosina, Fortini et al. 1996). En la primera, una polimerasa de ADN, con capacidad de liasa, remueve el producto 5'-dRP por β -eliminación usando una Lisina como catalizador. Después, la misma polimerasa incorpora el nucleótido correspondiente, para que finalmente una ligasa de ADN termine la reparación condensando el enlace fosfodiéster entre el nucleótido recién incorporado y la cadena ya existente. En el caso de la vía alterna del parche largo, es requerida una polimerasa de ADN con capacidad de desplazamiento de hebra. Así a diferencia de la vía del parche corto, la polimerasa de ADN implicada incorpora de 6 a 12 nucleótidos desplazando la hebra existente, generando de esta forma una estructura conocida como 5'-flap. Finalmente, una Flap-endonucleasa, como su nombre lo indica, remueve el producto 5'-flap para que la ligasa de ADN genere el enlace fosfodiéster requerido para completar la reparación. En la **figura 3** se muestra una representación esquemática de BER y sus vías alternas. Para información más detalla revisar (Frosina, Fortini et al. 1996, David, O'Shea et al. 2007).

Las moléculas anteriormente mencionadas son las principales protagonistas de la ruta de BER, sin embargo, es importante mencionar la participación de otras proteínas que actúan como accesorios y/o moduladores de la ruta. Por ejemplo, la proteína XRCC1, no realiza una reacción en particular, pero es capaz de reconocer sitios abásicos y reclutar elementos de BER para su procesamiento especialmente por la vía alterna de parche corto (Cappelli, Taylor et al. 1997). En contraparte, el homotrímero PCNA coordina la procesividad de la polimerasa de ADN durante la

polimerización asociada al desplazamiento de hebra y la ligación del ADN en la vía de parche largo (Levin, McKenna et al. 2000).

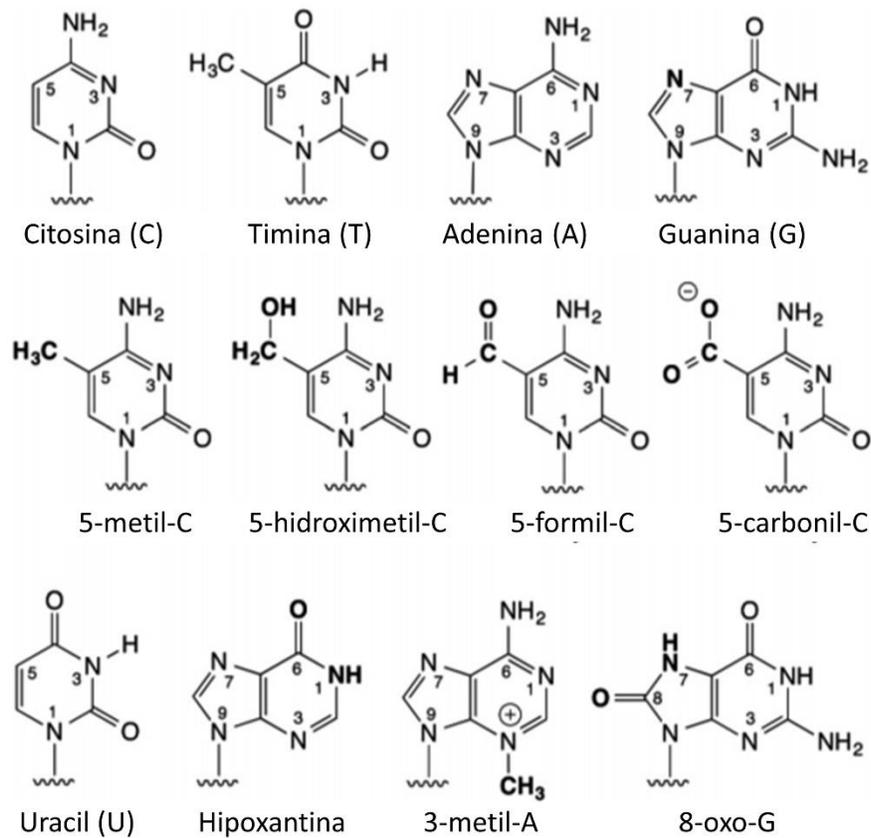


Figura 2. Bases canónicas del ADN y lesiones más frecuentes asociadas a la Ruta de Escisión de Bases (Modificado de Drohat y Maiti, 2014).

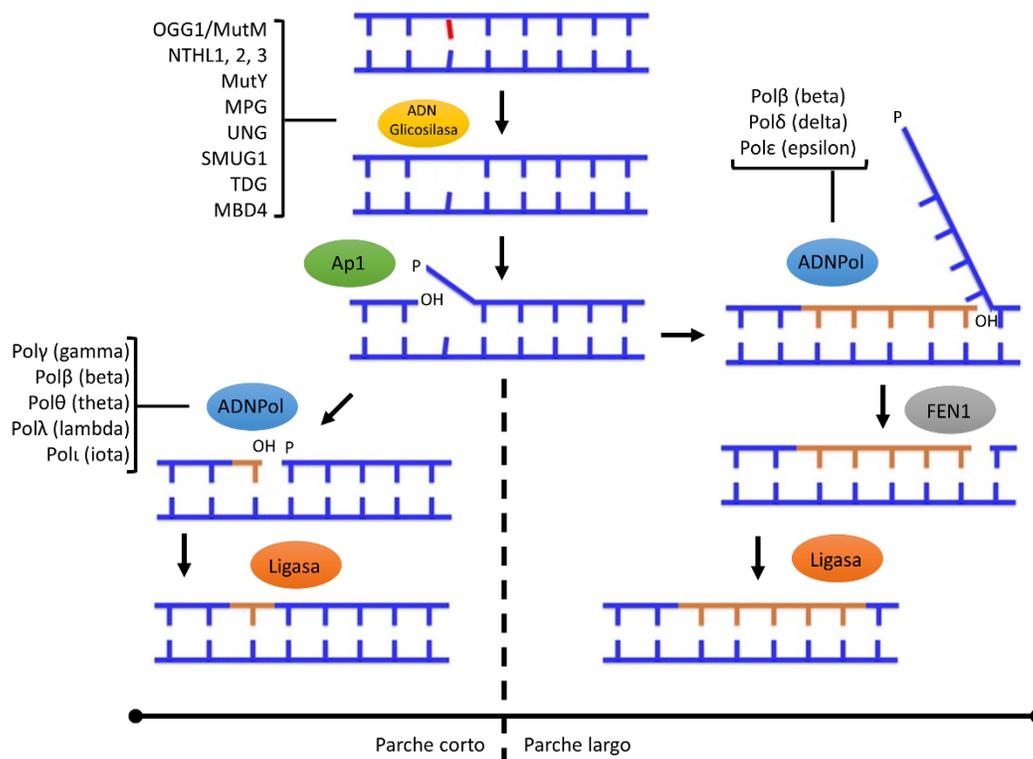


Figura 3. Representación esquemática de la vía BER. El proceso de reparación a través de la ruta BER se inicia con una glicosilasa de ADN que reconoce e hidroliza el enlace N-glucosídico entre la base dañada y la desoxirribosa, dejando un sitio abásico. Las glicosilasas de ADN se clasifican en glicosilasas monofuncionales y bifuncionales. El primero alberga actividad hidrolítica, mientras que el último, además de tener actividad hidrolítica, posee actividad de liasa rompiendo el enlace fosfodiéster en la posición 3'-OH del sitio Ab, produciendo fracciones 3'-OH y 5'-desoxirribosfosfato (5'-dRP). Si se requiere una glicosilasa de ADN monofuncional después de producir el sitio Ab, una endonucleasa Apurica/apirimidimica (AP) procesa el sitio abásicos que genera los productos 3'-OH y 5'-dRP. En este punto, la vía BER se bifurca en dos vías secundarias, la vía del parche corto y largo. En el parche corto se requiere una polimerasa de ADN con actividad de liasa para procesar el producto 5'-dRP e incorporar un único nucleótido correcto. Por último, una ligasa sella el intervalo o gap restante. En el caso del parche largo, se necesita una polimerasa de ADN con actividad desplazamiento de hebra para incorporar un conjunto de nucleótidos correctos produciendo una estructura 5'-Flap. La estructura 5'-Flap es eliminada por una Flap-endonucleasa posteriormente. Luego, una ligasa de ADN completa el proceso de reparación de Escisión de Bases.

1.4. Mecanismos de reparación de 8oxoG.

La oxidación de la guanina a 8-oxo-7,8-dihidroguanina (8oxoG) es una de las lesiones en el ADN con mayor prevalencia en la célula. Se calcula que en eucariontes se presentan de 1000 a 7000 lesiones/célula/día (Collins 1999). Si esta lesión no es reparada, puede derivar a la incorporación de una adenina junto a la base 8oxoG (lesión 8oxoG:A) durante un evento de replicación. Si esta lesión prevalece, dará lugar a una mutación por transversión GC→TA durante un segundo evento de replicación. Ese tipo de lesiones están relacionadas con enfermedades visuales, metabólicas y cáncer en humanos (Markkanen, Dorn et al. 2013).

Dada la importancia de la reparación de la lesión 8oxoG, existe un sistema especializado en su reparación, conocido como sistema GO (por guanina oxidada). En este sistema participan tres enzimas implicadas directamente con la lesión 8oxoG: MutT, MutM y MutY (Michaels and Miller 1992). MutT posee actividad de 8oxodGTPasa hidrolizando los trifosfato de deoxiguanina oxidados, evitando así que la polimerasa de ADN los incorpore durante la replicación (Abeygunawardana, Weber et al. 1995). En el caso de que se presente la lesión 8oxoG apareada con citosina (8oxoG:C) en el ADN ya sea por oxidación directa de la guanina o por incorporación del 8oxodGTP durante la replicación, la glicosilasa MutM (OGG1 en mamíferos) es la encargada de reparar este daño hidrolizando el enlace N-glucosídico de la 8oxoG, dejando un sitio Ab que posteriormente será reparado por las enzimas de BER (Sung, Zhang et al. 2013). No obstante, si la lesión 8oxoG:C no es corregida a tiempo y se presenta un evento replicativo, se genera la lesión 8oxoG:A. La glicosilasa MutY actúa sobre la bases mal apareada, mas no sobre la

base dañada, hidrolizando el enlace N-glucosídico de la adenina y dejando un sitio abásico que será procesado por enzimas que restauren el daño 8oxoG:C y finalmente se completará la reparación por la vía de MutM/BER (Michaels, Cruz et al. 1992) (En la **Figura 4** se describe el sistema GO).

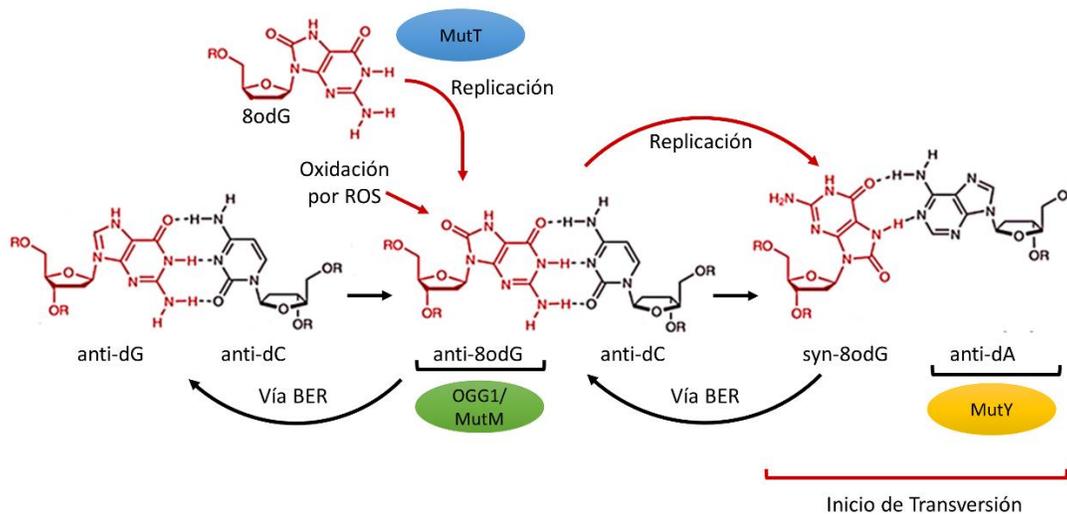


Figura 4. Representación esquemática del sistema GO. 8oxoG es generado por la oxidación de guanina por ROS ó RNS. Guanósina (dG) se hibrida con citosina (dC) en una conformación *anti*, pero cuando se presenta la lesión 8oxoG hay un cambio en la conformación a *syn*. En esta conformación *syn* las polimerasas de ADN incorporan un dA opuesto al 8odG que genera una transversión durante un evento replicativo. El sistema GO está constituido por MutT, MutM y MutY. El primero no pertenece a la vía BER pero es una 8oxodGTPasa a cargo de hidrolizar dGTP oxidado evitando su incorporación durante un evento replicativo. MutM es una glicosilasa de 8oxoG (bifuncional) que elimina 8oxoG dejando un producto 3'-dRP que se repara completamente a través de la vía BER. Cuando la lesión de 8oxoG no se repara a través de MutM y se genera un apareamiento de 8oG:A, la glicosilasa MutY (monofuncional) elimina la dA mal apareada dejando un sitio abásico. Este sitio se procesa mediante la vía BER y restaura la lesión inicial 8odG:dC que finalmente se repara a través de la vía MutM/BER.

1.5. MutY: Bioquímica y estructura.

La glicosilasa MutY ha sido una de las enzimas implicadas en la reparación del ADN más estudiadas desde el descubrimiento de su rol en la reparación a través de BER a finales de la década de los 80's. Inicialmente se encontró que la cepa de *Escherichia coli* AB1157-Y11 *mutY*⁻ presentaba un alto número de mutaciones por transversión GC→TA (Au, Cabrera et al. 1988). Posteriormente, se comprobó que el gen de *mutY* complementaba el fenotipo mutagénico. Este gen consistía de 1050 pb y codifica para una proteína de 39.1 kDa. Algo a resaltar es que esta MutY de *E. coli* tiene en sus primeros 218 amino ácidos un 63% de similitud (23% de identidad) con la Endonucleasa III (EndoIII), una glicosilasa encargada de la reparación de pirimidinas oxidadas (como glicol de timina, , glicol de uracilo, etc.) y que 4 cisteínas (C192, C199, C202 y C208) se encuentran conservadas entre estas enzimas, y en EndoIII se identificó que coordinan un clúster [4Fe-4S] (Michaels, Pham et al. 1990). Inicialmente se identificó a MutY como una enzima especializada en la reparación de lesiones 8oxoG:A y G:A, pero actualmente se han identificado hasta nueve lesiones diferentes sobre las que esta glicosilasa actúa (Markkanen, Dorn et al. 2013). La promiscuidad de esta enzima interesantemente depende del organismo, siendo la MutY de *E. coli* una de las más promiscuas reconociendo hasta 8 sustratos diferentes, mientras que la MutY de humano solamente reconoce 4 lesiones diferentes (Markkanen, Dorn et al. 2013). Además, se encuentran depositadas 16 estructuras en la base de datos Protein Data Bank, con estructuras de *E. coli*

(EcMutY), *Geobacillus stercophilus* (GsMutY) y humano (Brooks, Adhikary et al. 2013).

El mecanismo catalítico de MutY es similar al de otras glicosilasas de ADN, como MutM, Endo III y UDG. Primeramente, MutY tiene que identificar la lesión 8oxoG:A principalmente interactuando con la guanina oxidada. Este reconocimiento va a la par con la estabilización de la lesión 8oxoG por ocho aminoácidos a través de puentes de hidrógeno (Ver **figura 5**). La hidrólisis de la adenina se da en un sitio alejado de la ubicación inicial de 8oxoG:A (aproximadamente 15 Å de distancia) conocido como cavidad catalítica. Por lo tanto, para que la adenina llegue a la cavidad catalítica, la MutY extrude o empuja la adenina usando un residuo de Tyr como cuña para posicionar la lesión en la cavidad catalítica (Livingston, Kundu et al. 2005). En esta cavidad se ubican 4 moléculas de H₂O de las cuales 2 participan en la catálisis. Una primera molécula de H₂O auxiliada por un Glu estabiliza la adenina dentro de la cavidad y una segunda es activada por un Asp para realizar un ataque nucleofílico al carbono 1 de la deoxiribosa y así generar la hidrólisis. Se cree que la liberación del producto se da de manera asistida, ya que la afinidad de MutY por su producto abásico es tres veces mayor que su sustrato 8oxoG:A pero que esta afinidad MutY:Producto es reducida por la interacción con Ap-endonucleasa. Por lo tanto, se ha asociado el mecanismo de liberación de producto como un paso regulatorio de la catálisis, orquestando así los pasos subsecuentes de la reparación por AP-Endonucleasa, polimerasa de ADN y ligasas (Brunk, Arey et al. 2012).

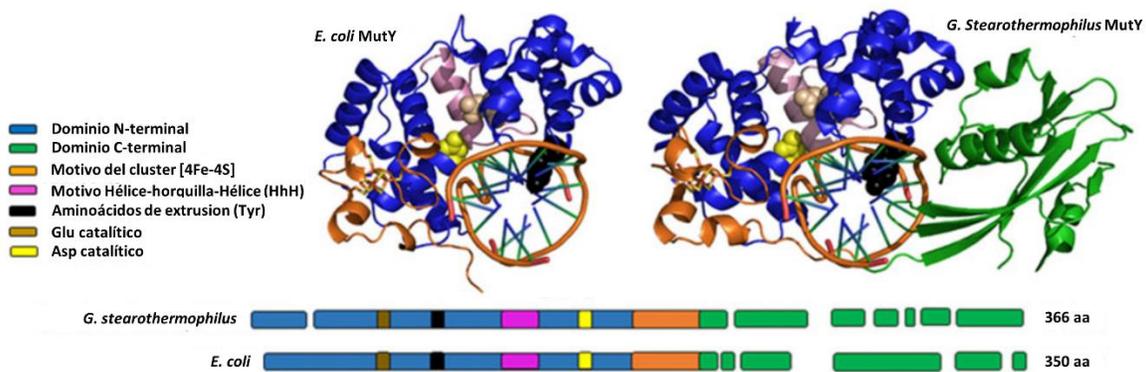


Figura 5. Estructura cristalográfica de MutY de *E. coli* y *G. stearothermophilus*. Resaltando sus diferentes motivos y dominios. En la parte de abajo se muestra un esquema del alineamiento entre MutYs de *E. coli*, humano y *B. stearothermophilus*. Modificado de David *et al*, 2007.

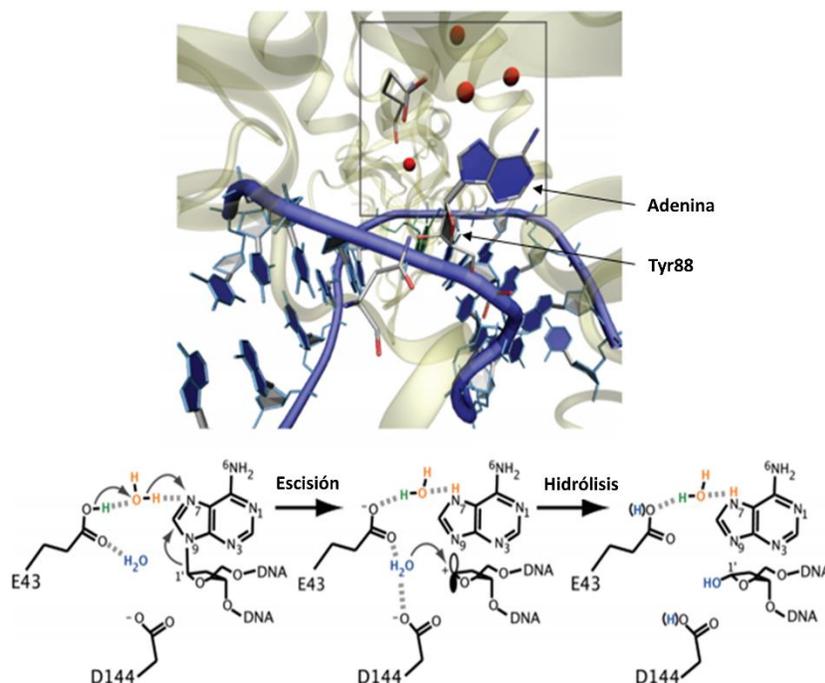


Figura 6. Mecanismo catalítico de MutY. En la figura de arriba se muestra el mecanismo de extrusión de la adenina por Tyr88 de MutY de *G. stearothermophilus*. En un recuadro se muestra el área de la cavidad catalítica y en esferas rojas las moléculas de agua. Modificado de Brunk *et. al*, 2012.

La MutY de *E. coli* puede ser dividida en dos dominios, un dominio N-terminal constituido por los primeros 225 aminoácidos (M1-K225, también llamado p26) y el dominio C-terminal que consiste de los últimos 125 amino ácidos (Q226-V350). Se sabe que el dominio N-terminal a pesar de carecer de 125 amino ácidos es estable y activa *in vitro* (Manuel, Czerwinski et al. 1996). Sin embargo, este dominio presenta una disminución en su afinidad por sustratos con lesiones como 8oxoG:A y G:A de hasta 12 veces comparando con la MutY completa. Estudios estructurales y bioquímicos mostraron que el dominio N-terminal de la MutY contiene algunos motivos importantes para su actividad de reconocimiento e interacción. Este dominio contiene un motivo conocido por su plegamiento como hélice-horquilla-hélice (motivo HhH por sus siglas en inglés), aminoácidos involucrados en la catálisis (Y82, E37 y D138), y un clúster [4Fe-4S] coordinado por cuatro cisteínas. Prácticamente esto es suficiente para mantener la estabilidad estructural y actividad enzimática de MutY.

En lo que concierne al dominio C-terminal se ha visto que esta evolutivamente relacionado con MutT, la 8oxodGTPasa encargada de remover los deoxinucleotidos trifosfatos de guanina oxidados. Si bien es difieren en secuencia (aproximadamente 12% de identidad), su estructura terciaria es bastante similar teniendo cinco láminas β y dos hélices α (Volk, House et al. 2000). Estudio bioquímicos y estructurales muestran un rol del dominio C-terminal en la especificidad de MutY sobre 8oxoG:A estabilizando la lesión 8oxoG para la eficiente extrusión de la adenina (Noll, Gogos et al. 1999) (**Figura 4**).

El motivo HhH es característico de un gran número de glicosilasas, es una firma estructural y evolutiva de un amplio número de glicosidasas. De hecho, casi todas la glucosidasas de la vía de Reparación por Escisión de Bases pertenecen a la superfamilia de ADN glicosilasas con motivo HhH (Denver, Swenson et al. 2003). El motivo HhH en glicosilasas tiene un rol general estabilizando la hélice de ADN en la cercanía del sitio activo. Esta estabilización se da particularmente con aminoácidos de la horquilla que interactúan con el esqueleto de fosfatos del ADN (Figura 7).

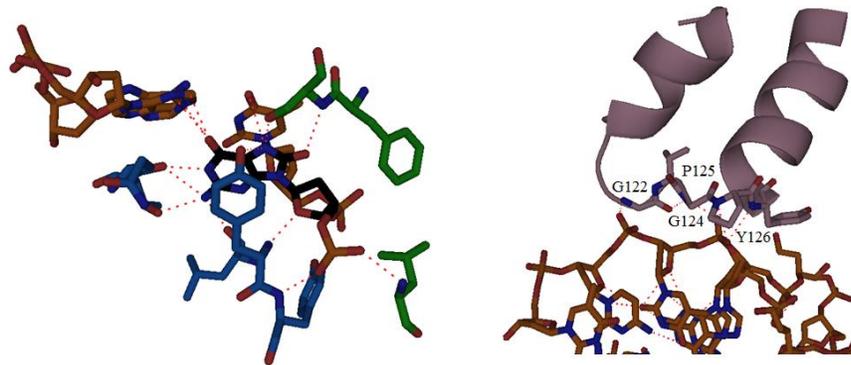


Figura 7. Izquierda; Aminoácidos involucrados en la estabilización del 8oxoG. En azul se muestran los residuos del dominio N-terminal, en verde los del C-terminal, en naranja el ADN y en negro la lesión 8oxoG. **Derecha;** Estabilización del ADN a través del motivo HhH. En palillos se resaltan los aminoácidos que contactan el esqueleto de fosfato del ADN.

1.6. Clúster [4Fe-4S]

El clúster de hierro es un cofactor de metaloproteínas que como su nombre lo indica, contienen átomos de hierro coordinados normalmente por cuatro ligandos de tiolato cedidos por cisteínas, aunque también en algunos casos puede ser

encontrado un ligando diferente a cisteína como histidina, glutamina, aspartato y serina (Moulis, Davasse et al. 1996). Este clúster en sus diferentes configuraciones ([1Fe], [2Fe-2S] y [4Fe-4S] y [8Fe-8S]) es uno de los cofactores con mayor presencia en la biología llevando a cabo una amplia variedad de roles, tales como andamio estructural, actividad redox, transducción de señales, etc. El Fe^{2+} reacciona con el oxígeno produciendo un radical hidroxilo a través de la reacción de Fenton, una de las especies reactivas de oxígeno más mutagénicas que se conocen. Es por eso que hasta hace algunos años, el clúster de hierro era considerado un componente raro de las proteínas que interactúan con los ácidos nucleicos (White and Dillingham 2012). Actualmente, el clúster de hierro se ha identificado en proteínas que interactúan con ADN con diferentes funciones como, reparación, replicación, transcripción, recombinación del ADN y reestructuración cromosomal. En reparación del ADN a través de la vía por BER hay varias enzimas con clúster de hierro, especialmente [4Fe-4S], entre ellas Endonucleasa III, MutY, uracil ADN glicosilasa, UV glicosilasa, y Mig glicosilasa, la cual repara timina mal apareada con guanina, la helicasa XPD y la nucleasa AddB. Las glicosilasas tienen una firma característica de ligandos de cisteína muy conservados ($\text{CX}_6\text{CX}_2\text{CX}_5\text{C}$ donde X puede ser cualquier aminoácido) sugiriendo que su ancestro común también contenía este motivo (White and Dillingham 2012).

En MutY de *G. steraothermophilus* (GsMutY) el motivo clúster [4Fe-4S] es coordinado por C198, C204, C208 y C214 que junto con 22 amino ácidos (196-218, 190-212 en GsMutY y MutY de *E. coli*, respectivamente) constituyen el dominio del asa del clúster [4Fe-4S] (FCL por sus siglas en inglés), que es estabilizado por 13

puentes de hidrógeno y que además está expuesto al solvente. Inicialmente se creía que el clúster [4Fe-4S] estaba involucrado en el rompimiento del enlace N-glucosídico de la adenina a través de reacciones de óxido-reducción. No obstante, una vez conocido el mecanismo catalítico de la MutY y dada la distancia física entre el clúster [4Fe-4S] y la cavidad catalítica de 19.01 Å, su rol dentro del mecanismo catalítico fue rechazado. Posteriormente se observó que al desnaturalizar (con 6 M de Gu-HCl) y renaturalizar la MutY con y sin los elementos necesarios para la formación del clúster [4Fe-4S], la estructura secundaria es esencialmente idénticas al igual que su estabilidad termodinámica. Sin embargo, las MutY sin clúster [4Fe-4S] no fueron activas (Porello, Cannon et al. 1998). Además, se ha encontrado que el efecto de la sustitución de los ligandos de cisteína depende de la cisteína sustituida y del tipo de sustitución, siendo los ligandos C208 y C204 lo más sensibles a la sustitución dando como resultado proteínas inestables estructuralmente (Golinelli, Chmiel et al. 1999). Otros estudios bioquímicos han ido más allá de la función de los ligandos de cisteínas en la estabilidad del dominio FCL, observándose que al mutar los aminoácidos de carga positiva como R194, K196 y K198 de *EcMutY* por alaninas o glicinas, disminuyeron su capacidad para unirse a su sustrato y para llevar a cabo la hidrólisis (Chepanoske, Golinelli et al. 2000). Resultados similares fueron obtenidos con la Uracil ADN glicosilasa de *Archaeoglobus fulgidus* (Engstrom, Partington et al. 2012).

Es importante mencionar que estos estudios se basan en las capacidades para reconocer e hidrolizar la adenina de la lesión 8oxoG:A, pero no toman en consideración las repercusiones estructurales del clúster [4Fe-4S]. Como

anteriormente se mencionó, el clúster [4Fe-4S] está físicamente alejado de la cavidad catalítica (19.01 Å), lo suficientemente alejado como para no interactuar con el sustrato. Sin embargo a pesar de esta lejanía, el ligando C198 de GsMutY (C192 en EcMutY) está a solo cinco aminoácidos del E192 (D186 en EcMutY), un aminoácido involucrado en la estabilización de la cavidad catalítica interactuando con los aminoácidos E188 y R26 (Fromme, Banerjee et al. 2004). Esto hace suponer, que sustituciones de los ligandos de cisteína puedan ocasionar una alteración del dominio FCL incidiendo a su vez en la estabilidad de la cavidad catalítica. Otro aspecto a discutir, es la sustitución de los aminoácidos positivamente cargados del dominio FLC por Gly y Ala hecha por Chepanoske y colaboradores (2000) ya que la Gly es alrededor de 3 veces más pequeño que la Arg y la Lys y la Ala es un aminoácido no polar, por lo que estas sustituciones también pudieron desestabilizar la cavidad catalítica y no necesariamente solo mermar la capacidad de reconocimiento del sustrato.

1.7. El Clúster [4Fe-4S] y transporte de carga del ADN.

Las glicosilasas de ADN llevan a cabo la primera defensa en contra de las lesiones en el ADN, identificando y removiendo las bases dañadas. Lo anterior significa que cargan sobre sus aminoácidos la tarea más complicada en el mantenimiento de la integridad del material genético. Primero, las bases modificadas o apareamientos no canónicos ocurren con baja frecuencia y están ocultos entre un mar de bases no dañadas (Lindahl 1993). Y segundo, algunas lesiones como la 8-oxoG:C generan alteraciones en la topología del ADN resultando un sustrato fácil de identificar para las glicosilasas, pero hay otras lesiones como la

8oxoG:A cuya topología es idéntica a la del ADN canónico (McAuley-Hecht, Leonard et al. 1994). Esto representa una complicación particular para MutY para identificar la lesión 8oxoG:A (**Figura 9**).

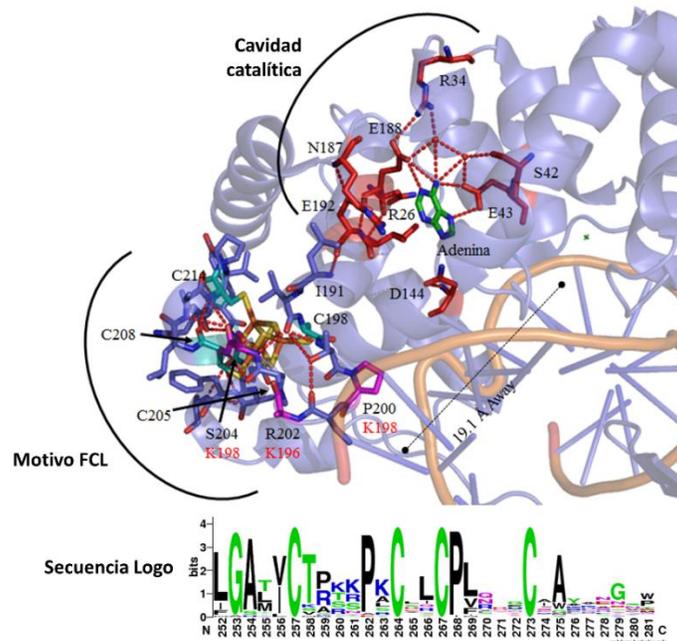


Figura 8. Relación entre el motivo del clúster [4Fe-4S] y la cavidad catalítica de MutY. Ubicación de los aminoácidos involucrados en la catálisis (Rojo) y pertenecientes al motivo FCL (Morado). En naranja con amarillo se resalta el clúster [4Fe-4S], en cian los ligandos de cisteínas y en rosa los aminoácidos del dominio FCL involucrados en el reconocimiento del sustrato. Los aminoácidos señalados en rojo corresponden a los de GsMutY. En verde se muestra la adenina. En la parte inferior se muestra la secuencia logo de la región de coordinación del clúster [4Fe-4S] de 88 secuencias de aminoácidos de MutY.

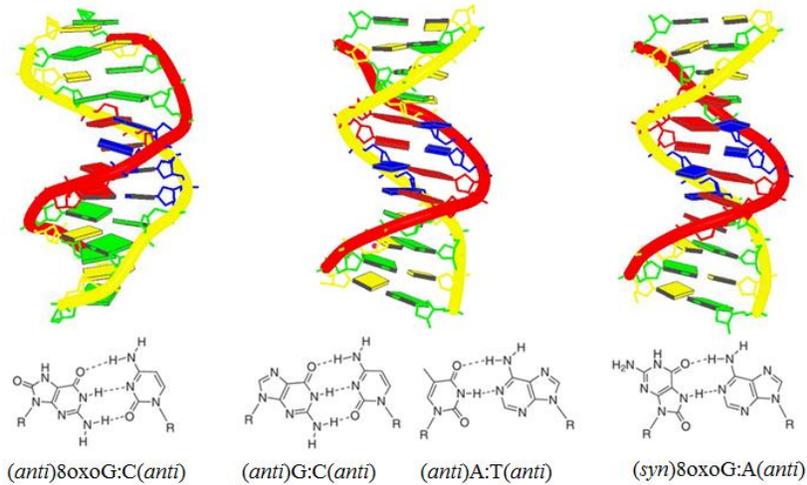


Figura 9. Estructura cristalográfica de ADN con lesión 8oxoG:C (izquierda), con apareamientos canónicos (medio) y con la lesión 8oxoG:A (derecha). McAuley-Hecht *et al*, 1994.

Para que la MutY reconozca su sustrato y empiece el proceso de reparación, hay dos mecanismos de búsqueda posibles. Uno de ellos es un mecanismo de búsqueda procesiva usando la difusión o desplazamiento a través de la hélice del ADN (Kolesov, Wunderlich *et al.* 2007) y la otra posible es una identificación puntual de la lesión para su rápida localización y reparación (Boal, Yavin *et al.* 2005). Como se puede entender, la primera requiere de una búsqueda constante y azarosa de la lesión, mientras que la segunda representa una opción más económica en tiempo y por ende más eficiente. Estudios han demostrado que la MutY usa ambos mecanismos para sobrellevar la complicada tarea de la identificación de 8oxoG:A (Boal, Yavin *et al.* 2005).

En primera instancia, en lo que concierne al mecanismo de búsqueda procesiva de MutY, se puede dividir en dos búsquedas independientes, la primera acoplada a la replicación donde se ha encontrado que la MutY de *Saccharomyces*

pombe y de humano interaccionan a través del asa o lazo interconector de dominios (IDCL por sus siglas en Inglés, 292-353 en *HsMutY* y 245-293 en *SpMutY*) con el antígeno nuclear de proliferación celular (PCNA por sus siglas en Inglés) (Parker, Gu et al. 2001, Chang and Lu 2002) y con la polimerasa de ADN λ (región 40-130) una polimerasa especializada en reparación del ADN. El PCNA es una homotrímero de forma toroidal que abraza la doble hélice del ADN y se desliza a través de él interactuando con proteínas como las polimerasas de ADN favoreciendo así su procesividad. A través del complejo MutY-PCNA-polimerasa, MutY puede desplazarse a través del ADN buscando lesiones 8oxoG:A para su reparación o también reparar los malos apareamientos G:A derivados de errores de la misma polimerasa de ADN. Por otro lado, también se sabe que la MutY interactúa con el complejo Rad9-Rad1-Hus1 (complejo 9-1-1) involucrado en el control del ciclo celular arrestándolo al presentarse daños oxidativo del ADN. En este contexto de arresto del ciclo celular donde la replicación se detiene, MutY se auxilia del complejo 9-1-1 para identificar su sustrato. El complejo 9-1-1 tiene una estructura similar a la del PCNA, permitiéndole abrazar la doble hélice del ADN en lugares donde hay bases dañadas. Evitando que esa región sea replicada a la vez que recluta enzimas para la reparación de las bases como en el caso de MutY. Estas interacciones (MutY-PCNA y MutY-Complejo 9-1-1) han sido identificadas *in vivo* usando técnicas de co-localización e *in vitro* (Chang and Lu 2002).

1.8. Localización puntual del sustrato por MutY y el clúster [4Fe-4S].

En la última década se ha descrito una característica sorprendente de la configuración de la doble hélice del ADN, refiriéndosele como un cable eléctrico con

la capacidad de transportar electrones a través de ella (Xie, Teramae et al. 2013). Esto se debe, principalmente, a los grupos amino de carga positiva de las bases nitrogenadas en el centro de la doble hélice y recubrimiento de fosfatos de carga negativa. Estudios de conductividad de la doble hélice del ADN han demostrado que esta capacidad de transporte de electrones (transporte de carga) a través de ella se puede dar *in vitro* a través de hasta 200 pares de bases y que puede ser perturbada por lesiones como la generación de sitios abásicos (Boal, Yavin et al. 2005, Xie, Teramae et al. 2013). El clúster [4Fe-4S] de EndoIII, MutY y UDG esta predominantemente en su estado de reducción 2+ en ausencias de ADN (potencial medio de +200 mV vs electrodo normalizado de hidrógeno; NHE) y no es fácilmente alterado por agentes reductores u oxidantes presentes en condiciones fisiológicas (Boon, Livingston et al. 2003). En presencia de ADN el clúster [4Fe-4S] de MutY tiene un potencial medio de +90 mV vs NHE lo que indica que el clúster [4Fe-4S]²⁺ se ha oxidado pasando a un estado 3+ (Gorodetsky, Boal et al. 2006). Aunado a esto, se ha encontrado que el estado oxidado [4Fe-4S]³⁺ es 3 veces más afín por el ADN que el estado reducido [4Fe-4S]²⁺ (Boal, Yavin et al. 2005).

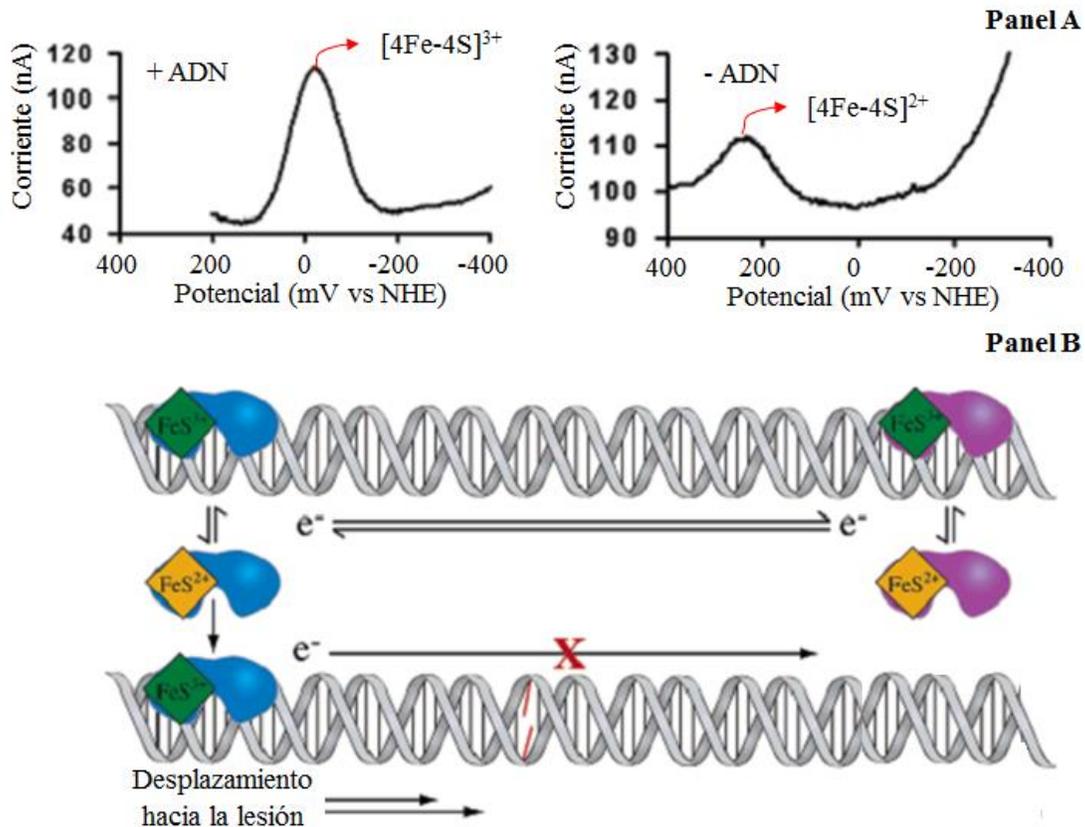


Figura 10. Arriba-izquierda se muestra los potenciales medios de MutY con ADN y a la Derecha sin ADN (vs NHE) Modificados de Gorodestky *et al*, 2006. Abajo se muestra un esquema del mecanismo de localización puntual de la lesión a través del transporte de carga del ADN y del clúster [4Fe-4S]. Modificado de Boal *et al*, 2005.

Estos datos juntos, con el conocimiento que la capacidad de transporte de carga se ve comprometida por lesiones en ADN, han permitido proponer un modelo de identificación puntual de las lesiones. Donde se estipula que hay una comunicación redox entre dos glicosilasas de ADN con [4Fe-4S]. Esta comunicación redox se da a través de la oxidación y reducción de dos clústeres $[4Fe-4S]^{2+/3+}$ de dos enzimas distantes, donde el electrón donado y aceptado pasa a través del ADN.

Esto genera que las enzimas se estén asociando y disociando del ADN al oxidarse y reducirse su clúster [4Fe-4S], respectivamente. Pero en el caso de presentarse una lesión, el transporte de electrones es truncado, por lo tanto, una de las enzimas queda permanentemente oxidada y unida al ADN, mientras que la segunda queda reducida y disociada del ADN. Esto genera que una de las enzimas se ubique a la proximidad de la lesión para su rápida localización y reparación.

1.9. Antecedentes Específicos.

García-Salcedo en sus tesis de maestría (2013) reportó la identificación del gen de MutY de *Entamoeba histolytica* (EhMutY) encontrando un gen de 924 pb que codifica para una proteína de 307 aminoácidos. Las MutY de procariontes son en promedio 150 aminoácidos más pequeñas que las de eucariota. La EhMutY es más similar en tamaño y secuencia a las de procariontes como *Clorobium ferrooxidans* y *Lactobacillus brevis* que a sus homólogos eucariontes. Una característica que llamó fuertemente la atención en nuestro grupo de investigación, es la carencia de los ligandos de cisteínas del clúster [4Fe-4S]. Lo mismo se encontró para las MutY de *Clorobium ferrooxidans* (CfMutY) y *Lactobacillus brevis* (LbMutY). Además, nuestro grupo encontró que la carencia de los ligandos típicos del clúster se debe a dos causas diferentes, en el caso de CfMutY al igual que en EhMutY la región completa de coordinación de clúster [4Fe-4S] ha sido eliminada y en el caso de LbMutY así como en otros organismos del género *Lactobacillus* los ligandos han sido sustituidos por aminoácidos como Met, Glu, Ser, Thr, Pro, Asn y Asp (García-Salcedo 2013).

Para entender más a profundidad la carencia de este importante cofactor, se realizó un modelado estructural por homología de LbMutY y EhMutY usando como

templado GsMutY y se encontró que los dominios N-terminal de estas enzimas son prácticamente igual con excepción que el dominio en GsMutY está constituido por 7 hélices- α , mientras que en *LbMutY* y *EhMutY* tienen 4 y 5 respectivamente. El dominio C-terminal de GsMutY está constituido principalmente por 7 láminas β y una hélice α , mientras que *LbMutY* tiene una hélice α más y en *EhMutY* muestra una configuración completamente diferente con solo una pequeña hélice α (**Figura 11**). Otras diferencias interesantes fueron encontradas en el aminoácido de la extrusión de la adenina hacia el sitio activo, viéndose que en GsMutY y *LbMutY* es una Tyr pero en *EhMutY* es una Phe. La Phe es un aminoácido similar a la Tyr careciendo un grupo hidroxilo, que también se ha visto involucrado en la actividad de extrusión en MutM (Sugahara, Mikawa et al. 2000). Como se esperaba, estas diferencias estructurales son congruentes con las diferencias en secuencia, ya que el dominio C-terminal es el más variable teniendo solamente un 13.2% de identidad. Por su parte el dominio N-terminal es bastante conservado en comparación con el C-terminal con un 40.4% de identidad. El motivo del clúster [4Fe-4S] es igual de variable que el dominio C-terminal, siendo la carencia de los ligandos de cisteína la característica más notoria de esta variabilidad.

Como ya se ha mencionado, un clúster [4Fe-4S] es coordinado por cuatro ligandos, donde al menos tres de ellos son cisteínas. Dado que *LbMutY* no contiene ni una cisteína y *EhMutY*, si bien tiene tres cisteínas, las mismas se encuentran lo suficientemente alejadas como para coordinar un clúster [4Fe-4S], por lo tanto, se puede inferir que carecen de este motivo. Con base en estos datos, se purificaron

la EhMutY y LbMutY y resultaron ser proteínas estables y activas sobre 8oxoG:A pese la carencia predicha del clúster [4Fe-4S].

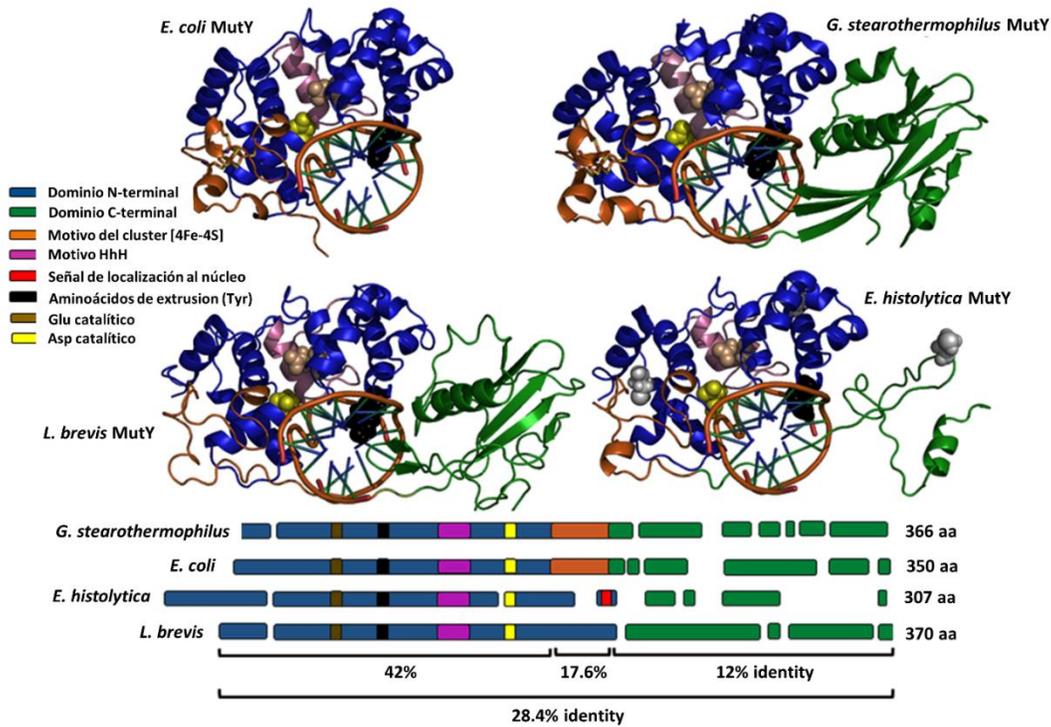


Figura 11. Estructura cristalográfica de GsMutY y EcMutY y modelados estructurales de MutY *E. histolytica* y *L. brevis*. Abajo se presenta el esquema del alineamiento resaltando los diferentes dominios y motivos. Modificado de Trasviña-Arenas *et al* (2016).

El hecho de que pueda haber ADN glicosilasas que carecen del clúster [4Fe-4S] abre nuevas preguntas, cuyas respuestas nos podrían ayudar a entender mejor la diversidad bioquímica-estructural de las enzimas involucradas en la reparación del ADN. Y sobre todo, entender su evolución nos permitirían entender como las MutY en particular, han podido solventar la carencia del clúster [4Fe-4S] para lidiar

con 8oxoG:A, mientras otras ADN glicosilasas de adenina son incapaces de prescindir de este cofactor.

2. OBJETIVO GENERAL.

Analizar evolutiva, bioquímica y estructuralmente los elementos pertenecientes a la ruta de reparación del ADN por la vía de Escisión de Bases de *Entamoeba histolytica*

2.1. Objetivos específicos.

2.1.1. Identificar los diferentes genes codificantes para los elementos de BER de *E. histolytica*.

2.1.2. Realizar análisis filogenéticos de las enzimas de BER para entender su origen evolutivo.

2.1.3. Generar modelos estructurales por homología para entender las adecuaciones estructurales y bioquímicas que puedan tener las glicosilasas de ADN de *E. histolytica*.

2.1.4. Caracterizar bioquímica la MutY de *E. histolytica* EhMutY.

2.1.4.1. Sobre expresar y purificar la MutY recombinante de *E. histolytica*.

2.1.4.2. Descartar la presencia del cofactor clúster [4Fe-4S] por métodos químicos.

2.1.4.3. Caracterizar bioquímicamente EhMutY y comparar el comportamiento bioquímico con la MutY canónica de *E. coli*.

2.1.4.4. Corroborar la funcionalidad *in vivo* de EhMutY.

2.1.4.5. Generar quimeras funcionales *in vivo* e *in vitro* de EcMutY-EhMutY.

2.1.4.6. Entender el mecanismo por el cual se prescinde del cofactor en EhMutY a través de los modelos estructurales.

3. MATERIALES Y MÉTODOS.

3.1. Identificación de genes de la Ruta de Escisión de bases de *E. histolytica*.

Para la identificación de genes pertenecientes a la Ruta de Escisión de Bases en el genoma de *Entamoeba histolytica* se hicieron búsquedas con el algoritmo de BLASTp en la base de datos de Amoeba Genomic Resources (<http://amoebadb.org/amoeba/>) (Aurrecoechea, Barreto et al. 2010). Los plantados usados fueron secuencias de aminoácidos de los ortólogos de la ruta de reparación pertenecientes de bacteria, arqueas y eucariontes. Los parámetros de búsqueda fueron un E-value de 10, Maximum description/Alignment (V=B) de 50, sin filtros de baja complejidad. Los ortólogos putativos encontrados fueron seleccionados con la consideración de tener un E-value menor a 1×10^{-3} y una cobertura de al menos 60%. Para corroborar la ortología se hizo el análisis dominios de las secuencias colectadas en la base de datos de Pfam de European Bioinformatics Institute (EMBL-EBI) (Bateman, Coin et al. 2004) con un E-value de 1.0. De la misma forma, la búsqueda de las enzimas fue hecha en los genomas de las especies de *Entamoeba nutalli*, *invadens* y *dispar*, así como en los protozoarios *Acanthamoeba castellanii* str. Neff y *Naegleria fowleri* ATCC 30863.

3.2. Análisis filogenéticos de la Ruta de Escisión de Bases.

Una vez identificados los genes de BER en cuestión, se colectaron homólogos pertenecientes a los tres dominios; arquea, bacteria y eucariota usando

el Algoritmo de BLASTp en el servicios del National Center for Biotechnology Information (NCBI). Los criterios de selección fueron un E-value menor a 1×10^{-3} y una cobertura de al menos 60%. Se procuró ser incluyente en cuanto a los phylums analizados. Para la construcción de las filogenias se partió de alineamientos de secuencia de aminoácidos múltiples con el algoritmo MUSCLE (Edgar 2004) con un número máximo de iteraciones de 8, en el software Geneious (Kearse, Moir et al. 2012) Una vez optimizado los alineamientos manualmente, se procedió a la construcción de los árboles filogenéticos en el software de Molecular Evolutionary Genetics Analysis (MEGA) versión 6.0 (Tamura, Stecher et al. 2013). El análisis filogenético fue realizado usando el algoritmo de Máxima verosimilitud (Guindon and Gascuel 2003) con un número de Bootstrap de 1000. Cada alineamiento fue sometido a un análisis de modelos de sustitución por Máxima verosimilitud. Los arboles derivados fueron editados para presentación final con el programa FigTree (Rambaut 2007).

3.3. Modelado estructural por homología.

Para entender las características estructurales de las enzimas de la ruta de BER de *E. histolytica* se construyeron modelos por homología en la suite de I-TASSER (Yang, Yan et al. 2015). Para esto se seleccionaron templados haciendo un Blastp contra la base de datos del Protein Data Bank (PDB) en el servidor del NCBI. I-TASSER para generar los modelos estructurales y su valoración genera alineamientos estructurales de miles de templados, valorando los diferentes templados en función del Z-score y la cobertura, que en término de iteraciones

representan el nivel de significancia estadística con la que se valora el templado y el porcentaje de residuos alineados entre el templado y la secuencia a estudiar. Valores de Z-score > 1 significan buen alineamiento. La suite genera un gran número de ensamblados de conformaciones estructurales y para seleccionar el mejor modelo se consideró el criterio de C-score que es calculado con base a la significancia de los alineamientos estructurales (Z-score) y la convergencia. Los C-score son típicamente reportados con valores de (-5 a 2) siendo el valor más alto el seleccionado. Los análisis estructurales finos fueron hechos en el software de PyMOL (DeLano 2002).

En particular, el modelado por homología de las MutYs fue hecho con el software Molecular Operating Environment (MOE) (ChemicalComputingGroup 2008). Como templado se usó la MutY de *G. stearothermophilus* en complejo con ADN de doble cadena con la lesión 8oxoG (número de acceso del PDB: IRRQ) (Fromme, Banerjee et al. 2004). La secuencia de aminoácidos de LbMutY y EhMutY (número de acceso del NCBI; WP_003554055.1 y XP_653060.1, respectivamente) fueron alineadas con la secuencia de MutY de *G. stearothermophilus*. Una serie de 30 modelos fueron construidos y los modelos con menos diferencias estructurales en comparaciones con el templado fueron seleccionados.

3.4. Identificación del Gen de MutY de *E. histolytica* y clonación.

El gen de la MutY de *E. histolytica* (EhMutY) fue identificado como anteriormente se menciona con base a la búsqueda en la base de datos de Amoeba Genomic Resources y su comparación en la Pfam usando modelos ocultos de Markov (HMMs por sus siglas en Inglés) (Eddy 1998). El resultado con mayor valor

correspondió al locus EHI_010700 y fue identificado como gen probable para la glicosilasa de Adenina. Debido a que la secuencia genómica contiene un intron, el ORF (marco de lectura abierto, por sus siglas en Inglés) fue ensamblado por SOEing (Empalme por Extensión por Superposición) PCR (Pogulis, Vallejo et al. 1996). El gen de MutY de *L. brevis* (LbY) fue identificado con el número de acceso del NCBI de WP_021741015. Este ORF bacteriano optimizado para su expresión en *E. coli* fue comprado de GeneScript (Piscataway, NJ, USA). Ambos ORFs fueron clonados en los vectores pCold y pET28b (modificado con sitio de corte de la proteasa PPS [PreScission Protease]). Los sitios de restricción para la clonación fueron de NdeI y BamHI para la LbMutY y BamHI y HindIII para EhMutY. Los vectores de pCold integran, entre otras características, un tracto de 6 histidinas en el extremo N-terminal, mientras que el pET28b puede incluir la misma cola en ambos extremos para pasos de purificación como cromatografía de afinidad a níquel. Los oligonucleótidos para esta parte experimental, así como la secuencia del gen sintético de LbMutY y los mencionados subsecuentemente son mostrados en **Apéndice 1**. La parte del Ensamblaje y clonación en el vector pCold fue hecha como parte de la Tesis de maestría de la M.C. Yarden Salcedo en el Laboratorio de Bioquímica Estructural del LANGEBIO.

3.5. Mutagénesis puntuales y ensamble de quimeras.

Las mutantes simples y dobles de los sitios activos correspondientes a EhMutY (E78A E179A) y LbMutY (E42A, D143A) fueron hechas de acuerdo al kit Quick Change Mutagenesis (Stratagene, La Jolla, CA, USA) usando como templado

los genes clonados de EhMutY y LbMutY en pCold con los oligonucleótidos especificados indicados en el **Apéndice 1**.

El diseño de quimeras (EcMutY-EhMutY y EcMutY-LbMutY) fue realizado de acuerdo al esquema mostrado en la **Figura 12**. Tres sets de quimeras fueron construidos fusionando parcial o totalmente el dominio catalítico de EcMutY y C-terminal de las variantes de EhMutY y LbMutY. La región de coordinación del clúster [4Fe-4S] fue excluida en los dos primeros sets de quimeras y complementada por la región correspondiente de EhMutY y LbMutY. Mientras que el tercer set de quimeras contenía el dominio catalítico completo de EcMutY y la región correspondiente del dominio C-terminal de EhMutY y LbMutY. Los genes de las quimeras fueron ensamblados por SOEing PCR. Todos los fragmentos para el ensamblaje de los genes fueron amplificados de los genes wild-type y clonados en el vector pCold1. Los genes para las quimeras EcY-EhY-2 y 3 fueron comprados como gBlocks (Integrated DNA Technologies, Coralville, IA, USA). Las secuencias de los gBlocks son mostradas en el **Apéndice 1**.

Los fragmentos utilizados para el ensamble de quimeras fueron amplificados independientemente usando los genes wild-type clonados en el vector pCold1 con los primers enlistados en el apéndice 1. La reacción de PCR se realizó de acuerdo a lo siguiente: 10 μ L de Solución amortiguadora phusion 10X, 4 μ L de dNTPs [2.5 mM], 2 oligonucleótido Rv/Fw [10 μ M], 4 μ L de MgSO₄ [25 mM], 2 μ L de ADN phusion, 5 μ L de plásmido, en una reacción de volumen final de 100 μ L. Las condiciones del PCR fueron las siguientes.

Paso	°C/min	Ciclos
1	95°C/5 min 95°C/1 min	x1
2	68°C/0.5 min 72°C/1.2 min	x35
3	72°C/5 min	x1
4	4°C/∞	x1

Los productos resultantes fueron purificados a partir de gel usando el kit GeneJet Gel Extraction (ThermoFisher). Para el ensamblaje de los productos resultantes y de los G-blocks para las quimeras Eh-EhY1 y 3 se realizó de la misma manera usando los oligos Nterm-EcYFw/Cterm-LbYRv para las quimeras Ec-LbY y Nterm-EcYFw/Cterm-EhYRv para las quimeras Ec-EhY. Los productos ensamblados fueron purificados, clonados en el vector pCold1 y secuenciados. Los resultados del ensamblaje de los genes quiméricos se muestran en el apéndice.

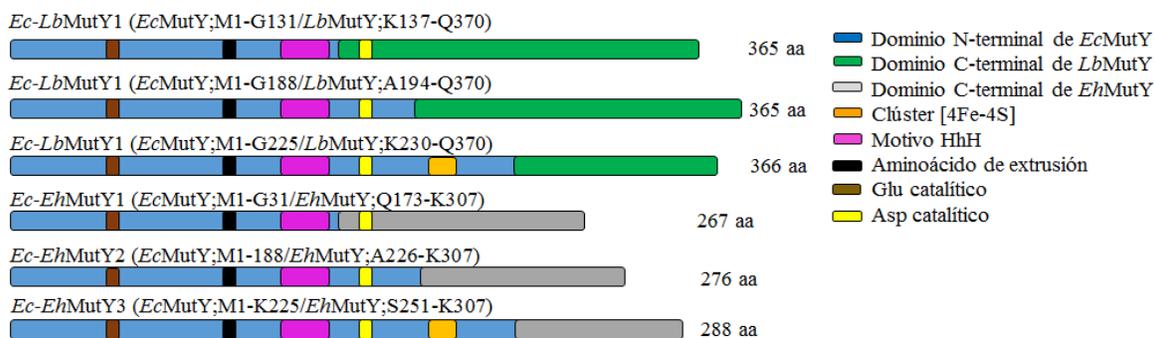


Figura 12. Esquema del diseño de las quimeras de MutYs.

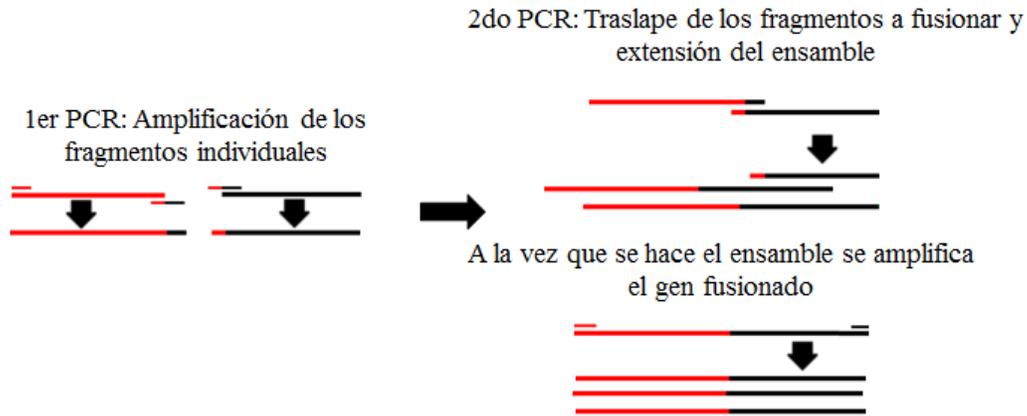


Figura 13. Esquema del PCR por SOEing utilizado para el ensamblaje de las quimeras.

3.6. Sobre expresión y Purificación de MutYs.

Para la expresión heteróloga de las variantes de MutY se usó la cepa de *E. coli* JM101 *muty::mini-tn10* cuyo gen de *mutY* está interrumpido por un transposón *tn10* que da resistencia a tetraciclina (Miller 1992). De esta forma evitamos contaminación endógena de la MutY de *E. coli* durante los ensayos bioquímicos. La cepa JM101 fue transformada con los plásmidos en cuestión por el protocolo convencional de choque-térmico. Para la expresión de cada construcción se inocularon 2 L de medio LB suplementado con 100 $\mu\text{g}\cdot\text{mL}^{-1}$ de ampicilina y 15 $\mu\text{g}\cdot\text{mL}$ de tetraciclina y se dejó crecer a una OD_{600} de 0.6 a 37 °C, para una vez alcanzada esta OD inducir con IPTG a una concentración final de 0.5 mM y se incubó durante 16 h a 15°C con agitación constante.

Una vez transcurrido el periodo de inducción, se colectaron células por centrifugación a 3000 rpm por 15 min a 4 °C. Después se resuspendieron con 30 mL de buffer lisis (NaCl 400 mM, Tris-HCl 15mM, Triton 0.1%; pH 8) y se le agregó

inhibidor de proteasas fluoruro de fenilmetilsulfonilo (PMSF, concentración final de 10 μM). El rompimiento celular se inició con la incubación de la solución celular con lisozima (1 mg mL⁻¹) por 40 min en hielo y por sonicación en el sonicador Branson Sonifier 450 (Branson Instruments, Co., Stamford Ca. USA) con 10 series de 20 pulso·s⁻¹ a una capacidad del 50%.

Después de una centrifugación a 15000 rpm por 45 min para remover el material celular no soluble, el sobrenadante fue cargado en una columna de 1 mL de HisTrap FF (GE Healthcare Life Science, Little Chalfont, UK) previamente cargada con Níquel y equilibrada con solución amortiguadora lisis. Los lavados fueron hechos con 50 mL de 15, 30 y 45 mM de imidazol (en solución amortiguadora lisis) y la proteína se eluyó con 10 mL de solución amortiguadora lisis con 500 mM de imidazol. La fracción de la elusión fue dializada en contra de solución amortiguadora de fosfatos (20 mM de fosfatos de sodio [pH 7.4], 75 mM NaCl, 1 mM DTT y 1 mM EDTA) durante toda la noche. La fracción dializada fue pasada a través de un segundo paso de purificación por cromatografía de intercambio catiónico en una columna de 1 mL HiTrapSP FF (GE Healthcare Life Science). La columna cargada con la proteína, previamente equilibrada con solución amortiguadora de fosfatos, fue lavada con 5 mL de 75, 100, 200, 300 mM NaCl y eluída con 500 y 1000 mM de NaCl (en solución amortiguadora de fosfatos). Las fracciones con las proteínas purificadas se juntaron y dializaron en solución amortiguadora de fosfatos con 150 mM NaCl y 50% de glicerol para cuantificarlas por espectrometría a 280 usando el coeficiente de extinción y por curva de BSA, para finalmente almacenarlas a -20 °C hasta su uso en los ensayos bioquímicos.

3.7. Identificación química de clúster [4Fe-4S]

La identificación del clúster [4Fe-4S] de las variantes de las MutYs fue realizado usando un enfoque quimioluminiscente semi-cuantitativo basado en la reacción de luminol con Fe en presencia de H₂O₂ (Specht 1937). El procedimiento de las reacciones quimioluminiscentes fue desarrollado de acuerdo a lo reportado (Högbom, Ericsson et al. 2005) con 30 y 150 μmol de enzima. La colecta de datos fue hecha en el lector de microplaca Infinite 1000 (Tecan, Maennedorf, Switzerland) en placa oscura de 96 pozos (Nunc, Roskilde, Denmark) con un tiempo de integración de 0.25 s. Los datos fueron reportados como Quimioluminiscencia relativa con respecto al blanco.

3.8. Preparación de sustratos para los ensayos de glucosidasas.

Para la evaluación de la actividad enzimática de las glicosilasas de ADN se usó un ensayo radioactivo, por lo que fue necesario realizar el marcaje, purificación e hibridación de los oligonucleótidos correspondientes. Para el marcaje un oligonucleótido de 45-mer conteniendo en la posición 30 cualquiera de las 4 bases nitrogenadas canónicas (A, C, T, ó G) fue marcado con ³²P en su extremo 5' con T4 Polinucleótido cinasa (New England Biolab) y purificado con Nucleotide Removal Kit (Quiagen, Valencia, CA, USA) de acuerdo con las instrucciones del proveedor. El oligonucleótido marcado fue hibridado con un exceso del 20% de oligonucleótido complementario conteniendo cualquiera de las bases canónicas o la lesión 8oxoG en la posición 30. Para la hibridación se incubaron los oligonucleótidos a 95 °C por 5 min y dejándolo enfriar lentamente hasta temperatura ambiente en solución

amortiguadora de hibridación (20 mM Tris-HCl [pH 7.4] 150 mM NaCl) Los oligonucleótidos usados se muestran en el apéndice 1.

3.9. Ensayos de actividad de glucosilasa.

Los ensayos enzimáticos de MutYs fue llevado a cabo usando diferentes concentraciones de sustrato de 10 a 1 nM en una solución de reacción (40 mM Tris-HCl [pH 7.5] 10 mM EDTA, 0.1 mg/mL BSA y 150 mM NaCl) a 37 o 21 °C. La concentración de enzima es especificada en la parte de resultados y discusión en cada apartado. Las reacciones fueron detenidas con 0.2 M de NaOH y calentadas a 90 °C por 5 min. Posteriormente, un volumen igual de solución amortiguadora de carga (99% formamida, 20 mM EDTA y 0.025% de azul de bromofenol) fue agregado a la reacción. Las muestras fueron resueltas por un gel desnaturizante (8 M Urea) de acrilamida (19:1) al 12 % y visualizadas en Storage phosphor autoradiography.

3.10. Determinación de las MutY como Monofuncionales o bifuncionales.

Las ADN glicosilasas según su actividad pueden ser divididas en monofuncionales si solo llevan a cabo la hidrólisis de la base nitrogenada o bifuncionales (Hidrolasa/Liasa) si llevan a cabo la escisión de la hebra con el sitio abásico concomitante a la actividad hidrolítica. Técnicamente, la actividad de las bifuncionales puede ser observada sin tratamiento con NaOH ya que ellas mismas llevan la escisión de la hebra con sitio abásico, mientras que en las monofuncionales no se observa actividad sin NaOH. Para esta determinación se llevaron a cabo una

reacción para cada enzima y se fueron tomando dos sets de alícuotas (10 μ L) para cada reacción por diferentes tiempos (5, 10, 20, 40, 80 min). Las reacciones de un set de alícuotas fueron detenidas con 0.2 M de NaOH y a 95 °C por 1 min (tratamiento 1), mientras que el otro set fue solamente con amortiguador de carga detenida a 95 °C por 1 min (tratamiento 2).

3.11. Identificación funcional del clúster [4Fe-4S].

Para descartar la presencia del clúster [4Fe-4S] de forma funcional se procedió a la construcción del cofactor *in vitro* de acuerdo con el protocolo clásico de David y colaboradores (Porello, Cannon et al. 1998) con ciertas modificaciones. El experimento en general se muestra esquemáticamente en la **Figura 31**. Primeramente, una solución de MutY a una concentración de 0.5 mg mL⁻¹ en solución amortiguadora A (50 mM Tris-HCl [pH 7.6], 200 mM NaCl, 0.1 mM 2-mercaptoetanol y 0.1 mM EDTA) fue desnaturalizado en presencia de 6 M Gu-HCl por al menos 1 h para liberar el cofactor. Subsecuentemente, el fierro fue quelado incubando la enzima en solución amortiguadora A con 6 M Gu-HCl y 5 mM EDTA por 6 h. Para el replegamiento de la enzima y reconstrucción del clúster [4Fe-4S], una muestra de la enzima fue renaturalizada en solución amortiguadora A con 0.2 mM de Sulfato de Amonio Ferroso (II) y 0.2 mM de sulfuro de amonio por 15 h. Estas muestras renaturalizadas con reconstrucción del cofactor fueron llamadas R^{Fe-S}, fueron dializadas extensivamente contra solución amortiguadora A durante 3 rounds de diálisis por 4 h cada round, para eliminar el exceso de iones en solución. Otra muestra de enzimas, llamadas R, fueron renaturalizadas en ausencia de fuentes Fe y S para evitar la reconstrucción de cofactor y dializadas extensivamente en solución

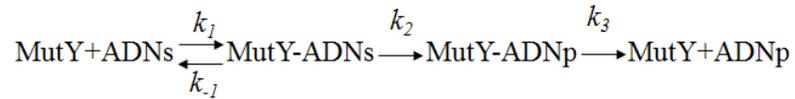
amortiguadora A. En paralelo, un control positivo para cada enzima, llamado Nd, fue preparado sin los pasos de desnaturalización, dializando contra solución amortiguadora A durante el mismo tiempo de los otros tratamientos. Todas las soluciones amortiguadoras usadas en estos experimentos fueron tratadas con Chelex-100 (Sigma, St. Louis, MO, USA) para eliminar la contaminación con metales. Al final, la actividad *in vitro* fue evaluada con 1 nM de duplex de ADN con la lesión 8oG:A y 500 nM de MutYs por 30 min a 37 °C. Adicionalmente, se realizó una detección cualitativa del cofactor por quimioluminiscencia usando 1.5 µM de enzima en placa, con solución de luminol y exposición a pantallas CL-Xposure (Thermo Scientific, Waltham, MA, USA).

3.12. Conceptualización de cinéticas de enfoque mínimo (no Michaeliano) para la determinación de constantes bioquímicas.

Para la determinación de las constantes catalíticas, primeramente, se realizó una titulación de los sitios activos para saber la concentración de enzima activa. Esto se hizo de acuerdo a Fersh (1998). En este ensayo se monitorea la actividad catalítica durante un periodo de tiempo usando una concentración de enzima igual o menor a la del sustrato para generar una producción rápida de producto. De esta forma se puede relacionar la concentración de enzima activa con la magnitud de la formación del producto.

Las glicosilasas como la MutY tienen una afinidad mayor por el producto de la reacción catalizada que por su sustrato. Por lo que es necesario definir dos constantes importantes en el proceso catalítico. Por lo que para el estudio

bioquímico de este tipo de enzimas no se puede proceder con un enfoque michaeliano, ya que la enzima se convierte en el reactivo limitante.



La constante de intercambio simple (k_2) es determinada en una cinética donde la concentración de sustrato es limitante (sustrato < enzima). Así se puede determinar la velocidad de solamente un ciclo catalítico (reconocimiento, catálisis y liberación del producto), es decir la velocidad de hidrólisis. Por otro lado, La constante de intercambio múltiple (k_3), donde la concentración de enzima es limitante (sustrato > enzima), nos permite entender la velocidad de ciclos catalíticos completos, especialmente la eficiencia de la liberación del producto. Para determinar la constante de afinidad de la MutY por su sustrato (k_1 también conocida como constante de disociación; k_d) se obtuvo por anisotropía. Esta técnica espectrofotométrica, se basa en que una molécula de ADN en solución gira a una determinada velocidad, pero al generar un complejo con una proteína, su tamaño aumenta y la rotación del complejo es más lenta que la molécula de ADN sola. Estas diferencias en rotación son proporcionales a la capacidad de polarización de la luz y a la afinidad. Las lecturas fueron tomadas a una longitud de onda de emisión y adsorción de 475 y 520 nm

3.13. Determinación de constantes k_2 y k_3 por cinéticas de intercambio múltiple y simple.

Para la determinación de las constantes cinéticas fueron usadas concentraciones de enzima activa. Para lo anterior, se hizo titulación de sitios activos por triplicado como ya se ha reportado (Porello, Leyes et al. 1998) usando una concentración absoluta de enzima de 2.5 nM y dúplex de ADN con la lesión 8oG:A de 10 nM a 21 °C.

Para la cinética de intercambio múltiple, 2.5 nM de enzima activa fueron incubadas con sustrato 8oG:A en solución amortiguadora de reacción. Se colectaron muestras de las reacciones a los 0.25, 0.5, 1, 2, 8, 15, 30, 60 y 120 min y se detuvieron con 0.2 M de NaOH y analizaron por Storage phosphor autoradiography como se describe anteriormente. Para la cinética de intercambio simple se usaron 10 nM de sustrato y 50 nM de enzima activa. Las concentraciones de los productos fueron cuantificadas usando el Software QuantityOne (Bio-Rad, Hércules, CA, USA). Las constantes de velocidad de la reacción enzimática (k_2) y de recambio (k_3) fueron calculadas analizando la producción de producto en función del tiempo en una regresión no-lineal usando el programa PRISM. Cada cinética fue evaluada por triplicado.

3.14. Determinación de constantes de afinidad ($[P_{1/2}]$).

La determinación de k_d 's fueron hechas por anisotropía con un dúplex de ADN de 45-mer, donde un oligonucleótido conteniendo un sitio abásico (Tetrahidrofunano, THF, Sigma-Aldrich) estaba marcado con fluoresceína en su

extremo 5'. Este fue hibridado como anteriormente se mencionó con otro oligonucleótido con la lesión 8oxoG apareando el sitio abásico La titulación del oligonucleótido fue hecho usando concentraciones crecientes de enzima activa que van desde 0.1-75 nM. Los experimentos de unión fueron conducidos como anteriormente se ha descrito (Owen and McMurray 2009). Los cambios en anisotropía fueron medidos usando un detector de intensidad de fluorescencia (Tecan Infinite M1000) con longitudes de onda de excitación y emisión de 495 y 520 nm, respectivamente. En caso de las mediciones con EcMutY las intensidades de fluorescencia decrecieron en función de la cantidad de enzima, por lo tanto los cálculos fueron ajustados para compensar el decrecimiento del “quantum-yield” del complejo proteína-sustrato relativos al sustrato, como se ha descrito anteriormente (Lee, Bai et al. 2004). Las curvas de anisotropía fueron ajustadas a modelos de unión específica-simple con ajuste de pendiente de Hill de acuerdo la siguiente fórmula:

$$Y = \frac{B_{\max} (X^h)}{[P]_{1/2}^h + X^h}$$

Donde X es la concentración de enzima, B_{\max} es valor máximo de unión específica, $[P]_{1/2}$ es la concentración de enzima necesitada para llevar a cabo la mitad de valor de unión en condiciones de equilibrio, Y es el valor de anisotropía ajustado a la pérdida del quantum-yield y h es el coeficiente de Hill. Las constantes fueron calculadas de isothermas de unión con el software PRISM, versión 6.07 (GraphPad, La Jolla, CA, USA). Todos los experimentos fueron realizados por triplicado.

3.15. Ensayos *in vivo*

En estos ensayo se usó la cepa de *E. coli* CC104 *muty::mini-tn10 mutm::mini-kan* cuyos genes de *muty* y *mutm* están interrumpidos por transposones. De esta forma la cepa presenta un fenotipo susceptible a lesiones 8oxoG. El fenotipo seleccionado para determinar la frecuencia de mutación (f) es la resistencia a rifampicina. Este antibiótico interactúa con la subunidad β de la RNA polimerasa deteniendo la transcripción y resultando letal para las bacterias. Cuando hay una alta frecuencia de mutación la posibilidad de que estas ocurran en región de interacción entre la RNA polimerasa y la rifampicina es alta. Si esto ocurre la configuración de la región de interacción es alterada, resultando en células resistentes. Por lo que al transformar la cepa CC104 carente parcialmente del sistema GO con una construcción con genes de MutY funcional se debe de reducir la frecuencia de mutación y la resistencia a rifampicina.

Para estos ensayos se transformó la cepa CC104 con los plásmidos pCold-EcMuty pCold-LbMuty, pCold-EhMuty y sus respectivas mutantes por choque térmico. Después se seleccionaron 8 colonias por cada gen a evaluar y se crecieron en 10 mL de medio LB y se dejó crecer a 37 °C durante 16 h. Posteriormente, se inocularon por duplicado 100 μ L de cultivo en una placa de LB-Agar suplementada con ampicilina y rifampicina y otras placas con 100 μ L de una dilución 1×10^{-7} en placas suplementadas solamente con ampicilina para determinar el número de células viables. Estas se dejaron crecer por 16 h a 37 °C. Finalmente se contaron las UFCs y la frecuencia de mutación fue determinada de la siguiente manera.

$$f = \frac{\text{mediana de las células crecidas en Rif}}{\text{promedio de células viables}}$$

4. RESULTADOS Y DISCUSIÓN

4.1. EVOLUCIÓN DE LA VÍA DE ESCISIÓN DE BASES EN *Entamoeba histolytica*.

Esta sección de resultados y discusión está enfocada a identificar los elementos presentes y faltantes de la ruta de BER de reparación del ADN de *Entamoeba histolytica*, así como entender el origen de los genes que componen dicha vía y sus características estructurales a través de modelados estructurales por homología con un especial enfoque en glicosilasas de ADN.

4.1.1. Análisis Evolutivo de BER de *E. histolytica*.

Durante su ciclo de vida *E. histolytica* está expuesta a diferentes entornos que amenazan la integridad de su información genética, como los subproductos metabólicos como ROS, RNS, las respuestas inmunológicas y la exposición a fármacos. Por lo tanto, este patógeno alberga vías de reparación de ADN altamente conservadas que reparan diferentes tipos de daños como BER, NER, reparación por recombinación homóloga (HRR), recombinación no homóloga (NHEJ) y reparación de malos apareamientos (MMR) (López-Camarillo, Lopez-Casamichana et al. 2009, Marchat, López-Casamichana et al. 2011). Lopez-Camarillo y colaboradores (2009) estudiaron los mecanismos de reparación del ADN de este patógeno, encontrando que la mayoría de sus elementos estaban presentes. Sin embargo, algunas piezas faltantes son descritas en la maquinaria de reparación. En HHR no se identificaron algunos parálogos de RAD51, HPR5 helicasa y el complejo 9-1-1. Las vías NHJE y NER carecen de KU80 y RAD14, respectivamente. Notablemente, esta revisión de las vías de reparación del ADN de *E. histolytica*

identificó a BER como la ruta más precaria dado que no se encontraron ortólogos de las enzimas MAG1, OGG1, MUS81, MMS4, APN1 y APN2.

La vía de BER inicia el proceso de reparación con el reconocimiento y eliminación de lesiones no voluminosas tales como oxidación, alquilación o deaminación de las bases. Este comienzo es realizado por una glicosilasa de ADN específica que después del reconocimiento de la lesión se encarga de hidrolizar el enlace N-glicosídico entre la base dañada y la deoxirribosa, dejando un sitio abásico (Ab). El sitio Ab puede ser procesado por la misma glicosilasa (en caso de una glicosilasa bifuncional con actividad hidrolasa/liasa) o una AP-endonucleasa que lleva a cabo la descomposición del enlace fosfodiéster próximo al extremo 5' del sitio Ab, produciendo un extremo 3'-OH y un producto 5'-dRP. Después de la producción del procesamiento de los sitios Ab, la ruta de BER se bifurca en dos subvías; subvía de parche corto y parche largo. El primero implica una polimerasa de ADN con actividad de liasa (ADN pol γ , β , θ , λ , ι) que elimina el producto 5'-dRP e incorpora la base correcta, para que después una ligasa de ADN genere la formación del enlace fosfodiéster entre el nucleótido recién incorporado y la cadena ya existente. El parche largo utiliza una polimerasa de ADN con capacidad de desplazamiento de hebra (ADN pol β , δ , ϵ). Esta actividad de desplazamiento de hebra produce una estructura 5'-Flap que se resuelve mediante una Flap-endonucleasa. Finalmente, una ligasa ADN sella la hebra de DNA (Fortini, Pascucci et al. 1998, Fortini, Parlanti et al. 1999, Fortini and Dogliotti 2007).

En nuestra búsqueda de miembros BER de *E. histolytica*, identificamos 18 componentes de esta vía reparadora (enlistados en la **Tabla 1**). Se pudo identificar

8 genes que codifican para glicosilasas de ADN como MutY, dos ortólogos de endonucleasa III, dos miembros de la superfamilia de UDG y tres ortólogos de la glicosilasa de 7-metil guanina. Este repertorio de glicosilasa cubre la mayoría de los requisitos de protección para BER, incluidos apareamientos erróneos, deaminación, oxidación y alquilación de bases. Anteriormente, se reportó la falta de un gen codificante para Ap-endonucleasa en el genoma de Entamoeba (López-Camarillo, Lopez-Casamichana et al. 2009). La falta de una Ap-endonucleasa es altamente perjudicial para la funcionalidad de BER ya que la mayoría de las glicosilasas son monofuncionales en este patógeno. Por lo tanto, el procesamiento del sitio Ab se basa solo en esta endonucleasa. Sin la actividad de la Ap-endonucleasa, la reparación del sitio Ab restante se canalizaría a otra vía de reparación como NER, o síntesis por translesión (Boiteux and Guillet 2004, Prakash, Johnson et al. 2005). Sin embargo, con nuestra metodología fuimos capaces de identificar en la base de datos de Pathema un gen que codifica para la Ap-endonucleasa (ID EHI_046670). Con este hallazgo, la mayoría de los miembros críticos para la función río-arriba de BER están presentes en este patógeno. Nuestro grupo de investigación también ha identificado 6 polimerasas de ADN pertenecientes a la familia A, B y Y, incluida la polimerasa α , δ , ζ , una polimerasa derivada de transposón y Rev1 (Pastor-Palacios, Azuara-Liceaga et al. 2010, Pastor-Palacios, López-Ramírez et al. 2012, Pastor-Palacios, López-Ramírez et al. 2015). Dos de estas polimerasas de ADN, una de la familia A y B2, mostraron actividad de translesión a través de los sitios de glicol de timina, 8oxoG y sitios Ab, así como la actividad de desplazamiento de hebra, fundamental para la vía secundaria de BER de parche largo. Curiosamente, ninguna de estas polimerasas ha sido reportada con actividad de liasa intrínseca, importante

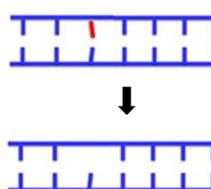
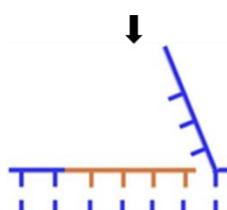
para el procesamiento del producto 5'-dRP durante la subvía del parche corto (Bebenek and Kunkel 2004). La falta de una polimerasa de ADN sin capacidad para eliminar los productos 5'-dRP sugiere fuertemente que toda la actividad de BER debe canalizarse a la subvía del parche largo. Esta noción de falta de la subvía de parche corto en *E. histolytica* está respaldada por la falta del gen XRCC1, un andamio funcional importante para la subvía de parche corto que coordina el proceso de reparación reclutando elementos de BER en el sitio Ab (Vidal, Boiteux et al. 2001, Petermann, Keil et al. 2006). Es de destacar que amibas relacionadas filogenéticamente con *E. histolytica* como *Acanthamoeba castellanii* y *Naegleria fowleri* albergan genes con Pathema ID de ACA_374140 y NF0032580 que codifican una proteína con un dominio N-terminal XRCC1 y un dominio C-terminal BRCT, sugiriendo que la vía de BER a través de un parche corto es ausente solamente en el género Entamoeba. Con respecto a la subvía del parche largo, *E. histolytica* alberga genes codificantes para una Flap-endonucleasa necesaria para procesar la estructura 5'-flap derivada de la actividad de desplazamiento de hebra de la ADN polimerasa. Además, de un gen codificante para una ligasa de ADN de la familia I única que genera la formación de enlaces fosfodiéster en el ADN. Así mismo contiene PCNA, un homotrímero que promueve el desplazamiento de la cadena durante el parche largo de BER (Levin, McKenna et al. 2000, Cardona-Felix, Lara-Gonzalez et al. 2011).

Además de la falta de polimerasas de ADN sin actividad liasa sobre 5'-dRP, y el andamio funcional XRCC1, se ha reportado que el genoma de *E. histolytica* carece de genes MAG1 y OGG1 (López-Camarillo, Lopez-Casamichana et al. 2009). El gen

MAG1 codifica para una glicosilasa de 3-metil-adenina, mientras que OGG1 es una glicosilasa de 8oxoG. La repercusión biológica de la falta de OGG1 o su análogo bacteriano MutM se discutirá más adelante. La glicosilasa Mag1 es activa sobre 3-metil adenina, 3-metil guanina y 7-metil guanina (Alseth, Osman et al. 2005). No obstante, pudimos identificar tres ortólogos de AlkD en este patógeno, que codifican para una glicosilasa de metil-purinas con la misma especificidad que Mag1 pero con diferente eficiencia catalítica (Alseth, Rognes et al. 2006). Por lo tanto, los ortólogos de AlkD soportan la pérdida de MAG1 en *E. histolytica* adecuadamente. Las predicciones bioquímico-estructurales y evolutivas de las AlkD de este patógeno se abordarán posteriormente. Otros elementos de BER que no se identificaron en este estudio se enumeran en la Tabla 2. La mayoría de las piezas faltantes de la ruta de BER en *E. histolytica* son ADN glicosilasas como MPG, SMUG, TDG y MBD4. Esta última no es una glicosilasa *per se*, sino una proteína con un dominio N-terminal que se asocia a las islas CpG metiladas y un dominio C-terminal con actividad glicosilasa de 5-metil citosina (Hendrich, Hardeland et al. 1999, Zhu, Zheng et al. 2000). Este gen se encontró en el genoma de *A. castellanii*, pero no en el género Entamoeba. La parte de glicosilasa de MBD4 actúa sobre malos apareamientos G/T, 5-metil citosina y deaminación de citosina. Dado que la 5-metil citosina es una marca epigenética, el papel de MBD4 es fundamental en eucariotas en la regulación transcripcional y la estabilidad del genoma (Sjolund, Senejani et al. 2013). La importancia de la metilación dentro y fuera de los sitios CpG en *E. histolytica* se ha estudiado previamente, demostrando la presencia de la maquinaria necesaria para la metilación de la citosina (Lavi, Isakov et al. 2006). Dado que la actividad epigenética a través de la interacción del genoma de metilación/demetilación tiene

papeles determinantes en la patogénesis y el ciclo de vida de este organismo, la falta de MBD4 resulta desconcertante. Y esto se ve agravado por la falta de TDG, otra glicosilasa de ADN que modula la actividad de DNMT3 inhibiendo la metilación de la citosina (Sjolund, Senejani et al. 2013). En *Entamoeba* debe de estar presente una vía de modulación de metilación independiente de MBD4/TDG o una versión altamente divergente de estas enzimas está actuando, que por lo tanto complica su identificación en nuestro estudio. Otra carencia es la glicosilasa SMUG que pertenece a la familia III de la superfamilia UDG. Esta glicosilasa de uracilo, como su nombre indica, actúa sobre uracilo en ADN de doble cadena y 5-hidroxi-metil uracilo (Wibley, Waters et al. 2003). No obstante, dos UDG pertenecientes a la familia I y VI pueden compensar esta pérdida en *Entamoeba*.

Tabla 1. Identificación de genes de BER en *E. histolytica*.

Gen	Nombre funcional	Sustratos	Pathema ID	Gene Bank ID	Paso de reparación	
MuT	Fosfatasa de 8-Oxo-guanine	8oGTP, 2-OH-ATP, 8-OH-ATP	EHI_054150	XP_651412.1	B	
MutY	glicosilasa de adenina	A mal apareada con G o 8oG	EHI_010700	XP_653060.1	B	<p>I) Reconocimiento y remoción de la lesión</p> 
Ntg1 Ntg2	Endonucleasa III	Tg, Ug, 5-6-OHT, β-UBA, 5-OH-6-HT, 5-OH-6-HU, 5,6-HU, 5-OHC, 5-OHU, Urea, Sitio Ab	EHI_118790 EHI_083460	XP_654116.1 XP_655025.1	E	
UNG Familia I	glicosilasa de uracilo	U, ssU, 5-FU	EHI_178910	XP_656294.1	E	
UDG Familia VI	glicosilasa de hipoxantina	U, ssU, I, 8oG	EHI_198640	XP_655177.1	B	
AlkD1 AlkD2 AlkD3	glicosilasa de 7-methyl Guanina	3-mA, 3-mG, 7-mG	EHI_087400 EHI_002000 EHI_141100	XP_651367.1 XP_653223.1 XP_654596.1	B	
Apn1	Endonucleasa Apurínica/Apirimidínica (AP)-1	Sitio Ab, 3-dRP	EHI_046670	XP_650532.2	E	<p>II) Procesamiento del sitio Ab</p> 
PCNA	Sliding clamp		EHI_128450	XP_651510.1	E	<p>III) Subvía del parche largo; Actividad de desplazamiento de hebra</p> 
ADNPol A [A]	Pol de traslesión		EHI_073640	XP_653960.2	E	
ADNPol α [B]	Pol Replicativa		EHI_151520	XP_657373.1	E	
ADNPol δ [B]	Pol Replicativa		EHI_006690	XP_654477.1	E	
ADNPol ζ [B]	Pol de traslesión	ADN de doble cadena con extremo 3'-OH	EHI_068010	XP_656768	E	
PolB2 [B]	Elemento Transposable		EHI_018010	XP_001914292.1	E	
Rev1 [Y]	Pol de traslesión		EHI_053480	XP_654241.1	E	

FEN-1	Flap Endonucleasa- 1	Estructura 5'-flap	EHI_099740	XP_651270.1	E	IV) Procesamiento de la estructura 5'- Flap y ligación del ADN
Ligasa de ADN	Ligasa I	ADN no ligado	EHI_111060	XP_657595.1	E	

Abreviaciones: 8oGTP, 8-oxo-7,8-dihidro-deoxiguanosina trifosfato; 2-OH-ATP, 2-hidoxi-deoxiadenosina trifosfato; 8-OH-ATP, 8-hidoxi-deoxiadenosine trifosfato; A, deoxiadenosina; G, deoxiguanosina; 8oG, 8-oxo-7,8-dihidro-deoxiguanosina; Tg, deoxitimidin glicol; Ug, deoxiuridin glicol; 5-6-OHT, 5,6-dihidro-deoxitimidina; 5-6-OHU, 5,6-dihidro-deoxiuridina ;β-UBA, ácido β-ureidoisobutirico; 5-OH-6-HT, 5-hidroxi-6-hidro-deoxitimidina; 5-OH-6HU, 5-hidroxi-6-hidro-deoxiuridina; 5-OHC, 5-hidroxi-deoxicitosina; 5-OHU, 5-hidroxi-deoxiuracil; Ab site, sitio abásico; ssU, ADN de cadena sencilla con deoxiuridina; 5-FU, 5-fluoro-deoxiuridina; I, Inosina; 3-mA, 3-metil-deoxiadenosina; 3-mG, 3-metil-deoxiguanosina; 7-mG, 7-metil-deoxyguanosina; 3'-dRP, ADN con extremo 3'-α,β-aldehído insaturado; E, Eucariote; B, Bacteria.

Tabla 2. Elementos faltantes de BER en *E. histolytica*.

Proteína	Función
OGG1/MutM	ADN glicosilasa bifuncional con actividad sobre 2,6-diamino-4-hidroxi-5-N-metilformamidopirimidina (FAPY-guanina), 4,6-diamino-5-formamidopirimidina (FAPY-adenina) y 8oG
Neil (EndoVIII)	Endonucleasa con actividad β- y δ-liase sobre glicol de timina, dihidrotimina, β-UBA y Urea.
Glicosilasa de 3-metil-purina (MPG)	MPG remueve una amplia gama de alquilaciones en el ADN incluyendo 3-metil-A, 7-metil-G y 3-metil-G.
SMUG (Family III)	Glicosilasas de uracilo activa sobre uracilo en ADN de doble y cadena sencilla.
TDG (MUG/TDG-Familia II)	Glicosilasa timina a cargo de la remoción de malos apareamientos T/G, C y T.
Proteína con dominio de unión a CpG metilado 4 (MBD4)	Proteína con dominio de unión a Metil-CpG y con dominio C-terminal con actividad de ADN glicosilasa involucrado en la remoción de la deaminación de 5-metil-citosina and citosina en sitios CpG.
ADN Pol γ, β, λ, ι, θ (con actividad de liasa 5'-dRP)	ADN polimerasas con capacidad de Liasa sobre el producto 5'-dRP e incorporación de nucleótido por la subvía del parche corto de BER.
XRCC1	Andamio funcional de la vía del parche corto de BER, con la capacidad de reclutar ADN ligasa y ADNPolβ.

4.1.2. El sistema GO de *E. histolytica*; una pieza perdida y dos extrañas.

La oxidación de guanina a 8oxoG por ROS es una de las lesiones de ADN que ocurre con más frecuencia en la célula, $\approx 90,000$ lesiones/células/día (Fraga, Shigenaga et al. 1990). El daño potencial de 8oxoG se observa durante un evento replicativo donde una adenina en lugar de una citosina se puede incorporar ante 8oxoG, comenzando una transversión G:C \rightarrow T:A (Shibutani, Takeshita et al. 1991, Brieba, Eichman et al. 2004). Dada la importancia de la reparación de esta lesión, existe una maquinaria de reparación especializada denominada sistema GO, que incluye las enzimas MutT, MutM y MutY (Michaels and Miller 1992). En la **Figura 14** se muestra un esquema del sistema GO. MutT no es un miembro de BER pero es importante para la desintoxicación de 8oxoG, ya que es una 8oxodGTPasa a cargo de hidrolizar dGTP oxidados evitando su incorporación durante la replicación (Ito, Hayakawa et al. 2005). Si se produce la incorporación de una oxidación de 8oxoG, MutM (cuyo ortólogo eucarionte es OGG1) es una glicosilasa de ADN bifuncional encargada de eliminar 8oxoG dejando un 5'-fosfato y 3'- α, β -aldehído insaturado (producto 3'-dRP) que se procesa vía BER. En caso de que la remoción de 8oxoG no se lleve a cabo antes de la replicación, y una adenina se aparee incorrectamente con 8oxoG, MutY una glicosilasa de adenina (monofuncional) elimina la adenina mal apareada, dejando un sitio Ab. Posteriormente, la vía BER restaura la lesión inicial 8oxoG:C para finalmente ser reparada mediante MutM/BER.

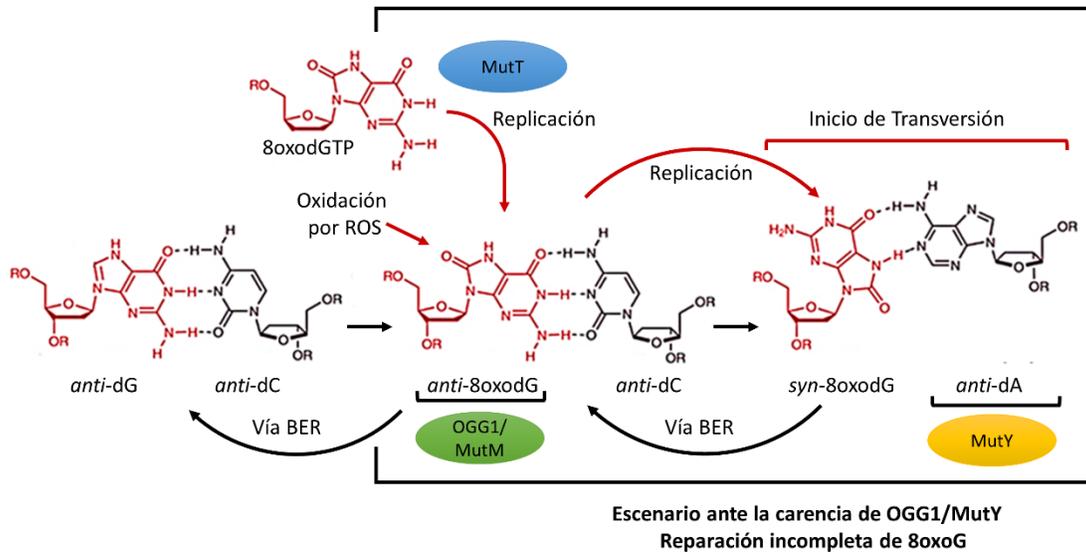


Figura 14. Representación esquemática del sistema GO de *E. histolytica*. 8oxoG es generado por la oxidación de guanina por ROS. La conformación de bases canónicas es *anti*, pero cuando hay una lesión de 8oxoG ocurre un cambio de conformación de *anti*→*syn* generando que durante un evento replicativo las ADN polimerasas incorporen una dA opuesta a 8oxoG comenzando una transversión. El sistema GO está compuesto por MutT, MutM y MutY. El primero no pertenece a la vía BER pero es una 8oxodGTPasa a cargo de hidrolizar dGTP oxidado evitando su incorporación durante un evento replicativo. MutM es una glicosilasa de 8oxoG (bifuncional) que elimina 8oxoG dejando el producto 3'-dRP que se repara completamente a través de la vía BER. Cuando la lesión de 8oxoG no se repara mediante MutM y se genera el mal apareamiento 8oxoG:dA, MutY (monofuncional) elimina la dA mal incorporada dejando un sitio Ab. El sitio Ab se procesa mediante la ruta BER restaurando la lesión inicial 8oxodG:dC que finalmente se repara a través de la vía MutM/BER. *E. histolytica* carece del gen MutM, por lo que las lesiones relacionadas con 8oxoG están atrapadas en un ciclo 8oxodG:dC ↔ 8oxodG:dA en el caso de una adecuada función de MutY. Este escenario de carencia de OGG1/MutM en *E. histolytica* (marco negro) produce una predisposición genómica para un bajo contenido de GC en el genoma.

Como se mencionó anteriormente, la ruta de BER en *E. histolytica* no tiene genes OGG1/MutM. Esta glicosilasa de ADN es determinante para evitar las transversiones dependientes de 8oxoG ya que elimina las lesiones de novo 8oxoG:C o derivadas de la actividad MutY/BER. La falta de las glicosilasas OGG1/MutM da como resultado una tarea de reparación incompleta, haciendo que la lesión oscile de 8oxoG:C↔8oxoG:A, en el caso de una actividad eficiente de la vía MutY/BER. Este hecho genera en *E. histolytica* una predisposición a tener un genoma con un alto contenido en AT. De hecho, su genoma es característico por un bajo contenido de GC (24.1%). Otras especies de Entamoeba como *E. invadens*, *E. nutalli* y *E. dispar* también carecen de OGG1/MutM teniendo un contenido de GC similar en sus genomas. En la **Tabla 3** se muestra un análisis comparativo de la relación entre la presencia de MutM/OGG1/MutY y el contenido de GC. En las bacterias, la pérdida de cualquier glicosilasa del sistema GO está relacionada con un bajo contenido de GC, que varía entre el 17 y el 34%, mientras que el sistema GO intacto produce un contenido de GC del 43-67% (Lind and Andersson 2008). La falta del gen MBD4 en *E. histolytica* podría ser otra causa que puede abonar al bajo contenido genómico de GC, ya que su inactivación está relacionada con transversiones C→T (Wong, Yang et al. 2002). El agente causante de la malaria *Plasmodium falciparum* es otro protista con contenido marginal de GC (23%). Sin embargo, este genoma patógeno tiene genes codificantes para OGG1 y MutY (PlasmoDB ID; PF3D7_0917100 y PF3D7_1129500, respectivamente). Por lo tanto, la razón del bajo contenido de GC en el genoma de *P. falciparum* debe ser una causa independiente de OGG1/MutY.

Con respecto a MutY de *E. histolytica*, solo se mencionarán algunas características de esta particular glicosilasa de ADN, ya que posteriormente se abundará en los detalles bioquímico-estructurales, así como en su evolución. Las características más destacadas de la MutY en este organismo fue la falta del clúster [4Fe-4S]. Este cofactor ha sido descrito como un componente indispensable para la actividad de MutY, independientemente de no participar en su catálisis (Porello, Cannon et al. 1998, Golinelli, Chmiel et al. 1999, Messick, Chmiel et al. 2002). La función del motivo del clúster [4Fe-4S] está asociada a procesos de búsqueda/reconocimiento de lesiones aprovechando la propiedad de transporte de carga del ADN de una manera redox-dependiente (Boon, Livingston et al. 2003, Boal, Yavin et al. 2005). Esta MutY sin clúster fue activa sobre A:G y A:8oxoG con un comportamiento cinético e *in vivo* similar a la MutY de *E. coli* que contiene clúster [4Fe-4S]. De acuerdo con un modelo estructural por homología el mecanismo en que la MutY de *E. histolytica* prescinde de este cofactor es a través de un efecto estérico en el que la cavidad dejada por la ausencia del clúster [4Fe-4S] ($\approx 550 \text{ \AA}^3$) es suplida con residuos grandes como Phe, Gln y Lys. Además, el origen de esta MutY es interesante, ya que el análisis filogenético sugiere que la MutY de este patógeno fue adquirida por LGT a partir de una bacteria anaeróbica (Trasviña-Arenas, Lopez-Castillo et al. 2016).

Tabla 3. Relación entre la presencia de OGG1/MutM/MutY y en el contenido de GC en los genomas de diferentes organismos.

Organismo	Contenido de GC (%)	OGG1/ MutM	MutY
<i>Homo sapiens</i>	45	+	+
<i>Saccharomyces cerevisiae</i>	38	+	+
<i>Arabidopsis thaliana</i>	36	+	+
<i>Entamoeba histolytica</i>	24	-	+
<i>Entamoeba dispar</i>	23	-	+
<i>Entamoeba invadens</i>	29	-	+
<i>Plasmodium falciparum</i>	23	+	+
<i>Escherichia coli</i>	50	+	+
<i>Bacillus subtilis</i>	43	+	+
<i>Deinococcus radiodurans</i>	67	+	+
<i>Streptomyces coelicolor</i>	72	+	+
<i>Carsonella ruddi</i>	17	-	-
<i>Wigglesworthia glossinidia</i>	22	-	-
<i>Borrelia burgdorferi</i>	28	-	-
<i>Fusobacterium nucleatum</i>	27	-	-
<i>Rickettsia prowazekii</i>	29	-	-
<i>Buchnera aphidicola</i>	22	-	+
<i>Blochmannia floridanus</i>	27	-	+
<i>Mycoplasma genitalium</i>	32	+	-
<i>Ehrlichia ruminantium</i>	28	+	-
<i>Baumannia cicadellinicola</i>	33	+	-
<i>Wolbachia pipientis</i>	34	+	-

MutT no es miembro de la vía BER pero pertenece al sistema GO. Es una 8oxodGTPasa a cargo de la remoción de dGTP oxidados del pool de dNTP. Su estructura terciaria y mecanismo catalítico están bien estudiados. El plegamiento MutT es un sándwich α - β - α . En la **Figura 15** se muestra un análisis estructural comparativo entre el modelo por homología de la MutT de *E. histolytica* y la estructura resuelta de la MutT de *E. coli*. La reacción de hidrólisis es la siguiente: $8odGTP+H_2O \rightarrow 8oxodGMP+PPi$. El mecanismo catalítico que conlleva tal reacción posee algunas características interesantes. En primer lugar, a pesar de que la

principal característica química es la oxidación la guanina en el C8, esta modificación no contribuye a la red de puentes de hidrógeno para el reconocimiento del sustrato, sino la conformación *syn* de 8oxodGTP es la característica clave para la discriminación de sustratos (Bessman, Frick et al. 1996, Nakamura, Meshitsuka et al. 2010). Otro aspecto catalítico a destacar es la reacción *per se*, ya que la hidrólisis se lleva a cabo mediante una sustitución nucleofílica con un ataque poco frecuente del β -fósforo del 8oxodGTP, dando como resultado un pirofosfato inorgánico y 8oxodGMP (Weber, Abeygunawardana et al. 1993). El sitio activo implica una red compleja de enlaces de hidrógeno que se estructura de forma sustrato-dependiente. El plegado sándwich α - β - α implica la conexión de varias asas. En especial el asa 1 es la involucrada en la catálisis albergando la mayoría de los elementos catalíticos como K39, implicado en la liberación del producto, F35 y E34 que participan como andamios importantes para las interacciones polares y de van der Waals (Nakamura, Meshitsuka et al. 2010). E57 está posicionado en un motivo muy conservado en MutTs llamado firma MutT. Este Glu se representa como un activador de la molécula de agua para el ataque nucleofílico, asistido por Mn^{+2} . El papel de los metales en la catálisis se ha descrito como crucial, ya que permite el desplazamiento nucleofílico desfavorable en el β -fósforo a través de la quelación de metales tanto por el β -fosforo como por el γ -fósforo del 8oxodGTP. Otro protagonista en la catálisis es K39 que facilita la salida de 8oxodGMP (Harris, Wu et al. 2000), que de acuerdo con a las características físico-químicas de K39 y el producto debe ser conducida por efecto electrostático.

El modelo por homología estructural de la MutT de *E. histolytica* (EhMutT) muestra que se conserva el plegamiento sándwich α - β - α . La firma MutT y la α -hélice correspondiente están altamente conservadas, así como también el asa 1 involucrada en la catálisis. Un aspecto particular para destacar es la presencia de un segmento no estructurado en la región N-terminal de 57 residuos. Esta extensión se conserva en el género *Entamoeba* y *Acanthamoeba*. La posición física de esta extensión no obstaculiza el sitio activo y tampoco lo hace en otras regiones. De hecho, la organización de esta región N-terminal no estructurada es independiente del resto de la proteína. Para entender la importancia de esta extensión en EhMutT se deben realizar estudios funcionales. El papel de las proteínas o dominios intrínsecamente no estructurados es amplio. Estudios estructurales en glicosilasas como MutY (Luncsford, Chang et al. 2010), NEIL (Doublíé, Bandaru et al. 2004) y TDG (Smet-Nocca, Wieruszkeski et al. 2008) han demostrado la presencia de regiones no estructuradas que están relacionadas con la interacción proteína-proteína que modula sus funciones. La naturaleza no estructurada de estas regiones permite un andamio versátil para las interacciones cuyo plegamiento depende de la proteína con la que está interactuando. Así mismo, el resto no estructurado en EhMutT podría funcionar como un modulador de actividad a través de un andamio para interacciones proteína-proteína. En lo que respecta a la catálisis, la mayoría de los residuos implicados en el reconocimiento del sustrato se conservan al menos en sus características físico-químicas, así como E59 catalítico (E90 en EhMutT). F35 se ha descrito como un residuo importante para la interacción con el sustrato a través de contactos no polares (Nakamura, Meshitsuka et al. 2010). En esa posición EhMutT tiene una Leu (L72). Aunque F35 no se conserva, el Leu puede suministrar

el entorno no polar para la interacción van der Waals. La H28 en EcMutT participa en el reconocimiento del sustrato promoviendo un enlace de hidrógeno entre fracción 3'-OH de la desoxirribosa y el imidazol. H28 es un cambio drástico en EhMutT, ya que presenta una Glu (E64) en su lugar. El cambio de His a Glu es químicamente contradictorio, ya que el grupo carboxilo de Glu no podría interactuar con la fracción 3'-OH de la desoxirribosa. Sin embargo, para comprender el alcance real de esta sustitución, se deben realizar estudios bioquímicos. La principal diferencia entre MutT de *E. coli* y de *E. histolytica* se puede encontrar en la K39 que está a cargo de la liberación del producto, posiblemente de forma electrostática. En esa posición, EhMutT tiene una Phe (posición 76), lo que sugiere que el producto podría ser liberado de una manera diferente, tal vez de manera estérica/mecánica y/o hidrofóbica.

Para entender la evolución de la MutT de *Entamoeba*, se realizó un análisis filogenético con secuencia de aminoácidos de MutT de los tres dominios utilizando el algoritmo de máxima verosimilitud. Todos los MutT del género de *Entamoeba* se agruparon en un clado único que incluye otros MutT del phylum bacteroidetes. Bacteroidetes ha sido el phylum que más contribución genética ha tenido a los genomas de *Entamoeba* a través del mecanismo de LGT (Alsmark, Sicheritz-Ponten et al. 2009, Romero, Cerritos et al. 2016). Este clado está altamente soportado con un Bootstrap de 92.5%. Este agrupamiento filogenético sugiere fuertemente que el gen EhMutT fue adquirido por LGT a partir de un ancestro bacteriodete. Además, todos los genes de MutT bacterianos de este clado son bacteroidetes anaeróbicos/microaerofílicos y patógenos (o se encuentran por lo menos en

muestras clínicas). Este contexto cladístico sugiere que el ancestro bacteriano donde se produjo el intercambio génico tenía la capacidad de adaptarse al entorno anaeróbico y la maquinaria necesaria para la patogenicidad. Además, que la adquisición del gen MutT era anterior a la especiación de *Entamoeba*. De hecho, los estudios sobre la ocurrencia-temporal de los eventos de LGT sugieren que la mayoría de los eventos de transferencia genética fueron previos a la especiación del género *Entamoeba* (Romero, Cerritos et al. 2016).

indican residuos implicados en el reconocimiento/estabilización del sustrato y la catálisis, respectivamente. También se muestra una representación esquemática de la estructura secundaria y la secuencia de consenso de la firma MutT.

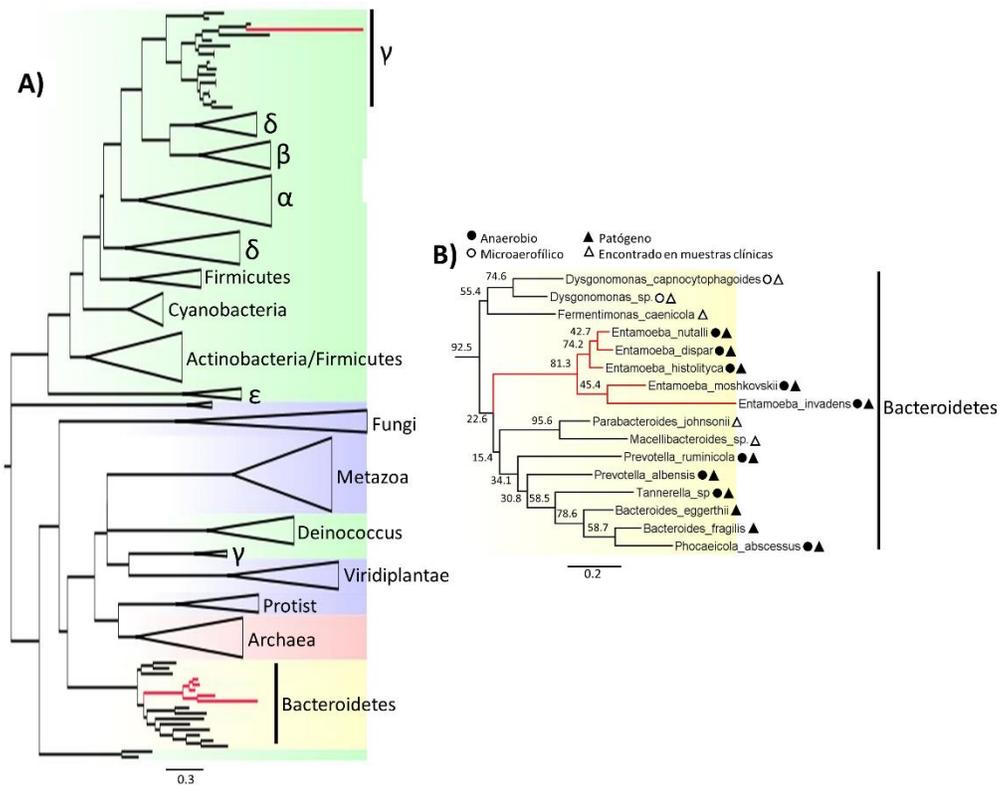


Figura 16. Análisis filogenético MutT. A) Análisis evolutivo de 128 secuencias de aminoácidos de MutT incluyendo Arquea (clado rojos), Bacteria (clados verdes) y Eucarionte (clados azules). El análisis filogenético se realizó con el algoritmo de máxima verosimilitud con 1000 réplicas de Bootstrap y LG+G como modelo de sustitución. En líneas rojas, se muestran los eventos LGT. La línea roja en clado γ representa la MutT de *Dictyostelium discoideum*. B) Vista ampliada del clado de Bacteroidetes que incluye MutTs de *Entamoeba* indicando otro evento de LGT. Los círculos blancos y negros son organismos con características anaeróbicas o microaerofílicas. Los triángulos blancos y negros indican organismos patógenos o al menos encontrados en muestras clínicas, respectivamente. El valor de bootstrap se indica en la base de cada nodo. La barra de escala también se muestra. Abreviaturas de clado: α , alfa-proteobacteria; β , beta-proteobacteria; γ , gamma-proteobacteria; δ , delta-proteobacteria; ϵ , epsilon-proteobacteria.

4.1.3. Glicosilasa AlkD: dos adquisiciones y una duplicación.

Otro tipo importante de lesión en el ADN que las células tienen que enfrentar es el daño por alquilación. La tasa de ocurrencia de esta lesión no se ha estimado hasta el momento. Sin embargo, el hecho de que las células tengan mecanismos de reparación especializados da una idea de su repercusión biológica. La alquilación de bases puede ser tanto citotóxica como genotóxica. Cada base tiene sus propias modificaciones por alquilación, como 7-metilguanina (7mG), 4-metilimidina (4mT), 3-metiladenina (3mA) y 5-metilcitosina (5mC). La primera es la alquilación más dominante en el ADN y está sujeto a una rápida transformación a sitios Ab, altamente citotóxicos Ab. Debido a la importancia de eliminar este tipo de lesiones, hay diferentes mecanismos de reparación como las vías NER, BER, MMR, HRR o mediante la reparación directa por metil-transferasas [Para mayor detalle revisar (Drabløs, Feyzi et al. 2004)].

La vía de BER tiene su propio repertorio glicosilasas de metilo presentes en los tres dominios (Wyatt, Allan et al. 1999, Denver, Swenson et al. 2003). La mayoría de estas enzimas reparadoras han evolucionado independientemente para resolver metilaciones, desarrollando diferentes especificidades de sustrato y mecanismo catalítico, (Wyatt, Allan et al. 1999, Hollis, Ichikawa et al. 2000, Mullins, Shi et al. 2015). Las diferentes glicosilasas de metilo son TAG, AlkA, AlkC, AlkD, MAG, ADPG AGG y sMAG. Todas las glicosilasas hasta ahora se han descrito como monofuncionales (Schärer, Nash et al. 1998), por lo tanto, la presencia de una AP-

endonucleasa funcional es fundamental para el proceso de reparación a través de la vía BER.

E. histolytica durante su ciclo de vida dentro del huésped está expuesto a diferentes agentes alquilantes, como fármacos (Drabløs, Feyzi et al. 2004)[84], ácidos biliares (Busby, Shuker et al. 1985) y su metabolismo en sí mismo (Drabløs, Feyzi et al. 2004). Un mecanismo que implica un solo paso para reparar metilaciones es a través de demetilasas como Ogt, Ada y AlkB. No obstante, no pudimos identificar los genes correspondientes en el genoma de *E. histolytica*. Por lo tanto, la reparación de metilaciones se basa fuertemente en la vía BER. Sin embargo, se encontraron homólogos para demetilasas en *A. catellanii* (PathemaID ACA1_237280) y *N. fowleri* (NF113890). Interesantemente, *E. histolytica* posee genes para metilasas como Dmnt2 y Meth (Fisher, Siman-Tov et al. 2004, Tovy, Tov et al. 2010). En este estudio pudimos identificar solo los genes que codifican para la glicosilasa AlkD con Pathema ID's EHI_087400, EHI_002000 y EHI_141100, denominados EhAlkD1, 2 y 3, respectivamente. AlkD es una glicosilasa de ADN casi exclusivamente bacteriana y de arquea. AlkD tiene una presencia reducida solo en eucariotas unicelulares. Para entender la historia evolutiva de estas AlkD de Entamoeba realizamos un análisis filogenético que incluía 124 secuencias de aminoácidos de bacterias, de arqueas y protistas. El árbol filogenético mostrado en la **Figura 17** separa claramente las AlkD bacterianas, de arqueas y protista por separado, excepto EhAlkD3 que está agrupada con otros homólogos de *Entamoeba* AlkD en un clado de bacterias patógenas, denominado clado de microbios patógeno (Figura **17C**). En este clado se agrupan α -, γ -, δ -protobacterias, así como

espiroquetas, bacteroidetes y homólogos de AlkD de *Entamoeba*. Este agrupamiento filogenético sugiere fuertemente una adquisición por LGT para los homólogos de AlkD3 de *Entamoeba* con un soporte de bootstrap del 98.6%. Esta composición mixta del clado de microbios patógenos es similar al clado anaeróbico reportada para MutY al que pertenece el ortólogo de *E. histolytica*, como se verá más adelante. Interesantemente, también la característica patogénica y anaeróbica/microaerófila está presente en este clado. Con respecto a EhAlkD1 y 2, estos genes están agrupados en un clado exclusivo de protista (con un 95% de soporte de bootstrap). De acuerdo a la topología de este clado, estos genes son parálogos (99,9% de soporte de bootstrap). El clado protista incluye al protista patógeno *Blastocystis sp*, género *Naegleria*, *A. castellanii*, y *Tetrahymena thermophila* (no patógena), *Reticulomyxia filosa* y *Paramecium tetraurelia* (**Figura 17B**). Si bien la evolución y taxonomía de los protistas es bastante desconcertante (Adl, Simpson et al. 2005, Koonin 2010, Moreira and López-García 2017), la explicación de la reducción del número de genes AlkD en eucariontes y a su presencia exclusiva en protistas es diversa. Una hipótesis es que estos genes tienen un origen vertical pero que no se fijaron en otros eucariotas debido a la presencia de otras glicosilasas de metilo. Otra explicación más parsimoniosa respaldada por nuestro análisis filogenético es que la AlkD se adquirió a partir de una bacteria o arquea en las primeras etapas de la evolución de los protistas. De hecho, en el árbol filogenético AlkD se pueden observar 3 eventos de LGT independientes. Se puede observar un evento de transferencia en el clado de amebozoa (incluyendo *Dictyostelium discoideum*, *D. fasciculatum* y *Acytostelium subglobosum*) que están agrupados con bacterias de la familia bacillaceae, listereaceae, estreptococos y β -

proteobacterias (**Figura 17A**). Otro evento LGT ocurrió en protistas y el hecho de la distribución reducida y exclusiva en este grupo puede ser explicada a un suceso ocurrido en etapas evolutivas muy temprana en este reino. Este evento de LGT está asociada a la duplicación genética que produce EhAlkD1 y 2. Además, ocurrió un tercer evento de transferencia que originó el gen bacteriano AlkD3 en el género *Entamoeba*.

Las primeras descripciones bioquímicas demostraron que AlkD de *Bacillus cereus* (BcAlkD) puede procesar 7mG, 3mG y 3mA (Alseth, Rognes et al. 2006). Excepcionalmente, AlkD es más eficiente en la eliminación de 7mG que AlkA y AlkC. Sin embargo, AlkD demostró tener una capacidad similar a AlkA y AlkC para reparar otros sustratos metilados. Por lo tanto, AlkD parece ser una glicosilasa de metil-purinas bastante versátil. Algunos estudios estructurales han revelado algunas características bioquímicas singulares de esta enzima. En primer lugar, su estructura sigue una arquitectura original de una matriz helicoidal de repeticiones HEAT, compuesta por 6 módulos de dos hélices α unidas por asas cortas (**Figura 18**) (Rubinson, Metz et al. 2008). Otra particularidad de esta enzima es que no utiliza el mecanismo altamente conservado en glicosilasas de ADN conocido como extrusión de bases (flipping-out), sino que realiza su catálisis dentro de la hélice del ADN (Mullins, Shi et al. 2015). Para entender el mecanismo catalítico en detalle se han realizado hermosos estudios bioquímico-estructurales. La catálisis hidrolítica se lleva a cabo a través de una molécula de agua activada por un carboxilato (D113 en BcAlkD) como en la mayoría de las glicosilasas de ADN (véase la **Figura 19**). Desde un punto de vista catalítico, AlkD reconoce el sustrato utilizando interacciones

polares, electrostáticas e hidrofóbicas. La principal contribución para los contactos polares está dada por Y27, cuya mutación reduce drásticamente hasta 2-5 veces la afinidad por el ADN con base metilada (Mullins, Shi et al. 2015). La interacción electrostática está dada por R148 a través de un puente salino con la desoxirribosa que contiene la base metilada. La mutación en este residuo afecta drásticamente la afinidad por el sustrato y la velocidad catalítica. Otro aspecto interesante relacionado con el reconocimiento del sustrato de AlkD es la interacción hidrofóbica en la que están implicadas dos Phe. La interacción hidrofóbica se produce a través de la interacción CH- π entre W109, W187 y el nucleótido metilado. Se estima que la mitad de la unión del sustrato y la estabilización estructural de los estados de transición en la reacción están dados por esta interacción (Mullins, Shi et al. 2015). Como se esperaba, las mutaciones en estos residuos comprometen la capacidad de unión al sustrato y las capacidades catalíticas (Mullins, Rubinson et al. 2014).

EhAlkDs conserva el plegado de repeticiones HEAT (**Figura 18**). Los seis módulos de hélices α están altamente conservados. La mayoría de los residuos catalíticos, así como los implicados en el reconocimiento del sustrato se conservan. Los parálogos EhAlkD1 y EhAlkD2 presentan una extensión C-terminal de 20 y 110 residuos de longitud, respectivamente. Esta extensión es exclusiva en el género *Entamoeba*. En cuanto al sitio activo (**Figura 19**), se conservan el Asp catalítico y la Arg implicados en la formación del puente salino. Asimismo, W109 se conserva en los tres ortólogos de *E. histolytica*. No obstante, W187 solo se conserva en EhAlkD3 (W196). Mientras tanto, EhAlkD1 y 2 tienen una Cisteína en esa posición (C193 y C201, respectivamente). Dada la importancia de este residuo para el

reconocimiento y la estabilización del sustrato, así como la naturaleza polar y el pequeño tamaño de la cisteína, este cambio puede comprometer la funcionalidad de estos parálogos. Este hecho y nuestro análisis filogenético nos permiten plantear el siguiente escenario evolutivo para EhAlkD. Una LGT inicial dio origen a un gen de AlkD en *E. histolytica*. Los residuos activos estaban presentes en este ancestro de AlkD. Entonces, la mutación del segundo triptófano se presentó seguido de la duplicación de genes produciendo EhAlkD1 y 2. Dadas las necesidades de reparación para hacer frente a las bases metiladas debido a la presencia de parálogos de AlkD defectuosos, otro evento LGT generó a EhAlkD3, una variante funcional que se fijó finalmente. Somos conscientes de la naturaleza especulativa de esta hipótesis, sin embargo, estudios funcionales de estos homólogos de AlkD de *E. histolytica* pueden ayudar a dar soporte a este escenario hipotético.

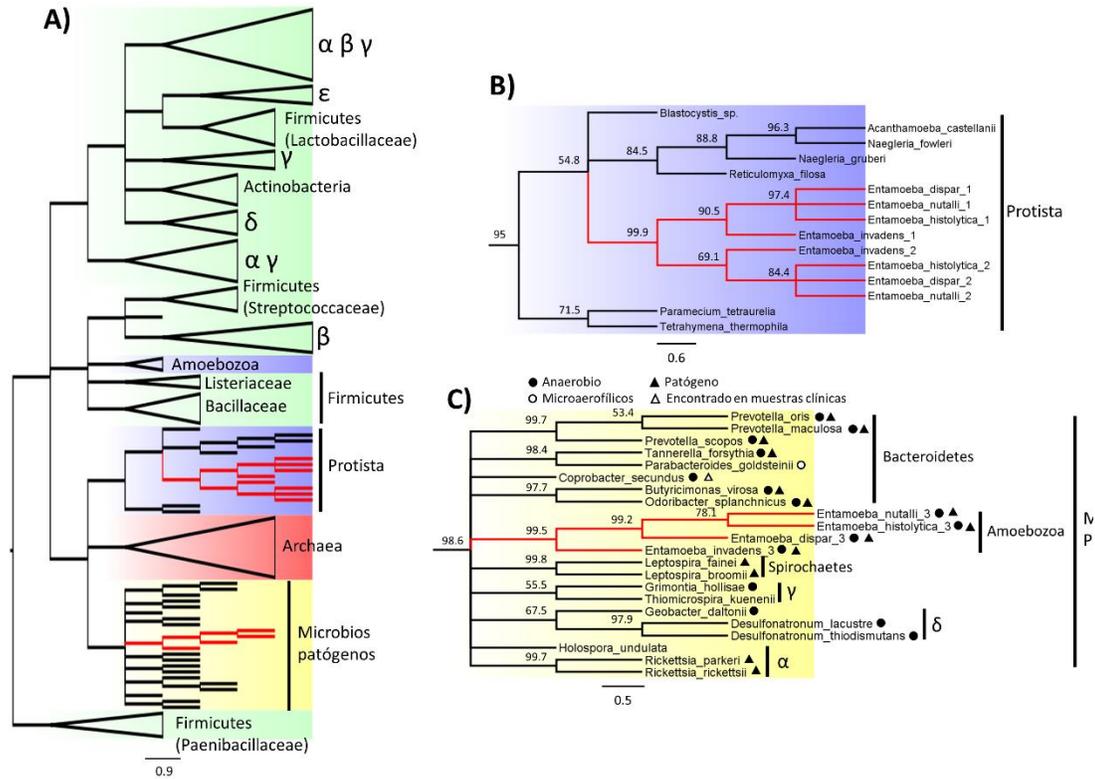


Figura 17. Análisis filogenético de AlkD. A) Análisis evolutivo de 124 secuencias de aminoácidos de AlkD incluyendo Arquea (clado rojo), Bacteria (clados verdes) y Eucarionte (clados azules). El análisis filogenético se realizó con el algoritmo de máxima verosimilitud con 1000 réplicas de bootstrap y LG+G como modelo de sustitución. AlkD de Entamoeba se muestran en líneas rojas. B) Vista ampliada del clado de Protista que incluye AlkD1 y AlkD2 del género Entamoeba. C) Vista ampliada del clado de microbios patógenos con AlkD3 de Entamoeba. Los círculos blancos y negros indican organismos con características anaeróbicas o microaerófilas. Los triángulos blancos y negros indican organismos patógenos o al menos encontrados en muestras clínicas, respectivamente. El porcentaje de bootstrap se indica en la base de cada nodo. La barra de escala también se muestra. Abreviaturas de clado: α , alfaproteobacteria; β , betaproteobacteria; γ , gammaproteobacteria; δ , deltaproteobacteria; ϵ , epsilonproteobacteria.

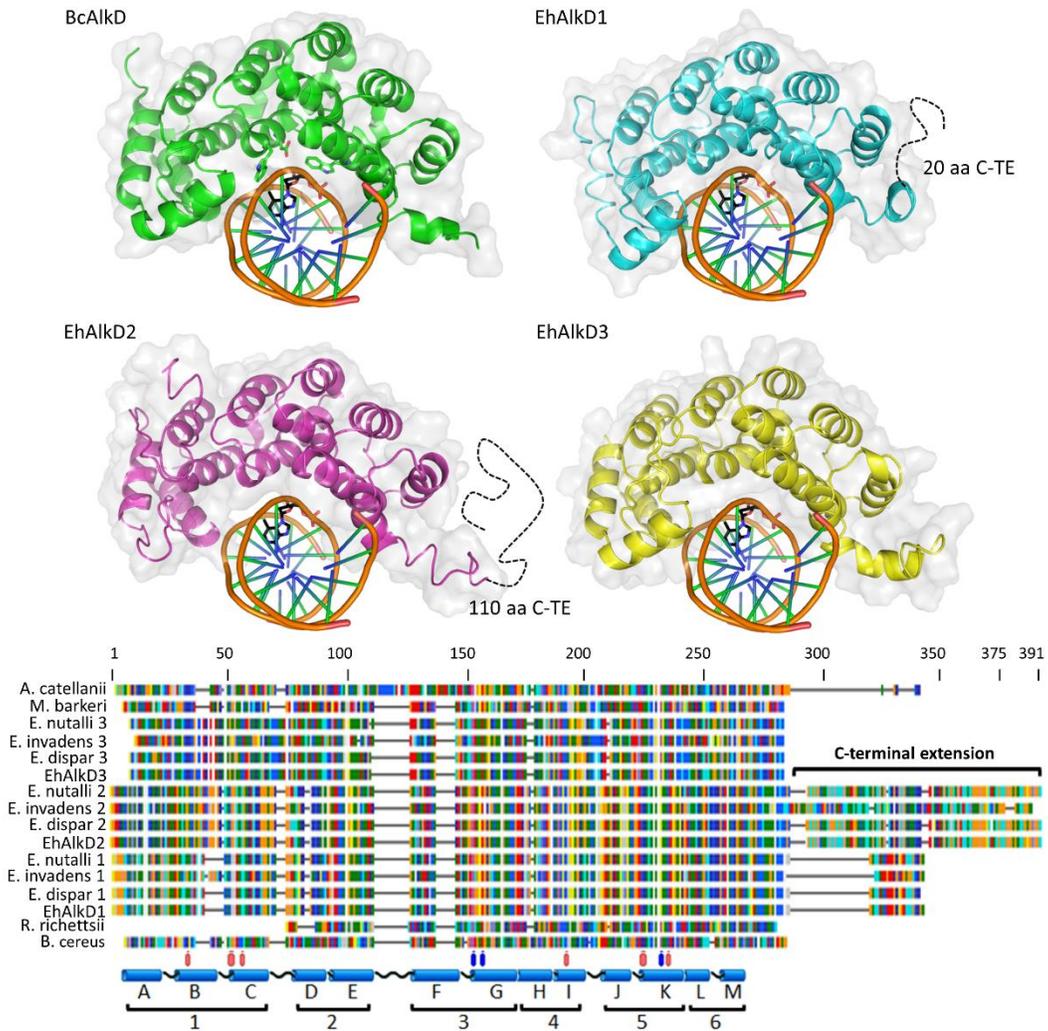


Figura 18. Análisis estructural de AlkD. Las AlkD de *E. histolytica* se modelaron usando como templado la AlkD de *Bacillus cereus* (PDB ID 3JXY) en el servidor I-TASSER. Para el análisis del modelo de EhAlkD1 y EhAlkD2 las extensiones C-terminal fueron eliminadas. Como se observa en el análisis estructural y en el alineamiento esquemático, la matriz helicoidal de repeticiones de HEAT, compuesta por 6 módulos hélices α unidas por asas cortas, está ampliamente conservada. En la alineación esquemática, las repeticiones de HEAT están indicadas en la alineación con números, los residuos implicados en la estabilización del sustrato y la catálisis en rectángulos rojo y azul, respectivamente.

4.1.4. Superfamilia UDG; La familia I y VI están presentes en *E. histolytica*.

Como se mencionó anteriormente, 8oxoG es una de las lesiones del ADN que ocurre con mayor frecuencia, produciendo transversiones GC→TA. Otra lesión importante es la formación de uracilo (o en su forma oxidada 5-OH-U) la cual se origina través de la degradación de citosina por oxidaciones. Estos productos de deaminaciones alteran la preferencia de apareamiento de bases causando transiciones CG→TA. Se ha determinado que el nivel de 5-OH-U y U en leucocitos humanos es de 7 y 20 lesiones por 10^7 bases, respectivamente (Kreutzer and Essigmann 1998). Sin embargo, no todo es malo acerca de los productos de deaminación de citosina, ya que juega un papel importante en un nivel epigenético e inmunológico, ya que está asociado al proceso de hipermutación de inmunoglobulinas y la plasticidad celular (Di Noia and Neuberger 2002, Morgan, Dean et al. 2004).

La superfamilia de glicosilasa de uracilo es la encargada de reparar lesiones relacionadas con uracilo en ADN de doble y cadena sencilla. Hasta ahora, se han identificado seis familias monofuncionales, que se diseminan en los tres dominios (Lee, Dominy et al. 2011, Lucas-Lledó, Maddamsetti et al. 2011). Estas glicosilasas de ADN tienen diferentes especificidades de sustrato, pero el uracilo es un denominador común entre ellas. De la familia I a la VI se tiene un origen evolutivo común. Curiosamente, la familia UDG VI (denominada familia de glicosilasa de Hipoxantina (Hx)) de acuerdo a análisis filogenético pertenece a la superfamilia UDG y, además de tener actividad sobre uracilo, también tiene una actividad clara

sobre el ADN con inosina. Este producto de desaminación de la adenina puede aparearse con citosina, timina o adenina, siendo el primero el apareamiento más estable (Alseth, Dalhus et al. 2014). Estas glicosilasas tienen un plegado α/β ampliamente conservado. El plegamiento altamente conservado de la superfamilia UDG, las similitudes catalíticas y la conservación de los sitios de unión del sustrato, sugiere un mecanismo catalítico general. Sin embargo, las sutiles diferencias en especificidad indican algunas posibles innovaciones catalíticas. El mecanismo de extrusión de la base fuera de la doble hélice del ADN, así como el carboxilato que activa una molécula de H₂O para el ataque nucleofílico se conservan (Slupphaug, Mol et al. 1996, Drohat, Jagadeesh et al. 1999, Drohat, Xiao et al. 1999, Aravind and Koonin 2000).

Nuestra búsqueda de miembros de la superfamilia UDG en el genoma de *E. histolytica* resultó en la identificación de dos genes que según nuestro análisis funcional son UDG (Pathema ID; EHI_178910 y EHI_198640). Con el fin de identificar a la familia a la que pertenecían estos UDG, realizamos un análisis filogenético que incluía 75 secuencias de aminoácidos de las seis familias de UDG (**Figura 20**). En el árbol filogenético se puede observar que todas las familias fueron claramente clasificadas. El gen con Pathema ID EHI_178910 fue agrupado en un clado que solo contiene miembros de la familia I. Por otro lado, el gen con Pathema ID EHI_198640 se agrupó en el clado bacteriano de la familia VI (HxADN glicosilasa). Estas glicosilasas de ahora en adelante se nombrarán EhUDG-FI y EhHxDG. El subclado en el que se agrupa EhUDG-FI incluye otras glicosilasas del género *Entamoeba* y *Acanthamoeba*, lo que sugiere un origen eucariota. Como se

mencionó, EhHxDG así como sus homólogos de Entamoeba son los elementos eucariotas únicos en el clado de la familia VI sugiriendo otro producto de LGT (con un soporte del 89%, **Figura 20C**). La característica anaeróbica/microaerófila y patogénica es un patrón conservado en este clado. Otra característica conservada es que la transferencia de genes se produjo antes de la especiación del género Entamoeba.

El análisis estructural de la familia I (**Figura 21**) muestra que el plegado α/β está conservado en EhUDG-FI. La estructura del sitio activo y los residuos implicados en la catálisis y el reconocimiento del uracilo también están ampliamente preservados. En UDG de *E. coli* la S88 y H187 (S128 y H226 en EhUDG-FI) participan en la estabilización del sustrato posicionando dos moléculas de agua que interactúan con el uracilo. G62, Q63 e Y66 (G102, Q103 e Y126 en EhUDG-FI) se identifican como residuos importantes para la especificidad enzimática que bloquea el ingreso de purinas al sitio catalítico y proporciona contactos específicos para el posicionamiento correcto del uracilo. N123 (N163 en EhUDG-FI) es un residuo determinante para distinguir entre uracilo y citosina ya que su grupo amino hace que la interacción de puentes de hidrógeno con N3 de uracilo y el carbonilo interactúa con O4 del uracilo (Parikh, Walcher et al. 2000, Pearl 2000). El carboxilato catalítico que activa una molécula de agua para el ataque nucleofílico de esta glicosilasa es el D64 (D104 en EhUDG-FI) (Xia, Liu et al. 2017). Este alto grado de conservación sugiere que EhUDG-I tiene una funcionalidad adecuada en este patógeno.

Para comprender las características estructurales de EhHxDG, construimos un modelo de homología estructural utilizando como template la UDG de

Methanosarcina acetivorans (PDB ID 2L3F) que según el análisis del motivo es una glicosilasa de hipoxantina (en lo sucesivo, MaHxDG). Estas HxDG comparten el 31.1% de identidad. Los estudios iniciales de la familia VI de UDG propusieron una relación con la superfamilia Hélice-horquilla-Hélice (HhH) a la que pertenecen Endonuclease III, MutY y MutM (Chung, Im et al. 2003, Lee, Dominy et al. 2011). Este motivo tiene la función de andamio de interacción que estabiliza el complejo ADN-proteína que interactúa con la estructura del fosfato de ADN de manera no específica (Doherty, Serpell et al. 1996, Fromme and Verdine 2003, Fromme, Banerjee et al. 2004). El análisis estructural de esta familia se muestra en la **Figura 22**. El plegado de MaHxDG resuelta por NMR (PDB ID 2L3F) y el modelo EhHxDG no muestran el plegamiento característico de la superfamilia HhH compuesta por un dominio construido por dos módulos de hélice α (Guan, Manuel et al. 1998). Como se puede observar, estas glicosilasas así como la glicosilasa de timina de humano (Familia II) presentan el plegado α/β clásico de la superfamilia UDG. De hecho, el motivo HhH detectado por Chung y colaboradores en estas enzimas presenta un plegado Hélice-horquilla-lámina (HhS, por sus siglas en inglés). De hecho, el motivo HhS se ha encontrado en el Factor regulador de interferón con una función de andamio de interacción similar al motivo HhH (Fujii, Shimizu et al. 1999). Por lo tanto, proponemos que la clasificación de la familia HxDG se debe revisar teniendo en cuenta los datos estructurales disponibles. Como la estructura MaHxDG se resolvió sin sustrato, realizamos el modelado con uracilo. En el sitio activo de EhHxDG identificamos que S23, N41 y S142 entran en contacto con el uracilo a través de puentes de hidrógeno. Estos residuos se conservan en secuencia y estructura en MaHxDG, la HxDG de *Methanosarcina barker* (MbHxDG) y enzima del

género *Entamoeba* (**Figura 23**). Los estudios bioquímicos hechos en MbHxDG muestran que N39 es un residuo crucial para la catálisis, proponiendo que esto podría estar involucrado como activador de una molécula de agua para el ataque nucleofílico (Lee, Dominy et al. 2011). Así mismo, una Asn está implicado en el ataque nucleofílico en TDG humano. Los residuos D74, D86 y N113 en MbHxDG parecen ser importantes para la actividad catalítica ya que sus versiones mutantes disminuyen notablemente su actividad glicosilica. Curiosamente, el mutante de residuo D86A en MaHxDG anula la actividad de glicosilasa sobre inosina apareada con C, A y T y disminuye 7.7 por sustrato con U:G. Estos cambios en la actividad de la mutante D86A sugieren fuertemente que tienen un papel directo durante la catálisis. No obstante, D86 y D88 en MaHxDG y EhHxDG están a 16 Å de distancia del sitio activo putativo. La distancia es lo suficientemente grande como para cuestionar el papel de este residuo en la catálisis, a menos que un cambio conformacional considerable esté implicado. Por lo tanto, sugerimos que el asa en el que se encuentra este residuo (Asa puerta) funciona como una puerta cuyas conformaciones abiertas/cerradas dependen de la presencia del sustrato. El modelado de EhHxDG con uracilo muestra una conformación abierta posiblemente a que su modelado se realizó usando un templado sin sustrato. Como se observa en la **Figura 23**, la conservación de estos residuos no es absoluta. Una restricción para este análisis es el pequeño número de secuencias disponibles y la reducción de los estudios bioquímicos/estructurales sobre la familia VI de UDG.

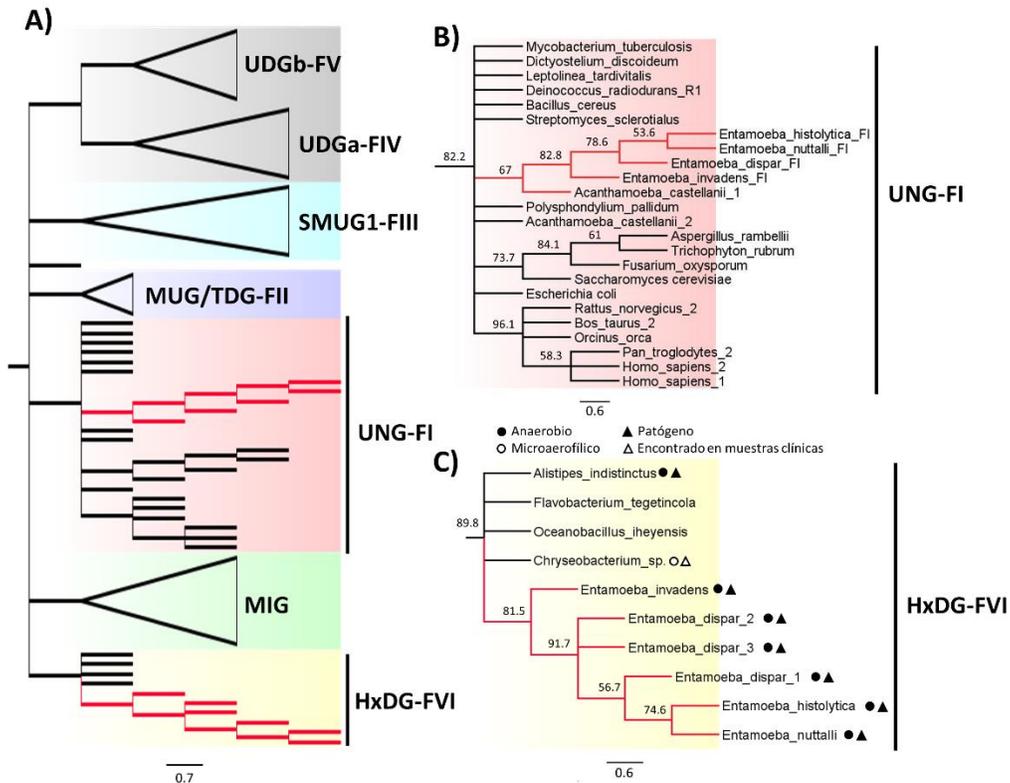


Figura 20. Análisis filogenético de la superfamilia UDG. A) Análisis evolutivo de 74 secuencias de aminoácidos de la superfamilia de UDG incluyendo Arquea, Bacteria y Eucariote. El análisis filogenético se realizó con el algoritmo de máxima verosimilitud con 1000 réplicas de Bootstrap y WAG+G como modelo de sustitución. Las familias de UDG están indicadas en el árbol filogenético con etiquetas y colores diferentes. Las UDG de Entamoeba se muestran en líneas rojas. B) Vista ampliada del clado de la UDG-FI. C) Vista ampliada del clado de patógeno microbiano que alberga a la HxDG's de Entamoeba. Los círculos blancos y negros son organismos indicados con características anaeróbicas o microaerófilicas. Los triángulos blancos y negros indican organismos patógenos o al menos encontrados en muestras clínicas, respectivamente. El porcentaje de soporte de bootstrap se indica en la base de cada nodo. La barra de escala también se muestra

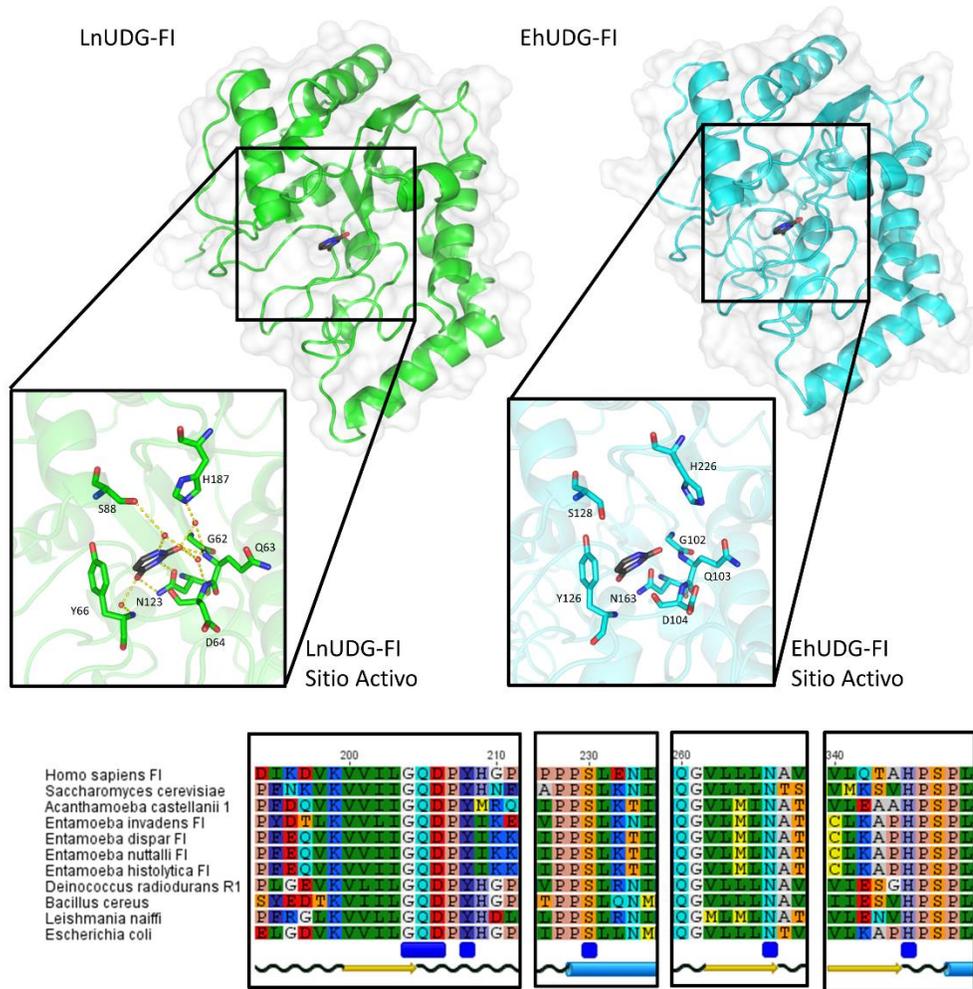


Figura 21. Análisis estructural de la familia UDG I. UDG-FI de *E. histolytica* se modeló con la UDG de *E. coli* (PDB ID 2EUG) como templado en el servidor I-TASSER. A la izquierda, se muestra la estructura de cristalografía UDG de *Leishmania naiffi* en complejo con 5-bromouracilo y una vista ampliada del sitio catalítico. A la derecha se muestra el análisis del modelo por homología estructural de UDG de *E. histolytica*. Abajo se presenta un alineamiento representativo resaltando el alto grado de conservación de residuos catalíticos (resaltados con rectángulos azules).

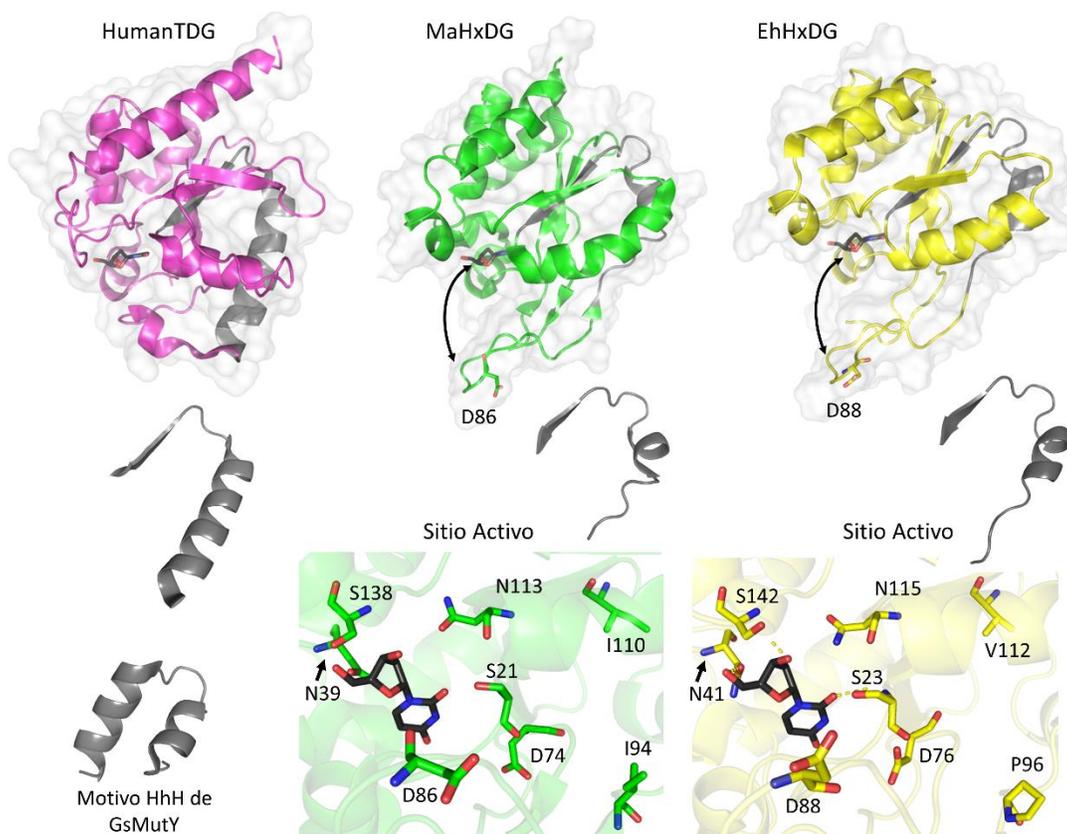


Figura 22. Análisis estructural de la glicosilasa de hipoxantina. Arriba se muestran la estructura cristalográfica TDG humana (PDB ID 2RBA), *Methanosarcina acetivorans* UDG (MaHxDG, ID PID 2L3F) resuelto por RMN y el modelo por homología estructural de *E. histolytica* hipoxantina ADN glicosilasa (EhHxDG). EhHxDG se modelaron con desoxiuracilo como ligando para identificar el sitio activo putativo utilizando como plantilla MaHxDG. Los datos bioquímicos sugieren que D86 y D88 de MaHxDG y EhHxDG, respectivamente, están involucrados en la catálisis. Por lo tanto, el Asa en los que están posicionados estos residuos presumiblemente tienen grandes cambios conformacionales que permitan interactuar con el sustrato. Este Asa puerta y los cambios conformacionales involucrados están indicados con flechas curvas. Se muestra una vista ampliada de los sitios activos MaHxDG y EhHxDG, presentando como barras los residuos catalíticos putativos. También el motivo Hélice-horquilla-Lámina (HhS) se presenta en gris y Hélice-horquilla-Hélice (HhH) de la MutY de *G. stearothermophilus*. EhHxDG se modeló con MaHxDG como templado en el servidor I-TASSER.

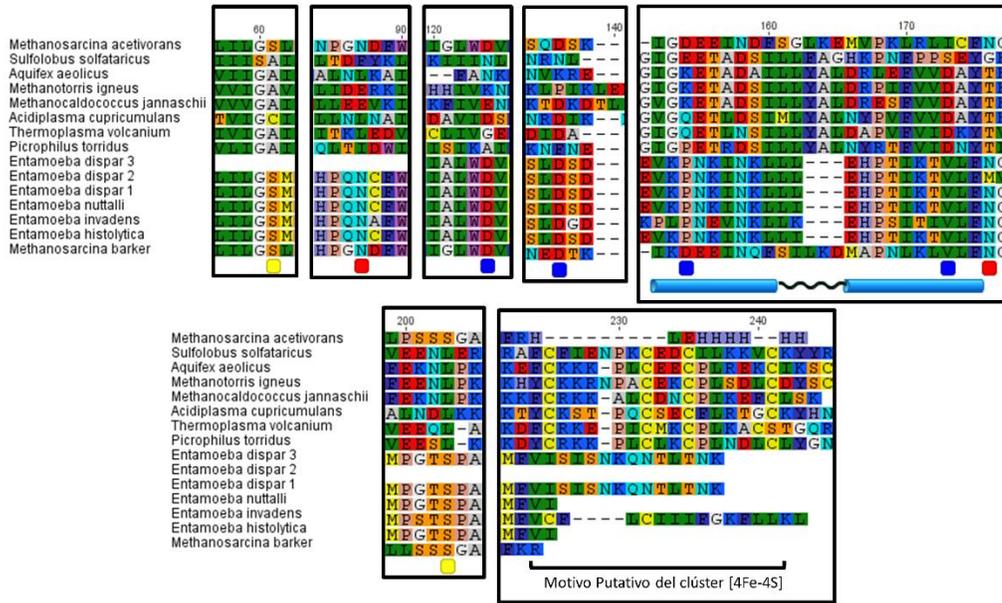


Figura 23. Alineamiento Representativo de glicosilasas de hipoxantina. En los cuadrados amarillos se indica la Ser que interactúan con uracilo a través de puentes de hidrógeno, los residuos catalíticos putativos están indicados en cuadrados rojos y los carboxilatos implicados en la reacción enzimática de acuerdo con los datos bioquímicos están indicados en cuadrados azules. El clúster [4Fe-4S] y motivo HhH también están indicados.

4.1.5. EndoIII; una historia de duplicación de genes.

La Endonucleasa III (EndoIII), codificada por el gen *nth* ó *ntg*, es una glicosilasa de ADN bifuncional a cargo de la eliminación de pirimidinas oxidadas como glicol de timina y uridina, 5-hidroxi citosina, 5-hidroxi-6-hidro uracilo y 5-hidroxi-6-hidro timina (Ver la Tabla 1 para más detalles sobre la especificidad de EndoIII). Esta glicosilasa que también pertenece a la superfamilia HhH, está relacionada estructuralmente con el núcleo catalítico de MutY, compartiendo una alta identidad de secuencia de aminoácidos. Además de compartir el motivo HhH, también se conserva el clúster [4Fe-4S], incluido el Asa FC, importante para la

estabilización del complejo ADN-proteína (Thayer, Ahern et al. 1995). Por otra parte, también se comparte el mecanismo de extrusión de base que arroja la base oxidada fuera de la doble hélice del ADN hacia el sitio catalítico (Ikeda, Biswas et al. 1998, Fromme and Verdine 2003). Sin embargo, el mecanismo catalítico *per se* es bastante diferente, ya que las glicosilasas monofuncionales requieren la activación de una molécula de agua por un carboxilato para llevar a cabo el ataque nucleofílico. EndoIII implica un Asp que desprotona una Lys que a su vez actúa como un nucleófilo con la formación de un intermediario entre la Lys y 5'-fosfato estabilizado en forma de base Schiff, que es concomitante con la escisión de cadena mediante β -eliminación produciendo un producto de 5'-fosfato y 3'-dRP (Liu and Roy 2001, Schärer and Jiricny 2001). Este producto 3'-dRP resultante es tóxico ya que bloquea la replicación del ADN, por lo que es procesado por la Ap-endonucleasa utilizando una actividad 3'-fosfodiesterasa (Boiteux and Guillet 2004).

El genoma de *E. histolytica* alberga dos genes con PathemaDB ID EHI_118790 y EHI_083460 denominados Nth1 y Nth2, respectivamente. Estos EndoIII comparten una identidad de aminoácidos del 31,2%. Para rastrear su historia evolutiva, realizamos un análisis filogenético que incluyó 132 secuencias de aminoácidos de los tres dominios. El árbol filogenético resultante, que se muestra en la **Figura 24**, separa con un soporte de bootstrap del 99.8% a EndoIII eucariótico de los homólogos de Archaea y Bacteria. Entamoeba y otros homólogos de Amoebozoa se agrupan en este clado eucariota. Especialmente, Las Nth1 y Nth2 de Entamoeba se agrupan en el mismo subclado con un 74.5% de soporte sugiriendo paralogía en estos genes. Todas las especies de *Entamoeba* tienen dos

parálogos, por lo tanto, la duplicación del gen ancestral de Nth en *Entamoeba* ocurrió antes de la especiación del género. Es de mencionar que el genoma de *E. dispar* tiene tres genes ortólogos. Los ortólogos 2 y 3 son producto de una segunda duplicación genética que es exclusiva en esta especie. Además, hemos detectado que ntg1 y ntg2 de *S. cerevisiae* también son parálogos.

En general, las Endonucleasas III de *Entamoeba* son bastante similares a los ortólogos Bacterianos y de Arqueas que comparten 30.5% de identidad en secuencia de aminoácidos. Estructuralmente, los homólogos de *E. histolytica* conservan el módulo de seis hélices α que incluye el motivo HhH y el clúster [4Fe-4S] como el núcleo catalítico de MutY (véase la **Figura 25**). Los Asp y Lys implicados en la catálisis también se conservan, así como los cuatro ligandos de cisteína necesarios para la coordinación del clúster [4Fe-4S]. El Asp catalítico implicado en la deprotonación de Lys es D168 y D169 en Nth1 y Nth2 de *Entamoeba*, respectivamente. La Lys que actúa como nucleófilo durante la catálisis después de la deprotonación es K149 y K150. La extrusión de la base dañada es llevada a cabo por una Leu en Bacteria, sin embargo, en Arqueas y Eucariontes una Phe parece estar involucrado. En la mayoría de nth1 de *Entamoeba* hay un Gly en esa posición. No obstante, una posición anterior a ella hay una Phe que presumiblemente puede estar involucrada en el mecanismo de extrusión de la base. Se ha descrito que tanto Leu como Phe pueden participar en la extrusión de bases (Livingston, Kundu et al. 2005). Los parálogos de Ntg en *S. cerevisiae* se son importantes para contrarrestar el estrés oxidativo y la exposición a los rayos UV. La expresión de estos genes es diversa, mientras que la expresión de Ntg1 es inducida

por agentes que dañan el ADN, Ntg2 se expresa constitutivamente. Estas enzimas tienen diferente especificidad de sustrato, lo que permite la protección contra un amplio espectro de pirimidinas oxidadas (Alseth, Eide et al. 1999). El clúster [4Fe-4S] está altamente conservado en EndoIII, sin embargo, la Ntg1 de *S. cerevisiae* tiene los ligandos característicos para este cofactor (You, Swanson et al. 1998). Los parálogos de Nth en *Entamoeba* pueden dar como resultado una expansión de la protección en contra del daño oxidativo del ADN ampliando sus especificidades como el caso de *S. cerevisiae*.

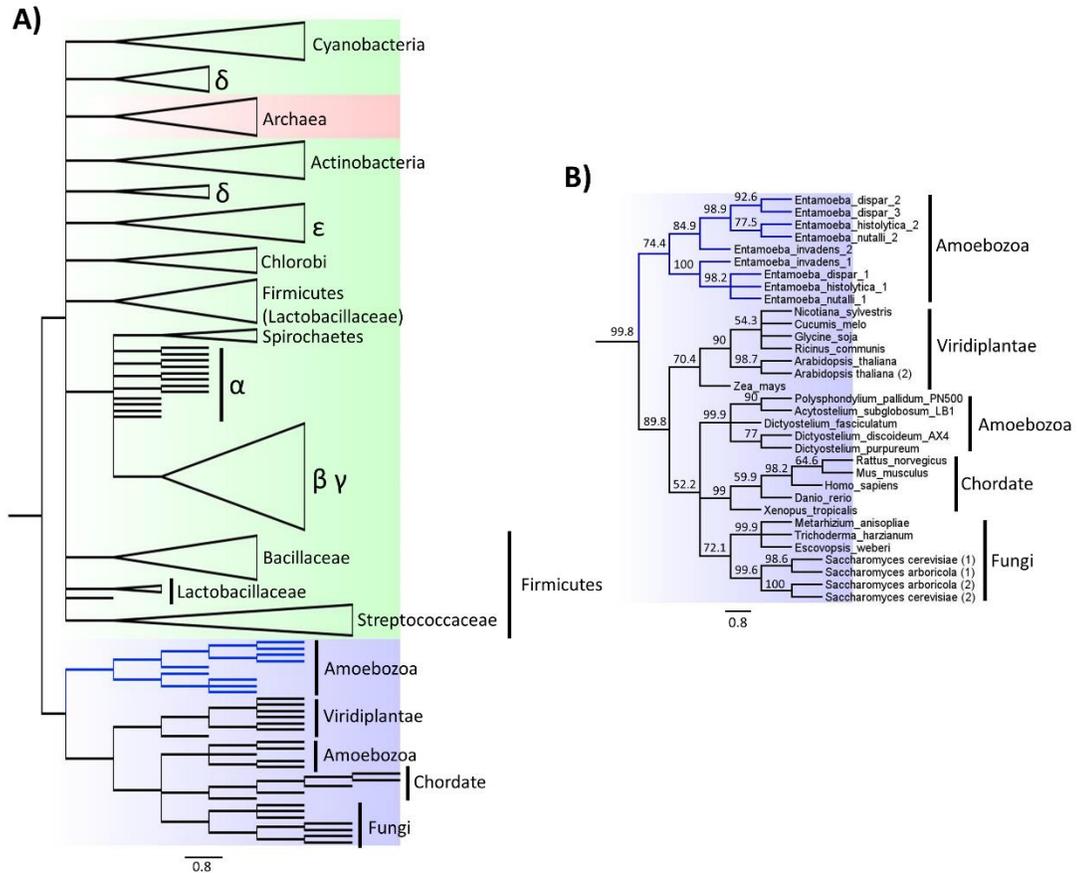


Figura 24. Análisis filogenético EndoIII. A) Análisis evolutivo de 132 secuencias de aminoácidos de EndoIII que incluyen Arquea (clado rojo), Bacteria (clados verdes) y Eucariote (clado azul). El análisis filogenético se realizó con el algoritmo de máxima verosimilitud con 1000 réplicas de Bootstrap y LG+G como modelo de sustitución. EndoIII de Entamoeba se muestran en líneas azules. B) Vista ampliada del clado eucariótico que alberga Nth1 y Nth2 de Entamoeba. El porcentaje de soporte de Bootstrap se indica en la base de cada nodo. La barra de escala también se muestra. Abreviaturas de clado: α , alfa-proteobacteria; β , beta-proteobacteria; γ , gamma-proteobacteria; δ , delta-proteobacteria; ϵ , epsilon-proteobacteria.

4.1.6. Observaciones finales y perspectivas del panorama evolutivo de BER de *Entamoeba histolytica*.

Las necesidades de reparación de ADN de *E. histolytica* están cubiertas por diferentes mecanismos, como las rutas BER, NER, HRR, NHEJ y MMR. En particular, la funcionalidad de la vía BER es determinante para reparar modificaciones sutiles de base con gran repercusión, como transversiones y transiciones. De acuerdo con nuestro análisis funcional de BER, *E. histolytica* es capaz de llevar a cabo su tarea solo a través de la subvía del parche corto, ya que no alberga polimerasas de ADN con actividad de liasa sobre productos 5'-dRP, ni proteína XRCC1, un andamio funcional para la subvía del parche largo. La ausencia de genes que codifican las glicosilasas OGG1/MutT y MDB4 podría ser responsable del bajo contenido de GC en el genoma de este patógeno. Nuestra reconstrucción evolutiva sugiere fuertemente que la mayoría del repertorio de glicosilasa de ADN de este patógeno está moldeado por eventos de LGT de origen bacteriano. Además, la duplicación de genes ha sido una fuerza que ha desempeñado un papel notorio en la formación del repertorio de glicosilasas de ADN en este *E. histolytica*.

Otra falta importante es la ausencia del modulador epigenético MDB4. La regulación epigenética en *E. histolytica* se ha asociado con la virulencia y el ciclo de vida de este patógeno (Lavi, Isakov et al. 2006, Mirelman, Anbar et al. 2006, Mirelman, Anbar et al. 2008, Huguenin, Bracha et al. 2010). De hecho, las metilasas como Dmmt2 y Meth relacionadas con regulación epigenética están presentes en este patógeno (Fisher, Siman-Tov et al. 2004, Tovy, Tov et al. 2010). Así, la presencia de enzimas que promueven la regulación epigenética mediante

metilación y la ausencia de MDB4 comprometen la comprensión de la modulación de este proceso regulador tan importante para *Entamoeba*. Posiblemente, otros mecanismos novedosos pueden estar implicados en la eliminación de marcas epigenéticas como 5-metil citosina en este protista o existen versiones altamente divergentes de MDB4 que no se han podido detectar hasta el momento.

En la Tabla 2, se muestran todos los componentes faltantes en la ruta de BER de *E. histolytica*. La mayoría de estas piezas faltantes son glicosilasas de ADN. La mayoría de las glicosilasas muestran una amplia especificidad de sustrato con una marcada redundancia. Esta redundancia permite la falta de algunas glicosilasas. Por ejemplo, la falta de TDG y SMUG se puede resolver con la presencia de UNG (perteneciente a la familia I de la superfamilia UDG) y UDG de la familia VI. Además, Las AlkD de origen bacteriano en *E. histolytica* puede cubrir la ausencia de genes que codifican para las glicosilasas como Mag1 y MPG especializadas en reparar metilaciones.

Otro aspecto a destacar de este apartado es que el patógeno humano *E. histolytica* tiene elementos de BER con una historia evolutiva y características estructurales-bioquímicas diferentes a las del huésped humano. En la última década, la maquinaria de reparación del ADN se ha propuesto como blancos para el diseño de fármacos (Fishel and Kelley 2007, Helleday, Petermann et al. 2008, Kelley and Fishel 2008). Por lo que se puede aprovechar las diferencias entre los mecanismos reparadores de *E. histolytica* y humanos para diseñar fármacos antiparásitos de nueva generación para el tratamiento de la amibiasis que afecta aproximadamente al 10% de la población mundial.

4.2. ESTUDIO BIOQUÍMICO-ESTRUCTURAL DE MUTY CON CARENCIA DE CLÚSTER [4FE-4S].

En esta sección de resultados y discusión se mostrarán resultados donde se demuestran que la MutY de *E. histolytica* (EhMutY) ha evolucionado para mantener su actividad de glicosilasa sin necesidad del clúster [4Fe-4S] y que el mecanismo estructural para prescindir de este cofactor parece estar dado por efecto estérico, además de demostrar el origen bacteriano de *E. histolytica*. De la misma forma el análisis estructural-bioquímico se realizó para la MutY de *Lactobacillus brevis* que comparte la característica de la carencia del cofactor.

4.2.1. Evolución de las novedosas MutYs carecientes del clúster [4Fe-4S].

El Sistema GO encargado de evitar y/o reparar las mutaciones relacionadas con la lesión 8oxoG, esta compuesto por los genes de MutM, MutY y MutT, virtualmente presente en los tres dominios (Eisen and Hanawalt 1999, Denver, Swenson et al. 2003, Baute and Depicker 2008, Jansson, Blomberg et al. 2010). MutY y MutM pertenecen a la superfamilia de glicosilasas de ADN con motivo hélice-horquilla-hélice (HHH) (Brooks, Adhikary et al. 2013). Otros miembros de la superfamilia son AlkA (Labahn, Schärer et al. 1996, Hollis, Ichikawa et al. 2000), Endonucleasa III (Cunningham, Asahara et al. 1989, Kuo, McRee et al. 1992) y la glicosilasa de 3-methyladenina (Drohat, Kwon et al. 2002). Miembros de esta superfamilia de glicosilasas muestran un plegado similar a pesar de compartir una baja identidad en aminoácidos.

Con la intención de entender la evolución de las soluciones estructurales alternativas para el clúster [4Fe-4S] en MutY, entre las cuales la EhMutY está presente, se llevó a cabo un análisis filogenético de 186 secuencias de aminoácidos de glicosilasas de Adenina incluyendo secuencias de arqueas, bacterias y eucariontes. De las MutYs analizadas, 140 secuencias conservaron los ligandos característicos de cisteínas para la coordinación del cofactor (CX₆CX₂CX₅C). No obstante, fueron encontradas algunas variantes en el primer y segundo espaciador donde hay una adición o delección de algún residuo en hongos, Arqueas, Euglenozoa y algunas bacterias (**Figura 26A**). 46 secuencias restantes de MutY no contenían las cisteínas de coordinación conservadas, haciendo suponer una incapacidad para coordinar el cofactor, Tales enzimas pertenecieron a los Phylums de *Lactobacillales*, *Chlorobi*, *Spirochaetes*, algunas *Deltaproteobacterias*, *Actinobacteria* y del protozoario patógeno de *Entamoeba*.

El análisis filogenético muestra que las MutYs con ausencia del clúster [4Fe-4S] están agrupadas en dos diferentes clados, nombrados clado de Lactobacillales y clado Anaeróbico (**Figura 27**). En el clado de Lactobacillales, como su nombre lo indica, solo se agrupan MutYs de bacterias pertenecientes al orden de *Lactobacillales*, mientras que el clado Anaeróbico solo se agrupan microorganismos anaeróbicos o microaerófilicos como; *Entamoeba*, *Chlorobi*, *Spirochaetes*, *Deltaproteobacterias* y *Actinobacteria*. Varios ejemplos se han reportado de la pérdida del clúster [Fe-4S], como en la fotoliasa (6-4) (Zhang, Scheerer et al. 2013), la helicasa-nucleasa AddAB (Yeeles, Cammack et al. 2009) y la sintetasa de triptophanil-tRNA (Stiebritz 2014). El hecho de que el clúster [4Fe-4S] este

conservado en los tres dominios de la vida provee una fuerte evidencia de que este motivo es un componente ancestral de la MutY y que este motivo ha sido perdido en dos eventos evolutivos independientes en los clados de *Lactobacillales* y Anaeróbico. La clase bacilli incluye los órdenes de *Bacillales* y *Lactobacillales*, la reconstrucción filogenética indica que la pérdida del cofactor es relativamente reciente y posterior al evento de divergencia de los *Lactobacillales*. Por otro lado, la composición mixta del clado Anaeróbico hace difícil entender la evolución de este grupo filogenético. Sin embargo, la presencia de MutYs no canónicas en el género *Entamoeba* es notorio. Otros organismos del phylum Amoebozoa, como *Acanthamoeba* y *Dictyotellium*, contienen MutYs con los ligandos característicos para el clúster [4Fe-4S] y están filogenéticamente distantes de las MutYs del género *Entamoeba* (*E. histolytica*, *E. dispar*, *E. Ginvadens*, y *E. nutalli*). Por lo tanto, es factible que los genes que codifican para esta glicosilasa carente de los ligandos en el género de *Entamoeba* hayan sido adquiridos por LGT a partir de una Bacteria. Se ha estimado que aproximadamente el 4% de los genes presentes en el genoma de *E. histolytica* fueron adquiridos por transferencia horizontal de linajes bacterianos (Loftus, Anderson et al. 2005, Grant and Katz 2014). Como se menciona anteriormente, *E. histolytica* también contiene varios genes involucrados en la reparación del ADN que están filogenéticamente relacionados a linajes bacterianos, como AlkD y la glicosilasas de Hipoxantina y la MutT que pertenece al sistema GO.

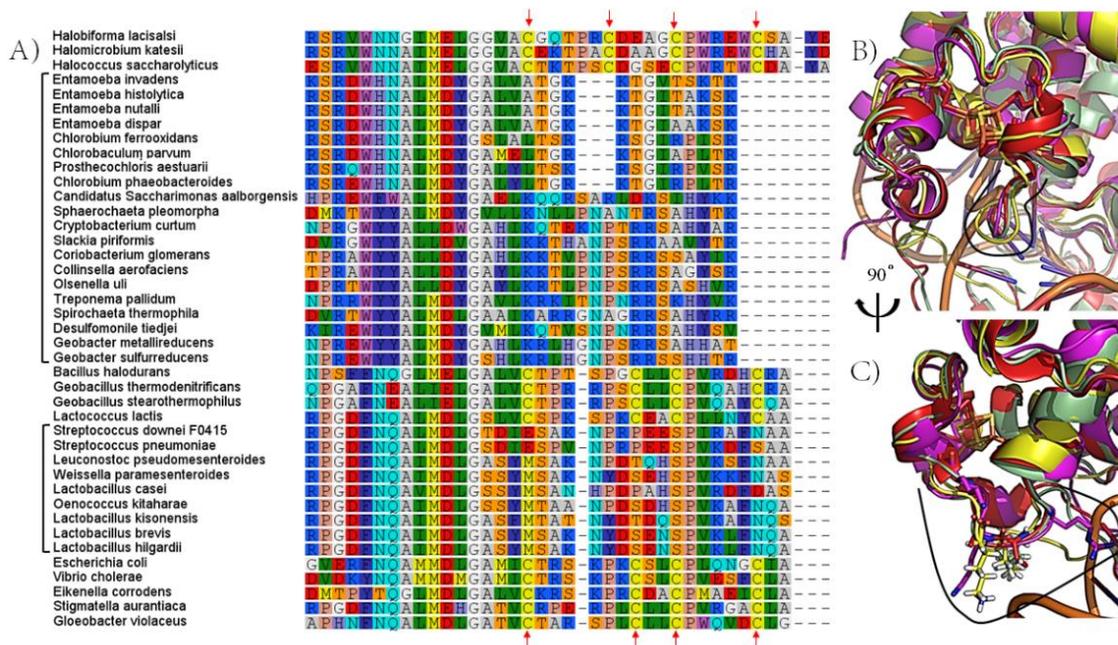


Figura 26. Alineamiento de secuencia y estructural de la región del motivo del clúster [4Fe-4S]. A) Alineamiento de aminoácidos que incluye algunas MutYs canónicas y soluciones alternativas al clúster [4Fe-4S]. Con flechas rojas están indicados los ligandos de cisteína. B) Alineamiento estructural de la región correspondiente del dominio del clúster [4Fe-4S] de la MutY de *E. coli* (rosa), *G. stearothermophilus* (rojo), *L. brevis* (verde) y *E. histolytica* (amarillo). C) Giro de 90° a la izquierda de la región correspondiente del motivo del clúster [4Fe-4S]. Los residuos positivos del asa del clúster [4Fe-4S] importantes para la interacción ADN-proteína están indicados en barras: R194, K196 y K198 de la MutY de *E. coli*; R201 y R202 de *G. stearothermophilus*; K200 para *L. brevis*; K232 y K233 de *E. histolytica*. El asa de clúster [4Fe-4S] se muestra con un contorno negro.

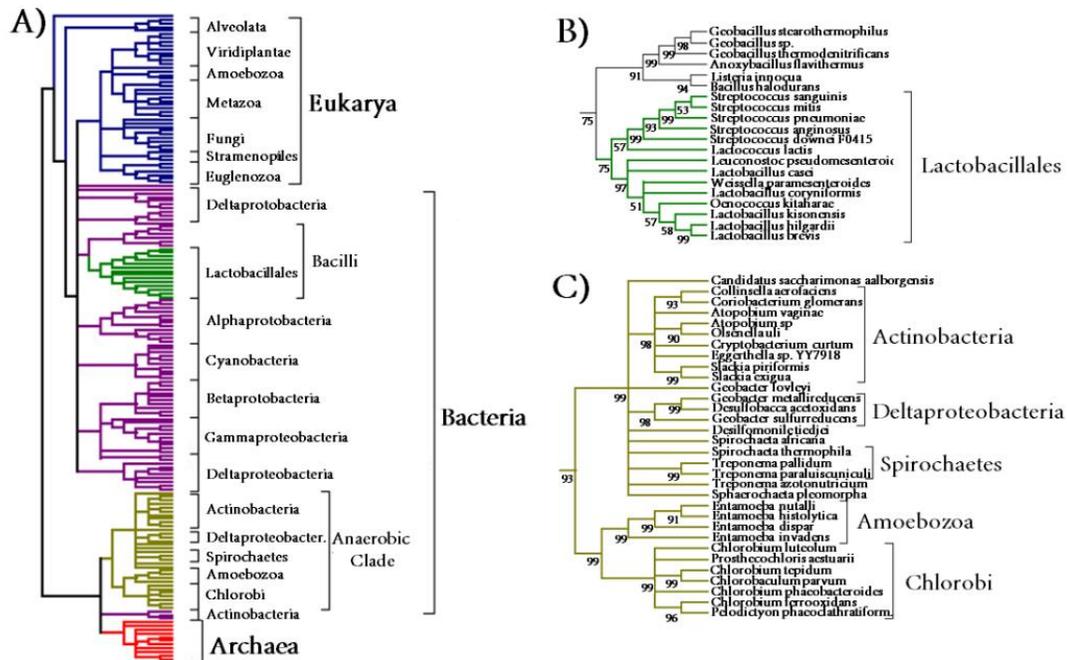


Figura 27. Análisis filogenético MutY. A) Análisis evolutivo de 186 secuencias de aminoácidos de MutY que incluyen Arquea, Bacteria y Eucarionte. La divergencia de las MutY novedosas sin clúster [4Fe-4S] se indica en verde para Lactobacillales y amarillo para el clado anaeróbico. El análisis filogenético se realizó utilizando el método de máxima verosimilitud con 1000 repeticiones de bootstrap. Los porcentajes de soporte de bootstrap están indicados en la base de cada nodo. B) Vista ampliada del clado de Bacilli. C) Vista ampliada del clado anaeróbico.

4.2.2. Análisis de secuencia y del modelado estructural de MutY.

Para probar la funcionalidad de las MutYs no canónicas y entender como estas glicosilasas mantienen su actividad sin la necesidad del clúster [4Fe-4S], se estudió a más profundidad EhMutY y a la MutY de la bacteria láctica *Lactobacillus brevis* (LbMutY) como representante del clado de *Lactobacillales*.

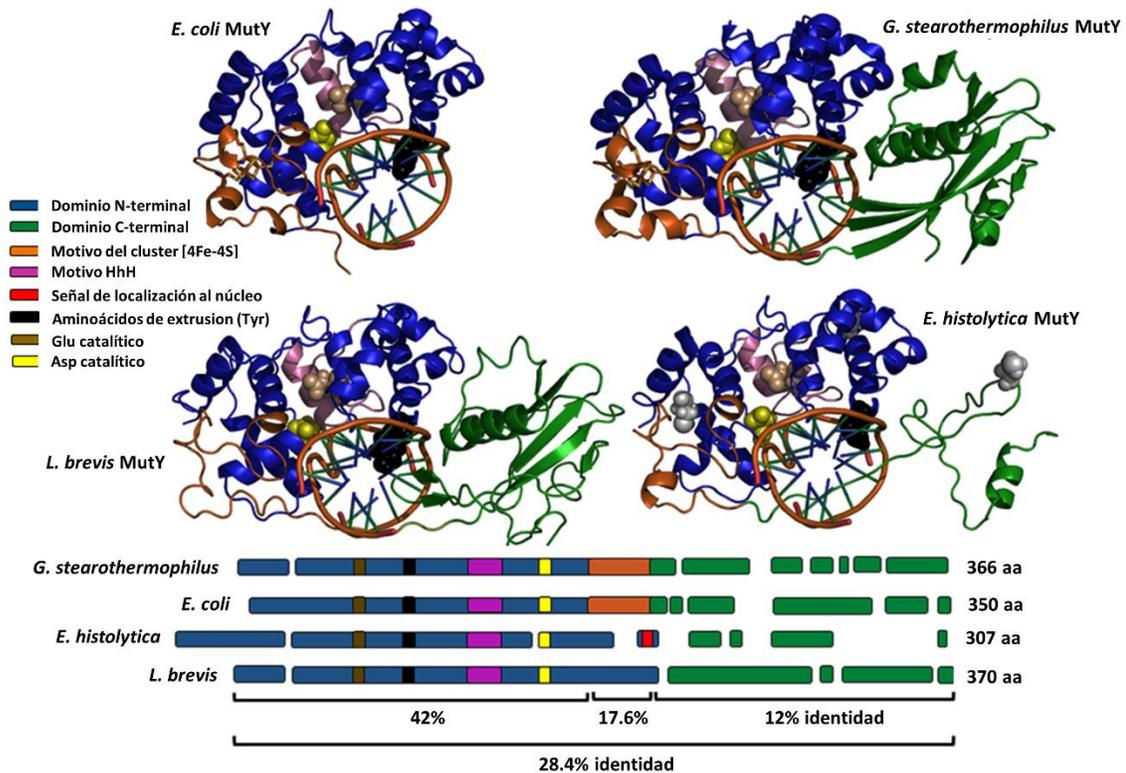


Figura 28. Comparación estructural y de secuencia de MutYs canónicas y novedosas. Parte superior: estructuras cristalográficas del dominio catalítico de MutY de *Escherichia coli* (PDB ID: 1MUY) y de *Geobacillus stearothermophilus* (PDB ID: 1RRQ). Parte media: modelado estructural por homología de la MutY de *Lactobacillus brevis* y de *Entamoeba histolytica*. Para el modelado por homología se utilizó como templado la MutY de *G. stearothermophilus*. Parte inferior: alineamiento esquemático de aminoácidos que indica el porcentaje de identidad entre regiones. El dominio catalítico se muestra en azul, el motivo del clúster [4Fe-4S] en naranja, el motivo HhH en rosa y el dominio C-terminal en verde. Las posiciones correspondientes a los residuos catalíticos de Glu y Asp están en café y amarillo, el residuo de extrusión en negro.

El núcleo catalítico de la MutY de *E. coli* (EcMutY) comprende los primeros 225 aminoácidos, que incluyen los carboxilatos catalíticos E37 y D138, un residuo de intercalación/extrusión Y82, un motivo hélice-horquilla-hélice y un motivo de clúster [4Fe-4S] (Manuel, Czerwinski et al. 1996, Manuel and Lloyd 1997, Guan,

Manuel et al. 1998). Estructuralmente, el núcleo catalítico consiste de seis hélices formando un barril y un motivo de clúster [4Fe-4S] (Guan, Manuel et al. 1998, Fromme, Banerjee et al. 2004). Los núcleos catalíticos de EhMutY y LbMutY comparten un 38.3% de identidad con respecto a EcMutY y *Geobacillus stearothermophilus* MutY (GsMutY) (**Figura 28**). Los residuos putativos del sitio activo son: E48 y D149 para LbMutY, y E78 y D179 para EhMutY. El residuo de intercalación/extrusión en LbMutY corresponde a Y87. Este residuo es también usado en endonucleasa VIII (Zharkov, Golan et al. 2002) y la glicosidasa AGG (Lau, Schärer et al. 1998). En el caso de la EhMutY, el residuo de intercalación/extrusión es el F123, como en la MutM (Gilboa, Zharkov et al. 2002). Los residuos aromáticos de Tyr y Phe están involucrados en la extrusión de la base de la doble hélice del ADN para posicionarla en la cavidad catalítica, participando a su vez en la afinidad y especificidad por el sustrato (Livingston, Kundu et al. 2005).

De acuerdo con el modelado estructural por homología, las principales diferencias de LbMutY y EhMutY están localizadas en el motivo del clúster [4Fe-4S] y en el dominio C-terminal. El motivo del clúster [4Fe-4S] es un elemento altamente conservado entre las MutY descritas hasta la fecha y las cisteínas de coordinación también conservan el patrón de espaciamiento bastante característico. Sin embargo, una particularidad de la LbMutY y la EhMutY es que estas glicosilasas no contienen los ligandos característicos para la coordinación del cofactor (**Figure 26A y 28**). Los grupos tioles de las cisteínas son los ligandos más comunes para este cofactor. No obstante, también pueden participar como ligandos los residuos de histidina, glutamina, serina y arginina (Moulis, Davasse et al. 1996, Golinelli, Chmiel

et al. 1999, Meyer 2008). En caso de que un ligando de cisteína sea sustituido, al menos 3 más deben de participar en la coordinación, aunque en algunos casos 2 ligandos de cisteínas pueden soportar la coordinación del cofactor (Moulis, Davasse et al. 1996, Moulis, Sieker et al. 1996, Sazanov and Hinchliffe 2006, Meyer 2008). LbMutY no contiene ninguna cisteína en su secuencia de aminoácidos, por lo tanto, es teóricamente imposible que pueda coordinar un clúster [4Fe-4S]. EhMutY contiene tres cisteínas, pero de acuerdo con el modelado estructural, estas están a más de 34 Å de distancia entre ellas, por lo que es físicamente imposible que la coordinación del cofactor se lleve a cabo (**Figura 28**). El asa del clúster [4Fe-4S] (asa FCL por sus siglas en inglés) es un elemento importante para el reconocimiento del sustrato, debido a que algunas mutaciones en los ligandos y en residuos positivamente cargados afectan la capacidad de unión al ADN (Golinelli, Chmiel et al. 1999, Chepanoske, Golinelli et al. 2000, Messick, Chmiel et al. 2002). A pesar de la baja identidad en secuencia de aminoácidos del motivo del clúster [4Fe-4S] (17.6%), el andamio global, y los residuos positivos importante del asa FCL están conservados en la EhMutY y LbMutY (**Figure 28B y C**). EcMutY contiene tres residuos positivos en el asa FCL (R194, K196 y K198) (Chepanoske, Golinelli et al. 2000). EhMutY por su parte presenta dos EhMutY (K232 and K233) y LbMutY un solo residuo cargado positivamente (K200).

El elemento más variable en LbMutY y EhMutY son sus dominios C-terminal. El dominio C-terminal de LbMutY está compuesto de 7 láminas β y una hélices α . En el caso de la EhMutY, solamente una pequeña fracción del dominio C-terminal es modelada como una hélice α , mientras que el resto es modelado como región

no estructurada. Es de notarse que el dominio C-terminal de EhMutY es considerablemente pequeño, comprendiendo 57 residuos. En LbMutY, GsMutY y EcMutY son 141, 136 y 126 aminoácidos, respectivamente (**Figure 28**). El dominio C-terminal de MutY esta estructural y evolutivamente relacionado con MutT. (Tajiri, Maki et al. 1995, Bessman, Frick et al. 1996). El dominio C-terminal de EcMutY juega un rol importante en la especificidad por sustrato en las lesiones 8oxoG:A y G:A (Manuel, Czerwinski et al. 1996, Manuel and Lloyd 1997, Noll, Gogos et al. 1999, Volk, House et al. 2000). La versión silvestre de EcMutY procesa 8oxoG:A y G:A 30 y 40 veces más eficiente que su contraparte deletional del dominio C-terminal (Manuel, Czerwinski et al. 1996, Manuel and Lloyd 1997). La estructura de GsMutY con sustrato muestra que la fracción amino de la S308 del dominio C-terminal interacciona con el O8 del 8oxoG (Fromme, Banerjee et al. 2004, Lee and Verdine 2009), propiciando así una región determinante para la identificación de la lesión.

4.2.3. Corroboración de la ausencia del clúster [4Fe-4S].

Los ligando de cisteínas (CX₆CX₂CX₅C) que coordinan el cofactor están ampliamente conservados en las MutYs canónicas y mutaciones en estos residuos disminuyen la actividad enzimática así como su estabilidad (Golinelli, Chmiel et al. 1999, Messick, Chmiel et al. 2002). Con la intención de hacer un estudio comparativo entre MutYs que contienen el clúster [4Fe-4S] con aquellas con las que se predice su ausencia, se llevó a cabo la clonación de los genes de EcMutY, LbMutY y EhMutY. Las enzimas silvestres fueron sobre expresadas en la cepa JM101 *muty*⁻ de *E. coli* (Qing, Ma et al. 2004), cuyo gen de *muty* está interrumpido.

De esta forma se evita la posibilidad de la presencia de una contaminación de la MutY endógena de *E. coli* en los ensayos bioquímicos. Las proteínas recombinantes fueron purificadas por cromatografía de afinidad a níquel y de intercambio catiónico. Las proteínas recombinantes puras se pueden observar en la figura **Figure 29A**.

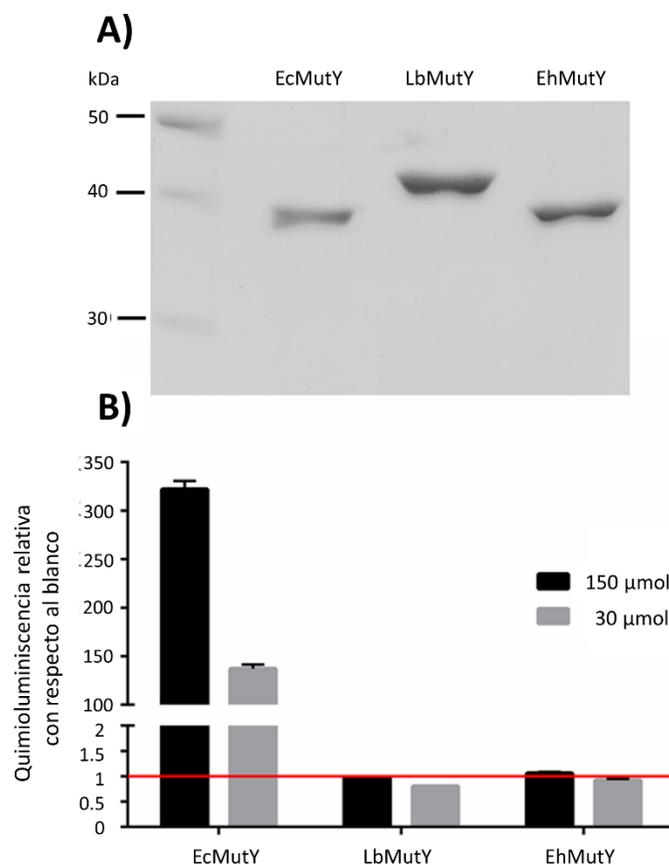


Figura 29. SDS-PAGE de MutYs purificadas y ensayo de quimioluminiscencia para la detección de clúster [4Fe-4S]. A) Análisis de gel de poliacrilamida al 12% teñido con azul de Coomassie después de cromatografía de afinidad a níquel y cromatografía de intercambio catiónico. B) Prueba de quimioluminiscencia para la detección de clúster [4Fe-4S] basada en la reacción de luminol-H₂O₂-Fe con 150 μmol y 30 μmol de MutYs. Todas las mediciones se dividieron sobre el fondo de quimioluminiscencia (reacción sin enzima) y se reportaron como relación señal/fondo. La línea roja indica el valor de fondo de quimioluminiscencia normalizado.

Para corroborar la ausencia o presencia del clúster [4Fe-4S] en las variantes de MutYs experimentalmente se usó un ensayo quimioluminiscente, principalmente debido a la posibilidad de que uno o más ligandos de cisteínas puedan ser reemplazados por un ligando distante y no conservado a través de un mecanismo de permuta de ligandos en la EhMutY (Kennedy, Kent et al. 1984, Golinelli, Akin et al. 1996). El clúster [4Fe-4S] puede ser detectados por diferentes enfoques experimentales, identificando una coloración amarilla en la solución proteica (Manuel and Lloyd 1997, Yeeles, Cammack et al. 2009, Engstrom, Partington et al. 2012), espectroscopia de los grupos tiol-Fe (III) de los ligandos de cisteína (Messick, Chmiel et al. 2002, Yeeles, Cammack et al. 2009), por el efecto cromogénico de agentes como sulfonato de batofenantrolina, fureno o luminol (Pieroni, Khalil et al. 2001, Högbom, Ericsson et al. 2005, Yeeles, Cammack et al. 2009) y por resonancia electrónica paramagnética (Messick, Chmiel et al. 2002, Boal, Yavin et al. 2005, Yeeles, Cammack et al. 2009). La solución purificada de LbMutY y EhMutY durante todo el proceso de purificación se observó incolora, además de que la señal espectroscópica a 410 nm característica del cofactor no fue observada. Para corroborar aún más la carencia del clúster se usó un ensayo quimioluminiscente que se basa en el hecho de que el luminol, en presencia de Fe y H₂O₂ produce quimioluminiscencia (Högbom, Ericsson et al. 2005). Una ventaja de este ensayo es la capacidad de detectar los diferentes estados de oxidación comunes del hierro (Fe²⁺ y Fe³⁺). Para incrementar la sensibilidad del ensayo, cada solución proteica fue tratada con 8 M de urea para inducir desnaturalización y liberación del Fe. El ensayo quimioluminiscente fue llevado a cabo con 30 y 150 μmol de enzima recombinante (El límite inferior de la sensibilidad de este enfoque experimental es

0.1 μmol). Los valores de quimioluminiscencia relativa con respecto al blanco (lectura sin proteína) para EcMutY fueron 321 y 137 para la detección con 150 y 30 nmol de proteína, respectivamente. Los valores relativos para LbMutY y EhMutY fueron similares al fondo (**Figure 29B**). Así, mediante el enfoque quimioluminiscente se pudo corroborar la presencia del clúster [4Fe-4S] en EcMutY y descartar su presencia en LbMutY y EhMutY.

4.2.4. Actividad *in vitro* y mutaciones del sitio activo.

El mecanismo catalítico de MutY involucra dos carboxilatos (E43 y D144) y un residuo aromático (Y82) para lanzar la adenina apareada con 8oxoG hacia la cavidad catalítica (Fromme, Banerjee et al. 2004, Livingston, Kundu et al. 2005, Lee and Verdine 2009, Brunk, Arey et al. 2012). La distancia entre el clúster [4Fe-4S] y la cavidad catalítica es de 19.1 Å. Aunque esta distancia impide la interacción física y química del clúster con el sustrato durante la catálisis, este cofactor es considerado un elemento importante para la catálisis (Porello, Cannon et al. 1998, Golinelli, Chmiel et al. 1999, Messick, Chmiel et al. 2002, Lu and Wright 2003). Análisis de secuencia de MutYs hechos en estudios anteriores sugieren la existencia de MutYs con carencia del clúster (Samrakandi and Pasta 2000, Messick, Chmiel et al. 2002, White and Dillingham 2012), pero un aspecto que no queda claro es si estas MutYs no canónicas mantienen su actividad de glicosilasas. Para abordar esta cuestión, se evaluó la actividad *in vitro* de las enzimas purificadas. El ensayo se realizó usando como sustrato de ADN de doble cadena con 8oxoG y G apareadas con bases canónicas. Para este ensayo, se usó un oligonucleótido radiactivamente marcado en el extremo 5' de 45-mer conteniendo A, G, C o T en la

posición 30. Este oligonucleótido marcado radioactivamente se alineo con su complementario donde en la posición 30 contenía una G o la lesión 8oxoG. Con la intención de observar claramente la actividad, el ensayo fue llevado a cabo usando un exceso de enzima (~200 nM) y 1 nM de sustrato de ADN. LbMutY, EhMutY, y EcMutY mostraron tener actividad sobre 8oxoG:A y G:A. En la **figura 30A** se puede observar la liberación del producto de 30-mer de la reacción completa en presencia de NaOH (0.2 M). Como anteriormente se ha reportado, ADN con la lesión 8oxoG:A y el mal apareamiento G:A son sustratos comunes para MutY (Michaels, Tchou et al. 1992, Lu, Tsai-Wu et al. 1995). Es importante resaltar que la LbMutY y la EhMutY fueron sobre expresadas en la cepa JM101 *muty*⁻ de *E. coli*, descartando así contaminación endógena de MutY de *E. coli*. Es así, que estos datos bioquímicos corroboran que EhMutY y LbMutY son glicosilasas de adenina capaces de reconocer específicamente y procesar la adenina mal apareada con 8oxoG o G sin necesidad de un clúster [4Fe-4S].

El mecanismo catalítico de la MutY es bien conocido, donde la adenina es separada por extrusión de la doble hélice de ADN y colocada en la cavidad catalítica. Una vez posicionada la adenina en cavidad catalítica, dos moléculas de H₂O estabilizadas por dos carboxilatos participan en la catálisis (Brinkmeyer, Pope et al. 2012, Brunk, Arey et al. 2012). Estos carboxilatos importantes son conservados en LbMutY (E48 y D149) y en EhMutY (E78 y D179). Para evaluar si estos residuos eran funcionalmente conservados, se realizaron las mutaciones simples y dobles a alanina. Las mutantes del sitio catalítico fueron expresadas como se realizó para las silvestres y se evaluó su actividad enzimática. Ninguna de las

mutantes evaluadas fueron activas en dúplex de ADN con la lesión 8oG:A (**Figura 30B**). Solamente la mutante E78A de EhMutY presentó una actividad residual, pero esta puede ser atribuida a una sutil variación en el mecanismo catalítico general, como el reclutamiento de otro carboxilato para asistir la reacción de forma ineficiente.

Las glucosilasas de ADN pueden ser clasificadas como monofuncionales y bifuncionales. Las glicosilasas bifuncionales, como la OGG1 y la endonucleasa III, usan un carboxilato y un grupo amino activo (comúnmente de una lisina) para realizar la hidrólisis del enlace N-glicosídico concomitante con la escisión del enlace fosfodiéster 3' mediante una reacción de liasa, mientras que las glicosilasa monofuncionales no poseen actividad liasa (Williams and David 1999, Williams and David 2000). Para medir la actividad *in vitro* de las glicosilasa monofuncionales es necesario añadir NaOH o piperidina al final de la reacción para escindir el enlace fosfodiéster a través de β y σ -eliminación. En caso de las glicosilasas bifuncionales la reacción de β -eliminación es llevada a cabo por la Lys activa, por lo que la actividad puede ser observada sin necesidad de añadir NaOH. Para identificar las características monofuncionales o bifuncionales de EhMutY y LbMutY, medimos su actividad *in vitro* con 8oxoG:A deteniendo la reacción con o sin NaOH (0.2 M). La escisión de ADN por EcMutY, LbMutY y EhMutY se observa claramente cuando se añade NaOH al final de la reacción. Sin embargo, cuando el NaOH está ausente, observando solamente una actividad residual (**Figura 30C**). Por lo tanto, se puede concluir que LbMutY y EhMutY son glicosilasa monofuncionales como es comúnmente reportado (Williams and David 1999, Williams and David 2000).

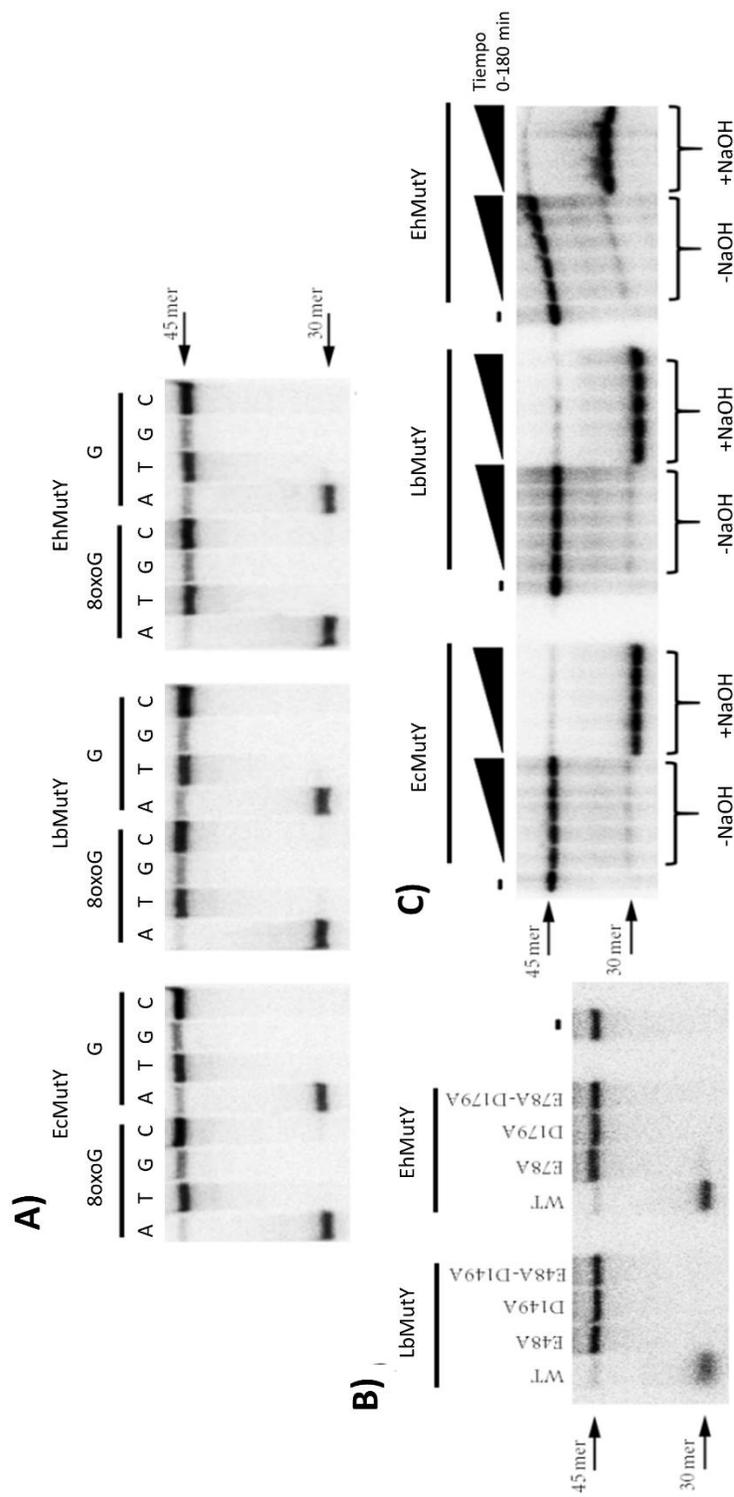


Figura 30. MutYs sin clúster [4Fe-4S] son glicosilasa activas *in vitro*. A) Electroforesis en gel desnaturalizante analizada a través de fosfoimágenes que ilustra la actividad de glicosilasa sobre 8oxG y G apareado con bases canónicas. EcMutY, LbMutY y EhMutY muestra una actividad sobre 8oxG:A y G:A. Los detalles experimentales se describen en la sección de materiales y métodos. C) Evaluación de la actividad *in vitro* de mutantes simples y dobles de los carboxilatos implicados en la catálisis. B) Determinación de actividad de liasa con 1 nM de sustrato que contiene 8oxG: A. Las LbMutY y EhMutY son glicosilasas monofuncionales. El sustrato intacto (45 mer) y el producto (30 mer) se indican con flechas.

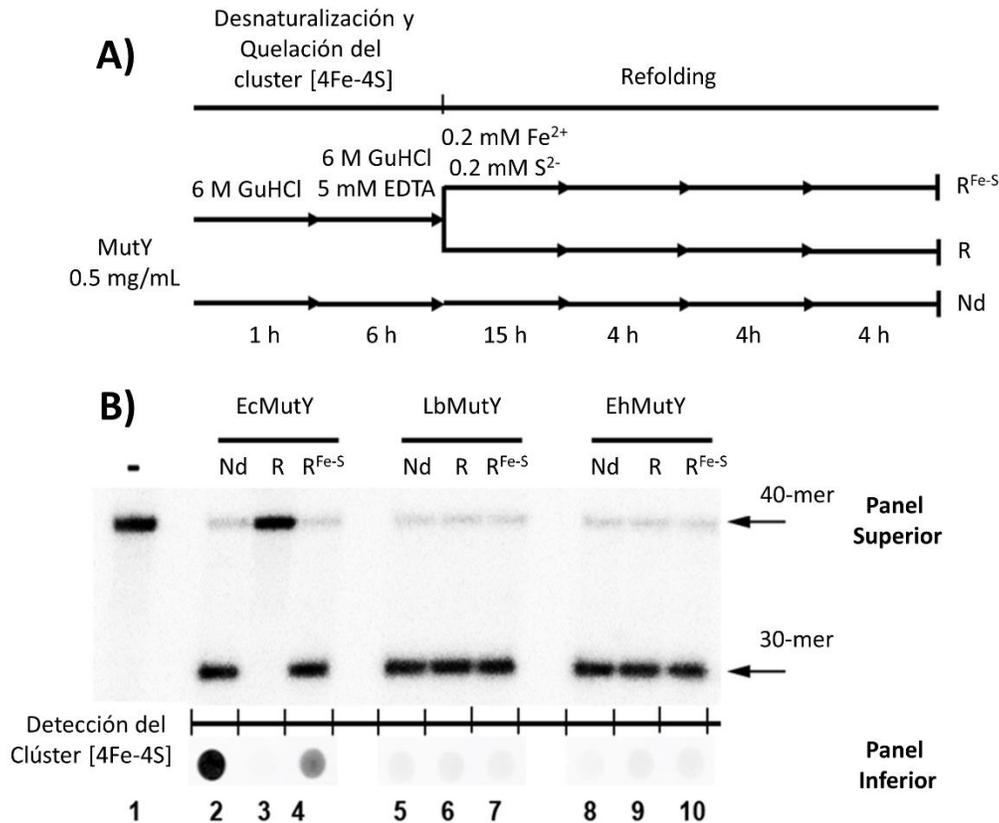


Figura 31. Experimento de reconstrucción *in vitro* del clúster [4Fe-4S]. A) Representación esquemática del experimento de reconstrucción *in vitro* del clúster [4Fe-4S] diseñado por David y colaboradores. La duración de cada paso experimental se indica con una flecha. La desnaturalización y la quelación del clúster [4Fe-4S] se realizó en solución amortiguadora A (Tris-HCl 50 mM [pH 7,6], NaCl 200 mM, 2-mercaptoetanol 2 mM y EDTA 0,1 mM) con la adición de los componentes indicados encima de cada flecha. Las muestras de MutY sometidas a procedimientos de quelación/replegamiento en ausencia de los iones de azufre y hierro necesarios para la reconstrucción del clúster [4Fe-4S] se indican como R. Las muestras que se volvieron replegar en presencia de iones ferrosos y de sulfuro se indican como R^{Fe-S} y un control positivo no desnaturalizado pero sometido a los pasos de diálisis en solución amortiguadora A como Nd. B) Actividad de glicosilasa *in vitro* (panel superior) y detección de quimioluminiscencia cualitativa (panel inferior) del clúster [4Fe-4S] derivado de un experimento de reconstrucción. La actividad de la glicosilasa se llevó a cabo con 1 nM de dúplex de ADN de con lesión 8oG:A y 500 nM de MutY durante 30 min a 37 °C. La actividad glicosilasa está presente en el control EcMutY en los tratamientos Nd y R^{Fe-S}, pero no en el tratamiento R (carriles 2 a 4). La actividad glicosilasa está presente en todos los tratamientos para EhMutY y LbMutY (carriles 5 a 10).

La cantidad de MutY expresada de forma recombinante ensamblada con un clúster [4Fe-4S] puede variar debido a factores de cultivo y de purificación (Nakamura, Saeki et al. 1999, Gupta, Aggarwal et al. 2006, Jaganaman, Pinto et al. 2007). Por lo tanto, las MutYs expresados pueden contener un clúster [4Fe-4S] en una cantidad por debajo de los límites de detección de los métodos de quimioluminiscencia y espectroscopia. Para investigar esta posibilidad, se realizó un ensayo de reconstrucción *in vitro* del clúster [4Fe-4S] diseñado por el grupo de la Dra. David y colaboradores (Porello, Cannon et al. 1998). En este experimento las MutYs son desnaturalizadas y los componentes del clúster [4Fe-4S] se eliminan por quelación y diálisis. El protocolo de reconstrucción de clúster se muestra esquemáticamente en la **Figura 31A** (Porello, Cannon et al. 1998). Si la actividad MutY depende del clúster [4Fe-4S], la actividad enzimática se agotará con este procedimiento. Cuando EcMutY se somete al proceso de desnaturalización/quelación/replegamiento (tratamiento denominado R), su actividad enzimática se pierde, mientras que LbMutY y EhMutY mantienen su actividad de glicosilasa de ADN (**Figura 31B panel superior**). En todas las muestras del tratamiento R el clúster [4Fe-4S] no es detectado por el ensayo de quimioluminiscencia, lo que indica su ausencia desde el principio o que el hierro se retiró correctamente (**Figura 31B panel inferior**). Cuando se replegó EcMutY en presencia de iones de azufre y hierro (tratamiento denominado R^{Fe-S}) se recuperó su actividad y detección, indicando que el protocolo es adecuado para obtener una población de EcMutY correctamente plegada y con cofactor. Estos datos apoyan la

idea de que el clúster [4Fe-4S] no es necesario para la actividad catalítica de EhMutY y LbMutY.

4.2.5. Cinética de intercambio simple y múltiple

Las glicosilasas monofuncionales permanecen unidas al producto (sitio abásico) después de la hidrólisis. Este hecho hace que la liberación del complejo enzima-producto sea un paso limitante durante la catálisis, lo que indica que los parámetros de Michaelis-Menten K_m y K_{cat} no son adecuados para el análisis cinético. Por lo tanto, en este estudio se obtuvieron constantes cinéticas en condiciones de intercambio simple y múltiple. Con este enfoque es posible determinar la constante de velocidad para la reacción enzimática (k_2) y la liberación del producto o recambio (k_3) (Porello, Leyes et al. 1998). Para la cinética de intercambio simple y múltiple, se utilizó como sustrato un dúplex de ADN con apareamiento 8oxoG:A. Decidimos llevar a cabo nuestros ensayos cinéticos a 21 °C para obtener datos medibles con 8oG:A, ya que la catálisis de la EcMutY a 37°C genera un consumo de sustrato total antes de 1 min de reacción, lo que complica su medición.

Para evaluar las propiedades de unión de la MutY al ADN se realizó un ensayo de anisotropía de fluorescencia utilizando un oligonucleótido con un sitio abásico (THF) etiquetado con fluoresceína e hibridado con un oligonucleótido que contenía 8oxoG. Los datos se analizaron mediante regresión no lineal con ajuste de pendiente de Hill para observar si hay un comportamiento cooperativo. Dado que los procesos cooperativos resultan en múltiples valores de K_d , preferimos referir el parámetro de afinidad de unión como $[P]_{1/2}$ en lugar de K_d . Las mediciones de

afinidad de unión de EcMutY dieron como resultado una disminución de hasta 40% en la señal detectada comparando con EhMutY y LbMutY (datos no mostrados). Este fenómeno ha sido observado previamente para EcMutY y es asociado a la presencia del clúster [4Fe-4S] ya que puede atenuar la emisión de fluorescencia (Lee, Bai et al. 2004). Por lo tanto, los datos de anisotropía se ajustaron para compensar la disminución del quantum-yield del complejo de proteína-sustrato en relación con el sustrato libre (Dandliker, Hsu et al. 1981, Lee, Bai et al. 2004).

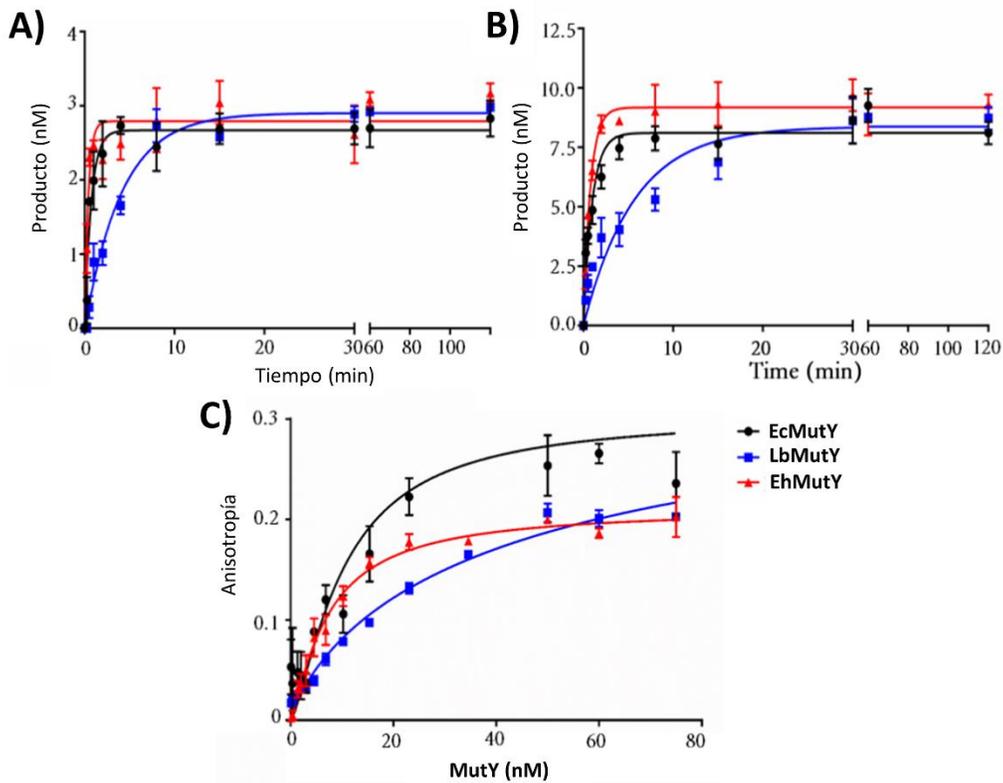


Figura 32. Cinética de intercambio múltiple y simple sobre 8oxoG:A e isotermas afinidad. A) Cinética de intercambio múltiple. Para este experimento se incubaron 10 nM de sustrato y 2,5 nM de MutY activa a 21 ° C en Tris 40 mM [pH 7,5], EDTA 10 mM, BSA 0,1 mg / ml y NaCl 150 mM. B) Cinética de intercambio simple. Para este experimento se utilizaron 50 nM de enzima activa. Para cada experimento se tomaron alícuotas de 8 μ L de reacción a 0.25, 0.5, 1, 2, 4, 8, 15, 30, 60 y 120 min. Todos los experimentos se realizaron por triplicado. Las reacciones se detuvieron con NaOH 0,2 M y se calentaron a 90 ° C durante 5 min. La concentración del producto se representa en función del tiempo. C) Isotermas de unión obtenidas por titulación de 1 nM de dúplex marcado con fluoresceína con sitio abásico (THF) apareado con 8oxoG con diferentes concentraciones de MutY que varían de 0,1 a 75 nM.

Las constantes de afinidad de EcMutY, LbMutY y EhMutY fueron obtenidas usando un dúplex de ADN con 8oxoG:THF. Para lo cual, 1 nM de sustrato fue titulado con diferentes concentraciones de MutY activa que oscilaban entre 0.1-75 nM. Los isotermas de unión (**Figura 32C**) se ajustaron a un modelo de unión específica

de sitio único con ajuste a la pendiente de Hill para determinar los valores de $[P]_{1/2}$. EhMutY tuvo la mayor afinidad por el el dúplex 8oxoG:THF, con un $[P]_{1/2}$ de 6.83 nM, seguido de EcMutY con 12.59 y LbMutY con 41.37 nM. Se han descrito valores de afinidad similares para MutY de *Schizosaccharomyces pombe* y UDG de *E. coli* con un ADN doble cadena conteniendo sitio abásico (Grippon, Zhao et al. 2010, Luncsford, Chang et al. 2010). Sin embargo, usando métodos radioactivos más sensibles se puede lograr obtener valores de afinidad <45 pM para el dúplex abásico:8oxoG (Schärer, Nash et al. 1998, Chmiel, Golinelli et al. 2001, Pope and David 2005).

Para determinar las constantes de velocidad, se midió el porcentaje de enzimas activas por la amplitud de la aceleración catalítica en condiciones de intercambio múltiples. El porcentaje de MutYs activas varió de 28 a 66% (**Tabla 4**). La constante de velocidad k_3 se derivó de la pendiente de la porción lineal de las curvas en condiciones de intercambio múltiple con 2,5 nM de MutYs activas y 10 nM de sustrato con la lesión 8oG:A. Todos los valores cinéticos se muestran en la **Tabla 4**. Las constantes de velocidad k_3 fueron de 0,002, 0,006 y 0,005 min^{-1} para EcMutY, LbMutY y EhMutY, respectivamente. Estos valores lentos de k_3 indican que los eventos de intercambio de estas enzimas es limitado y la afinidad por el producto de MutY es más alta que por la del sustrato. Se han descrito datos similares para EcMutY y su homólogo humano (Porello, Leyes et al. 1998, Chmiel, Golinelli et al. 2001, Livingston, Kundu et al. 2005, Kundu, Brinkmeyer et al. 2009).

Para evaluar la velocidades de reacción (k_2) se realizaron cinéticas de intercambio simple con 10 nM de sustrato duplex de 8oG:A y 50 nM de enzima

activa a 21°C. Bajo estas condiciones, los valores de k_2 resultantes fueron 1.324, 0.323 y 1.485 min^{-1} para EcMutY, LbMutY y EhMutY, respectivamente. Es digno de mención que realizar el ensayo a 21 °C nos permitió obtener valores k_{obs} factibles durante las cinéticas de intercambio múltiple y simple. Livingston y sus colegas llevaron a cabo un estudio cinético comparativo de EcMutY con diferentes temperaturas y encontraron que a 25 °C el valor de k_2 para 8oxoG:A era 5.7 min^{-1} , mostrando una reducción de 3 veces la k_2 con respecto los valores a 37 °C (Livingston, Kundu et al. 2005). Con respecto a esto, observamos una reducción de al menos un orden de magnitud para la constante de velocidad en comparación con la obtenida a 37 °C (EcY $k_{\text{obs}} > 10 \text{ min}^{-1}$) (Porello, Leyes et al. 1998, Chmiel, Golinelli et al. 2001, Livingston, Kundu et al. 2005). Como se puede ver en la **Figura 32**, el producto máximo liberado para LbY se alcanza en aproximadamente a 10 min, que a diferencia de EcMutY y EhMutY donde se alcanza a aproximadamente 4 min. La constante k_2 es una expresión de las capacidades catalíticas y de unión. Por lo tanto, la baja tasa catalítica de LbY se puede entender prestando atención a los valores k_{obs} y $[P]_{1/2}$. Como puede observarse, la constante de disociación de LbMutY es 2.11 veces mayor que el valor de $[P]_{1/2}$ de EcMutY, por lo que la baja velocidad catalítica para LbMutY podría explicarse por una afinidad reducida por sustrato. Curiosamente, a pesar de que EhMutY carece del clúster [4Fe-4S], muestra una mayor afinidad de sustrato que EcMutY o LbMutY, y una tasa catalítica similar a EcMutY

Table 4. Constantes cinéticas para MutY con 8oxoG:A y constante de afinidad por 8oxoG:THF

		Cinetica de intercambio simple ^b			Cinetica de intercambio multiple ^a	
MutY	Enzima activa (%) ^a	k_2 (min ⁻¹)	Relativo /EcMut Y	k_{obs} (min ⁻¹)	k_3 (min ⁻¹)	[P] ^{1/2} (nM) ^b
EcMutY	38.7, 66	1.324±0.34	1	1.057±0.2	0.002±0.001	12.593±2
LbMutY	28.1, 47.4	0.323±0.02	0.2	0.177±0.0	0.006±0.002	41.370±8
EhMutY	51.8, 42.8	1.485±0.17	1.1	1.306±0.1	0.005±0.001	6.837±2

Hay algunos factores que pueden afectar la actividad de MutY. Por ejemplo, la delección del dominio C-terminal genera una disminución en la velocidad catalítica de al menos dos órdenes de magnitud (Chmiel, Golinelli et al. 2001). Otro aspecto importante es la integridad del motivo del clúster [4Fe-4S]. Un ligando de cisteína puede ser sustituido por residuos de His o Ser sin variación importante de los parámetros catalíticos, sin embargo el efecto de estas sustituciones se exagera en ensayos *in vivo* (Golinelli, Chmiel et al. 1999, Messick, Chmiel et al. 2002). Los residuos R194, K196 y K198 son elementos importantes para el reconocimiento del ADN, y las mutaciones a Ala de estos residuos reducen la afinidad por el sustrato (Chepanoske, Golinelli et al. 2000). Según el modelo estructural por homología (**Figura 26**), el FCL de la LbMutY está estructuralmente conservado, pero sólo alberga un residuo cargado positivamente (K200), mientras que EhMutY muestra dos (K232 y K233). Las principales diferencias en la secuencia y en estructura se centran en la región del clúster [4Fe-4S] y en el dominio C-terminal. De hecho, el dominio C-terminal para EhMutY se muestra como una región intrínsecamente no

estructurada. Sin embargo, estas diferencias no proporcionan ninguna desventaja catalítica. De forma excepcional, la falta del clúster [4Fe-4S] no tiene ninguna repercusión en los parámetros catalíticos en estas MutYs novedosas carentes del clúster [4Fe-4S].

4.2.6. Ensayo de Frecuencia de mutación: ensayo *in vivo*.

La actividad *in vitro* de LbMutY y EhMutY es clara, no obstante, es necesario corroborar si estas enzimas son activas *in vivo*. Para hacer frente a esto, se midió la frecuencia de mutación con base a un ensayo de resistencia a rifampicina en *E. coli* (Golinelli, Chmiel et al. 1999, Chmiel, Golinelli et al. 2001, Wolff, Kim et al. 2004). La rifampicina interactúa con una región específica de la subunidad β de la ARN polimerasa (β -Rpo), inhibiendo la transcripción (Hartmann, Honikel et al. 1967, Severinov, Soushko et al. 1993, Garibyan, Huang et al. 2003). Por lo tanto, cuando hay una alta frecuencia de mutaciones (dada en nuestro caso por la falta de *muty* o la presencia de una variante defectuosa) la probabilidad de que las mutaciones se produzcan en la región de interacción Rpo-rifampicina aumenta. Así, la actividad de la ADN glicosilasa *in vivo* se determina cuantificando la aparición de colonias resistentes a rifampicina y expresado como frecuencia de mutación (f). Los ensayos de resistencia a rifampicina se realizaron en cepa mutante de *E. coli* CC104 *muty*⁻ *mutm*⁻. Esta cepa es sensible a las mutaciones dependientes de 8oxoG, ya que carece de casi toda la maquinaria necesaria para repararlas. Los valores de f se presentan en la **Tabla 5**. Como el vector de expresión pCold expresa proteínas recombinantes preferentemente a bajas temperaturas (por debajo de 18 °C) y los ensayos *in vivo* se llevaron a cabo a 37 °C, se utilizó como control el gen de la

EcMutY clonado en un pUC (PKK223-3) (Brosius and Holy 1984, Manuel and Lloyd 1997). Las cepas de *E. coli* CC104 *muty*⁻ *mutm*⁻ sin transformación y transformadas con vector vacío se usaron como controles negativos. El control positivo pKKEcY y pColdEcMutY presentaron una reducción similar en sus valores de *f*, por debajo del 98% con respecto a los controles negativos. Por lo tanto, las MutYs recombinantes clonadas en un vector pCold son adecuadas para medir la actividad *in vivo* a 37 °C. Las construcciones pColdLbMuTY y pColdEhMuTY fueron capaces de disminuir la frecuencia de mutación en la misma medida que pColdEcY y pKKEcY. Curiosamente, LbMuTY, a pesar de tener una reducción en la velocidad de hidrólisis *in vitro* (k_2), disminuyó los valores *f* en casi 2 veces mejor en comparación con pColdEcY. Para corroborar que las mutaciones de β -Rpo eran transversiones G:C a T:A, se amplificó por PCR la región de los clústeres 1 y 2 de la β -Rpo (de 1410 a 2235 pb) (Severinov, Soushko et al. 1993). Se encontró que más del 85% de las colonias analizadas tenían mutación en la región 1 o 2 de Rpo- β y todas las mutaciones eran transversiones (Ver apéndice), apoyando la noción de que EhMutY y LbMutY a pesar de carecer el clúster [4Fe-4S] son ADN glicosilasa funcionales.

Tabla 5. Ensayo de resistencia a rifampicina en la cepa CC104 *muty mutm*.

Plásmido	$f(\times 10^{-6})^b$	Incremento/ pCold-EcY	Disminución relativa (%) al control (pCold)
None	8.104±3.74	76.7	21.9
pCold	10.386±4.36	98.4	0
pKKEcY	0.043±0.02	0.4	99.5
pColdEcY	0.106±0.08	1	98.9
pColdLbY	0.068±0.04	0.6	99.3
pColdEhY	0.146±0.10	1.3	98.5

4.2.7. La ausencia del clúster [4Fe-4S] puede ser mediada estructuralmente por efecto estérico.

Se ha reportado que la integridad del clúster [4Fe-4S] es esencial para la unión al ADN, el reconocimiento de las lesiones y la actividad en varias ADN glicosilasa (Porello, Cannon et al. 1998, Golinelli, Chmiel et al. 1999, Chepanoske, Golinelli et al. 2000, Hinks, Evans et al. 2002, Messick, Chmiel et al. 2002, Lu and Wright 2003, Boal, Yavin et al. 2005, Engstrom, Partington et al. 2012). Por lo tanto, si el clúster [4Fe-4S] es un componente esencial para las ADN glicosilasas lo albergan, ¿por qué EhMutY y LbMutY son funcionales? La glicosilasa AlkA, un miembro de la superfamilia HhH que repara daños de alquilación, podría haber perdido el clúster [4Fe-4S] durante su historia evolutiva, con un reordenamiento de motivos, compensando estructuralmente su pérdida (Labahn, Schärer et al. 1996). Esta glucosidasa explora las lesiones en el ADN mediante un mecanismo combinado de difusión facilitada y salto (Sliding-hopping) en el que los aminoácidos

cargados positivamente son relevantes para localizar las lesiones (Hedglin and O'Brien 2010, Hedglin, Zhang et al. 2015, Zhang and O'Brien 2015). Otros ejemplos de proteínas implicadas en el metabolismo del ácido nucleico que han perdido este cofactor son la fotoliasa (6-4) (Zhang, Scheerer et al. 2013), tryptophanyl-tRNA sintetasa (Stiebritz 2014) y la helicasa-nucleasa AddAB (Yeeles, Cammack et al. 2009) indicando que este motivo puede ser sustituido estructuralmente. Otro ejemplo de la dispensabilidad de dominios estructurales específicos en ADN glicosilasa es la ausencia del motivo de dedo de zinc de la horquilla β importante para la actividad de glicosilasa en la endonucleasa VII humana (Doubl  , Bandaru et al. 2004). Por lo tanto, el reordenamiento del motivo podr  a ser una soluci  n estructural en estas nuevas MutYs. Una posible pista es la variabilidad en la secuencia primaria encontrada en la regi  n donde el cl  ster [4Fe-4S] deber  a de estar (**Figuras 26 y 28**). Cuando se comparan estas MutYs novedosas con las can  nicas, la secuencia primaria del dominio catal  tico es altamente conservada y la variabilidad en la secuencia de amino  cidos aumenta en la regi  n que corresponde al motivo del cl  ster y el dominio C-terminal (**Figura 28**). El volumen de van der Waals de un cl  ster [4Fe-4S] es $\sim 550 \text{ \AA}^3$ (Mitra, George et al. 2013). Comparando la secuencia de amino  cidos nos permitimos plantear la hip  tesis de que el cl  ster [4Fe-4S] podr  a ser necesario para preservar plegamiento de la regi  n adyacente al motivo.

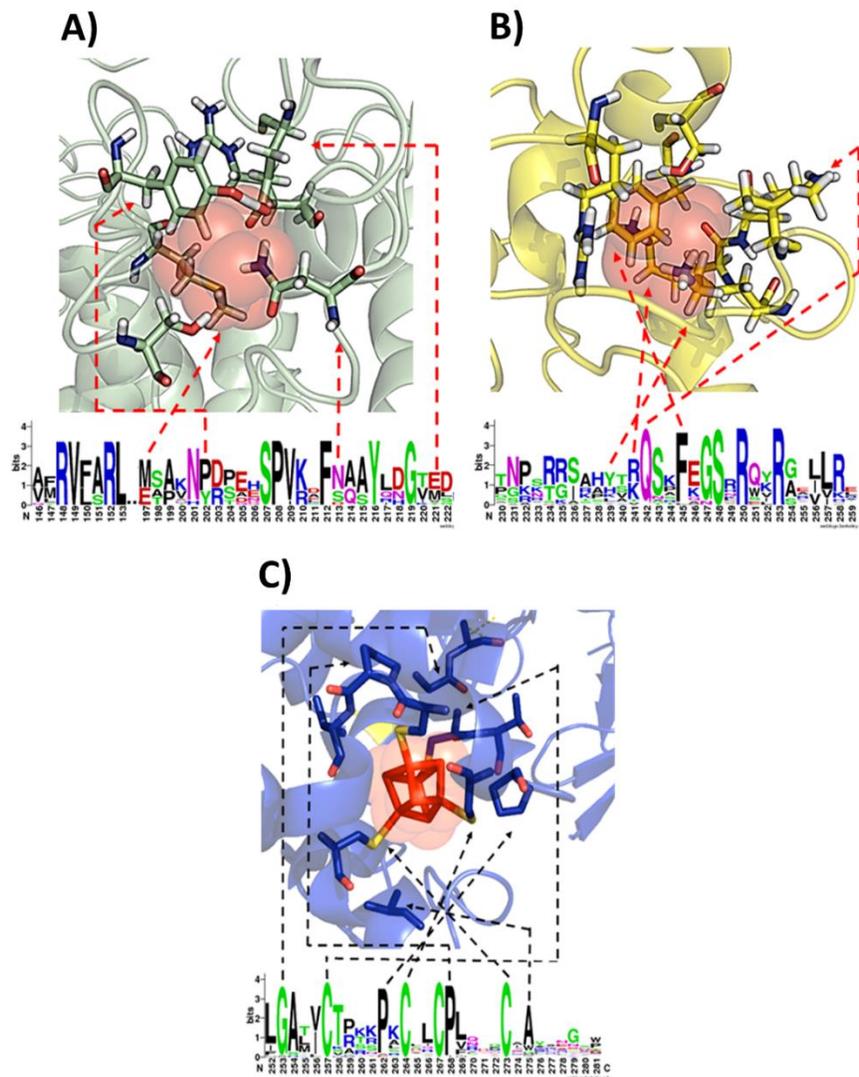
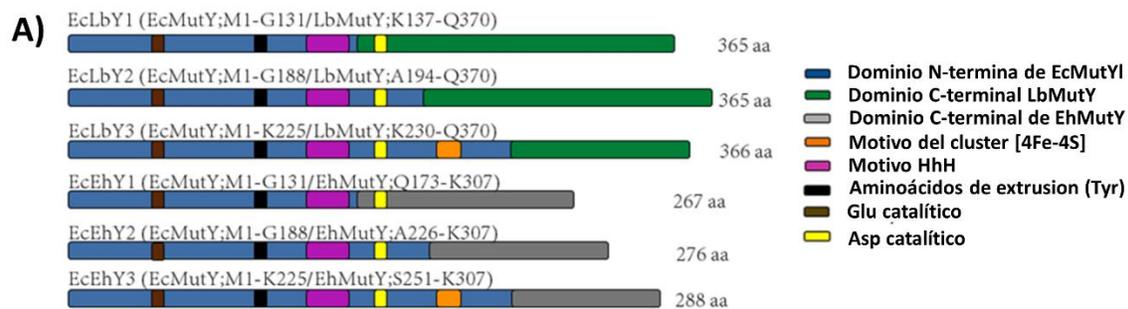


Figura 33. Comparación estructural y de secuencia de logo de la región correspondiente al motivo del clúster [4Fe-4S] entre el clado de Lactobacillales, anaeróbico y MutYs canónicos. La región correspondiente del motivo del clúster [4Fe-4S] de MutY de *L. brevis* (A), *E. histolytica* (B) y *E. coli* (C). Los supuestos residuos implicados en la estabilización estérica del pseudo motivo del clúster [4Fe-4S] en la MutY de *L. brevis* (M197, Y202, N213 y E221) y *E. histolytica* (K239, K24, Q214 y F245) se muestran en forma de bastones y flechas rojas. El volumen de van der Waals del clúster [4Fe-4S] (550 Å³) derivado de la alineación estructural se muestra en esferas rojas. Los residuos altamente conservados del motivo del clúster [4Fe-4S] de EcMutY se indican con flechas negras

Para investigar esta hipótesis, hemos prestado atención a la secuencia de aminoácidos de LbMutY y EhMutY en la región que corresponden al clúster [4Fe-4S]. Si se elimina el clúster [4Fe-4S], la cavidad dejada por su carencia podría ser llenada por aminoácidos voluminosos. Cuatro aminoácidos voluminosos, como Arg, Tyr, Thr, Met o His, cuyos volúmenes de van der Waals son $\sim 140 \text{ \AA}^3$ serían suficiente para llenar la cavidad creada por la carencia del clúster [4Fe-4S]. De hecho, encontramos que los residuos volumétricos son posicionados en el interior de la cavidad dejada por el cofactor en ambas ADN glicosilasa. Los residuos M197, Y202, N213 y E221 y los residuos K239, K24, Q214 y F245 posiblemente estabilizan localmente la región correspondiente en LbMutY y EhMutY, respectivamente (**Figura 33**). La mayor parte de las MutYs del clado de Lactobacillales y del clado anaeróbicos conservan la característica de la cadena lateral grande de los residuos que estabilizan la cavidad dejada por la carencia del clúster [4Fe-4S]. Estas observaciones sugieren que la sustitución del clúster [4Fe-4S] puede llevarse a cabo mediante sustituciones de aminoácidos con cadenas laterales voluminosas que conducen la estabilización del motivo por efecto estérico.



B)

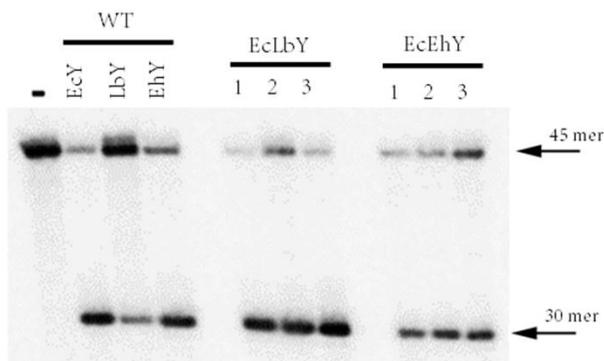


Figura 34. Representación esquemática del diseño de quimeras y su correspondiente actividad de glicosilasa *in vitro* sobre 8oG:A. A) Diseño esquemático de glicosilasa quimérica. Los dos primeros conjuntos de glucosilasas quiméricas carecen del clúster [4Fe-4S]. El tercer conjunto (EcLbY3 y EcEhY3) incluye el dominio catalítico de la mutY de *E. coli* (M1-K225) y se complementa con 140 y 56 residuos del dominio C-terminal de la MutY de *L. brevis* o *E. histolytica*. El primer conjunto de quimeras (EcLbY1 y EcEhY1) carece de 94 residuos del extremo carboxilo del dominio catalítico de EcMutY, que incluye D138 y la región correspondiente del motivo del clúster [4Fe-4S], y se complementa con 233 y 134 residuos de C-terminal dominio de LbMutY y EhMutY. El segundo conjunto de quimeras (EcLbY2 y EcLbY) carece de los últimos 37 residuos del dominio catalítico de *E. coli* incluido el dominio de coordinación del clúster [4Fe-4S] y se complementa con 176 y 81 residuos del dominio C-terminal de LbMutY y EhMutY. B) Ensayo *in vitro* de las MutYs quiméricas (con 1 nM de sustrato 8oG:A a 37 ° C). El sustrato (45 mer) y el producto (30 mer) se indican con flechas negras.

Si la observación de la estabilización por efecto estérico es correcta, debería ser posible sustituir el clúster [4Fe-4S] en MutYs canónicos mediante la construcción de genes quiméricos. Por lo que se diseñó tres conjuntos de genes quiméricos de EcMutY con partes complementarias del dominio C-terminal de LbMutY y EhMutY. En los dos primeros sets de quimeras, el motivo del clúster [4Fe-4S] de EcMutY fue excluido y complementado por la región correspondiente de LbMutY o EhMutY, mientras que en el tercer set se mantuvo el dominio catalítico de EcY (incluido el motivo de clúster [4Fe-4S]) intacto y se añadió el dominio C-terminal de LbMutY y EhMutY. La **Figura 34A** muestra una representación esquemática del diseño de la ADN glicosilasa quiméricas. Los genes correspondientes a las ADN glicosilasas quiméricas se clonaron en un vector de expresión pCold1, se sobreexpresaron en la cepa de *E. coli* de JM101 *muty*⁻ y se purificaron como las MutYs silvestres. Sorprendentemente, todas las MutYs quiméricas fueron activas sobre un sustrato 8oxoG:A (**Figura 34B**). Empero, cuando probamos la actividad *in vivo* de las ADN glicosilasa quiméricas utilizando una cepa mutante de *E. coli* *muty*⁻ *mutM*⁻ CC104, la disminución en la frecuencia de mutación se mantuvo en 42,8 a 61,8%, mientras que EcY intacta redujo el valor de *f* al 98,9% (Tabla 6). La quimera EcLbY3 fue el único constructo que presentó actividad de glicosilasa similar a la silvestre *in vitro* e *in vivo*, sin embargo, contiene un clúster [4Fe-4S]. En general, estos resultados indican que la ausencia del clúster [4Fe-4S] puede ser mitigado por complementación de los dominios C-terminal de LbMutY o EhMutY. Consideramos importante resaltar que estos resultados no descartan la importancia del clúster [4Fe-4S] en el reconocimiento de lesiones en las MutYs canónicas.

Tabla 6. Ensayo de resistencia a rifampicina en la cepa CC104 *muty mutm* de las MutYs quiméricas.

Plásmido	$f(\times 10^{-6})$	Incremento/Cold-EcY	Decremento relativo al control (pCold)
pCold	10.386±4.36	98.4	0
pColdEcY	0.106±0.08	1	98.9
pColdEcLbY1	4.737±1.69	44.8	54.3
pColdEcLbY2	5.935±2.73	56.2	42.8
pColdEcLbY3	0.188±0.05	1.7	98.1
pColdEcEhY1	4.480±1.23	42.4	56.8
pColdEcEhY2	5.219±1.64	49.4	49.7
pColdEcEhY3	3.962±1.44	37.5	61.8

4.2.8. Observaciones finales y perspectivas los estudios bioquímicos-estructurales de MutYs carentes de clúster [4Fe-4S] en *E. histolytica* y *L. brevis*.

En el presente trabajo hemos demostrado la funcionalidad *in vitro* e *in vivo* de ADN glicosilasas de adenina MutYs sin necesidad de un clúster [4Fe-4S]. Una búsqueda exhaustiva de homólogos nos permite encontrar una serie de MutYs que, aparentemente, tienen que lidiar con la lesión 8oxoG sin este motivo importante. Un análisis filogenético indica que la pérdida del clúster [4Fe-4S] se ha producido al menos dos veces durante la evolución de esta enzima, en un clado mixto compuesto

por microorganismos anaeróbicos y microaerófilos y en otro exclusivo de Lactobacillales. También se mostró evidencia de que los dominios C-terminales en estos nuevos homólogos juegan un papel importante para prescindir del clúster [4Fe-4S], probablemente por el reordenamiento del motivo impulsado por efectos estéricos. Sin embargo, la obtención de la estructura cristalográfica de estas novedosas variantes y estudios en sus mecanismos para el escaneo de lesiones nos ayudará a entender más claramente el mecanismo que permite que estos nuevos MutYs sean funcionales sin clúster [4Fe-4S].

5. CONCLUSIONES.

- 5.1. El análisis de los elementos constituyentes de la ruta de reparación de BER de *E. histolytica* mostró contener una vía incompleta solamente capaz de la reparación del ADN por la vía de parche largo.
- 5.2. La carencia de la ADN glicosilasa OGG1/MutY por si sola explica el bajo contenido de GC en el genoma del patógeno.
- 5.3. Las ADN glicosilasas MutY, MutT, UDG-FVI así como los ortólogos de AlkD son de origen bacteriano.
- 5.4. La EndoIII y UDG-F1 son las únicas glicosilasas de este organismo con origen eukariote.
- 5.5. BER carece de glicosilasas eucariotes como; MBD4, MPG, OGG1/MutY, EndoVIII, SMUG, TDG, ADN polimerasas con capacidad de liasa y XRCC4.
- 5.6. La MutY de *E. histolytica* es una enzima funcional *in vitro* e *in vivo* a pesar de carecer del cofactor del cúster [4Fe-4S].
- 5.7. La forma en que prescinde de este cofactor es mediante un rearrreglo estructural conducido por un efecto estérico.
- 5.8. Quimeras de EcY-EhY muestran que el dominio C-terminal de EhY esta involucrado la estabilización del motivo ante la carencia del clúster [4Fe-4S].

REFERENCIAS BIBLIOGRÁFICAS.

- Abeygunawardana, C., D. J. Weber, A. G. Gittis, D. N. Frick, J. Lin, A.-F. Miller, M. J. Bessman and A. S. Mildvan (1995). "Solution structure of the MutT enzyme, a nucleoside triphosphate pyrophosphohydrolase." Biochemistry **34**(46): 14997-15005.
- Adl, S. M., A. G. Simpson, M. A. Farmer, R. A. Andersen, O. R. Anderson, J. R. Barta, S. S. Bowser, G. Brugerolle, R. A. Fensome and S. Fredericq (2005). "The new higher level classification of eukaryotes with emphasis on the taxonomy of protists." Journal of Eukaryotic Microbiology **52**(5): 399-451.
- Alseth, I., B. Dalhus and M. Bjørås (2014). "Inosine in DNA and RNA." Current opinion in genetics & development **26**: 116-123.
- Alseth, I., L. Eide, M. Pirovano, T. Rognes, E. Seeberg and M. Bjørås (1999). "The *Saccharomyces cerevisiae* homologues of endonuclease III from *Escherichia coli*, Ntg1 and Ntg2, are both required for efficient repair of spontaneous and induced oxidative DNA damage in yeast." Molecular and cellular biology **19**(5): 3779-3787.
- Alseth, I., F. Osman, H. Korvald, I. Tsaneva, M. C. Whitby, E. Seeberg and M. Bjørås (2005). "Biochemical characterization and DNA repair pathway interactions of Mag1-mediated base excision repair in *Schizosaccharomyces pombe*." Nucleic acids research **33**(3): 1123-1131.
- Alseth, I., T. Rognes, T. Lindbäck, I. Solberg, K. Robertsen, K. I. Kristiansen, D. Mainieri, L. Lillehagen, A. B. Kolstø and M. Bjørås (2006). "A new protein superfamily includes two novel 3-methyladenine DNA glycosylases from *Bacillus cereus*, AlkC and AlkD." Molecular microbiology **59**(5): 1602-1609.
- Alsmark, U. C., T. Sicheritz-Ponten, P. G. Foster, R. P. Hirt and T. M. Embley (2009). "Horizontal gene transfer in eukaryotic parasites: a case study of *Entamoeba histolytica* and *Trichomonas vaginalis*." Horizontal Gene Transfer: Genomes in Flux: 489-500.
- Aravind, L. and E. V. Koonin (2000). "The α/β fold uracil DNA glycosylases: a common origin with diverse fates." Genome biology **1**(4): research0007. 0001.
- Au, K., M. Cabrera, J. Miller and P. Modrich (1988). "*Escherichia coli* mutY gene product is required for specific AG---CG mismatch correction." Proceedings of the National Academy of Sciences **85**(23): 9163-9166.
- Aurrecoechea, C., A. Barreto, J. Brestelli, B. P. Brunk, E. V. Caler, S. Fischer, B. Gajria, X. Gao, A. Gingle and G. Grant (2010). "AmoebaDB and MicrosporidiaDB: functional genomic resources for Amoebozoa and Microsporidia species." Nucleic acids research **39**(suppl_1): D612-D619.
- Ayala, P., J. Samuelson, D. Wirth and E. Orozco (1990). "*Entamoeba histolytica*: physiology of multidrug resistance." Experimental parasitology **71**(2): 169-175.
- Bateman, A., L. Coin, R. Durbin, R. D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon and E. L. Sonnhammer (2004). "The Pfam protein families database." Nucleic acids research **32**(suppl_1): D138-D141.
- Baute, J. and A. Depicker (2008). "Base excision repair and its role in maintaining genome stability." Critical Reviews in Biochemistry and Molecular Biology **43**(4): 239-276.
- Bebenek, K. and T. A. Kunkel (2004). "Functions of DNA polymerases." Advances in protein chemistry **69**: 137-165.
- Bessman, M. J., D. N. Frick and S. F. O'Handley (1996). "The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes." Journal of Biological Chemistry **271**(41): 25059-25062.

Boal, A. K., E. Yavin, O. A. Lukianova, V. L. O'Shea, S. S. David and J. K. Barton (2005). "DNA-bound redox activity of DNA repair glycosylases containing [4Fe-4S] clusters." Biochemistry **44**(23): 8397-8407.

Boiteux, S. and M. Guillet (2004). "Abasic sites in DNA: repair and biological consequences in *Saccharomyces cerevisiae*." DNA repair **3**(1): 1-12.

Boon, E. M., A. L. Livingston, N. H. Chmiel, S. S. David and J. K. Barton (2003). "DNA-mediated charge transport for DNA repair." Proceedings of the National Academy of Sciences **100**(22): 12543-12547.

Brieba, L. G., B. F. Eichman, R. J. Kokoska, S. Doublé, T. A. Kunkel and T. Ellenberger (2004). "Structural basis for the dual coding potential of 8-oxoguanosine by a high-fidelity DNA polymerase." The EMBO journal **23**(17): 3452-3461.

Brinkmeyer, M. K., M. A. Pope and S. S. David (2012). "Catalytic contributions of key residues in the adenine glycosylase MutY revealed by pH-dependent kinetics and cellular repair assays." Chemistry & biology **19**(2): 276-286.

Brooks, S. C., S. Adhikary, E. H. Rubinson and B. F. Eichman (2013). "Recent advances in the structural mechanisms of DNA glycosylases." Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics **1834**(1): 247-271.

Brosius, J. and A. Holy (1984). "Regulation of ribosomal RNA promoters with a synthetic lac operator." Proceedings of the National Academy of Sciences **81**(22): 6929-6933.

Brunk, E., J. S. Arey and U. Rothlisberger (2012). "Role of environment for catalysis of the DNA repair enzyme MutY." Journal of the American Chemical Society **134**(20): 8608-8616.

Busby, W. F., D. E. Shuker, G. Charnley, P. M. Newberne, S. R. Tannenbaum and G. N. Wogan (1985). "Carcinogenicity in rats of the nitrosated bile acid conjugates N-nitrosoglycocholic acid and N-nitrosotaurocholic acid." Cancer research **45**(3): 1367-1371.

Cappelli, E., R. Taylor, M. Cevasco, A. Abbondandolo, K. Caldecott and G. Frosina (1997). "Involvement of XRCC1 and DNA ligase III gene products in DNA base excision repair." Journal of Biological Chemistry **272**(38): 23970-23975.

Cardona-Felix, C. S., S. Lara-Gonzalez and L. G. Brieba (2011). "Structure and biochemical characterization of proliferating cellular nuclear antigen from a parasitic protozoon." Acta Crystallographica Section D: Biological Crystallography **67**(6): 497-505.

Cardona, C. and L. Brieba (2009). "Biochemical Characterization of a Dna ligase I from *Entamoeba histolytica*." Protein Science **18**: 172.

Cedillo-Rivera, R., B. Chávez, A. González-Robles, A. TAPIA and L. YÉPEZ-MULIA (2002). "In vitro effect of nitazoxanide against *Entamoeba histolytica*, *Giardia intestinalis* and *Trichomonas vaginalis* trophozoites." Journal of Eukaryotic Microbiology **49**(3): 201-208.

Collins, A. R. (1999). "Oxidative DNA damage, antioxidants, and cancer." Bioessays **21**(3): 238-246.

Cunningham, R. P., H. Asahara, J. F. Bank, C. P. Scholes, J. C. Salerno, K. Surerus, E. Munck, J. McCracken, J. Peisach and M. H. Emptage (1989). "Endonuclease III is an iron-sulfur protein." Biochemistry **28**(10): 4450-4455.

Chang, D.-Y. and A.-L. Lu (2002). "Functional interaction of MutY homolog with proliferating cell nuclear antigen in fission yeast, *Schizosaccharomyces pombe*." Journal of Biological Chemistry **277**(14): 11853-11858.

ChemicalComputingGroup, M. (2008). Molecular Operating Environment, Chemical Computing Group Montreal, Quebec, Canada.

Chepanoske, C. L., M.-P. Golinelli, S. D. Williams and S. S. David (2000). "Positively Charged Residues within the Iron-Sulfur Cluster Loop of *E. coli* MutY Participate in Damage Recognition and Removal." Archives of biochemistry and biophysics **380**(1): 11-19.

Chmiel, N. H., M.-P. Golinelli, A. W. Francis and S. S. David (2001). "Efficient recognition of substrates and substrate analogs by the adenine glycosylase MutY requires the C-terminal domain." *Nucleic acids research* **29**(2): 553-564.

Chung, J. H., E. K. Im, H. Y. Park, J. H. Kwon, S. Lee, J. Oh, K. C. Hwang, J. H. Lee and Y. Jang (2003). "A novel uracil-DNA glycosylase family related to the helix-hairpin-helix DNA glycosylase superfamily." *Nucleic acids research* **31**(8): 2045-2055.

Dandliker, W. B., M.-L. Hsu, J. Levin and B. R. Rao (1981). "Equilibrium and kinetic inhibition assays based upon fluorescence polarization." *Methods in enzymology* **74**: 3.

David, S. S., V. L. O'Shea and S. Kundu (2007). "Base excision repair of oxidative DNA damage." *Nature* **447**(7147): 941.

DeLano, W. L. (2002). "The PyMOL molecular graphics system." <http://pymol.org>.

Denver, D. R., S. L. Swenson and M. Lynch (2003). "An evolutionary analysis of the helix-hairpin-helix superfamily of DNA repair glycosylases." *Molecular biology and evolution* **20**(10): 1603-1611.

Dhar, S. K., N. R. Choudhury, A. Bhattacharaya and S. Bhattacharya (1995). "A multitude of circular DNAs exist in the nucleus of Entamoeba histolytica." *Molecular and biochemical parasitology* **70**(1-2): 203-206.

Di Noia, J. and M. S. Neuberger (2002). "Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase." *Nature* **419**(6902): 43.

Dobrindt, U., J. Zdziarski, E. Salvador and J. Hacker (2010). "Bacterial genome plasticity and its impact on adaptation during persistent infection." *International Journal of Medical Microbiology* **300**(6): 363-366.

Doherty, A. J., L. C. Serpell and C. P. Ponting (1996). "The helix-hairpin-helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA." *Nucleic acids research* **24**(13): 2488-2497.

Doublié, S., V. Bandaru, J. P. Bond and S. S. Wallace (2004). "The crystal structure of human endonuclease VIII-like 1 (NEIL1) reveals a zincless finger motif required for glycosylase activity." *Proceedings of the National Academy of Sciences of the United States of America* **101**(28): 10284-10289.

Drabløs, F., E. Feyzi, P. A. Aas, C. B. Vaagbø, B. Kavli, M. S. Bratlie, J. Peña-Díaz, M. Otterlei, G. Slupphaug and H. E. Krokan (2004). "Alkylation damage in DNA and RNA—repair mechanisms and medical significance." *DNA repair* **3**(11): 1389-1407.

Drohat, A. C., J. Jagadeesh, E. Ferguson and J. T. Stivers (1999). "Role of electrophilic and general base catalysis in the mechanism of Escherichia coli uracil DNA glycosylase." *Biochemistry* **38**(37): 11866-11875.

Drohat, A. C., K. Kwon, D. J. Krosky and J. T. Stivers (2002). "3-Methyladenine DNA glycosylase I is an unexpected helix-hairpin-helix superfamily member." *Nature Structural & Molecular Biology* **9**(9): 659-664.

Drohat, A. C., G. Xiao, M. Tordova, J. Jagadeesh, K. W. Pankiewicz, K. A. Watanabe, G. L. Gilliland and J. T. Stivers (1999). "Heteronuclear NMR and crystallographic studies of wild-type and H187Q Escherichia coli uracil DNA glycosylase: electrophilic catalysis of uracil expulsion by a neutral histidine 187." *Biochemistry* **38**(37): 11876-11886.

Eddy, S. R. (1998). "Profile hidden Markov models." *Bioinformatics (Oxford, England)* **14**(9): 755-763.

Edgar, R. C. (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput." *Nucleic acids research* **32**(5): 1792-1797.

Edwards, D. I. (1993). "Nitroimidazole drugs-action and resistance mechanisms I. Mechanism of action." *Journal of Antimicrobial Chemotherapy* **31**(1): 9-20.

Eisen, J. A. and P. C. Hanawalt (1999). "A phylogenomic study of DNA repair genes, proteins, and processes." Mutation Research/DNA Repair **435**(3): 171-213.

Engstrom, L. M., O. A. Partington and S. S. David (2012). "An Iron–Sulfur Cluster Loop Motif in the Archaeoglobus fulgidus Uracil–DNA Glycosylase Mediates Efficient Uracil Recognition and Removal." Biochemistry **51**(25): 5187-5197.

Engstrom, L. M., O. A. Partington and S. S. David (2012). "An iron sulfur cluster loop motif in the Archaeoglobus fulgidus uracil-DNA glycosylase mediates efficient uracil recognition and removal." Biochemistry **51**(25): 5187.

Field, J., B. Rosenthal and J. Samuelson (2000). "Early lateral transfer of genes encoding malic enzyme, acetyl-CoA synthetase and alcohol dehydrogenases from anaerobic prokaryotes to Entamoeba histolytica." Molecular microbiology **38**(3): 446-455.

Fishel, M. L. and M. R. Kelley (2007). "The DNA base excision repair protein Ape1/Ref-1 as a therapeutic and chemopreventive target." Molecular aspects of medicine **28**(3): 375-395.

Fisher, O., R. Siman-Tov and S. Ankri (2004). "Characterization of cytosine methylated regions and 5-cytosine DNA methyltransferase (Ehmet) in the protozoan parasite Entamoeba histolytica." Nucleic acids research **32**(1): 287-297.

Fortini, P. and E. Dogliotti (2007). "Base damage and single-strand break repair: mechanisms and functional significance of short-and long-patch repair subpathways." DNA repair **6**(4): 398-409.

Fortini, P., E. Parlanti, O. M. Sidorkina, J. Laval and E. Dogliotti (1999). "The type of DNA glycosylase determines the base excision repair pathway in mammalian cells." Journal of Biological Chemistry **274**(21): 15230-15236.

Fortini, P., B. Pascucci, E. Parlanti, R. W. Sobol, S. H. Wilson and E. Dogliotti (1998). "Different DNA polymerases are involved in the short-and long-patch base excision repair in mammalian cells." Biochemistry **37**(11): 3575-3580.

Fraga, C. G., M. K. Shigenaga, J.-W. Park, P. Degan and B. N. Ames (1990). "Oxidative damage to DNA during aging: 8-hydroxy-2'-deoxyguanosine in rat organ DNA and urine." Proceedings of the National Academy of Sciences **87**(12): 4533-4537.

Fromme, J. C., A. Banerjee, S. J. Huang and G. L. Verdine (2004). "Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase." Nature **427**(6975): 652-657.

Fromme, J. C., A. Banerjee, S. J. Huang and G. L. Verdine (2004). "Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase." Nature **427**(6975): 652.

Fromme, J. C. and G. L. Verdine (2003). "Structure of a trapped endonuclease III–DNA covalent intermediate." The EMBO journal **22**(13): 3461-3471.

Frosina, G., P. Fortini, O. Rossi, F. Carrozzino, G. Raspaglio, L. S. Cox, D. P. Lane, A. Abbondandolo and E. Dogliotti (1996). "Two pathways for base excision repair in mammalian cells." Journal of Biological Chemistry **271**(16): 9573-9578.

Fujii, Y., T. Shimizu, M. Kusumoto, Y. Kyogoku, T. Taniguchi and T. Hakoshima (1999). "Crystal structure of an IRF-DNA complex reveals novel DNA recognition and cooperative binding to a tandem repeat of core sequences." The EMBO journal **18**(18): 5028-5041.

García-Salcedo, Y. (2013). Soluciones alternativas para el cluster hierro-azufre de MutY. Maestría CINVESTAV-IPN.

Garibyan, L., T. Huang, M. Kim, E. Wolff, A. Nguyen, T. Nguyen, A. Diep, K. Hu, A. Iverson and H. Yang (2003). "Use of the rpoB gene to determine the specificity of base substitution mutations on the Escherichia coli chromosome." DNA repair **2**(5): 593-608.

Ghosh, S., J. M. Chan, C. R. Lea, G. A. Meints, J. C. Lewis, Z. S. Tovian, R. M. Flessner, T. C. Loftus, I. Bruchhaus and H. Kendrick (2004). "Effects of bisphosphonates on the growth of Entamoeba

histolytica and Plasmodium species in vitro and in vivo." Journal of medicinal chemistry **47**(1): 175-187.

Gilboa, R., D. O. Zharkov, G. Golan, A. S. Fernandes, S. E. Gerchman, E. Matz, J. H. Kycia, A. P. Grollman and G. Shoham (2002). "Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA." Journal of Biological Chemistry **277**(22): 19811-19816.

Golinelli, M.-P., L. A. Akin, B. R. Crouse, M. K. Johnson and J. Meyer (1996). "Cysteine ligand swapping on a deletable loop of the [2Fe-2S] ferredoxin from Clostridium pasteurianum." Biochemistry **35**(27): 8995-9002.

Golinelli, M.-P., N. H. Chmiel and S. S. David (1999). "Site-directed mutagenesis of the cysteine ligands to the [4Fe-4S] cluster of Escherichia coli MutY." Biochemistry **38**(22): 6997-7007.

Golinelli, M.-P., N. H. Chmiel and S. S. David (1999). "Site-directed mutagenesis of the cysteine ligands to the [4Fe-4S] cluster of Escherichia coli MutY." Biochemistry **38**(22): 6997-7007.

Gorodetsky, A. A., A. K. Boal and J. K. Barton (2006). "Direct electrochemistry of endonuclease III in the presence and absence of DNA." Journal of the American Chemical Society **128**(37): 12082-12083.

Grant, J. R. and L. A. Katz (2014). "Phylogenomic study indicates widespread lateral gene transfer in Entamoeba and suggests a past intimate relationship with parabasalids." Genome biology and evolution **6**(9): 2350-2360.

Grippon, S., Q. Zhao, T. Robinson, J. J. Marshall, R. J. O'Neill, H. Manning, G. Kennedy, C. Dunsby, M. Neil and S. E. Halford (2010). "Differential modes of DNA binding by mismatch uracil DNA glycosylase from Escherichia coli: implications for abasic lesion processing and enzyme communication in the base excision repair pathway." Nucleic acids research: gkq913.

Guan, Y., R. C. Manuel, A. S. Arvai, S. S. Parikh, C. D. Mol, J. H. Miller, R. S. Lloyd and J. A. Tainer (1998). "MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily." Nature structural biology **5**(12).

Guan, Y., R. C. Manuel, A. S. Arvai, S. S. Parikh, C. D. Mol, J. H. Miller, R. S. Lloyd and J. A. Tainer (1998). "MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily." Nature Structural & Molecular Biology **5**(12): 1058-1064.

Guindon, S. and O. Gascuel (2003). "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood." Systematic biology **52**(5): 696-704.

Gupta, P., N. Aggarwal, P. Batra, S. Mishra and T. K. Chaudhuri (2006). "Co-expression of chaperonin GroEL/GroES enhances in vivo folding of yeast mitochondrial aconitase and alters the growth characteristics of Escherichia coli." The international journal of biochemistry & cell biology **38**(11): 1975-1985.

Harris, T. K., G. Wu, M. A. Massiah and A. S. Mildvan (2000). "Mutational, kinetic, and NMR studies of the roles of conserved glutamate residues and of lysine-39 in the mechanism of the MutT pyrophosphohydrolase." Biochemistry **39**(7): 1655-1674.

Hartmann, G., K. O. Honikel, F. Knüsel and J. Nüesch (1967). "The specific inhibition of the DNA-directed RNA synthesis by rifamycin." Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis **145**(3): 843-844.

Hedglin, M. and P. J. O'Brien (2010). "Hopping enables a DNA repair glycosylase to search both strands and bypass a bound protein." ACS Chem Biol **5**(4): 427-436.

Hedglin, M., Y. Zhang and P. J. O'Brien (2015). "Probing the DNA structural requirements for facilitated diffusion." Biochemistry **54**(2): 557-566.

Helleday, T., E. Petermann, C. Lundin, B. Hodgson and R. A. Sharma (2008). "DNA repair pathways as targets for cancer therapy." Nature reviews. Cancer **8**(3): 193.

Hendrich, B., U. Hardeland, H.-H. Ng, J. Jiricny and A. Bird (1999). "The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites." nature **401**(6750): 301.

Hinks, J. A., M. C. Evans, Y. de Miguel, A. A. Sartori, J. Jiricny and L. H. Pearl (2002). "An iron-sulfur cluster in the family 4 uracil-DNA glycosylases." Journal of Biological Chemistry **277**(19): 16936-16940.

Högbom, M., U. B. Ericsson, R. Lam, E. Kuznetsova, P. Nordlund and D. B. Zamble (2005). "A high throughput method for the detection of metalloproteins on a microgram scale." Molecular & Cellular Proteomics **4**(6): 827-834.

Hollis, T., Y. Ichikawa and T. Ellenberger (2000). "DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase, Escherichia coli AlkA." The EMBO journal **19**(4): 758-766.

Huguenin, M., R. Bracha, T. Chookajorn and D. Mirelman (2010). "Epigenetic transcriptional gene silencing in Entamoeba histolytica: insight into histone and chromatin modifications." Parasitology **137**(4): 619-627.

Ikeda, S., T. Biswas, R. Roy, T. Izumi, I. Boldogh, A. Kurosky, A. H. Sarker, S. Seki and S. Mitra (1998). "Purification and Characterization of Human NTH1, a Homolog of Escherichia coli Endonuclease III DIRECT IDENTIFICATION OF LYS-212 AS THE ACTIVE NUCLEOPHILIC RESIDUE." Journal of Biological Chemistry **273**(34): 21585-21593.

Ito, R., H. Hayakawa, M. Sekiguchi and T. Ishibashi (2005). "Multiple enzyme activities of Escherichia coli MutT protein for sanitization of DNA and RNA precursor pools." Biochemistry **44**(17): 6670-6674.

Jaganaman, S., A. Pinto, M. Tarasev and D. P. Ballou (2007). "High levels of expression of the iron-sulfur proteins phthalate dioxygenase and phthalate dioxygenase reductase in Escherichia coli." Protein expression and purification **52**(2): 273-279.

Jansson, K., A. Blomberg, P. Sunnerhagen and M. Alm Rosenblad (2010). "Evolutionary loss of 8-oxo-G repair components among eukaryotes." Genome Integr **1**(1): 12.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz and C. Duran (2012). "Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data." Bioinformatics **28**(12): 1647-1649.

Kelley, M. R. and M. L. Fishel (2008). "DNA repair proteins as molecular targets for cancer therapeutics." Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents) **8**(4): 417-425.

Kennedy, M. C., T. Kent, M. Emptage, H. Merkle, H. Beinert and E. Münck (1984). "Evidence for the formation of a linear [3Fe-4S] cluster in partially unfolded aconitase." Journal of Biological Chemistry **259**(23): 14463-14471.

Kolesov, G., Z. Wunderlich, O. N. Laikova, M. S. Gelfand and L. A. Mirny (2007). "How gene order is influenced by the biophysics of transcription regulation." Proceedings of the National Academy of Sciences **104**(35): 13948-13953.

Koonin, E. V. (2010). "The origin and early evolution of eukaryotes in the light of phylogenomics." Genome biology **11**(5): 209.

Kreutzer, D. A. and J. M. Essigmann (1998). "Oxidized, deaminated cytosines are a source of C→T transitions in vivo." Proceedings of the National Academy of Sciences **95**(7): 3578-3582.

Krokan, H. E., R. Standal and G. Slupphaug (1997). "DNA glycosylases in the base excision repair of DNA." Biochemical Journal **325**(1): 1-16.

Kundu, S., M. K. Brinkmeyer, A. L. Livingston and S. S. David (2009). "Adenine removal activity and bacterial complementation with the human MutY homologue (MUTYH) and Y165C, G382D, P391L and Q324R variants associated with colorectal cancer." DNA repair **8**(12): 1400-1410.

Kuo, C.-F., D. E. McRee, C. L. Fisher, S. F. O'Handley, R. P. Cunningham and J. A. Tainer (1992). "Atomic structure of the DNA repair [4Fe-4S] enzyme endonuclease III." Science **258**(5081): 434-440.

Labahn, J., O. D. Schärer, A. Long, K. Ezaz-Nikpay, G. L. Verdine and T. E. Ellenberger (1996). "Structural basis for the excision repair of alkylation-damaged DNA." *Cell* **86**(2): 321-329.

Lau, A. Y., O. D. Schärer, L. Samson, G. L. Verdine and T. Ellenberger (1998). "Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision." *Cell* **95**(2): 249-258.

Lavi, T., E. Isakov, H. Harony, O. Fisher, R. Siman-Tov and S. Ankri (2006). "Sensing DNA methylation in the protozoan parasite *Entamoeba histolytica*." *Molecular microbiology* **62**(5): 1373-1386.

Lee, C.-Y., H. Bai, R. Houle, G. M. Wilson and A.-L. Lu (2004). "An *Escherichia coli* MutY mutant without the six-helix barrel domain is a dimer in solution and assembles cooperatively into multisubunit complexes with DNA." *Journal of Biological Chemistry* **279**(50): 52653-52663.

Lee, H.-W., B. N. Dominy and W. Cao (2011). "New family of deamination repair enzymes in uracil-DNA glycosylase superfamily." *Journal of Biological Chemistry* **286**(36): 31282-31287.

Lee, S. and G. L. Verdine (2009). "Atomic substitution reveals the structural basis for substrate adenine recognition and removal by adenine DNA glycosylase." *Proceedings of the National Academy of Sciences* **106**(44): 18497-18502.

Levin, D. S., A. E. McKenna, T. A. Motycka, Y. Matsumoto and A. E. Tomkinson (2000). "Interaction between PCNA and DNA ligase I is critical for joining of Okazaki fragments and long-patch base-excision repair." *Current Biology* **10**(15): 919-912.

Lind, P. A. and D. I. Andersson (2008). "Whole-genome mutational biases in bacteria." *Proceedings of the National Academy of Sciences* **105**(46): 17878-17883.

Lindahl, T. (1993). "Instability and decay of the primary structure of DNA." *nature* **362**(6422): 709-715.

Lioutas, C., C. Schmetz and E. Tannich (1995). "Identification of various circular DNA molecules in *Entamoeba histolytica*." *Experimental parasitology* **80**(2): 349-352.

Liu, X. and R. Roy (2001). "Mutation at active site lysine 212 to arginine uncouples the glycosylase activity from the lyase activity of human endonuclease III." *Biochemistry* **40**(45): 13617-13622.

Livingston, A. L., S. Kundu, M. Henderson Pozzi, D. W. Anderson and S. S. David (2005). "Insight into the roles of tyrosine 82 and glycine 253 in the *Escherichia coli* adenine glycosylase MutY." *Biochemistry* **44**(43): 14179-14190.

Loftus, B., I. Anderson, R. Davies and U. C. M. Alsmark (2005). "The genome of the protist parasite *Entamoeba histolytica*." *Nature* **433**(7028): 865.

Loftus, B., I. Anderson, R. Davies, U. C. M. Alsmark, J. Samuelson, P. Amedeo, P. Roncaglia, M. Berriman, R. P. Hirt and B. J. Mann (2005). "The genome of the protist parasite *Entamoeba histolytica*." *Nature* **433**(7028): 865-868.

López-Camarillo, C., M. Lopez-Casamichana, C. Weber, N. Guillen, E. Orozco and L. A. Marchat (2009). "DNA repair mechanisms in eukaryotes: special focus in *Entamoeba histolytica* and related protozoan parasites." *Infection, Genetics and Evolution* **9**(6): 1051-1056.

Lu, A.-L., J.-J. Tsai-Wu and J. Cillo (1995). "DNA determinants and substrate specificities of *Escherichia coli* MutY." *Journal of Biological Chemistry* **270**(40): 23582-23588.

Lu, A.-L. and P. M. Wright (2003). "Characterization of an *Escherichia coli* mutant MutY with a cysteine to alanine mutation at the iron-sulfur cluster domain." *Biochemistry* **42**(13): 3742-3750.

Lucas-Lledó, J. I., R. Maddamsetti and M. Lynch (2011). "Phylogenomic analysis of the uracil-DNA glycosylase superfamily." *Molecular biology and evolution* **28**(3): 1307-1317.

Luncsford, P. J., D.-Y. Chang, G. Shi, J. Bernstein, A. Madabushi, D. N. Patterson, A.-L. Lu and E. A. Toth (2010). "A structural hinge in eukaryotic MutY homologues mediates catalytic activity and Rad9–Rad1–Hus1 checkpoint complex interactions." *Journal of molecular biology* **403**(3): 351-370.

Manuel, R. C., E. W. Czerwinski and R. S. Lloyd (1996). "Identification of the structural and functional domains of MutY, an Escherichia coli DNA mismatch repair enzyme." Journal of Biological Chemistry **271**(27): 16218-16226.

Manuel, R. C. and R. S. Lloyd (1997). "Cloning, overexpression, and biochemical characterization of the catalytic domain of MutY." Biochemistry **36**(37): 11140-11152.

Marchat, L. A., M. López-Casamichana, E. Orozco and C. López-Camarillo (2011). DNA repair in pathogenic eukaryotic cells: insights from comparative genomics of parasitic protozoan. Selected Topics in DNA Repair, InTech.

Markkanen, E., J. Dorn and U. Hübscher (2013). "MUTYH DNA glycosylase: the rationale for removing undamaged bases from the DNA." Frontiers in genetics **4**.

McAuley-Hecht, K. E., G. A. Leonard, N. J. Gibson, J. B. Thomson, W. P. Watson, W. N. Hunter and T. Brown (1994). "Crystal structure of a DNA duplex containing 8-hydroxydeoxyguanine-adenine base pairs." Biochemistry **33**(34): 10266-10270.

Meneses, E., H. Cárdenas, S. Zárate, L. G. Brieba, E. Orozco, C. López-Camarillo and E. Azuara-Liceaga (2010). "The R2R3 Myb protein family in Entamoeba histolytica." Gene **455**(1): 32-42.

Messick, T. E., N. H. Chmiel, M.-P. Golinelli, M. R. Langer, L. Joshua-Tor and S. S. David (2002). "Noncysteiny coordination to the [4Fe-4S] 2+ cluster of the DNA repair adenine glycosylase MutY introduced via site-directed mutagenesis. Structural characterization of an unusual histidinyl-coordinated cluster." Biochemistry **41**(12): 3931-3942.

Meyer, J. (2008). "Iron-sulfur protein folds, iron-sulfur chemistry, and evolution." JBIC Journal of Biological Inorganic Chemistry **13**(2): 157-170.

Michaels, M. and J. H. Miller (1992). "The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7, 8-dihydro-8-oxoguanine)." Journal of bacteriology **174**(20): 6321.

Michaels, M. L., C. Cruz, A. P. Grollman and J. H. Miller (1992). "Evidence that MutY and MutM combine to prevent mutations by an oxidatively damaged form of guanine in DNA." Proceedings of the National Academy of Sciences **89**(15): 7022-7025.

Michaels, M. L., L. Pham, Y. Nghiem, C. Cruz and J. H. Miller (1990). "MutY, an adenine glycosylase active on GA mispairs, has homology to endonuclease III." Nucleic acids research **18**(13): 3841-3845.

Michaels, M. L., J. Tchou, A. P. Grollman and J. H. Miller (1992). "A repair system for 8-oxo-7, 8-dihydrodeoxyguanine." Biochemistry **31**(45): 10964-10968.

Miller, J. H. (1992). A] short course in bacterial genetics: a laboratory manual and handbook for Escherichia coli and related bacteria.

Mirelman, D., M. Anbar and R. Bracha (2008). "Trophozoites of Entamoeba histolytica epigenetically silenced in several genes are virulence-attenuated." Parasite **15**(3): 266-274.

Mirelman, D., M. Anbar, Y. Nuchamowitz and R. Bracha (2006). "Epigenetic silencing of gene expression in Entamoeba histolytica." Archives of medical research **37**(2): 226-233.

Mitra, D., S. J. George, Y. Guo, S. Kamali, S. Keable, J. W. Peters, V. Pelmeshnikov, D. A. Case and S. P. Cramer (2013). "Characterization of [4Fe-4S] cluster vibrations and structure in nitrogenase Fe protein at three oxidation levels via combined NRVs, EXAFS, and DFT analyses." Journal of the American Chemical Society **135**(7): 2530-2543.

Moreira, D. and P. López-García (2017). "Protist Evolution: Stealing Genes to Gut It Out." Current Biology **27**(6): R223-R225.

Morgan, H. D., W. Dean, H. A. Coker, W. Reik and S. K. Petersen-Mahrt (2004). "Activation-induced cytidine deaminase deaminates 5-Methylcytosine in DNA and is expressed in pluripotent tissues implications for epigenetic reprogramming." Journal of Biological Chemistry **279**(50): 52353-52360.

Moulis, J.-M., V. Davasse, M.-P. Golinelli, J. Meyer and I. Quinkal (1996). "The coordination sphere of iron-sulfur clusters: lessons from site-directed mutagenesis experiments." JBIC Journal of Biological Inorganic Chemistry **1**(1): 2-14.

Moulis, J. M., L. C. Sieker, K. S. Wilson and Z. Dauter (1996). "Crystal structure of the 2 [4Fe-4S] ferredoxin from *Chromatium vinosum*: Evolutionary and mechanistic inferences for [3/4Fe-4S] ferredoxins." Protein science **5**(9): 1765-1775.

Mullins, E. A., E. H. Rubinson and B. F. Eichman (2014). "The substrate binding interface of alkylpurine DNA glycosylase AlkD." DNA repair **13**: 50-54.

Mullins, E. A., R. Shi, L. A. Kotsch and B. F. Eichman (2015). "A New Family of HEAT-Like Repeat Proteins Lacking a Critical Substrate Recognition Motif Present in Related DNA Glycosylases." PLoS one **10**(5): e0127733.

Mullins, E. A., R. Shi, Z. D. Parsons, P. K. Yuen, S. S. David, Y. Igarashi and B. F. Eichman (2015). "The DNA glycosylase AlkD uses a non-base-flipping mechanism to excise bulky lesions." Nature **527**(7577): 254.

Nakamura, M., K. Saeki and Y. Takahashi (1999). "Hyperproduction of recombinant ferredoxins in *Escherichia coli* by coexpression of the ORF1-ORF2-iscS-iscU-iscA-hscB-hscA-fdx-ORF3 gene cluster." Journal of Biochemistry **126**(1): 10-18.

Nakamura, T., S. Meshitsuka, S. Kitagawa, N. Abe, J. Yamada, T. Ishino, H. Nakano, T. Tsuzuki, T. Doi and Y. Kobayashi (2010). "Structural and dynamic features of the MutT protein in the recognition of nucleotides with the mutagenic 8-oxoguanine base." Journal of Biological Chemistry **285**(1): 444-452.

Nixon, J. E., A. Wang, J. Field, H. G. Morrison, A. G. McArthur, M. L. Sogin, B. J. Loftus and J. Samuelson (2002). "Evidence for lateral transfer of genes encoding ferredoxins, nitroreductases, NADH oxidase, and alcohol dehydrogenase 3 from anaerobic prokaryotes to *Giardia lamblia* and *Entamoeba histolytica*." Eukaryotic Cell **1**(2): 181-190.

Noll, D. M., A. Gogos, J. A. Granek and N. D. Clarke (1999). "The C-Terminal domain of the Adenine-DNA glycosylase MutY confers specificity for 8-Oxoguanine[⊖] adenine mispairs and may have evolved from MutT, an 8-Oxo-dGTPase." Biochemistry **38**(20): 6374-6379.

Owen, B. and C. McMurray (2009). "Rapid method for measuring DNA binding to protein using fluorescence anisotropy." Protocol Exchange.

Parikh, S. S., G. Walcher, G. D. Jones, G. Slupphaug, H. E. Krokan, G. M. Blackburn and J. A. Tainer (2000). "Uracil-DNA glycosylase–DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects." Proceedings of the National Academy of Sciences **97**(10): 5083-5088.

Parker, A., Y. Gu, W. Mahoney, S.-H. Lee, K. K. Singh and A.-L. Lu (2001). "Human homolog of the MutY repair protein (hMYH) physically interacts with proteins involved in long patch DNA base excision repair." Journal of Biological Chemistry **276**(8): 5547-5555.

Pastor-Palacios, G., E. Azuara-Liceaga and L. G. Brieba (2010). "A nuclear family A DNA polymerase from *Entamoeba histolytica* bypasses thymine glycol." PLoS neglected tropical diseases **4**(8): e786.

Pastor-Palacios, G., V. López-Ramírez, C. S. Cardona-Felix and L. G. Brieba (2012). "A transposon-derived DNA polymerase from *Entamoeba histolytica* displays intrinsic strand displacement, processivity and lesion bypass." PLoS one **7**(11): e49964.

Pastor-Palacios, G., V. López-Ramírez, C. S. Cardona-Félix, E. A. Liceaga, S. Lara-Gonzalez and L. G. Brieba (2015). Archetypical and Specialized DNA Replication Proteins in *Entamoeba histolytica*. Amebiasis, Springer: 393-408.

Pearl, L. H. (2000). "Structure and function in the uracil-DNA glycosylase superfamily." Mutation Research/DNA Repair **460**(3): 165-181.

Perez, D. G., C. Gómez, E. López-Bayghen, E. Tannich and E. Orozco (1998). "Transcriptional analysis of the EhPgp5 promoter of *Entamoeba histolytica* multidrug-resistant mutant." Journal of Biological Chemistry **273**(13): 7285-7292.

Petermann, E., C. Keil and S. L. Oei (2006). "Roles of DNA ligase III and XRCC1 in regulating the switch between short patch and long patch BER." DNA repair **5**(5): 544-555.

Pieroni, L., L. Khalil, F. Charlotte, T. Poynard, A. Piton, B. Hainque and F. Imbert-Bismut (2001). "Comparison of bathophenanthroline sulfonate and ferene as chromogens in colorimetric measurement of low hepatic iron concentration." Clinical chemistry **47**(11): 2059-2061.

Pogulis, R. J., A. N. Vallejo and L. R. Pease (1996). "In vitro recombination and mutagenesis by overlap extension PCR." In vitro mutagenesis protocols: 167-176.

Pope, M. A. and S. S. David (2005). "DNA damage recognition and repair by the murine MutY homologue." DNA repair **4**(1): 91-102.

Porello, S. L., M. J. Cannon and S. S. David (1998). "A substrate recognition role for the [4Fe-4S] 2+ cluster of the DNA repair glycosylase MutY." Biochemistry **37**(18): 6465-6475.

Porello, S. L., A. E. Leyes and S. S. David (1998). "Single-turnover and pre-steady-state kinetics of the reaction of the adenine glycosylase MutY with mismatch-containing DNA substrates." Biochemistry **37**(42): 14756-14764.

Prakash, S., R. E. Johnson and L. Prakash (2005). "Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function." Annu. Rev. Biochem. **74**: 317-353.

Qing, G., L. C. Ma, A. Khorchid, G. V. Swapna, T. K. Mal, M. M. Takayama, B. Xia, S. Phadtare, H. Ke, T. Acton, G. T. Montelione, M. Ikura and M. Inouye (2004). "Cold-shock induced high-yield protein production in *Escherichia coli*." Nat Biotechnol **22**(7): 877-882.

Rambaut, A. (2007). "FigTree, a graphical viewer of phylogenetic trees." See <http://tree.bio.ed.ac.uk/software/figtree>.

Romero, M., R. Cerritos and C. Ximenez (2016). "Horizontal Gene Transfers from Bacteria to *Entamoeba* Complex: A Strategy for Dating Events along Species Divergence." Journal of parasitology research **2016**.

Rosenthal, B., Z. Mai, D. Caplivski, S. Ghosh, H. de la Vega, T. Graf and J. Samuelson (1997). "Evidence for the bacterial origin of genes encoding fermentation enzymes of the amitochondriate protozoan parasite *Entamoeba histolytica*." Journal of Bacteriology **179**(11): 3736-3745.

Rubinson, E. H., A. H. Metz, J. O'Quin and B. F. Eichman (2008). "A new protein architecture for processing alkylation damaged DNA: the crystal structure of DNA glycosylase AlkD." Journal of molecular biology **381**(1): 13-23.

Samarawickrema, N., D. Brown, J. Upcroft, N. Thammapalerd and P. Upcroft (1997). "Involvement of superoxide dismutase and pyruvate: ferredoxin oxidoreductase in mechanisms of metronidazole resistance in *Entamoeba histolytica*." The Journal of antimicrobial chemotherapy **40**(6): 833-840.

Samrakandi, M. M. and F. Pasta (2000). "Hyperrecombination in *Streptococcus pneumoniae* Depends on an Atypical mutY Homologue." Journal of bacteriology **182**(12): 3353-3360.

Sazanov, L. A. and P. Hinchliffe (2006). "Structure of the hydrophilic domain of respiratory complex I from *Thermus thermophilus*." science **311**(5766): 1430-1436.

Schärer, O. D. and J. Jiricny (2001). "Recent progress in the biology, chemistry and structural biology of DNA glycosylases." Bioessays **23**(3): 270-281.

Schärer, O. D., H. M. Nash, J. Jiricny, J. Laval and G. L. Verdine (1998). "Specific binding of a designed pyrrolidine abasic site analog to multiple DNA glycosylases." Journal of Biological Chemistry **273**(15): 8592-8597.

Severinov, K., M. Soushko, A. Goldfarb and V. Nikiforov (1993). "Rifampicin region revisited. New rifampicin-resistant and streptolydigin-resistant mutants in the beta subunit of Escherichia coli RNA polymerase." Journal of Biological Chemistry **268**(20): 14820-14825.

Shibutani, S., M. Takeshita and A. P. Grollman (1991). "Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG." nature **349**(6308): 431-434.

Sjolund, A. B., A. G. Senejani and J. B. Sweasy (2013). "MBD4 and TDG: multifaceted DNA glycosylases with ever expanding biological roles." Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis **743**: 12-25.

Slupphaug, G., C. D. Mol, B. Kavli, A. S. Arvai, H. E. Krokan and J. A. Tainer (1996). "A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA." Nature **384**(6604): 87.

Smet-Noocca, C., J.-M. Wieruszkeski, V. Chaar, A. Leroy and A. Benecke (2008). "The Thymine- DNA Glycosylase Regulatory Domain: Residual Structure and DNA Binding." Biochemistry **47**(25): 6519-6530.

Smith, M. W., D.-F. Feng and R. F. Doolittle (1992). "Evolution by acquisition: the case for horizontal gene transfers." Trends in biochemical sciences **17**(12): 489-493.

Specht, W. (1937). "Die Chemiluminescenz des Hämins, ein Hilfsmittel zur Auffindung und Erkennung forensisch wichtiger Blutspuren." International Journal of Legal Medicine **28**(1): 225-234.

Stanley, S. L. (2003). "Amoebiasis." The Lancet **361**(9362): 1025-1034.

Stiebritz, M. T. (2014). "A role for [Fe4S4] clusters in tRNA recognition—a theoretical study." Nucleic acids research **42**(9): 5426-5435.

Sugahara, M., T. Mikawa, T. Kumasaka, M. Yamamoto, R. Kato, K. Fukuyama, Y. Inoue and S. Kuramitsu (2000). "Crystal structure of a repair enzyme of oxidatively damaged DNA, MutM (Fpg), from an extreme thermophile, *Thermus thermophilus* HB8." The EMBO journal **19**(15): 3857-3869.

Sung, R.-J., M. Zhang, Y. Qi and G. L. Verdine (2013). "Structural and biochemical analysis of DNA helix invasion by the bacterial 8-oxoguanine DNA glycosylase MutM." Journal of Biological Chemistry **288**(14): 10012-10023.

Tajiri, T., H. Maki and M. Sekiguchi (1995). "Functional cooperation of MutT, MutM and MutY proteins in preventing mutations caused by spontaneous oxidation of guanine nucleotide in *Escherichia coli*." Mutation Research/DNA Repair **336**(3): 257-267.

Tamura, K., G. Stecher, D. Peterson, A. Filipiński and S. Kumar (2013). "MEGA6: molecular evolutionary genetics analysis version 6.0." Molecular biology and evolution **30**(12): 2725-2729.

Tannich, E., I. Bruchhaus, R. D. Walter and R. D. Horstmann (1991). "Pathogenic and nonpathogenic *Entamoeba histolytica*: identification and molecular cloning of an iron-containing superoxide dismutase." Molecular and biochemical parasitology **49**(1): 61-71.

Tekwani, B. L. and R. K. Mehlotra (1999). "Molecular basis of defence against oxidative stress in *Entamoeba histolytica* and *Giardia lamblia*." Microbes and infection **1**(5): 385-394.

Thayer, M., H. Ahern, D. Xing, R. Cunningham and J. Tainer (1995). "Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure." The EMBO journal **14**(16): 4108.

Torres-Guerrero, H., D. A. Peattie and I. Meza (1991). "Chromatin organization in *Entamoeba histolytica*." Molecular and biochemical parasitology **45**(1): 121-130.

Tovy, A., R. S. Tov, R. Gaentzsch, M. Helm and S. Ankri (2010). "A new nuclear function of the *Entamoeba histolytica* glycolytic enzyme enolase: the metabolic regulation of cytosine-5 methyltransferase 2 (Dnmt2) activity." PLoS pathogens **6**(2): e1000775.

Trasviña-Arenas, C. H., C. S. Cardona-Felix, E. Azuara-Liceaga, C. Díaz-Quezada and L. G. Brieba (2017). "Proliferating cell nuclear antigen restores the enzymatic activity of a DNA ligase I deficient in DNA binding." FEBS open bio **7**(5): 659-674.

Trasviña-Arenas, C. H., L. M. Lopez-Castillo, E. Sanchez-Sandoval and L. G. Brieba (2016). "Dispensability of the [4Fe-4S] cluster in novel homologues of adenine glycosylase MutY." The FEBS journal **283**(3): 521-540.

Upcroft, P. and J. A. Upcroft (2001). "Drug targets and mechanisms of resistance in the anaerobic protozoa." Clinical microbiology reviews **14**(1): 150-164.

Van Der Giezen, M., S. Cox and J. Tovar (2004). "The iron-sulfur cluster assembly genes *iscS* and *iscU* of *Entamoeba histolytica* were acquired by horizontal gene transfer." BMC evolutionary biology **4**(1): 7.

Vidal, A. E., S. Boiteux, I. D. Hickson and J. P. Radicella (2001). "XRCC1 coordinates the initial and late stages of DNA abasic site repair through protein–protein interactions." The EMBO journal **20**(22): 6530-6539.

Volk, D. E., P. G. House, V. Thiviyanathan, B. A. Luxon, S. Zhang, R. S. Lloyd and D. G. Gorenstein (2000). "Structural similarities between MutT and the C-terminal domain of MutY." Biochemistry **39**(25): 7331-7336.

Wassmann, C., A. Hellberg, E. Tannich and I. Bruchhaus (1999). "Metronidazole resistance in the protozoan parasite *Entamoeba histolytica* is associated with increased expression of iron-containing superoxide dismutase and peroxiredoxin and decreased expression of ferredoxin 1 and flavin reductase." Journal of Biological Chemistry **274**(37): 26051-26056.

Weber, D. J., C. Abeygunawardana, M. J. Bessman and A. S. Mildvan (1993). "Secondary structure of the MutT enzyme as determined by NMR." Biochemistry **32**(48): 13081-13088.

White, M. F. and M. S. Dillingham (2012). "Iron–sulphur clusters in nucleic acid processing enzymes." Current opinion in structural biology **22**(1): 94-100.

Wibley, J. E., T. R. Waters, K. Haushalter, G. L. Verdine and L. H. Pearl (2003). "Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1." Molecular cell **11**(6): 1647-1659.

Williams, S. D. and S. S. David (1999). "Formation of a Schiff base intermediate is not required for the adenine glycosylase activity of *Escherichia coli* MutY." Biochemistry **38**(47): 15417-15424.

Williams, S. D. and S. S. David (2000). "A single engineered point mutation in the adenine glycosylase MutY confers bifunctional glycosylase/AP lyase activity." Biochemistry **39**(33): 10098-10109.

Wolff, E., M. Kim, K. Hu, H. Yang and J. H. Miller (2004). "Polymerases leave fingerprints: analysis of the mutational spectrum in *Escherichia coli* *rpoB* to assess the role of polymerase IV in spontaneous mutation." Journal of bacteriology **186**(9): 2900-2905.

Wong, E., K. Yang, M. Kuraguchi, U. Werling, E. Avdievich, K. Fan, M. Fazzari, B. Jin, A. M. Brown and M. Lipkin (2002). "Mbd4 inactivation increases C→T transition mutations and promotes gastrointestinal tumor formation." Proceedings of the National Academy of Sciences **99**(23): 14937-14942.

Wyatt, M. D., J. M. Allan, A. Y. Lau, T. E. Ellenberger and L. D. Samson (1999). "3-methyladenine DNA glycosylases: structure, function, and biological importance." Bioessays **21**(8): 668-676.

Xia, B., Y. Liu, J. Guevara, J. Li, C. Jilich, Y. Yang, L. Wang, B. N. Dominy and W. Cao (2017). "Correlated Mutation in the Evolution of Catalysis in Uracil DNA Glycosylase Superfamily." Scientific Reports **7**.

Xie, P., H. Teramae, K. Liu and Y. Aoki (2013). "Electronic states of mixed base pairs systems of DNA and the effect of base composition and sequences on the band structures using screw axis translational symmetry." International Journal of Quantum Chemistry **113**(4): 489-496.

Yang, J., R. Yan, A. Roy, D. Xu, J. Poisson and Y. Zhang (2015). "The I-TASSER Suite: protein structure and function prediction." Nature methods **12**(1): 7-8.

Yeeles, J. T., R. Cammack and M. S. Dillingham (2009). "An iron-sulfur cluster is essential for the binding of broken DNA by AddAB-type helicase-nucleases." Journal of Biological Chemistry **284**(12): 7746-7755.

You, H. J., R. L. Swanson and P. W. Doetsch (1998). "Saccharomyces cerevisiae possesses two functional homologues of Escherichia coli endonuclease III." Biochemistry **37**(17): 6033-6040.

Zhang, F., P. Scheerer, I. Oberpichler, T. Lamparter and N. Krauß (2013). "Crystal structure of a prokaryotic (6-4) photolyase with an Fe-S cluster and a 6, 7-dimethyl-8-ribityllumazine antenna chromophore." Proceedings of the National Academy of Sciences **110**(18): 7217-7222.

Zhang, Y. and P. J. O'Brien (2015). "Repair of Alkylation Damage in Eukaryotic Chromatin Depends on Searching Ability of Alkyladenine DNA Glycosylase." ACS Chem Biol.

Zharkov, D. O., G. Golan, R. Gilboa, A. S. Fernandes, S. E. Gerchman, J. H. Kycia, R. A. Rieger, A. P. Grollman and G. Shoham (2002). "Structural analysis of an Escherichia coli endonuclease VIII covalent reaction intermediate." The EMBO journal **21**(4): 789-800.

Zhu, B., Y. Zheng, H. Angliker, S. Schwarz, S. Thiry, M. Siegmann and J.-P. Jost (2000). "5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence." Nucleic acids research **28**(21): 4157-4165.

APÉNDICES.

Apéndice 1. Oligonucleótidos, genes sintéticos y g-Blocks utilizados en este trabajo.

Gen sintético de MutY de *Lactobacillus brevis* (5'-3').

ATGGTTGAATGGACGCCGGAAAAAATCGTGGCCTTCCAAGAAACGCTGCTGA
AATGGTACGATAATAATAAACGCAATCTGCCGTGGCGTCGCGATCATGACCC
GTATCACATTTGGATCAGTGAAATTATGCTGCAGCAAACGCAGGTGCAAACCG
TTATCCCGTATTACGAACGTTTTATGAAACTGTTCCCGACCGTGCAGGCACTG
GCTAGTGC GGATGAAGCAATTCTGATGAAAGCTTGGGAAGGCCTGGGTTATT
ACTCCCGTGCACGTAACCTGCAGAAAGCAGCACAGCAAATTGTCAACGATTAT
AATGGCCAGTGGCCGACCACGGTGAAGAAGTCAAGAAGTGTGAGGTATCG
GTCCGTACACGGCAGGTGCAATTGCTTCGATCGCCTTTAACAAACCGGTCCC
GGCGGTGGATGGTAATGCCCTGCGTGTTCGACGCCTGCTGGAAATTGAT
GAAGACATCGCAAACCGCAGACCCGTAAACTGTTGAAAACATCATCAAAAA
ACTGATGCCGAAAAACCGCCCGGGCGATTTCAATCAAGCTATTATGGACCTG
GGTGCTTCATATATGTCGGCGAAAAACTACGATAGCGAAAATTCTCCGGTTAA
ACTGTTTAACCAGGCATATCTGGATGGCGTTGAAGACAATTACCCGGTCAAAA
CCAAGAAAAAACGTCCGATTCCGGTCAACTATTTTGGTCTGCTGATCCATTCT
CAGGATGACTACCTGTTGAAACGTCGCCCGAATAGTGGCATTCTGTCCCGCT
TTTGGATGTTCCCGCTGATTAAGGTGATGACATCCAGACGAAAAAAGATGCC
AGCGAAGATGACGTGCTGCGTGCCTGGAAGCCCAGTTTCTGTCTACCTATC
AACTGGAAATTCACGTGAAGAAAATTGGCGGCCCGCCGGTTACCCATACGTT
CACCCACCAGAAATGGCAAATCACCTGCTGGAAGCGGAACTGAACAATAGC
TCTGATCTGAGCTACTTTCCGGGTAAATGGATTAGTGAATCCGACTTCCGCGA
AATCGCCTTCACGAAAGTCCAGACGAAAATGTGGGAACGCTACCAGCAGCAG
AAAGAACAATAAGGATCC

G-Blocks para el ensamblaje de los genes quiméricos EcEhY1 y 3.

Genes optimizados del dominio catalítico de EcMutY. Resaltado estas las regiones de traslape para el ensamble.

CATATGCAGGCAAGTCAATTCTCTGCGCAAGTATTGGACTGGTACGATAAATA
TGGTCGTAAAACCCTGCCCTGGCAGATTGATAAACTCCTTACAAAGTGTGGC
TGTCTGAGGTTATGCTGCAACAAACGCAGGTTGCTACCGTTATCCCATATTT
GAGCGTTTTATGGCGCGTTTTCCCAACTGTCACTGATCTGGCTAATGCGCCGC
TGGATGAAGTCCTGCATCTTTGGACCGGCCTCGGCTATTATGCACGGGCGCG
CAACCTGCATAAAGCAGCGCAGCAGGTGGCGACATTACATGGTGGGAAATTC
CCCGAAACATTCGAAGAAGTTGCTGCGCTG**CCGGGTGTGGGTGCTCTACC**
GCGGGAGCAATTCTCTCACTTAGTCTGGGT

Genes optimizados del dominio C-terminal de EhMutY para el ensamblaje de la quimera EcEhY1. Resaltado estas las regiones de traslape para el ensamble.

CCGGGTGTGGGTCGCTCTACCGCGGGAGCAATTCTCTCACTTAGTCTGGGT
CAAGATATTGCAACAGTAGACGCTAATATTAGAAGAATTCTTATTCACGAACTT
TCTCTTGATGAAAAACACCTGAAAATATATTATTCGCTTATGCAAATAAATGT
CTCCCTTTGGGAAGAAGTCGAGATTGGCATAACGCTTTAATGGATTATGGTGC
TCTTGTTGCAACTGGTAAAAAACTGGAATTACAGCAAATCAAAACAGAGTAA
ATTCGAAGGATCAAGAAGATCATATAGAGCAGAGTTGTTGAGAAGAATTAGAA
AGAATAATGGTTCTTTATCAATAGAGAATTGTGAAGAAGTGATCAAGGGAACT
AAGTACACTTTAGAGGAAATCATTGAAGGATTAGAAAAAGATGGAATTGTTTCAT
AAAGAAGGAGATTTTATAGTTTTGAATAAATAAGGATCC

Genes optimizados del dominio C-terminal de EhMutY para el ensamblaje de la quimera EcEhY3. Resaltado estas las regiones de traslape para el ensamble.

CCGGGTGTGGGTCGCTCTACCGCGGGAGCAATTCTCTCACTTAGTCTGGGT
AAACATTTTCCGATTTTGGATGGTAATGTAAAACGGGTTCTGGCACGTTGTTAT
GCGGTGTCGGGCTGGCCGGGCAAGAAGGAAGTAGAAAACAACTCTGGAGC
CTGTCTGAACAAGTGACGCCCGCAGTCGGGGTTGAACGCTTCAACCAGGCTA
TGATGGATTTAGGTGCAATGATCTGCACCCGCAGCAAACCGAAATGCTCGCT
GTGTCCTTTACAGAACGGTTGCATTGCGGCCGCGAATAACTCCTGGGCACTG
TATCCGGGGAAAAAGCCGAAATCATATAGAGCAGAGTTGTTGAGAAGAATTAG
AAAGAATAATGGTTCTTTATCAATAGAGAATTGTGAAGAAGTGATCAAGGGAA
CTAAGTACACTTTAGAGGAAATCATTGAAGGATTAGAAAAAGATGGAATTGTTTC
ATAAAGAAGGAGATTTTATAGTTTTGAATAAATAAGGATCC

Apendice I. Información de Oligonucleótidos usados en este trabajo

Oligonucleótido	Secuencia (5'-3')	Observación	Uso
EhYNtermFw	GGTTGGGGATCCATGAAAAGAAGATGGGAACAAC	Sitio BamHI	Clonación y ensamble del gen EhMutY
EhYNtermRv	CTTCAAGAACATCAGCAAGATTTGCTTTTGCTAAGTCTTGAATAGTTG		Ensamble del gen EMuthY
EhYCtermFw	CCAATCTTGCTGATGTTGTTGAAG		Ensamble del gen EhMutY
EhYCtermRv	GGTTGGAAGCTTTTATTTATTCAAAACTATAAAATCTCC	Sitio HindIII	Ensamble del gen EhMutY
LbYFw	GGAGCTGACATATGGTTGAATGGACGCCGGAA	Sitio NdeI	Clonación de LbMutY
LbYRv	AGGTACGGATCCTTATTGTTCTTTCTGCTGCTG	Sitio BamHI	Clonación de LbMutY
EcYFw	ACGCCTGGGATCCATGCAAGCGTCGCAATTTTCAGCCCT	Sitio BamHI	Clonación de EcMutY
EcYRv	AGTCCGCAAGCTTCTAAACCGGCGCCAGTGCCTAACT	Sitio HindIII	Clonación de EcMutY
Nterm-EcYFw	GGTCTTCATATGCAAGCGTCGCAATTTTCAGCC	Sitio NdeI	Clonación y ensamble de quimera
EcFrag1Rv	ACCCAGAGAAAGCGAGAGAATCGCGCCTGCG		Ensamble de quimera EcLbY/EhY1
EcFrag2Rv	ACCCAAATCCATCATCGCCTGATTAAACCGT		Ensamble de quimera EcLbY/EhY2
EFrag3Rv	TTTCGGTTTTTTGCCCCGATAAAGCGCCCAG		Ensamble de quimera EcLbY/EhY3
EcLbY1Fw	CGCGATTCTCTCGCTTTCTCTGGGTAAACCGTCCC GGCGGTG		Ensamble de quimera EcLbY1
EcLbY2Fw	TAATCAGGCGATGATGGATTGGGTGCTTCATATATGTCGGC		Ensamble de quimera EcLbY2
EcLbY3Fw	GCTTTATCCGGGCAAAAAACCGAAAAAAAACGTCCGATTCCG		Ensamble de quimera EcLbY3
Cterm-LbYRv	ATTAGCGGATCCTTATTGTTCTTTCTGCTGCTG	Sitio BamHI	Ensamble de quimera EcLbY
EcEhY2Fw	TAATCAGGCGATGATGGATTGGGTGCTCTTGTTGCAACTGGTA		Ensamble de quimera EcEhY2
Cterm-EhYRv	AACTGAGGATCCTTATTTATTCAAAACTATAAAATCTC	Sitio BamHI	Ensamble de quimera EcEhY
LbYE42AFw	TATCACATTTGGATCAGTGCAATTATGCTGCAGCAAACGC	codón Ala	Mutante LbY E42A
LbYE42ARv	GCGTTTGTGCAGCATAATTGCACTGATCCAAATGTGATA	codón Ala	Mutante LbY E42A
LbYD143AFw	CAAACCGGTCCC GGCGGTGGCTGGTAATGCCCTGCGTGTTC	codón Ala	Mutante LbY D143A
LbYD143ARv	GAAAAACGCAGGGCATTACCAGCCACCGCGGGACCGGTTTG	codón Ala	Mutante LbY D143A
EhYE78AFw	ATATAAAATACTTGTAAGTGCAAGTAATGTTACAACAAACAC	codón Ala	Mutante EhY E78A

EhYE78ARv	GTGTTTGTGTAACATTACT <u>UGC</u> ACTTACAAGTATTTTATAT	codón Ala	Mutante EhY E78A
EhYD179AFw	TCAAGATATTGCAACAGTAG <u>CCG</u> CTAATATTAGAAGAATTCTT	codón Ala	Mutante EhY D179A
EhYD179ARv	AAGAATTCTTCTAATATTAG <u>CGG</u> CTACTGTTGCAATATCTTGA	codón Ala	Mutante EhY D179A
RpoCluster2Fw	CGTGTAGAGCGTGCGGTGAAA		Secuenciación del cluster II de Rpo
RpoCluster2Rv	TTCACCCGGATACATCTCGTC		Secuenciación del cluster II de Rpo
Substrate complementary strand	CCTTGCCACTAGCGCYATATCAGTTAGGTGGGCAGGTCCGCTGCG	Y=8oG, G, A, C ó T	Actividad y anisotropía
Substrate template strand	CGCAGCGGACCTGCCCACCTAACTGATATXGCGCTAGTGCCAAGG	X= G, A, C ó T	Actividad
Fluorescein-abasic oligonucleotide	CGCAGCGGACCTGCCCACCTAACTGATATXGCGCTAGTGCCAAGG	X=Sitio Ab. 5'-marcaje con fluoresceina	Anisotropía

Resultados de la secuenciación del gen de Rpo- β de las colonias resistentes a rifampicina del ensayo *in vivo*.

	Mutación		base
	Tipo de mutación	aminoácido	
pCold 1	G→T	D516Y	1546
pCold 2	C→A	H526N	1576
pCold 3	None	---	---
pCold 4	C→A	S531Y	1592
pCold 5	C→A	H526N	1576
pCold-EcY 1	C→A	H526N	1576
	T→A	L680Stop	2039
pCold-EcY 2	C→A	S531Y	1592
pCold-EcY 3	G→T	D516Y	1546
pCold-EcY 4	None	---	---
pCold-EcY 5	C→A	H526N	1576
pCold-LbY 1	C→A	H526N	1576
pCold-LbY 2	C→A	S512Y	1535
pCold-LbY 3	C→A	S522Y	1565
pCold-LbY 4	C→A	S522Y	1565
pCold-LbY 5	C→A	S522Y	1565
pCold-EhY 1	None	---	---
pCold-EhY 2	C→A	H526N	1576
pCold-EhY 3	C→A	H526N	1576
pCold-EhY 4	C→A	S531Y	1592
pCold-EhY 5	C→A	S522Y	1565

Mutation				
Amino acid	Type of mutation	Site	Incidence (n)	Incidence (%)
S512Y	C→A	1532	1	4.7
D516Y*	G→T	1546	2	9.5
S522Y	C→A	1565	4	19
H526N	C→A	1576	7	33.3
S531Y*	C→A	1592	3	14.2
L680Stop	T→A	2039	1	4.7
No mutation found	---	---	3	14.2

*Aminoácidos que interaccionan con rifampicina (F513, R540, Q512 D516 and S531).

Ensamblaje y clonación de quimiras Ec-LbY y Ec-EhY.

Para el ensamblaje los genes quiméricos se uso el método de SOEing PCR. Para la mayoría de las quimeras se partio de la amplificación de los productos independientes a partir de los genes wild-type clonados en el vector pCold1. Para el ensamblaje de las quimeras Ec-EhY1 y 2 se partieron de G-blocks. Los fragmentos esperados para la región N-terminal de EcMuY son de tamaño de 393, 596 y 675 pb para los fragmentos 1, 2 y 3, respectivamente. Los tamaños correspondientes para los fragmentos C-terminal de LbMutY on 753, 581 y 468 pb para los fragmentos 1, 2 y 3, respectivamente. En lo concerniente al fragmento C-terminal de EhMutY su tamaño es 246 pb. Los fragmentos amplificados se muestran en la **Figura A1**. Estos fueron purificados y usados como templados para el ensamblaje por SOEing

PCR. Los amplicones ensamblados dan un tamaño de 1125 pb para Ec-LbY1 y 2 y 1128 pb para Ec-LbY3. Los tamaños de los genes ensamblados para Ec-EhY1, 2 y 3 fueron 827, 840 y 876 pb, respectivamente (**Figura A2**). Estos ensamblados fueron clonados en el vector pCold1 y secuenciados. Estas MutYs quiméricas fueron expresados y purificadas como las wild-type.

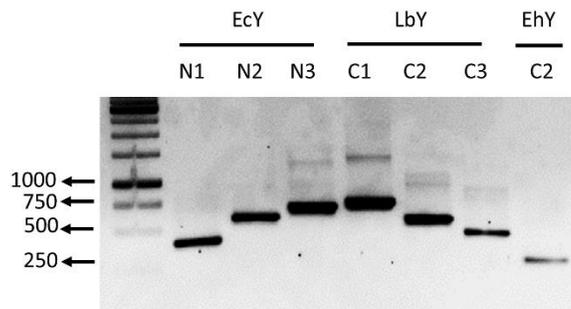


Figura A1. Amplificación de los fragmentos individuales para los ensamblajes de las mutYs quiméricas.

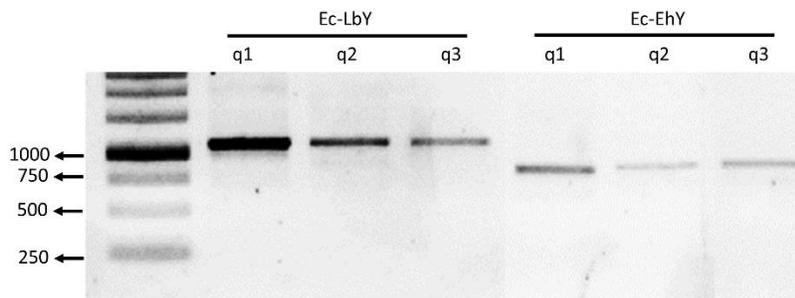


Figura A2. Ensamblaje de quimeras por SOEing PCR:

1er producto de investigación

The **FEBS**
Journal



Dispensability of the [4Fe-4S] cluster in novel homologues of adenine glycosylase MutY

Carlos H. Trasviña-Arenas, Laura M. Lopez-Castillo, Eugenia Sanchez-Sandoval and Luis G. Briebe

Laboratorio Nacional de Genómica para la Biodiversidad, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, Guanajuato, México

Dispensability of the [4Fe-4S] cluster in novel homologues of adenine glycosylase MutY

Carlos H. Trasviña-Arenas, Laura M. Lopez-Castillo, Eugenia Sanchez-Sandoval and Luis G. Brieba

Laboratorio Nacional de Genómica para la Biodiversidad, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, Guanajuato, México

Keywords

[4Fe-4S] cluster; adenine glycosylase; DNA repair; domain rearrangement

Correspondence

L. G. Brieba, Laboratorio Nacional de Genómica para la Biodiversidad, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Km 9.6 Libramiento Norte Carretera Irapuato-León 36821, Irapuato, Guanajuato, CP 36500, México
Fax: +52(462)6245846
Tel: +52(462)1663007
E-mail: lgrieba@langebio.cinvestav.mx

(Received 30 May 2015, revised 15 November 2015, accepted 24 November 2015)

doi:10.1111/febs.13608

7,8-Dihydro-8-deoxyguanine (8oG) is one of the most common oxidative lesions in DNA. DNA polymerases misincorporate an adenine across from this lesion. Thus, 8oG is a highly mutagenic lesion responsible for G:C→T:A transversions. MutY is an adenine glycosylase, part of the base excision repair pathway that removes adenines, when mispaired with 8oG or guanine. Its catalytic domain includes a [4Fe-4S] cluster motif coordinated by cysteinyl ligands. When this cluster is absent, MutY activity is depleted and several studies concluded that the [4Fe-4S] cluster motif is an indispensable component for DNA binding, substrate recognition and enzymatic activity. In the present study, we identified 46 MutY homologues that lack the canonical cysteinyl ligands, suggesting an absence of the [4Fe-4S] cluster. A phylogenetic analysis groups these novel MutYs into two different clades. One clade is exclusive of the order Lactobacillales and another clade has a mixed composition of anaerobic and microaerophilic bacteria and species from the protozoan genus *Entamoeba*. Structural modeling and sequence analysis suggests that the loss of the [4Fe-4S] cluster is compensated by a convergent solution in which bulky amino acids substitute the [4Fe-4S] cluster. We functionally characterized MutYs from *Lactobacillus brevis* and *Entamoeba histolytica* as representative members from each clade and found that both enzymes are active adenine glycosylases. Furthermore, chimeric glycosylases, in which the [4Fe-4S] cluster of *Escherichia coli* MutY is replaced by the corresponding amino acids of LbY and EhY, are also active. Our data indicates that the [4Fe-4S] cluster plays a structural role in MutYs and evidences the existence of alternative functional solutions in nature.

Introduction

DNA is subject to physical and chemical insults that alter its coding potential. Thus, organisms have evolved DNA repair mechanisms to maintain their genome integrity [1]. The low redox potential of deoxyguanine (G) makes it prone to oxidation. The oxidation product of G by reactive oxygen species yields the formation of

the most common DNA lesion: 7,8-dihydro-8-oxo-2'-deoxyguanine (8oG) [2]. The chemical repercussion of 8oG is a change in the equilibrium from the *anti* to the *syn* conformation. During replication, 8oG in a *syn* conformation is prone to pair with deoxyadenine, resulting in the accumulation of mutations [2–5]. As a

Abbreviations

8oG, 7,8-dihydro-8-deoxyguanine; AP, apurinic/aprimidinic; CT, charge transport; EcY, *Escherichia coli* MutY; EhY, *Entamoeba histolytica* MutY; FCL, iron-sulfur cluster loop; GsY, *Geobacillus stearothermophilus* MutY; HhH, helix-hairpin-helix; LbY, *Lactobacillus brevis* MutY; PDB, Protein Data Bank.

result of the high incidence of 8oG, a molecular mechanism for its avoidance is pivotal.

The mechanism by which organisms cope with 8oG is the GO repair system that includes at least three enzymes: (a) MutT, a nucleoside triphosphate pyrophosphorylase in charge of removing the 8odGTP from the cellular dNTP pool; (b) MutM, a DNA glycosylase that hydrolyses 8oG paired with C; and (c) MutY, an adenine glycosylase that acts on 8oG:A, removing the misincorporated A [6,7].

Structural and biochemical studies show that MutY is divided into two domains: an N-terminal domain or catalytic core [M1-K225 in *Escherichia coli* MutY (EcY)] and a C-terminal domain (Q226-V350) [8]. The catalytic core *per se* displays *in vitro* activity but with reduced substrate affinity [9,10]. This catalytic domain contains a [4Fe-4S] cluster, a widespread motif among proteins involved in nucleic acid metabolism. The [4Fe-4S] cluster plays diverse roles in proteins, such as a structural scaffold, DNA binding/translocation, lesion detection, protein assembly, redox sensing and enzymatic activity [11,12]. DNA glycosylases with [4Fe-4S] clusters encompass endonuclease III [13,14], UDG4 [15], Mig glycosylase [16], DME-class DNA glycosylase [17] and MutY [18]. In MutY and endonuclease III, the [4Fe-4S] cluster is coordinated by highly conserved cysteines (C₁₉₂X₆C₁₉₉X₂C₂₀₂X₅C₂₀₈) [14,19]. David and coworkers demonstrated that MutY activity *in vitro* is dependent on the [4Fe-4S] cluster, and that mutations of some of its cysteinyl ligands affect DNA binding [20–22].

In the last decade, the DNA charge transport (CT) property and the [4Fe-4S] cluster redox activity have been postulated as pivotal elements that account for a direct searching mechanism for DNA lesions. The positive base pair π -stack and negative phosphate backbone of DNA double helix form a molecular wire in which electrons can be transported [23–25]. However, when a DNA lesion or mismatch is present, the transport of electrons is severely reduced [23–25]. Electron paramagnetic resonance spectroscopy and electrochemistry measurements with MutY, UDG and endonuclease III indicate that DNA controls the [4Fe-4S] cluster redox state. In the absence of DNA, the oxidation state of the [4Fe-4S] cluster is 2+ and, when DNA is present, the oxidation state is 3+ [14,23,25–28]. The DNA-dependent oxidation of the [4Fe-4S] cluster is concomitant with a three order of magnitude increase of affinity for DNA [28]. These findings indicate that proteins containing a [4Fe-4S] cluster have distal redox communication. First, one protein with a [4Fe-4S]²⁺ cluster is oxidized with the concomitant association to DNA. The electrons resulting from the cluster oxida-

tion are passed through DNA and reduce the [4Fe-4S] cluster of another protein at a distal location, and so on. This redox communication carries on until a DNA lesion is encountered; then, a protein with a [4Fe-4S]³⁺ is located near the lesion allowing DNA lesion recognition [23–28].

Despite the importance of the [4Fe-4S] cluster for DNA lesion searching, there are several examples in which this motif is absent or modified. First, several DNA glycosylases such as MutM [29], endonuclease VIII [30], uracil DNA glycosylase [31], 3-methyladenine glycosylase I and II [32,33], and alkyladenine DNA glycosylase [34] lack the [4Fe-4S] cluster and efficiently recognize DNA lesions. By contrast to the DNA charge transport mechanism postulated for MutYs, human alkyladenine DNA glycosylase and uracil DNA glycosylase scan for DNA lesions using concerted mechanisms of facilitated diffusion and hopping [35–37]. In the uracil DNA glycosylase from *Mycobacterium smegmatis*, one of the cysteinyl ligands is replaced by a His without compromising the [4Fe-4S] cluster assembly [38]. *Saccharomyces cerevisiae* harbors two functional homologous genes for endonuclease III: *Nth1* and *Nth2*. *Nth2* preserves the four cysteines that coordinate the [4Fe-4S] cluster, whereas *Nth1* lacks these ligands and it is assumed that the cluster is not present [39–41]. In enzymes with important roles in metabolic reactions such as ferredoxin, a loss of the [4Fe-4S] cluster has been reported in *Thermotoga maritima* [42], *Desulfobibrio gigas* [43] and *Bacillus thermoproteolyticus* [44]. Previous studies identified the presence of MutYs lacking the [4Fe-4S] cluster in *Treponema pallidum*, *Lactobacillus gasseri* and *Streptococcus* [11,21,45]; however, these proteins were not characterized. In the present study, we identified a wide number of MutY homologues that lack the cysteinyl ligands for the [4Fe-4S] cluster in bacteria such as *Lactobacillales*, *Chlorobi*, *Spirochaetes*, *Deltaproteobacterias* and *Actinobacteria*, as well as in the protozoan genus *Entamoeba*. We demonstrated that MutYs lacking the [4Fe-4S] cluster are active *in vitro* and *in vivo* and our structural homology modelling and phylogenetic analysis indicate that the loss of the motif was driven by a convergent solution of domain rearrangement.

Results and Discussion

Evolution of novel MutY without a [4Fe-4S] cluster

The three enzymes that constitute the GO repair system: MutM, MutY and MutT, are present in the three

domains of life [46–49]. MutY and MutM belong to the helix-hairpin-helix (HhH) glycosylase superfamily [50]. Other members of the superfamily are AlkA [51,52], endonuclease III [13,14] and 3-methyladenine glycosylase [32]. Members of this superfamily of DNA glycosylases display a similar fold, despite sharing low amino acid sequence identity.

To understand the evolution of alternative structural solutions for the [4Fe-4S] cluster in MutY, we carried out a phylogenetic analysis of 186 putative adenine DNA glycosylases covering Archaea, Eukarya and Bacteria. From the analyzed DNA glycosylases, 140 conserved the characteristic [4Fe-4S] cluster cysteinyl ligands (C₁₉₂X₆C₁₉₉X₂C₂₀₂X₅C₂₀₈). Nonetheless, some variations were found in the first and second spacer regions where there is the addition or deletion of one residue in fungi, Archaea, euglenozoa and bacteria (Figs 1A, S1 and S2). The remaining 46 MutY sequences do not harbor the archetypical cysteinyl ligand pattern. These putative MutYs are present in bacteria, such as *Lactobacillae*, *Chlorobi*, *Spirochaetes*,

some *Deltaproteobacterias* and *Actinobacteria*, as well as in protozoan *Entamoeba*. The absence of the cysteinyl ligands suggests the loss of the [4Fe-4S] cluster motif in these MutYs [11,53]. Our phylogenetic analysis shows that MutYs in which the [4Fe-4S] cluster is absent are harbored in two separated clades: Lactobacillales and the anaerobic clade (Fig. 2). In the Lactobacillales clade, only MutYs from the order *Lactobacillales* are included, whereas the anaerobic clade includes MutYs from anaerobic or microaerophilic organisms, such as *Entamoeba*, *Chlorobi*, *Spirochaetes*, *Deltaproteobacterias* and *Actinobacteria*. Examples of [4Fe-4S] cluster replacement occur in other enzymes, such as (6-4) photolyase [54], AddAB helicase-nuclease [55] and tryptophanyl-tRNA synthetase [56]. The fact that the [4Fe-4S] cluster is conserved in the three domains of life provides strong evidence that this motif is an original component of MutY and it is plausible that this motif has been replaced independently in the Lactobacillales and anaerobic clades. Because the bacilli class includes the

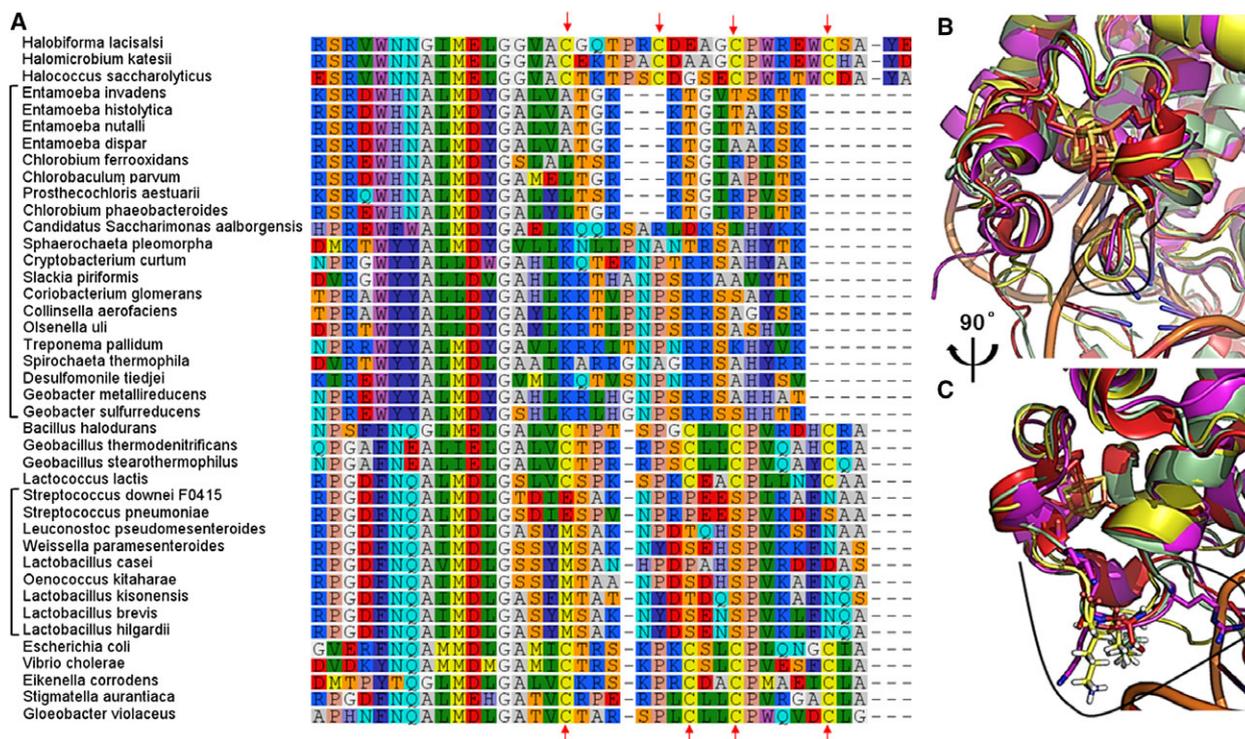


Fig. 1. Sequence and structural alignment of the [4Fe-4S] cluster motif. (A) Amino acid alignment, including canonical MutYs and MutYs with alternative solutions to the [4Fe-4S] cluster motif. Red arrows indicate the cysteinyl ligands. MutYs with alternative structural solution for the [4Fe-4S] cluster are indicated by square brackets. (B) Structural alignment for the [4Fe-4S] cluster domain of EcY (pink; PDB code: [1MUU](#)) [19], GsY (red; PDB code: [1RRQ](#)) [59], LbY (green) and EhY (yellow). LbY and EhY structures are derived from structural homology modeling. (C) 90° view of corresponding region of the [4Fe-4S] cluster domain. Positive residues of the [4Fe-4S] cluster loop are indicated by sticks: R194, K196 and K198 of EcY; R201 and R202 of GsY; K200 of LbY; and K232 and K233 of EhY. The [4Fe-4S] cluster loop is represented by a black outline.



Fig. 2. MutY phylogenetic analysis. (A) Evolutionary analysis of 186 amino acid sequences of MutY covering Archaea, Bacteria and Eukarya. The divergence of novel MutYs without a [4Fe-4S] cluster is indicated in green for Lactobacillales and yellow for the Anaerobic clade. The phylogenetic analysis was performed using the maximum likelihood method with 1000 replicates bootstrap resampling. Percentages of bootstrap support are indicated at the base of each node. (B) Magnified view of bacilli clade. (C) Magnified view of the Anaerobic clade.

orders Bacillales and Lactobacillales, the phylogenetic reconstruction indicates that the replacement of the [4Fe-4S] cluster is a relatively recent event after the divergence of *Lactobacillales*. On the other hand, the mixed composition of the anaerobic clade makes it difficult to understand its evolution. The presence of noncanonical MutYs in the genus *Entamoeba* is noteworthy. Others organism of the Amoebozoa phylum, such as *Acanthamoeba* and *Dictyotellium*, contain MutYs that harbor characteristic ligands for the [4Fe-4S] cluster and are phylogenetically distant from *Entamoeba* MutYs (*Entamoeba histolytica*, *Entamoeba dispar*, *Entamoeba invadens* and *Entamoeba nutalli*). Hence, it is plausible that *Entamoeba* acquired their noncanonical MutYs from bacteria. It is estimated that ~1% of the genes present in *E. histolytica* were acquired by lateral gene transfer from bacterial lineages [57]. For example, Loftus *et al.* [57] reported that the GO repair mutT has a bacterial origin in *E. histolytica*. *E. histolytica* also contains several DNA repair genes phylogenetically related to bacterial lin-

eages, such as Ung1, G:T/U glycosylase, Ape1 and specialized polymerases with *in vitro* DNA lesion bypass capabilities (data not shown) [6,58].

Sequence analysis and structural homology modeling

To confirm the functionality of noncanonical MutYs and understand how DNA glycosylase activity is maintained without the [4Fe-4S] cluster, we selected MutYs from the lactic acid bacteria *Lactobacillus brevis* (LbY) and from the pathogenic protozoan *Entamoeba histolytica* (EhY) as representative members of each clade.

The catalytic core of *E. coli* MutY comprises its first 225 amino acids, including the active site amino acids E37 and D138, the intercalating/extrusion residue Y82, the helix-turn-helix and [4Fe-4S] cluster motifs [8,9,19]. Structurally, the catalytic core consists of a six-helix barrel and a [4Fe-4S] cluster motif [19,59]. In LbY and EhY, their putative catalytic core presents 38.3% iden-

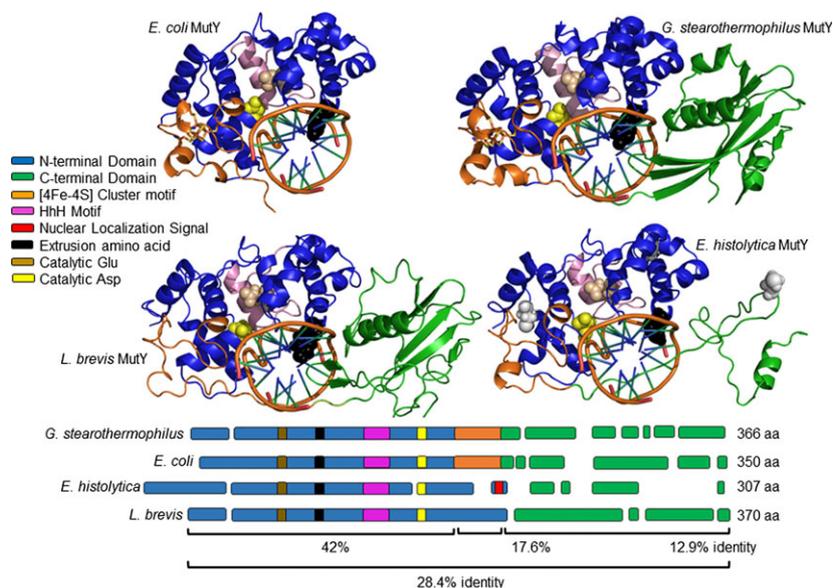
tity with respect to EcY and *Geobacillus stearothermophilus* MutY (GsY) (Fig. 3). Conserved residues include the putative active site amino acids E48 and D149 for LbY, and E78 and D179 for EhY. The intercalating/extrusion amino acid in LbY corresponds to Y87. This residue is also conserved in endonuclease VIII [60] and alkyladenine DNA glycosylases [34]. In the case of EhY, the putative intercalation/extrusion residue is F123, as in MutM [61]. The aromatic Tyr and Phe are involved in base flipping, participating in substrate affinity and specificity [62].

According to structural homology modeling, the main structural differences in LbY and EhY are localized at the [4Fe-4S] cluster motif and C-terminal domain. The [4Fe-4S] cluster motif is highly conserved in MutYs, and the cysteines that coordinate it present a conserved spacing pattern. However, an important feature of LbY and EhY is that these putative MutYs do not bear these cysteinyl ligands (Figs 1A and 3). Cysteine thiolates are the most common ligands for iron-sulfur clusters. Nevertheless, histidine, glutamine, serine and arginine are also ligands [20,53,63]. If a cysteinyl ligand is substituted, three cysteines are involved in cluster coordination and, in some cases, two cysteinyl ligands can coordinate [4Fe-4S] clusters [53,63–65]. LbY does not harbor cysteines in its amino acid sequence; hence, it is theoretically impossible to support a [4Fe-4S] cluster. EhY contains three cysteines but, according to the structural homology modeling, they are separated by more than 34 Å, which is too far for cluster coordination (Fig. 3). The [4Fe-4S] cluster loop (FCL) is an important element for substrate

recognition because some mutations on cysteine ligands and positively-charged residues affect DNA binding [20–22]. Despite the low identity of the [4Fe-4S] cluster motif (17.6%), its global scaffold and important positive residues of FCL for substrate affinity are preserved in the putative FCL of EhY and LbY (Fig. 1B, C). EcY contains three positive amino acids at its FCL (R194, K196 and K198), and mutations of these residues decrease substrate affinity [22]. EhY presents two Lys at its putative FCL (K232 and K233), whereas LbY presents one positively-charged residue (K200).

The most variable component between LbY and EhY is their C-terminal domain. The C-terminal domain of LbY is composed of seven β sheets and one α helix. In the case of EhY, one small moiety in its C-terminal domain is modeled as an α -helix, whereas the rest remain nonstructured. It is noteworthy that the C-terminal domain of EhY is considerably smaller, comprising 57 residues compared to 141, 136 and 126 residues for LbY, GsY and EcY, respectively (Fig. 3). The C-terminal domain of MutY and endonuclease III is distantly related to MutT, a pyrophosphohydrolase that acts on 2'-deoxy-8-oxoguanine triphosphate [66,67]. The C-terminal domain of EcY plays an important role in substrate specificity for 8oG and G paired with A [8,9,68,69]. Wild-type EcY processes 8oG:A and G:A substrates 30- and four-fold more efficiently than a C-terminal domain deletion [8,9]. The structure of GsY with its substrate shows that the amine moiety of the C-terminal residue S308 interacts with the O₈ of 8oG [59,70].

Fig. 3. Structural and sequence comparison of canonical and novel MutYs. Top: Crystal structures of the catalytic domain of EcY (PDB ID: [1MUJ](#) [19]) and full GsY (PDB code: [1RRQ](#)) [59]. Middle: Structural homology modeling of LbY and EhY. Homology modeling was performed using GsY as template. Bottom: Schematic amino acid alignment indicating percentage of identity among regions. The catalytic domain is colored in blue, the [4Fe-4S] cluster motif is in orange, the HhH motif is in pink, and the C-terminal domain is in green. The positions corresponding to Glu and Asp catalytic residues are in brown and yellow, and the extrusion residue in black.



Corroboration of the absence of the [4Fe-4S] cluster in LbY and EhY

The cysteinyl ligands (C₁₉₂X₆C₁₉₉X₂C₂₀₂X₅C₂₀₈) that coordinate the [4Fe-4S] cluster are highly conserved in canonical MutYs (Fig. S1) and mutations in these residues decrease enzymatic activity and stability [20,21]. To carry out a comparative biochemical study between MutYs in which the [4Fe-4S] cluster is predicted to be absent, as well as MutYs with a canonical [4Fe-4S] cluster, we cloned and purified LbY, EhY and EcY proteins. Wild-type proteins were overexpressed in a JM101 *muty*⁻ *E. coli* strain [71]. The recombinant MutYs were purified by nickel affinity and cationic exchange chromatography (Fig. 4A). We aimed to corroborate the absence or presence of the [4Fe-4S] cluster in LbY and EhY experimentally, mainly because of the possibility that one or more conserved cysteinyl ligands could be replaced by a distant nonconserved

ligand through a swapping mechanism [72,73]. The [4Fe-4S] cluster can be detected by different approaches, such as the solution turning a distinctive yellow [9,55,74], spectroscopy of the cysteinyl thiolates-to-Fe (III) [21,55], the chromogenic effect of agents such as bathophenanthroline sulfonate, ferene or luminol [55,75,76], and electron paramagnetic resonance [21,27,55]. The solutions with purified LbY and EhY remained colorless during the whole purification process, and the characteristic peak of the [4Fe-4S] cluster at 410 nm was not observed in the ultraviolet-visible spectrum (data not shown). To further investigate the fate of the [4Fe-4S] cluster, we used a chemiluminescence assay that relies on luminol producing chemiluminescence in the presence of iron and hydrogen peroxide [76]. One advantage of this procedure is the ability to detect Fe²⁺ and Fe³⁺, the common oxidation states of the [4Fe-4S] cluster in solution. To increase the sensitivity of the reaction, we treated each protein with 8 M urea to induce denaturation and iron release. The chemiluminescence assay was carried out with 30 and 150 μmol of recombinantly expressed proteins (0.1 nmol is the lowest sensitivity limit for this method). The relative chemiluminescence values with respect to the blank (no protein present) for EcY were 321 and 137 when 150 and 30 nmol of protein were added to the reaction respectively. For LbY and EhY, the observed chemiluminescence values were similar to the background (Fig. 4B). Thus, the chemiluminescence method corroborates the presence of the [4Fe-4S] cluster in EcY and indicates its absence in LbY and EhY.

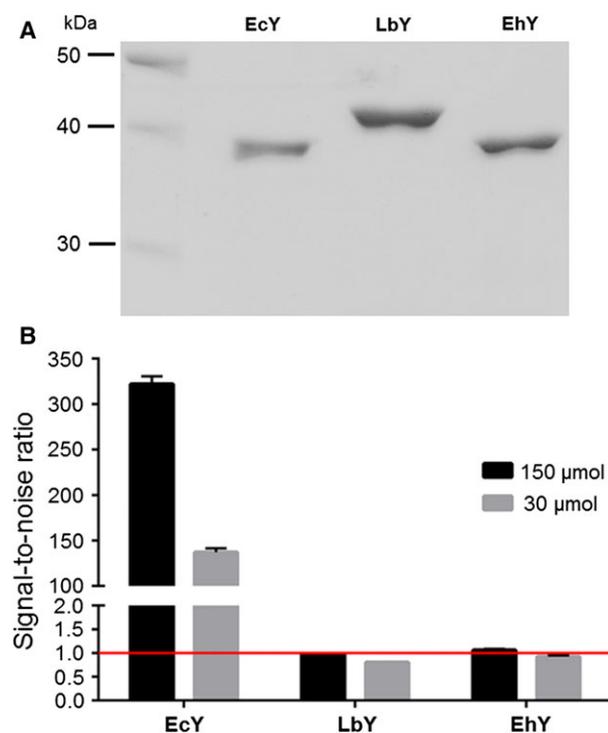


Fig. 4. SDS/PAGE of purified MutYs and chemiluminescence assay for the [4Fe-4S] cluster detection. (A) 12% SDS/PAGE analysis stained with Coomassie blue after nickel affinity and cationic exchange chromatography. The bands represent the purified DNA glycosylases. (B) Chemiluminescence assay for [4Fe-4S] cluster detection based on the luminol-H₂O₂-Fe reaction with 150 μmol and 30 μmol of MutY. All measurements were divided by the chemiluminescence background (reaction without enzyme) and reported as the signal-to-noise ratio. A red line indicates the normalized chemiluminescence background value.

In vitro activity and active site mutations

The catalytic mechanism of MutY involves two carboxylates, E43 and D144, and an aromatic residue (Y82) that separates the adenine from 8oG [59,62,70,77]. Although the distance of 19.1 Å between the [4Fe-4S] cluster and the catalytic pocket hinders physical and chemical interaction with adenine during catalysis, the [4Fe-4S] cluster is considered as an important factor for enzymatic activity [18,20,21,78]. Sequence analysis suggests the existence of MutYs without the [4Fe-4S] cluster [11,21,45], although one aspect that is unclear is whether these putative MutYs maintain DNA glycosylase activity. To tackle this question, we evaluated the *in vitro* activity of purified LbY and EhY with 8oG and G paired with canonical DNA. For this assay, we used a 5' radiolabeled 45-mer oligonucleotide in which the nucleotide at position 30 was either A, G, C or T. The radiolabeled oligonucleotides were paired with a complementary 45-mer

oligonucleotide containing 8oG or G adjacent to position 30. To observe a clear activity signal, the assay was performed using an excess of enzyme (~200 nM) over 1 nM DNA substrate. LbY, EhY and EcY show DNA glycosylase activity over 8oG:A and G:A base pairs because the cleavage of the 45-mer substrate releases a 30-mer product (Fig. 5A) as reported previously for canonical MutYs [7,79]. It is important to highlight that LbY and EhY were overexpressed in a JM101 *muty*⁻ *E. coli* strain discarding contamination with endogenous MutY. Thus, the biochemical data corroborate that EhY and LbY are DNA glycosylases able to specifically recognize and process 8oG or G paired with A.

The catalytic mechanism of MutY is well known where adenine is separated from the DNA double helix and placed in the catalytic pocket. Once in the catalytic pocket, two water molecules participate in the catalysis activated by two carboxylates, E37 and D138, in EcY [77,80]. These residues are conserved in LbY (E48 and D149) and EhY (E78 and D179). To determine whether they are functionally conserved, we generated single and double mutants of these residues to alanine. The MutY mutants were expressed and purified as their wild-type counterparts. None of the

mutants were active over 8oG:A (Fig. 5B). Only the E78A mutant of EhY presented residual activity, although this can be attributed to subtle variations in the general chemical mechanism for glycosylase activity between EcY and EhY.

DNA glycosylases can be grouped into monofunctional and bifunctional glycosylases. Bifunctional glycosylases, such as OGG1 and endonuclease III, use an active amino group (commonly a Lys) to perform N-glycosidic bond hydrolysis concomitant with 3' phosphodiester cleavage, whereas monofunctional glycosylases do not possess DNA lyase activity [81,82]. To measure the *in vitro* activity of monofunctional glycosylases, it is necessary to add NaOH or piperidine at the end of the reaction to cleave the phosphodiester bond through β and σ -elimination. To identify the monofunctional or bifunctional features of EhY and LbY, we measured their activity *in vitro* with 8oG:A, stopping the reaction with or without NaOH (0.2 M). DNA cleavage by EcY, LbY and EhY is clearly observed when NaOH is added at the end of the reaction. However, when NaOH is absent, only residual activity is observed (Fig. 5C). Thus, it can be concluded that, similar to EcY, LbY and EhY are canonical monofunctional glycosylases [81,82].

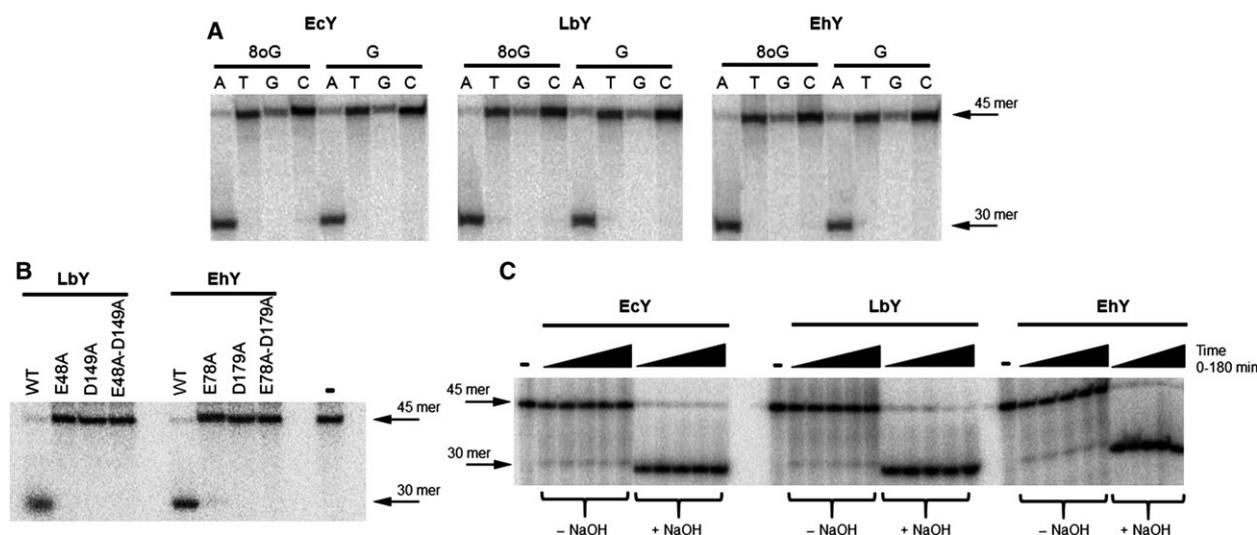


Fig. 5. MutYs without a [4Fe-4S] cluster motif present DNA glycosylase activity. (A) Denaturing gel electrophoresis analyzed through phosphorimager illustrating the glycosylase activity over 8oG and G paired with different bases and mismatches. EcY, LbY, and EhY are clearly active over 8oG:A and G:A. For the assays, 1 nM of 5'-end labeled substrate with ³²P containing A, G, C and T paired with 8oG and G was incubated with at least 200 nM of enzyme. The reactions were stopped with 0.2 M of NaOH and heated at 90 °C for 5 min after 1 h. (C) *in vitro* activity evaluation of single and double mutant of the carboxylates involved in the catalysis, indicating that active sites are functionally conserved. (B) AP lyase determination with 1 nM of substrate containing 8oG:A. An alternative structural solution for the [4Fe-4S] cluster MutY in LbY and EhY is monofunctional glycosylases. To identify AP, the reaction was stopped by adding only running buffer (99% formamide, 20 mM EDTA and 0.025% bromophenol blue) and heating at 90 °C for 1.5 min (treatment shown as -NaOH). To identify monofunctional activity, the reaction was stopped conventionally (treatment shown as +NaOH). The reactions for AP-lyase activity were monitored from 5 to 80 min. All reactions were carried out at 37 °C. The intact substrate (45 mer) and product (30 mer) are indicated by arrows.

The amount of recombinantly expressed MutY assembled with a [4Fe-4S] cluster can vary as a result of culture and purification factors [83–85]. Thus, the expressed MutYs may contain a [4Fe-4S] cluster at an amount below the detection limits of chemiluminescence and spectroscopy methods. To investigate this possibility, we performed an *in vitro* [4Fe-4S] cluster reconstruction experiment designed by Porello *et al.* [18]. In this experiment MutYs are denatured and the components of their [4Fe-4S] cluster are removed by chelation and dialysis. The cluster reconstruction protocol is shown schematically in Fig. 6A [18]. If MutY activity is dependent on the [4Fe-4S] cluster, the enzymatic activity will be depleted with this procedure. When EcY is subjected to the denaturation/cluster chelation/refolding process (treatment dubbed R), its enzymatic activity is lost, whereas LbY and EhY retain their glycosylase activity (Fig. 6B, upper). In all R samples, the [4Fe-4S] cluster is not detected by the chemiluminescence assay, indicating its absence from the beginning or that iron was removed correctly

(Fig. 6B, lower). When EcY was refolded in the presence of sulfur and iron (treatment called R^{Fe-S}), its activity and detection were recovered, indicating that the protocol is suitable for obtaining a population of correctly folded EcY. These data support the notion that the [4Fe-4S] cluster is not necessary for the catalytic activity of EhY and LbY.

Single and multiple turnover kinetics

Monofunctional glycosylases remain linked to the product (abasic site) after hydrolysis. This makes the release of enzyme–product complex a limiting step during catalysis, indicating that standard Michaelis–Menten parameters, K_m and K_{cat} , are unsuitable for kinetic analysis. Hence, in the present study, kinetic constants were obtained under single and multiple turnover conditions. With this approach, it is possible to determine the rate constant for chemical reaction (k_2) and product release (k_3) [86]. For single and multiple turnover kinetics, a duplex DNA containing ade-

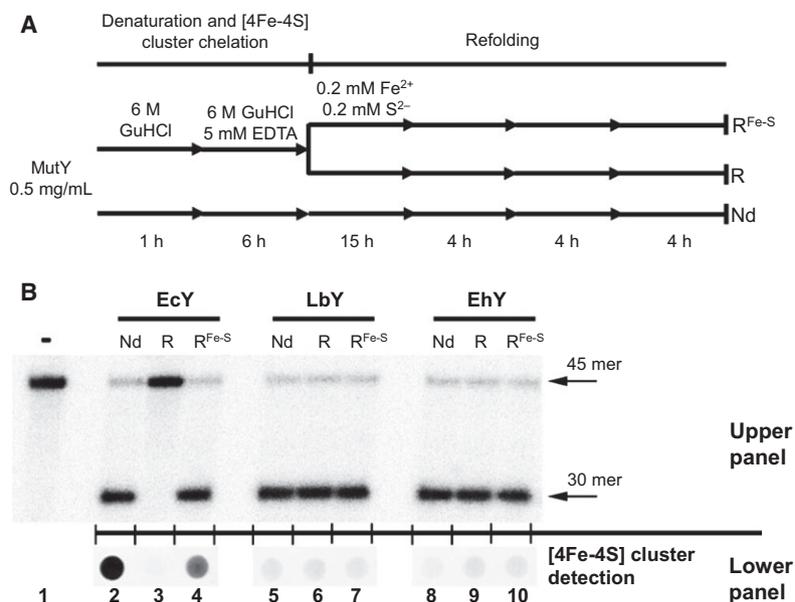


Fig. 6. *In vitro* [4Fe-4S] cluster reconstruction experiment. (A) Schematic representation of the *in vitro* [4Fe-4S] cluster reconstruction experiment designed by Porello *et al.* [18]. The duration of each experimental step is indicated by an arrow. The denaturation and [4Fe-4S] cluster chelation was performed in buffer A (50 mM Tris-HCl, pH 7.6, 200 mM NaCl, 2 mM 2-mercaptoethanol and 0.1 mM EDTA) with the addition of the components indicated above each arrow. MutY samples subjected to denaturation/[4Fe-4S] cluster chelation/refolding procedures are indicated as R. The samples that were refolded in the presence of ferrous and sulfide ions were labeled R^{Fe-S} and a nondenaturing positive control (Nd) was also present. (B) *In vitro* glycosylase activity (top) and qualitative chemiluminescence detection (bottom) of the [4Fe-4S] cluster derived from a reconstruction experiment. ADN glycosylase activity was carried out with 1 nM of radiolabeled 45-mer duplex containing 8oG:A lesion and 500 nM of MutY for 30 min at 37 °C. Glycosylase activity is present in the control EcMutY in the Nd and R^{Fe-S} treatments, but not in the R treatment (lanes 2–4). Glycosylase activity is present in all treatments for EhY and LbY (lanes 5–10). The qualitative chemiluminescence detection analysis of each MutY samples was assayed using 1.5 μM of each MutY subject to the R, R^{Fe-S} and Nd treatments (bottom). A Chemiluminescence signal above the background is observed for R^{Fe-S} and Nd EcMutY samples (lanes 2 and 4).

nine opposite 8oG was used as substrate. We decided to carry out our kinetic assays at 21 °C to obtain measurable data with 8oG:A duplex.

To evaluate MutY binding, a fluorescence anisotropy assay was performed using a fluorescein-labelled DNA duplex with an abasic site (THF) paired to 8oG to determine the half-maximal binding protein concentrations ($[P]_{1/2}$). The measurements of EcY binding affinity resulted in a decrease up to 40% in the detected signal compared to EhY and LbY (data not shown). This phenomenon has been previously observed for EhY and associated with the [4Fe-4S] cluster presence as a quencher of fluorescence emission [87]. Therefore, anisotropy data were adjusted to compensate the decrease of quantum yield of protein-substrate complex relative to free substrate [87,88].

The binding affinities of EcY, LbY and EhY for an 8oG:abasic site-containing duplex were obtained by titration of 1 nM labeled duplex with different MutY concentrations ranging from 0.1 to 75 nM. The binding isotherms (Fig. 7C) were fitted to a 'one site specific binding with Hill slope model' to determine the $[P]_{1/2}$ values. EhY had the highest binding affinity, with a $[P]_{1/2}$ of 6.83 nM, followed by EcY with 12.59 and LbY with 41.37 nM. Similar binding affinity values have been reported for *Schizosaccharomyces pombe*

MutY and *E. coli* UDG with an abasic-containing duplex [89,90]. Binding affinity values < 45 pM for radiolabel duplex containing the abasic:8oG pair have been reported using electrophoretic mobility shift assay [91–93].

To determine rate constants, we measured the percentage of active enzymes by the amplitude of the burst under multiple turnover condition and measured single and multiple turnover kinetics of active MutYs (Fig. S3). The percentage of active MutYs varied from 28% to 66% (Table 1). The rate constant k_3 was derived from the slope of the linear portion of the curves under multiple turnover condition with 2.5 nM active MutYs and 10 nM 8oG:A-containing substrate. All kinetic values are listed in Table 1. The rate constant k_3 was calculated to be 0.002, 0.006 and 0.005 min^{-1} for EcY, LbY and EhY, respectively. These slow k_3 values indicate that enzymatic turnover is limited and product affinity of MutY is high. Similar data for EcY and its human homologue have been reported [62,86,91,94].

Single turnover kinetics were performed with 10 nM 8oG:A duplex substrate and 50 nM active enzyme at 21 °C. Under these conditions, the resulting k_2 values were 1.324, 0.323 and 1.485 min^{-1} for EcY, LbY and EhY, respectively. Performing the assay at 21 °C

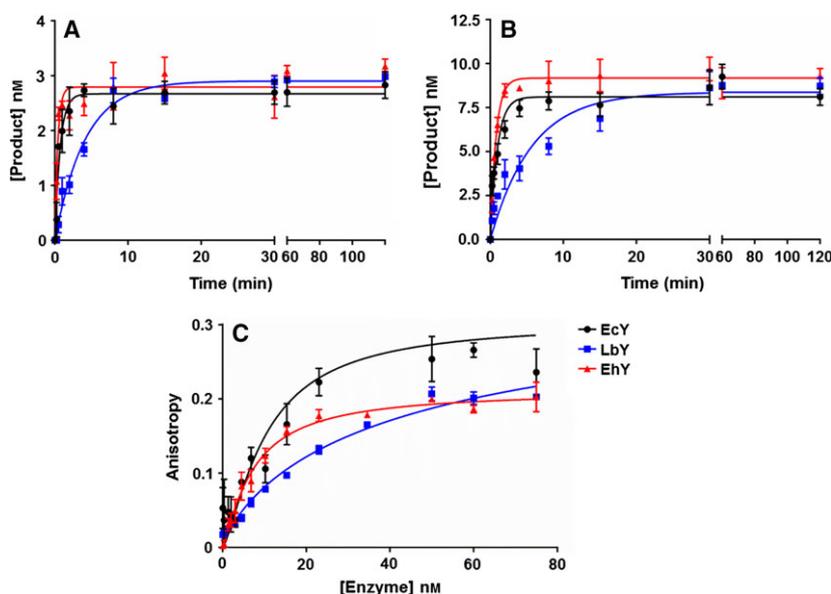


Fig. 7. Multiple and single turnover kinetics with duplex containing 8oG:A and binding isotherms for DNA duplex containing 8oG:abasic site. (A) Multiple turnover kinetics. In multiple turnover experiments, 10 nM of substrate and 2.5 nM of active MutYs were incubated at 21 °C in 40 mM Tris (pH 7.5), 10 mM EDTA, 0.1 $\text{mg}\cdot\text{mL}^{-1}$ BSA and 150 mM NaCl. (B) Single turnover kinetics. Single turnover experiments were carried out as FOR multiple turnover kinetics using 50 nM of active enzyme. Aliquots (8 μL) were taken from the reaction mixture at 0.25, 0.5, 1, 2, 4, 8, 15, 30, 60 and 120 min. All experiments were performed in triplicate, quenched with 0.2 M NaOH and heated at 90 °C for 5 min. The product concentration is plotted as function of time. (C) Binding isotherms obtained by titration of 1 nM of fluorescein-labelled duplex with abasic site (THF) paired to 8oG with different MutY concentrations ranging from 0.1 to 75 nM.

Table 1. Rate constants for adenine glycosylase activity with 8oG:A and dissociation constant for 8oG:THF. Percentage of active enzyme from different protein purifications was determined in multiple turnover conditions. Single turnover kinetics were carried out with 50 nM of active enzyme and 10 nM of substrate at 21 °C. For determination of k_2 , the $[P]_{1/2}$ values were considered as K_d . Multiple turnover conditions were carried out with 2.5 nM of active enzyme and 10 nM of substrate at 21 °C. $[P]_{1/2}$ was determined by anisotropy with 1 nM of abasic site (THF) paired with 8oG titrated with 0.1–75 nM MutY at 26 °C. All data were obtained in triplicate and the error reported corresponds to the SD.

MutY	Active enzyme (%)	Single turnover kinetics			Multiple turnover kinetics	
		k_2 (min ⁻¹)	Relative to EcY k_2	k_{obs} (min ⁻¹)	k_3 (min ⁻¹)	$[P]_{1/2}$ (nM)
EcY	38.7, 66	1.324 ± 0.34	1	1.057 ± 0.27	0.002 ± 0.001	12.593 ± 2.57
LbY	28.1, 47.4	0.323 ± 0.02	0.2	0.177 ± 0.01	0.006 ± 0.002	41.370 ± 8.69
EhY	51.8, 42.8	1.485 ± 0.17	1.1	1.306 ± 0.15	0.005 ± 0.001	6.837 ± 2.12

allowed us to obtain feasible k_{obs} values for multiple and single turnover kinetics with an 8oG:A-containing duplex. Livingston *et al.* [62] carried out a comparative kinetic study of EcY with different temperatures, and found that, at 25 °C, the k_2 value for 8oG:A was 5.7 min⁻¹, a reduction of three times the k_2 at 37 °C [62]. Accordingly, we observed a reduction of at least one order of magnitude for the rate constant compared to that obtained at 37 °C (EcY k_{obs} > 10 min⁻¹) [62,86,91]. As shown in Fig. 7, the maximum product released for LbY is reached at ~ 10 min, unlike EcY and EhY, where it is reached at ~ 4 min. The k_2 constant is an expression of catalytic and binding capabilities. Therefore, the low catalytic rate of LbY can be understood by paying attention to k_{obs} and $[P]_{1/2}$ values. As noted, the dissociation constant for LbY is 2.11-fold higher than the EcY $[P]_{1/2}$ value; hence, the low catalytic rate for LbY could be explained by a reduced substrate affinity. Interestingly, despite EhY lacking a [4Fe-4S] cluster, it displays a higher substrate binding affinity than EcY or LbY, and a catalytic rate similar to EcY.

There are some factors that can affect the activity of MutY. For example, deletion of the C-terminal domain generates a drop in the catalytic rate of at least two orders of magnitude [91]. Another important aspect is the integrity of the [4Fe-4S] motif. One cysteinyl ligand can be substituted by Hys or Ser residues without important variation of catalytic parameters, although the effect of these substitutions is intensified *in vivo* [20,21]. Within the [4Fe-4S] cluster loop in EcY, residues R194, K196 and K198 are important elements for DNA recognition, and mutations to Ala were shown to reduce substrate affinity [22]. According to structural homology modeling (Fig. 3), LbY with a structurally conserved FCL motif only harbors one positively-charged residue (K200), whereas EhY shows two (K232 and K233). The major differences in sequence and structure are centered at the [4Fe-4S] cluster region and C-terminal domain. Indeed, the C-

terminal for EhY remains unstructured. Nonetheless, these differences do not supply any catalytic disadvantage. Outstandingly, lack of the [4Fe-4S] cluster does not have any repercussion on the catalytic parameters in these novel MutYs.

Mutation frequency: *in vivo* assay

The *in vitro* activity of LbY and EhY is clear; however, it is necessary to corroborate whether these enzymes are active *in vivo*. To tackle this, we measured the mutation rate to rifampicin resistance [20,91,95]. Rifampicin interacts with a specific region of the β subunit of RNA polymerase (β -Rpo), inhibiting transcription [96–98]. Thus, when there is a high frequency of mutations (given in our case by lack of *muty* or the presence of a defective variant), the likelihood that mutations occur in the Rpo–rifampicin interaction region increases. Hence, a proxy for DNA glycosylase activity *in vivo* is determined by quantifying the appearance of rifampicin resistant colonies and expressed as the mutation frequency (f). The rifampicin resistance assays were performed in the CC104 *muty*⁻ *mutm*⁻ *E. coli* strain. This strain is sensitive to 8oG-dependent mutations because it lacks almost all of the machinery needed to repair them. The f values are presented in Table 2. Because the pCold expression vector preferentially expresses recombinant proteins at low temperatures (below 18 °C) and the *in vivo* assays were carried out at 37 °C, as a control, we used EcY cloned into a pUC derived vector (pKK223-3) [9,99]. The CC104 *muty*⁻ *mutm*⁻ *E. coli* strains without transformation (referred as none) and transformed with empty vector were used as negative controls. The pKKEcY positive control and pColdEcY presented a similar reduction in their f values, below 98% relative to negative controls. Therefore, recombinant MutYs cloned into a pCold vector are adequate for measuring *in vivo* activity at 37 °C. The pColdLbY and pColdEhY constructs were able to decrease the mutation

Table 2. Rifampicin resistance assay data in CC104 *muty⁻ mutm⁻* *Escherichia coli*. The *rpoB* mutation frequency (*f*) per cell was calculated by dividing the median number of mutants by the average number of cells in a series of cultures. *f* is obtained from at least eight separate experiments and errors correspond to one SD from the average.

Plasmid	<i>f</i> ($\times 10^{-6}$)	Increase (fold) over pCold-EcY	Relative decrease (%) to control (pCold)
None	8.104 \pm 3.74	76.7	21.9
pCold	10.386 \pm 4.36	98.4	0
pKKEcY	0.043 \pm 0.02	0.4	99.5
pColdEcY	0.106 \pm 0.08	1	98.9
pColdLbY	0.068 \pm 0.04	0.6	99.3
pColdEhY	0.146 \pm 0.10	1.3	98.5

frequency to the same extent as pColdEcY and pKKEcY. Interestingly, LbY, despite having a reduced *in vitro* hydrolysis rate, decreased the *f* values by almost two-fold compared to pColdEcY. To corroborate that mutations of β -Rpo were G:C to T:A transversions, we PCR amplified the region of clusters 1 and 2 of β -Rpo (from 1410 to 2235 bp) [98] and sequenced them. We found that more than 85% of analyzed colonies had a mutation in the cluster 1 or 2 of β -Rpo and all mutations were transversions (data not shown), supporting the notion that MutYs with alternative solutions for [4Fe-4S] clusters are functional DNA glycosylases.

Absence of the [4Fe-4S] cluster could be driven by steric effects

The integrity of the [4Fe-4S] cluster is reported to be essential for DNA binding, lesion recognition and activity in several DNA glycosylases [15,18,20–22,27,74,78]. Thus, if the [4Fe-4S] cluster is an essential component for DNA glycosylases that harbor it, why are EhY and LbY functional? The glycosylase AlkA, a member of the HhH-containing superfamily that repairs alkylation damages, could have lost the [4Fe-4S] cluster during its evolutionary history, with a motif rearrangement structurally compensating for its loss [51]. This glycosidase scans for DNA lesions by a combined mechanism of facilitated diffusion and hopping in which positively-charged amino acids are relevant to locate the lesions [35,36,100]. Other examples of proteins involved in nucleic acid metabolism that have lost the [4Fe-4S] cluster are (6-4) photolyase [54], tryptophanyl-tRNA synthetase [56] and AddAB helicase-nucleases [55], indicating that this domain can be structurally replaced. Another example of the dispensability of specific structural domains in DNA glycosy-

lases is the absence of the β -hairpin zinc finger motif important for glycosylase activity in human endonuclease VII [101]. Thus, motif rearrangement could be a structural solution in these novel MutYs. A possible clue is the variability in primary sequence found in the region where the [4Fe-4S] cluster should be (Figs 1 and 3). When novel MutYs are compared with canonical ones, the primary sequence of the catalytic domain is highly conserved and the variability in amino acid sequence increases at the region that corresponds to the [4Fe-4S] cluster and the C-terminal domain of canonical MutYs (Fig. 3). Because the van der Waals volume of a [4Fe-4S] cluster is $\sim 550 \text{ \AA}^3$ [102], comparing the amino acid sequence allowed us to hypothesize that the [4Fe-4S] cluster might be needed to preserve protein folding. To further investigate this hypothesis, we paid attention to the amino acid sequence of LbY and EhY in the region that corresponds to the [4Fe-4S] cluster. We dubbed this region a pseudo [4Fe-4S] cluster motif. If the [4Fe-4S] cluster is removed, the cavity could be filled by bulky amino acids. Four bulky amino acids, such as Arg, Tyr, Thr, Met or His, whose van der Waals volumes are $\sim 140 \text{ \AA}^3$ would be sufficient to fill the cavity created by the elimination of the [4Fe-4S] cluster. We found that volumetric residues are placed in the inner of the cavity left by the [4Fe-4S] cluster in both glycosylases. The residues M197, Y202, N213 and E221, as well as residues K239, K24, Q214 and F245, possibly stabilize the pseudo [4Fe-4S] cluster in LbY and EhY, respectively (Fig. 8). Most of the novel MutYs in Lactobacillales and the anaerobic clades conserve the large side chain feature of the pseudo [4Fe-4S] cluster. These observations suggest that the replacement of the [4Fe-4S] cluster can be carried out by amino acid substitutions with bulky side chains driving motif stabilization by steric effects. If this observation is correct, it should be possible to substitute the [4Fe-4S] cluster in canonical MutYs by constructing chimeric genes. Thus, we designed three sets of chimeric genes of EcY with complementary parts of LbY and EhY at its C-terminal domain. In the first two sets, the [4Fe-4S] cluster motif of EcY was excluded and complemented by the corresponding region of LbY or EhY and, in the third set, the EcY catalytic domain (including the [4Fe-4S] cluster motif) was kept intact and the C-terminal domain of LbY and EhY was added. Figure 9A shows a schematic representation of the chimeric glycosylase design. The genes corresponding to the chimeric glycosylases were cloned into a pCold1 expression vector, overexpressed in JM101 *muty⁻* *E. coli* strain, and purified as wild-type MutYs. Remarkably, all chimeric MutY DNA glycosylases were active on an 8oG:A sub-

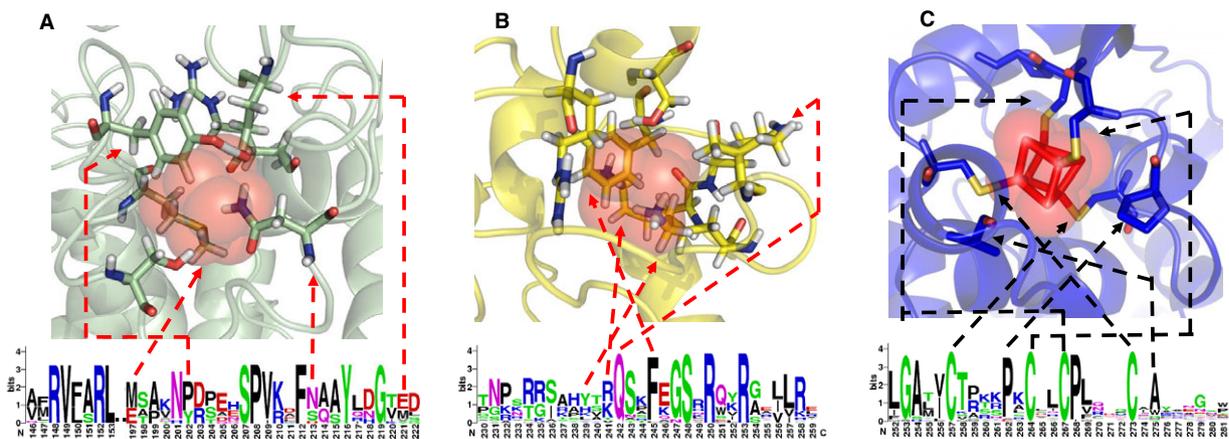


Fig. 8. Structural comparison and logo sequence of corresponding region of the [4Fe-4S] motif among Lactobacillales, Anaerobic clade and canonical MutYs. Corresponding region of the [4Fe-4S] motif of (A) LbY, (B) EhY and (C) EcY. The sequence logo of this region of Lactobacillales, Anaerobic clade and canonical MutYs is shown below. The putative residues involved in steric stabilization of the pseudo [4Fe-4S] cluster motif in LbY (M197, Y202, N213 and E221) and EhY (K239, K24, Q214 and F245) are shown as sticks and red arrows. The [4Fe-4S] cluster van der Waals volume (550 Å³) derived from structural alignment is shown as red spheres and sticks. The highly conserved residues of the [4Fe-4S] cluster motif of EcY are indicated by black arrows.

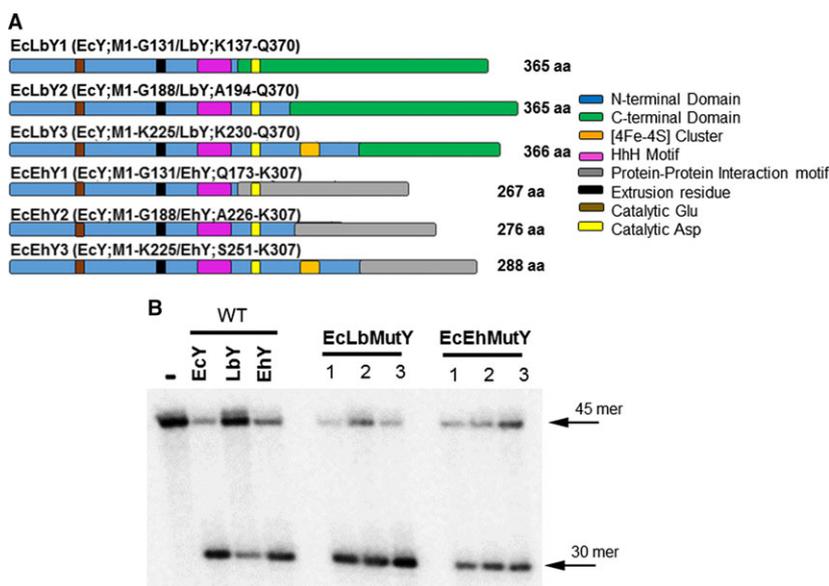


Fig. 9. Schematic design of chimeric glycosylase and glycosylase activity over 80G:A-containing duplex. (A) Schematic design of chimeric glycosylase. The two-first sets of chimeric glycosylases lack the [4Fe-4S] cluster. The third set (EcLbY3 and EcEhY3) includes the *Escherichia coli* catalytic domain (M1-K225) and is complemented with 140 and 56 residues of the C-terminal domain from LbY or EhY. The first set of chimeras (EcLbY1 and EcEhY1) lacks 94 residues from the carboxyl end of *E. coli* catalytic domain, including D138 and the corresponding region of the [4Fe-4S] cluster motif, and is complemented by 233 and 134 residues of the C-terminal domain from LbY and EhY. The second set of chimeras (EcLbY2 and EcLbY) lacks the last 37 residues of the *E. coli* catalytic domain, including the [4Fe-4S] cluster domain, and is complemented by 176 and 81 residues of the C-terminal domain from LbY and EhY. (B) Denaturing gel electrophoresis analysis illustrating the *in vitro* activity of chimeric glycosylases with 1 nM of 80G-containing substrate at 37 °C. The substrate (45 mer) and product (30 mer) are indicated by black arrows.

strate (Fig. 9B). When we tested the *in vivo* activity of the chimeric DNA glycosylases using a CC104 *muty*⁻ *mutm*⁻ *E. coli* strain, the decrease in mutation fre-

quency ranged from 42.8% to 61.8%, whereas the intact EcY reduced the *f* value to 98.9% (Table 3). The EcLbY3 chimera was the only construct that pre-

Table 3. Rifampicin resistance assay data of chimeric glycosylases in CC104 *muty⁻ mutm⁻ Escherichia coli*. The *rpoB* mutation frequency (*f*) per cell was calculated by dividing the median number of mutants by the average number of cells in a series of cultures. *f* is obtained from at least eight separate experiments and errors correspond to one SD from the average.

Plasmid	<i>f</i> ($\times 10^{-6}$)	Increase (fold) over pCold-EcY	Relative decrease (%) to control (pCold)
pCold	10.386 \pm 4.36	98.4	0
pColdEcY	0.106 \pm 0.08	1	98.9
pColdEcLbY1	4.737 \pm 1.69	44.8	54.3
pColdEcLbY2	5.935 \pm 2.73	56.2	42.8
pColdEcLbY3	0.188 \pm 0.05	1.7	98.1
pColdEcEhY1	4.480 \pm 1.23	42.4	56.8
pColdEcEhY2	5.219 \pm 1.64	49.4	49.7
pColdEcEhY3	3.962 \pm 1.44	37.5	61.8

sented wild-type glycosylase activity *in vitro* and *in vivo*; nonetheless, it contains a [4Fe-4S] cluster. Overall, these results indicate that the absence of the [4Fe-4S] cluster can be mitigated by the pseudo [4Fe-4S] cluster motif from LbY or EhY. We consider it important to emphasize that these results do not discard the importance of the [4Fe-4S] cluster in lesion recognition in canonical MutYs.

In the present study, we have demonstrated the *in vitro* and *in vivo* functionality of adenine glycosylases MutYs without requiring a [4Fe-4S] cluster. An exhaustive search of homologues enabled us to identify a number of MutYs that apparently have to deal with 8oG without this important motif. A phylogenetic analysis indicates that the loss of the [4Fe-4S] cluster has occurred at least twice during evolution in a mixed clade composed of anaerobic and microaerophilic microorganisms and in *Lactobacillales*. We also provide evidence showing that the C-terminal domains in these novel homologues play an important role in dispensing with the [4Fe-4S] cluster, probably by motif rearrangement driven by steric effects. However, further studies focused on the structure and the mechanisms for lesion scanning will help clarify the mechanism that allows these novel MutYs to be functional without a [4Fe-4S] cluster.

Materials and methods

Gene identification and cloning of MutYs

The gene for EhY was identified in the *Entamoeba histolytica* genome using hidden Markov models [103]. The result with the highest score corresponded to the locus EHI_010700 and was identified as a putative adenine glyco-

sylase. Because the EhY genomic sequence bears an intron, the ORF was assembled by overlap extension PCR with *E. histolytica* genomic DNA and confirmed by RT-PCR [104]. (All oligonucleotide information is given in Fig. S4. For LbY, the protein sequence with NCBI accession number WP_021741015 was identified as an adenine glycosylase and the synthetic optimized ORF was purchased from GeneScript (Piscataway, NJ, USA). The EhY and LbY ORFs were cloned into the *NdeI* and *BamHI* cloning sites of the pColdI vector. pKKYEco containing wild-type and untagged EcY was kindly provided by Sheila David (Department of Chemistry, University of California Davis) [105]. This gene was amplified with EcMutYF and EcMutYR oligonucleotides and subcloned into a pColdI vector. All recombinant MutYs expressed and purified from the pColdI vector contain a 6-His track at their N-terminal.

Structural-homology modeling of LbY and EhY

Homology models for LbY and EhY were constructed using the MOLECULAR OPERATING ENVIRONMENT (MOE) software [106]. The template structure was the MutY adenine glycosylase from *G. stearothermophilus* in complex with DNA containing a modified adenine paired with 8oG [Protein Data Bank (PDB) code: [1RRQ](#)] [59]. The amino acid sequences of LbY and EhY (NCBI Reference No. WP_003554055.1 and XP_653060.1, respectively) were aligned with the amino acid sequence present in the crystal structure of MutY DNA glycosylase from *G. stearothermophilus*. A series of 30 models were constructed and the models with less structural differences in comparison to the template were selected.

Site-directed mutagenesis of active site

Single and double mutants were constructed at position E42A, D143A for LbY, and E78A, D179A for EhY, in accordance with a Quick Change mutagenesis kit (Stratagene, La Jolla, CA, USA) using as template LbY and EhY genes cloned into pColdI with the oligonucleotides listed in Fig. S4.

Overexpression and purification

JM101 *muty::mini-tn10 E. coli* strain [107] used for protein expression was generously shared by Sheila David. JM101 *muty::mini-tn10* strain was transformed with pColdI plasmid containing EcY, LbY and EhY and chimeric glycosylase genes using standard heat-shock procedures. Two liters of LB medium containing 100 $\mu\text{g}\cdot\text{mL}^{-1}$ ampicillin and 15 $\mu\text{g}\cdot\text{mL}^{-1}$ tetracycline were grown until D_{600} of 0.6 at 37 °C was reached and protein expression was induced by the addition of a final concentration of 0.5 mM IPTG, and then cultured overnight at 16 °C and harvested by centrifugation. Harvested cells were resuspended in 60 mL of lysis

buffer (15 mM Tris-HCl, pH 8, 400 mM NaCl, 0.1% triton) supplemented with 1 mM PMSF and stored at -80°C . Cells were thawed at room temperature and treated with lysozyme ($1\text{ mg}\cdot\text{mL}^{-1}$) on ice for 45 min. After treatment with lysozyme, a second cell lysis step was performed using sonication with 10 series of 20 pulses $\cdot\text{s}^{-1}$ using a Branson sonifier 450 (Branson Instruments, Co., Stamford, CT, USA) at 50% of power.

After centrifugation to remove cellular debris, the supernatant was loaded onto a 1-mL HisTrap FF column (GE Healthcare Life Science, Little Chalfont, UK) with nickel and previously equilibrated with buffer lysis, and washed with 50 mL of 15, 30 and 45 mM imidazole (in buffer lysis), and the elution fraction was 10 mL of 500 mM imidazole. The elution fraction was dialyzed against phosphate buffer (20 mM phosphates, pH 7.4, 75 mM NaCl, 1 mM DTT, 1 mM EDTA). The dialyzed protein solutions were passed through a second step of purification using cationic exchange chromatography with 1-mL HiTrapSP FF column (GE Healthcare Life Science). Once the protein solution was loaded onto the cationic exchange column and equilibrated with phosphate buffer, the column was washed with 5 mL of 75, 100, 200 and 300 mM of NaCl and eluted with 5 mL of 500 mM and 1000 mM NaCl (in phosphate buffer). Elution fractions with purified proteins were placed together and dialyzed against phosphate buffer. Protein purity was confirmed by SDS/PAGE and, after the protein concentration had been determined, aliquots were stored at -20°C with 50% of glycerol.

[4Fe-4S] cluster detection

The [4Fe-4S] cluster was detected using a semi-quantitative chemiluminescent approach with luminol reaction [108]. The procedure was performed in accordance with the protocol described by Högbom *et al.* [76] with 30 and 150 μmol of enzyme. Data collection was carried out with a microplate reader Infinite 1000 (Tecan, Maennedorf, Switzerland) in a black 96-well plate (Nunc, Roskilde, Denmark) and an integration time of 0.25 s. Data are reported as relative chemiluminescence emission and are expressed as the signal-to-noise ratio.

Substrate preparation

45-mer template oligonucleotides ($5'$ -CGC AGC GGA CCT GCC CAC CTA ACT GAT ATX GCG CTA GTG CCA AGG- $3'$; where X is A, G, T or C) were ^{32}P - $5'$ -end-labeled with T4 polymerase kinase (Q) and purified with nucleotide removal kit (Qiagen, Valencia, CA, USA) in accordance with the manufacturer's instructions. $5'$ -end-labeled template oligonucleotides were hybridized at a final concentration of 100 nM with 20% excess of complementary oligonucleotide ($5'$ -CCT TGG CAC TAG CGC YAT

ATC AGT TAG GTG GGC AGG TCC GCT GCG- $3'$; where Y is 8oG, G, A, C or T) by heating at 95°C for 5 min in annealing buffer (20 mM Tris, pH 7.4, 150 mM NaCl) and then slowly cooled to room temperature overnight.

Glycosylase assay

The glycosylase assay was carried out using different concentrations of substrate ranging from 10 to 1 nM in reaction buffer (40 mM Tris, pH 7.5, 10 mM EDTA, $0.1\text{ mg}\cdot\text{mL}^{-1}$ and 150 mM NaCl) at 37 or 21°C . The protein concentration (active protein concentration; see below) used for these assays is specified in the Results and Discussion, as appropriate. The reactions were stopped with 0.2 M NaCl and heated at 90°C for 5 min. An equal volume of loading buffer (99% formamide, 20 mM EDTA and 0.025% bromophenol blue) was added to the reactions. The samples were resolved on a 12% (19 : 1) denaturing polyacrylamide gel and visualized by storage phosphor autoradiography.

[4Fe-4S] cluster reconstitution assay

The procedure of [4Fe-4S] cluster reconstitution was carried out in accordance with the classical protocol described by David and coworkers with certain modifications. First, an enzyme solution containing $0.5\text{ mg}\cdot\text{mL}^{-1}$ MutY in buffer A (50 mM Tris-HCl, pH 7.6, 200 mM NaCl, 0.1 mM 2-mercaptoethanol and 0.1 mM EDTA) was denatured in the presence of 6 M Gu-HCl for at least 1 h to release the [4Fe-4S] cluster. Subsequently, the iron was chelated incubating the enzyme solution in buffer A containing 6 M Gu-HCl and 5 mM EDTA for 6 h. To refold the enzyme and reconstitute the [4Fe-4S] cluster, an enzyme sample of each MutY was renatured in buffer A in the presence of 0.2 mM iron(II) ammonium sulfate and 0.2 mM ammonium sulfide for 15 h. These refolded samples with [4Fe-4S] cluster reconstitution, dubbed $\text{R}^{\text{Fe-S}}$, were dialyzed extensively against buffer A with three rounds of dialysis of 4 h each to eliminate excess of ions in solution. Another enzyme samples, dubbed R, were refolded in the absence of Fe and S sources to avoid [4Fe-4S] assemble and dialyzed extensively against buffer A. In parallel, a positive controls for each enzyme, named Nd, was prepared without denaturation procedure, dialyzing against buffer A for the same period as the other samples. All the buffers used for this experiment was treated with chelex-100 (Sigma, St Louis, MO, USA) to minimize metal contamination. At the end, *in vitro* activity was tested with 1 nM radiolabeled 45-mer duplex containing 8oG:A lesion and 500 nM MutY for 30 min at 37°C . The qualitative chemiluminescence detection was performed with 1.5 μM of MutY in plate wells with luminol solution and exposing to a CL-Xposure film (Thermo Scientific, Waltham, MA, USA). A scheme of the

[4Fe-4S] cluster reconstitution experiment is shown in Fig. 6A.

Fluorescence anisotropy

The binding affinity constants were measured by fluorescence anisotropy using 1 nM 45-mer duplex containing 8oG lesion paired with abasic site (tetrahydrofuran nucleotide, THF; Sigma-Aldrich) with an active protein concentrations range (0.1–75 nM). The binding experiments were conducted as described previously [109]. The change in anisotropy was measured using a fluorescence intensity detector (Tecan Infinite M1000) with excitation and emission wavelengths of 495 and 520 nm, respectively. In the case of EcY, the fluorescence intensity was decreased as the protein concentration increased; therefore, anisotropy for each binding reaction was adjusted to compensate for the decrease of quantum yield of protein–substrate complex relative to substrate, as proposed previously [87,88]. The anisotropy curves were fitted to a single-specific binding with Hill slope model according to the formula:

$$Y = \frac{B_{\max}(X^h)}{[P]_{1/2}^h + X^h}$$

where X is the enzyme concentration, B_{\max} is maximum specific binding, $[P]_{1/2}$ is the enzyme concentration needed to achieve a half-maximum binding at equilibrium, Y is quantum yield corrected anisotropy and h is the Hill coefficient. Constants were calculated from the binding isotherms using PRISM, version 6.07 (GraphPad, La Jolla, CA, USA). All experiments were performed in triplicate.

Kinetic constant determination by a minimal kinetic approach

For determinations of MutY kinetic constants, active protein concentrations were used. For determination of the active protein concentration of EcY, LbY and EhY, an active site titration assay was performed in triplicate, as reported previously [86], using absolute protein concentrations of 2.5 and 10 nM A:8oG at 21 °C.

For multiple turnover kinetics, 2.5 nM active protein and 10 nM substrate containing lesion of 8oG:A was used. Samples were collected at 0.25, 0.5, 1, 2, 8, 15, 30, 60 and 120 min and quenched as described above. For single turnover kinetics, 10 nM substrate and 50 nM active protein were used. The product concentration was quantified using QUANTITY ONE (Bio-Rad, Hercules, CA, USA). The chemical reaction (k_2) and turnover rate (k_3) were determined according to a minimal kinetics approach, fitting data to nonlinear regression as proposed by David and coworkers using PRISM. Each kinetics experiment was carried out in triplicate.

Glycosylase/apurinic/aprimidinic (AP) lyase activity

To test whether LbY and EhY have AP lyase activity, the reactions containing 1 nM 8oG:A duplex were stopped in two different conditions. To investigate glycosylase activity, the reactions were heat stopped as mentioned above (90 °C for 5 min) with 0.2 M NaOH and then loading buffer was added and heated again for 1.5 min. For AP lyase activity, the reactions were stopped adding only loading buffer and heating for 1.5 min to avoid breaking strands by heat. This assay was carried out using 1 nM A:8oG and an excess of enzyme (> 200 nM) at 37 °C. The reactions were monitored at 5, 10, 20, 40 and 80 min.

Measurement of mutation frequency

To evaluate MutY activity *in vivo*, we employed a rifampicin resistance approach [95–98] using CC104 *muty::mini-tn10 mutm::mini-kan E. coli* strain (a donation from Sheila David). The CC104 *muty::mini-tn10 mutm::mini-kan* strain was transformed with wild-type MutY genes and chimeric genes. Additionally, empty pCold1 vector and *E. coli* MutY gene cloned into pKK223-3 were used in these assays as negative and positive controls. A series of overnight cultures (eight for each gene tested) were grown in LB medium containing 100 µg·mL⁻¹ ampicillin and 15 µg·mL⁻¹ tetracycline antibiotics. To determine the number of rifampicin resistant cells, 100 µL of each individual culture was plated on LB agar media supplemented with 100 µg·mL⁻¹ rifampicin, 100 µg·mL⁻¹ ampicillin and 15 µg·mL⁻¹ tetracycline; to determine the number of viable cells, 100 µL of a 10⁷ dilution of each culture was plated onto LB plates (with ampicillin and tetracycline). Each culture was plated in duplicate and incubated at 37 °C overnight. The mutation frequency (f) was calculated with the median number of colonies from the rifampicin-containing plates divided by the average of viable cells.

Chimera design and gene assembly

Chimera design was performed according to the scheme shown in Fig. 9A. Three sets of chimeras were constructed fusing the partial or whole catalytic domain of EcY with the C-terminal domain of LbY or EhY. The [4Fe-4S] cluster coordination region of EcY was excluded in the two first sets of chimeras, 1 (EcLbY1 and EcEhY1) and 2 (EcLbY2 and EcEhY2), from G133 and G188, and complemented with different fragments of the C-terminal domain of LbY and EhY. The third set of chimeras consisted of the whole catalytic domain of EcY, including the [4Fe-4S] cluster and the C-terminal domain of LbY and EhY.

The chimeric genes were assembled by splicing via overlap extension PCR. All individual fragments, except EcEhY1 and EcEhY3, were amplified from wild-type genes

and cloned into a pCold1 vector. The EcEhY2 and EcEhY3 were acquired as gBlocks gene fragments (Integrated DNA Technologies, Coralville, IA, USA), where an overlapping region of 50 bp was considered (Fig. S1). All chimeras were confirmed by sequencing.

Phylogenetic analysis

For the phylogenetic analysis, 186 adenine glycosylases were surveyed, including MutY from Archaea, Bacteria and Eukarya domains. A multiple sequence alignment of complete deduced amino acid sequences of MutY was performed using the MUSCLE algorithm [110]. The phylogenetic analysis was performed with the maximum likelihood method with 1000 bootstrap replicates, starting with a tree built using the neighbor-joining method. All of the bioinformatics analysis was conducted using MEGA, version 5 [111].

Acknowledgements

We thank Susana Campos, Corina Díaz and Yardenia Garcia for their help with subcloning and mutagenesis. We thank Jenny Gómez for the initial EhMutY characterization. We thank Dr Elisa Azuara for cDNA from *E. histolytica* and Dr Sheila David for the plasmids and strains. We thank Dr Luis Delaye for his advice concerning the evolutionary analysis and Cei Abreu Goodger for critically reading the manuscript. This work was supported by CONACYT – *Problemas Nacionales* grant # 213915 to LGB.

Author contributions

CHTA and LGB conceived the project, analyzed data and wrote the manuscript. CHTA, LMLC and ESS performed the experiments.

References

- Friedburg E, Walker G, Siede W, Schultz R & Ellenberger T (2005) *DNA Repair and Mutagenesis*, 2 edn. ASM Press, Washington, DC.
- Burrows CJ & Muller JG (1998) Oxidative nucleobase modifications leading to strand scission. *Chem Rev* **98**, 1109–1152.
- David SS, O'Shea VL & Kundu S (2007) Base-excision repair of oxidative DNA damage. *Nature* **447**, 941–950.
- Briebe LG, Eichman BF, Kokoska RJ, Doublet S, Kunkel TA & Ellenberger T (2004) Structural basis for the dual coding potential of 8-oxoguanosine by a high-fidelity DNA polymerase. *EMBO J* **23**, 3452–3461.
- Hsu GW, Ober M, Carell T & Beese LS (2004) Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase. *Nature* **431**, 217–221.
- Pastor-Palacios G, Azuara-Liceaga E & Briebe LG (2010) A nuclear family A DNA polymerase from *Entamoeba histolytica* bypasses thymine glycol. *PLoS Negl Trop Dis* **4**, e786.
- Michaels ML, Tchou J, Grollman AP & Miller JH (1992) A repair system for 8-oxo-7, 8-dihydrodeoxyguanine. *Biochemistry* **31**, 10964–10968.
- Manuel RC, Czerwinski EW & Lloyd RS (1996) Identification of the structural and functional domains of MutY, an *Escherichia coli* DNA mismatch repair enzyme. *J Biol Chem* **271**, 16218–16226.
- Manuel RC & Lloyd RS (1997) Cloning, overexpression, and biochemical characterization of the catalytic domain of MutY. *Biochemistry* **36**, 11140–11152.
- Gogos A, Cillo J, Clarke ND & Lu A-L (1996) Specific recognition of A/G and A/7, 8-dihydro-8-oxoguanine (8-oxoG) mismatches by *Escherichia coli* MutY: removal of the C-terminal domain preferentially affects A/8-oxoG recognition. *Biochemistry* **35**, 16665–16671.
- White MF & Dillingham MS (2012) Iron–sulphur clusters in nucleic acid processing enzymes. *Curr Opin Struct Biol* **22**, 94–100.
- Lukianova OA & David SS (2005) A role for iron–sulfur clusters in DNA repair. *Curr Opin Chem Biol* **9**, 145–151.
- Kuo C-F, McRee DE, Fisher CL, O'Handley SF, Cunningham RP & Tainer JA (1992) Atomic structure of the DNA repair [4Fe-4S] enzyme endonuclease III. *Science* **258**, 434–440.
- Cunningham RP, Asahara H, Bank JF, Scholes CP, Salerno JC, Surerus K, Munck E, McCracken J, Peisach J & Emptage MH (1989) Endonuclease III is an iron-sulfur protein. *Biochemistry* **28**, 4450–4455.
- Hinks JA, Evans MC, de Miguel Y, Sartori AA, Jiricny J & Pearl LH (2002) An iron-sulfur cluster in the family 4 uracil-DNA glycosylases. *J Biol Chem* **277**, 16936–16940.
- Yang H, Fitz-Gibbon S, Marcotte EM, Tai JH, Hyman EC & Miller JH (2000) Characterization of a thermostable DNA glycosylase specific for U/G and T/G mismatches from the hyperthermophilic archaeon *Pyrobaculum aerophilum*. *J Bacteriol* **182**, 1272–1279.
- Mok YG, Uzawa R, Lee J, Weiner GM, Eichman BF, Fischer RL & Huh JH (2010) Domain structure of the DEMETER 5-methylcytosine DNA glycosylase. *Proc Natl Acad Sci* **107**, 19225–19230.
- Porello SL, Cannon MJ & David SS (1998) A substrate recognition role for the [4Fe-4S]²⁺ cluster of the DNA repair glycosylase MutY. *Biochemistry* **37**, 6465–6475.
- Guan Y, Manuel RC, Arvai AS, Parikh SS, Mol CD, Miller JH, Lloyd RS & Tainer JA (1998) MutY

- catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily. *Nat Struct Mol Biol* **5**, 1058–1064.
- 20 Golinelli M-P, Chmiel NH & David SS (1999) Site-directed mutagenesis of the cysteine ligands to the [4Fe-4S] cluster of *Escherichia coli* MutY. *Biochemistry* **38**, 6997–7007.
- 21 Messick TE, Chmiel NH, Golinelli M-P, Langer MR, Joshua-Tor L & David SS (2002) Noncysteinyll coordination to the [4Fe-4S] 2+ cluster of the DNA repair adenine glycosylase MutY introduced via site-directed mutagenesis. Structural characterization of an unusual histidinyl-coordinated cluster. *Biochemistry* **41**, 3931–3942.
- 22 Chepanoske CL, Golinelli M-P, Williams SD & David SS (2000) Positively charged residues within the Iron-Sulfur cluster loop of *E. coli* MutY participate in damage recognition and removal. *Arch Biochem Biophys* **380**, 11–19.
- 23 Boon EM, Livingston AL, Chmiel NH, David SS & Barton JK (2003) DNA-mediated charge transport for DNA repair. *Proc Natl Acad Sci* **100**, 12543–12547.
- 24 Kelley SO, Boon EM, Barton JK, Jackson NM & Hill MG (1999) Single-base mismatch detection based on charge transduction through DNA. *Nucleic Acids Res* **27**, 4830–4837.
- 25 Sontz PA, Mui TP, Fuss JO, Tainer JA & Barton JK (2012) DNA charge transport as a first step in coordinating the detection of lesions by repair proteins. *Proc Natl Acad Sci* **109**, 1856–1861.
- 26 Yavin E, Stemp ED, O'Shea VL, David SS & Barton JK (2006) Electron trap for DNA-bound repair enzymes: a strategy for DNA-mediated signaling. *Proc Natl Acad Sci USA* **103**, 3610–3614.
- 27 Boal AK, Yavin E, Lukianova OA, O'Shea VL, David SS & Barton JK (2005) DNA-bound redox activity of DNA repair glycosylases containing [4Fe-4S] clusters. *Biochemistry* **44**, 8397–8407.
- 28 Gorodetsky AA, Boal AK & Barton JK (2006) Direct electrochemistry of endonuclease III in the presence and absence of DNA. *J Am Chem Soc* **128**, 12082–12083.
- 29 Fromme JC & Verdine GL (2002) Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. *Nat Struct Mol Biol* **9**, 544–552.
- 30 Jiang D, Hatahet Z, Blaisdell JO, Melamede RJ & Wallace SS (1997) *Escherichia coli* endonuclease VIII: cloning, sequencing, and overexpression of the nei structural gene and characterization of nei and nei nth mutants. *J Bacteriol* **179**, 3773–3782.
- 31 Barrett TE, Savva R, Panayotou G, Barlow T, Brown T, Jiricny J & Pearl LH (1998) Crystal structure of a G: T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. *Cell* **92**, 117–129.
- 32 Drohat AC, Kwon K, Krosky DJ & Stivers JT (2002) 3-Methyladenine DNA glycosylase I is an unexpected helix-hairpin-helix superfamily member. *Nat Struct Mol Biol* **9**, 659–664.
- 33 Yamagata Y, Kato M, Odawara K, Tokuno Y, Nakashima Y, Matsushima N, Yasumura K, Tomita K-I, Ihara K & Fujii Y (1996) Three-dimensional structure of a DNA repair enzyme, 3-methyladenine DNA glycosylase II, from *Escherichia coli*. *Cell* **86**, 311–319.
- 34 Lau AY, Schärer OD, Samson L, Verdine GL & Ellenberger T (1998) Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell* **95**, 249–258.
- 35 Zhang Y & O'Brien PJ (2015) Repair of alkylation damage in eukaryotic chromatin depends on searching ability of alkyladenine DNA glycosylase. *ACS Chem Biol* **10**, 2606–2615.
- 36 Hedglin M & O'Brien PJ (2010) Hopping enables a DNA repair glycosylase to search both strands and bypass a bound protein. *ACS Chem Biol* **5**, 427–436.
- 37 Schonhoft JD, Kosowicz JG & Stivers JT (2013) DNA translocation by human uracil DNA glycosylase: role of DNA phosphate charge. *Biochemistry* **52**, 2526–2535.
- 38 Sang PB, Srinath T, Patil AG, Woo EJ & Varshney U (2015) A unique uracil-DNA binding protein of the uracil DNA glycosylase superfamily. *Nucleic Acids Res* **43**, 8452–8463.
- 39 You HJ, Swanson RL & Doetsch PW (1998) *Saccharomyces cerevisiae* possesses two functional homologues of *Escherichia coli* endonuclease III. *Biochemistry* **37**, 6033–6040.
- 40 Alseth I, Eide L, Pirovano M, Rognes T, Seeberg E & Bjørås M (1999) The *Saccharomyces cerevisiae* homologues of endonuclease III from *Escherichia coli*, Ntg1 and Ntg2, are both required for efficient repair of spontaneous and induced oxidative DNA damage in yeast. *Mol Cell Biol* **19**, 3779–3787.
- 41 Sentürker S, Dizdaroglu M, van der Kemp PA, You HJ, Doetsch PW & Boiteux S (1998) Substrate specificities of the Ntg1 and Ntg2 proteins of *Saccharomyces cerevisiae* for oxidized DNA bases are not identical. *Nucleic Acids Res* **26**, 5270–5276.
- 42 Macedo-Ribeiro S, Darimont B, Sterner R & Huber R (1996) Small structural changes account for the high thermostability of 1 [4Fe-4S] ferredoxin from the hyperthermophilic bacterium *Thermotoga maritima*. *Structure* **4**, 1291–1301.
- 43 Kissinger CR, Sieker LC, Adman ET & Jensen LH (1991) Refined crystal structure of ferredoxin II from *Desulfovibrio gigas* at 1.7 Å. *J Mol Biol* **219**, 693–715.
- 44 Fukuyama K, Okada T, Kakuta Y & Takahashi Y (2002) Atomic resolution structures of oxidized [4Fe-

- 4S] ferredoxin from *Bacillus thermoproteolyticus* in two crystal forms: systematic distortion of [4Fe-4S] cluster in the protein. *J Mol Biol* **315**, 1155–1166.
- 45 Samrakandi MM & Pasta F (2000) Hyperrecombination in *Streptococcus pneumoniae* depends on an atypical mutY homolog. *J Bacteriol* **182**, 3353–3360.
- 46 Baute J & Depicker A (2008) Base excision repair and its role in maintaining genome stability. *Crit Rev Biochem Mol Biol* **43**, 239–276.
- 47 Jansson K, Blomberg A, Sunnerhagen P & Alm Rosenblad M (2010) Evolutionary loss of 8-oxo-G repair components among eukaryotes. *Genome Integr* **1**, 12.
- 48 Eisen JA & Hanawalt PC (1999) A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat Res* **435**, 171–213.
- 49 Denver DR, Swenson SL & Lynch M (2003) An evolutionary analysis of the helix-hairpin-helix superfamily of DNA repair glycosylases. *Mol Biol Evol* **20**, 1603–1611.
- 50 Brooks SC, Adhikary S, Rubinson EH & Eichman BF (2013) Recent advances in the structural mechanisms of DNA glycosylases. *Biochim Biophys Acta* **1834**, 247–271.
- 51 Labahn J, Schärer OD, Long A, Ezaz-Nikpay K, Verdine GL & Ellenberger TE (1996) Structural basis for the excision repair of alkylation-damaged DNA. *Cell* **86**, 321–329.
- 52 Hollis T, Ichikawa Y & Ellenberger T (2000) DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase *Escherichia coli* AlkA. *EMBO J* **19**, 758–766.
- 53 Meyer J (2008) Iron-sulfur protein folds, iron-sulfur chemistry, and evolution. *J Biol Inorg Chem* **13**, 157–170.
- 54 Zhang F, Scheerer P, Oberpichler I, Lamparter T & Krauß N (2013) Crystal structure of a prokaryotic (6-4) photolyase with an Fe-S cluster and a 6, 7-dimethyl-8-ribityllumazine antenna chromophore. *Proc Natl Acad Sci* **110**, 7217–7222.
- 55 Yeeles JT, Cammack R & Dillingham MS (2009) An iron-sulfur cluster is essential for the binding of broken DNA by AddAB-type helicase-nucleases. *J Biol Chem* **284**, 7746–7755.
- 56 Stiebritz MT (2014) A role for [Fe4S4] clusters in tRNA recognition—a theoretical study. *Nucleic Acids Res* **42**, 5426–5435.
- 57 Loftus B, Anderson I, Davies R, Alsmark UCM, Samuelson J, Amedeo P, Roncaglia P, Berriman M, Hirt RP & Mann BJ (2005) The genome of the protist parasite *Entamoeba histolytica*. *Nature* **433**, 865–868.
- 58 Pastor-Palacios G, López-Ramírez V, Cardona-Felix CS & Briebe LG (2012) A transposon-derived DNA polymerase from *Entamoeba histolytica* displays intrinsic strand displacement, processivity and lesion bypass.
- 59 Fromme JC, Banerjee A, Huang SJ & Verdine GL (2004) Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. *Nature* **427**, 652–656.
- 60 Zharkov DO, Golan G, Gilboa R, Fernandes AS, Gerchman SE, Kycia JH, Rieger RA, Grollman AP & Shoham G (2002) Structural analysis of an *Escherichia coli* endonuclease VIII covalent reaction intermediate. *EMBO J* **21**, 789–800.
- 61 Gilboa R, Zharkov DO, Golan G, Fernandes AS, Gerchman SE, Matz E, Kycia JH, Grollman AP & Shoham G (2002) Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA. *J Biol Chem* **277**, 19811–19816.
- 62 Livingston AL, Kundu S, Henderson Pozzi M, Anderson DW & David SS (2005) Insight into the roles of tyrosine 82 and glycine 253 in the *Escherichia coli* adenine glycosylase MutY. *Biochemistry* **44**, 14179–14190.
- 63 Moulis J-M, Davasse V, Golinelli M-P, Meyer J & Quinkal I (1996) The coordination sphere of iron-sulfur clusters: lessons from site-directed mutagenesis experiments. *J Biol Inorg Chem* **1**, 2–14.
- 64 Sazanov LA & Hinchliffe P (2006) Structure of the hydrophilic domain of respiratory complex I from *Thermus thermophilus*. *Science* **311**, 1430–1436.
- 65 Moulis JM, Sieker LC, Wilson KS & Dauter Z (1996) Crystal structure of the 2 [4Fe-4S] ferredoxin from *Chromatium vinosum*: evolutionary and mechanistic inferences for [3/4Fe-4S] ferredoxins. *Protein Sci* **5**, 1765–1775.
- 66 Bessman MJ, Frick DN & O’Handley SF (1996) The MutT proteins or “Nudix” hydrolases, a family of versatile, widely distributed”, housecleaning” enzymes. *J Biol Chem* **271**, 25059–25062.
- 67 Tajiri T, Maki H & Sekiguchi M (1995) Functional cooperation of MutT, MutM and MutY proteins in preventing mutations caused by spontaneous oxidation of guanine nucleotide in *Escherichia coli*. *Mutat Res* **336**, 257–267.
- 68 Noll DM, Gogos A, Granek JA & Clarke ND (1999) The C-Terminal domain of the Adenine-DNA glycosylase MutY confers specificity for 8-Oxoguanine⊖ adenine mispairs and may have evolved from MutT, an 8-Oxo-dGTPase. *Biochemistry* **38**, 6374–6379.
- 69 Volk DE, House PG, Thivyanathan V, Luxon BA, Zhang S, Lloyd RS & Gorenstein DG (2000) Structural similarities between MutT and the C-terminal domain of MutY. *Biochemistry* **39**, 7331–7336.
- 70 Lee S & Verdine GL (2009) Atomic substitution reveals the structural basis for substrate adenine

- recognition and removal by adenine DNA glycosylase. *Proc Natl Acad Sci* **106**, 18497–18502.
- 71 Qing G, Ma LC, Khorchid A, Swapna GV, Mal TK, Takayama MM, Xia B, Phadtare S, Ke H, Acton T *et al.* (2004) Cold-shock induced high-yield protein production in *Escherichia coli*. *Nat Biotechnol* **22**, 877–882.
- 72 Kennedy MC, Kent T, Emptage M, Merkle H, Beinert H & Münck E (1984) Evidence for the formation of a linear [3Fe-4S] cluster in partially unfolded aconitase. *J Biol Chem* **259**, 14463–14471.
- 73 Golinelli M-P, Akin LA, Crouse BR, Johnson MK & Meyer J (1996) Cysteine ligand swapping on a deletable loop of the [2Fe-2S] ferredoxin from *Clostridium pasteurianum*. *Biochemistry* **35**, 8995–9002.
- 74 Engstrom LM, Partington OA & David SS (2012) An Iron-Sulfur cluster loop motif in the archaeoglobus fulgidus Uracil-DNA glycosylase mediates efficient uracil recognition and removal. *Biochemistry* **51**, 5187–5197.
- 75 Pieroni L, Khalil L, Charlotte F, Poynard T, Piton A, Hainque B & Imbert-Bismut F (2001) Comparison of bathophenanthroline sulfonate and ferene as chromogens in colorimetric measurement of low hepatic iron concentration. *Clin Chem* **47**, 2059–2061.
- 76 Högbom M, Ericsson UB, Lam R, Kuznetsova E, Nordlund P & Zamble DB (2005) A high throughput method for the detection of metalloproteins on a microgram scale. *Mol Cell Proteomics* **4**, 827–834.
- 77 Brunk E, Arey JS & Rothlisberger U (2012) Role of environment for catalysis of the DNA repair enzyme MutY. *J Am Chem Soc* **134**, 8608–8616.
- 78 Lu A-L & Wright PM (2003) Characterization of an *Escherichia coli* mutant MutY with a cysteine to alanine mutation at the iron-sulfur cluster domain. *Biochemistry* **42**, 3742–3750.
- 79 Lu A-L, Tsai-Wu J-J & Cillo J (1995) DNA determinants and substrate specificities of *Escherichia coli* MutY. *J Biol Chem* **270**, 23582–23588.
- 80 Brinkmeyer MK, Pope MA & David SS (2012) Catalytic contributions of key residues in the adenine glycosylase MutY revealed by pH-dependent kinetics and cellular repair assays. *Chem Biol* **19**, 276–286.
- 81 Williams SD & David SS (2000) A single engineered point mutation in the adenine glycosylase MutY confers bifunctional glycosylase/AP lyase activity. *Biochemistry* **39**, 10098–10109.
- 82 Williams SD & David SS (1999) Formation of a Schiff base intermediate is not required for the adenine glycosylase activity of *Escherichia coli* MutY. *Biochemistry* **38**, 15417–15424.
- 83 Jaganaman S, Pinto A, Tarasev M & Ballou DP (2007) High levels of expression of the iron-sulfur proteins phthalate dioxygenase and phthalate dioxygenase reductase in *Escherichia coli*. *Protein Expr Purif* **52**, 273–279.
- 84 Gupta P, Aggarwal N, Batra P, Mishra S & Chaudhuri TK (2006) Co-expression of chaperonin GroEL/GroES enhances in vivo folding of yeast mitochondrial aconitase and alters the growth characteristics of *Escherichia coli*. *Int J Biochem Cell Biol* **38**, 1975–1985.
- 85 Nakamura M, Saeki K & Takahashi Y (1999) Hyperproduction of recombinant ferredoxins in *Escherichia coli* by coexpression of the ORF1-ORF2-iscS-iscU-iscA-hscB-hscA-fdx-ORF3 gene cluster. *J Biochem* **126**, 10–18.
- 86 Porello SL, Leyes AE & David SS (1998) Single-turnover and pre-steady-state kinetics of the reaction of the adenine glycosylase MutY with mismatch-containing DNA substrates. *Biochemistry* **37**, 14756–14764.
- 87 Lee C-Y, Bai H, Houle R, Wilson GM & Lu A-L (2004) An *Escherichia coli* MutY mutant without the six-helix barrel domain is a dimer in solution and assembles cooperatively into multisubunit complexes with DNA. *J Biol Chem* **279**, 52653–52663.
- 88 Dandliker WB, Hsu M-L, Levin J & Rao BR (1981) Equilibrium and kinetic inhibition assays based upon fluorescence polarization. *Methods Enzymol* **74**, 3.
- 89 Luncsford PJ, Chang D-Y, Shi G, Bernstein J, Madabushi A, Patterson DN, Lu A-L & Toth EA (2010) A structural hinge in eukaryotic MutY homologues mediates catalytic activity and Rad9–Rad1–Hus1 checkpoint complex interactions. *J Mol Biol* **403**, 351–370.
- 90 Gripon S, Zhao Q, Robinson T, Marshall JJ, O'Neill RJ, Manning H, Kennedy G, Dunsby C, Neil M & Halford SE (2010) Differential modes of DNA binding by mismatch uracil DNA glycosylase from *Escherichia coli*: implications for abasic lesion processing and enzyme communication in the base excision repair pathway. *Nucleic Acids Res*, **39**, 2593–2603.
- 91 Chmiel NH, Golinelli M-P, Francis AW & David SS (2001) Efficient recognition of substrates and substrate analogs by the adenine glycosylase MutY requires the C-terminal domain. *Nucleic Acids Res* **29**, 553–564.
- 92 Pope MA & David SS (2005) DNA damage recognition and repair by the murine MutY homologue. *DNA Repair* **4**, 91–102.
- 93 Schärer OD, Nash HM, Jiricny J, Laval J & Verdine GL (1998) Specific binding of a designed pyrrolidine abasic site analog to multiple DNA glycosylases. *J Biol Chem* **273**, 8592–8597.
- 94 Kundu S, Brinkmeyer MK, Livingston AL & David SS (2009) Adenine removal activity and bacterial complementation with the human MutY homologue (MUTYH) and Y165C, G382D, P391L and Q324R

- variants associated with colorectal cancer. *DNA Repair* **8**, 1400–1410.
- 95 Wolff E, Kim M, Hu K, Yang H & Miller JH (2004) Polymerases leave fingerprints: analysis of the mutational spectrum in *Escherichia coli* rpoB to assess the role of polymerase IV in spontaneous mutation. *J Bacteriol* **186**, 2900–2905.
- 96 Hartmann G, Honikel KO, Knüsel F & Nüesch J (1967) The specific inhibition of the DNA-directed RNA synthesis by rifamycin. *Biochim Biophys Acta* **145**, 843–844.
- 97 Garibyan L, Huang T, Kim M, Wolff E, Nguyen A, Nguyen T, Diep A, Hu K, Iverson A & Yang H (2003) Use of the rpoB gene to determine the specificity of base substitution mutations on the *Escherichia coli* chromosome. *DNA Repair* **2**, 593–608.
- 98 Severinov K, Soushko M, Goldfarb A & Nikiforov V (1993) Rifampicin region revisited. New rifampicin-resistant and streptolydigin-resistant mutants in the beta subunit of *Escherichia coli* RNA polymerase. *J Biol Chem* **268**, 14820–14825.
- 99 Brosius J & Holy A (1984) Regulation of ribosomal RNA promoters with a synthetic lac operator. *Proc Natl Acad Sci* **81**, 6929–6933.
- 100 Hedglin M, Zhang Y & O'Brien PJ (2015) Probing the DNA structural requirements for facilitated diffusion. *Biochemistry* **54**, 557–566.
- 101 Doublé S, Bandaru V, Bond JP & Wallace SS (2004) The crystal structure of human endonuclease VIII-like 1 (NEIL1) reveals a zincless finger motif required for glycosylase activity. *Proc Natl Acad Sci USA* **101**, 10284–10289.
- 102 Mitra D, George SJ, Guo Y, Kamali S, Keable S, Peters JW, Pelmenschikov V, Case DA & Cramer SP (2013) Characterization of [4Fe-4S] cluster vibrations and structure in nitrogenase Fe protein at three oxidation levels via combined NRVS, EXAFS, and DFT analyses. *J Am Chem Soc* **135**, 2530–2543.
- 103 Brinkac LM, Davidsen T, Beck E, Ganapathy A, Caler E, Dodson RJ, Durkin AS, Harkins DM, Lorenzi H, Madupu R *et al.* (2010) Pathema: a clade-specific bioinformatics resource center for pathogen research. *Nucleic Acids Res* **38**, D408–D414.
- 104 Horton RM, Cai ZL, Ho SN & Pease LR (1990) Gene-splicing by overlap extension - tailor-made genes using the polymerase chain-reaction. *Biotechniques* **8**, 528.
- 105 Michaels ML, Cruz C, Grollman AP & Miller JH (1992) Evidence that MutY and MutM combine to prevent mutations by an oxidatively damaged form of guanine in DNA. *Proc Natl Acad Sci USA* **89**, 7022–7025.
- 106 Molecular Operating Environment (MOE), 2012.10, Chemical Computing Group Inc 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2012.
- 107 Miller JH (1992) A Short Course in Bacterial Genetics. Cold Spring Harbor, New York.
- 108 Specht W (1937) Die Chemilumineszenz des Hämins, ein Hilfsmittel zur Auffindung und Erkennung forensisch wichtiger Blutspuren. *Dtsch Z Gesamte Gerichtl Med* **28**, 225–234.
- 109 Owen B & McMurray C (2009) Rapid method for measuring DNA binding to protein using fluorescence anisotropy.
- 110 Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797.
- 111 Tamura K, Peterson D, Peterson N, Stecher G, Nei M & Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731–2739.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site:

Fig. S1. Multiple sequence alignment of MutYs. The alignment was performed using the MUSCLE algorithm, including MutYs from Archaea (11), Bacteria (127) and Eukarya (48).

Fig. S2. Complete phylogenetic tree built using the maximum likelihood method with 186 amino acid sequences of MutY. Bootstrap resampling was performed with 1000 replicates and a 50% cut-off threshold. Percentages of bootstrap support are indicated at the base of each node.

Fig. S3. Multiple and single turnover experiments. Storage phosphor autodiagrams of single and multiple turnover kinetic of LbY and EhY with 8oG:A are shown. For the assays, 10 nM 5'-end-labeled substrate with ³²P was incubated with 50 nM (single turnover) and 2.5 nM (multiple turnover) active LbY and EhY at 21 °C. An aliquot (8 μL) of the reaction was collected at 0.25, 0.5, 1, 2, 4, 8, 15, 30 60 and 120 min and stopped with 0.2 M NaOH and heating at 90 °C for 5 min.

Fig. S4. List of oligonucleotides and synthetic genes.