



Artículo de investigación

Predicción estructural y funcional de las ADN glicosilasas así como su relación filogenética por métodos bioinformáticos

Structural and functional prediction of DNA glycosylases as well as their phylogenetic relationship by bioinformatic methods



Estrella Alejandra Pinkney Rivas^{ID}, Marco Antonio Popoca Cuaya^{ID}

Facultad de Ciencias Químico Biológicas, Universidad Autónoma de Campeche. Av. Ing. Humberto Lanz Cárdenas s/n Col. Ex-Hacienda Kalá. C.P.24085, Campeche, Campeche, México

Autor de correspondencia: Marco Antonio Popoca Cuaya, Facultad de Ciencias Químico Biológicas, Universidad Autónoma de Campeche. Correo electrónico: mapopoca@uacam.mx. ORCID: 0000-0001-9315-8349

Recibido: 29 de Agosto del 2024 **Aceptado:** 15 de Octubre del 2024 **Publicado:** 28 de Octubre del 2024

Resumen. - Las bases nitrogenadas que conforman a los nucleótidos del ADN pueden ser alteradas por factores externos e internos. El mecanismo de reparación por escisión de bases (BER) se encarga de remover las bases dañadas a través de un conjunto de enzimas. En este trabajo realizamos un análisis *in silico* de las secuencias de los genes y proteínas de las glicosilasas encargadas de eliminar las bases alteradas: MPG, OGG1, NEIL1, MUTYH y NTHL1 que participan en la reparación por el mecanismo de BER de *Homo sapiens*. Utilizamos diferentes softwares bioinformáticos con el objetivo de caracterizar el contenido de guanina y citosina (G≡C) de los genes, las estructuras secundarias y terciaria de las glicosilasas, los motivos en las proteínas, así como la relación filogenética entre las glicosilasas. Las secuencias de los genes y de los aminoácidos se descargaron del GeneBank, se utilizaron los softwares en línea GENSCAN, Gor4, phyre2, InterPro y MEGA. El contenido G≡C obtenido en porcentaje fueron de 63.80%, 63.50%, 61.33%, 60.48% y 59.20% para MPG, NTHL1, NEIL1, MUTYH y OGG1 respectivamente. La estructura secundaria de las proteínas mostró que NTHL1 tiene el porcentaje más alto (43.42%) de alfa hélice, OGG1 (16.23%) en la estructura de cadena extendida y NEIL1 en el plegamiento aleatorio (57.69%). Adicionalmente se realizó la predicción de la estructura terciaria y de los dominios en las proteínas, el dominio HhH está presente en OGG1, MUTYH y NTHL1. El árbol filogenético mostró la relación evolutiva entre los genes estudiados, siendo el gen OGG1 el ancestro común. Los resultados de las predicciones son importantes para comprender la estructura molecular de las glicosilasas, además la información generada puede ser utilizados tanto en estudios experimentales, biotecnológicos y en la función evolutiva durante la reparación del ADN y en el diseño de estrategias terapéuticas en las cuales están involucradas las glicosilasas.

Palabras clave: Estrés oxidativo; Genotoxicidad; Reparación; ADN; *in silico*.

Abstract. - Nitrogenous bases are a component of DNA nucleotides and can be altered by both external and internal factors. The base excision repair (BER) mechanism is responsible for removing damaged bases through the action of various enzymes. In this study, we performed an *in-silico* analysis of the gene and protein sequences of glycosylases responsible for eliminating altered bases: MPG, OGG1, NEIL1, MUTYH, and NTHL1, which participate in the BER mechanism of *Homo sapiens*. We used various bioinformatics tools to characterize the guanine and cytosine (G≡C) content of the genes, the secondary and tertiary structures of the glycosylases, protein motifs, and the phylogenetic relationships between the glycosylases. Gene and amino acid sequences were downloaded from GenBank, and the online software tools GENSCAN, Gor4, Phyre2, InterPro, and MEGA were used. The G≡C content percentages obtained were 63.80%, 63.50%, 61.33%, 60.48%, and 59.20% for MPG, NTHL1, NEIL1, MUTYH, and OGG1, respectively. Secondary structure analysis of the proteins showed that NTHL1 has the highest percentage (43.42%) of alpha helix, OGG1 has the highest percentage (16.23%) of extended chain structure, and NEIL1 has the highest percentage of random coil (57.69%). Additionally, we performed the prediction of tertiary structure and domains in proteins, where the HhH domain was observed in OGG1, MUTYH, and NTHL1. The phylogenetic tree revealed the evolutionary relationships among the studied genes, with the OGG1 gene being the common ancestor. These findings are important for understanding the molecular structure of glycosylases and provide valuable information that can be utilized in both experimental and biotechnological studies, as well as in understanding the evolutionary function of DNA repair and in the design of therapeutic strategies involving glycosylases.

Keywords: Oxidative stress; Genotoxicity; Repair; DNA; *in silico*.



1. Introducción

En las células de los mamíferos, el ADN continuamente está expuesto a diversos agentes genotóxicos tanto exógenos como endógenos. Entre los agentes exógenos se encuentra la

radiación ionizante, luz ultravioleta, especies reactivas de oxígeno (ROS, por sus siglas en inglés), exposición a diversas sustancias químicas [1].

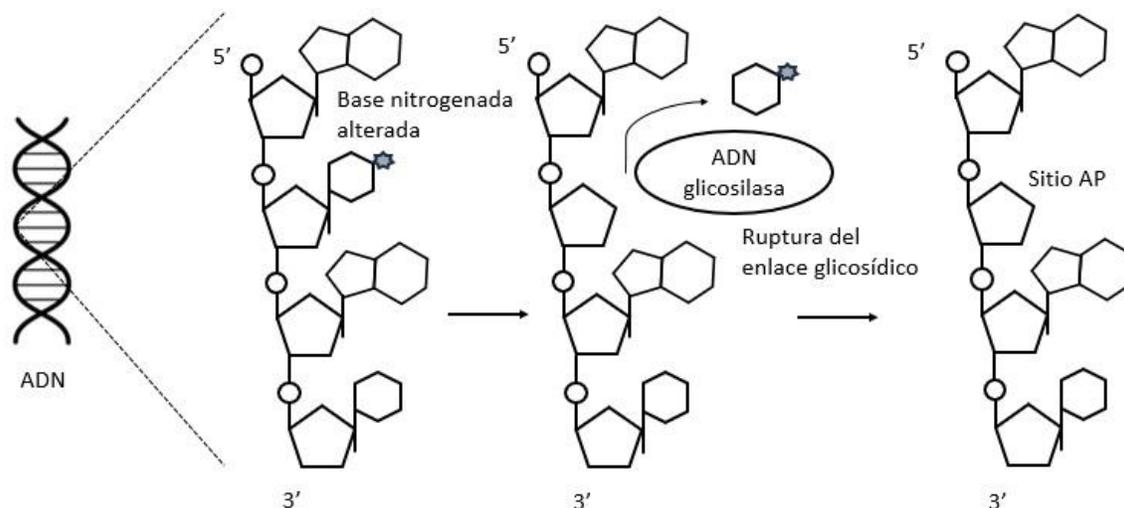


Figura 1. Escisión de bases dañadas por las glicosilasas, al inicio de la reparación del ADN por el mecanismo de BER, generando un sitio AP, el cual es sustrato para las endonucleasas y continuar con la reparación.

Por otra parte, las fuentes endógenas se encuentra el acortamiento de los telómeros, así como las ROS producidas durante el metabolismo celular [2]. Las ROS son moléculas inestables que contienen oxígeno altamente reactivo por lo que su interacción con el ADN puede producir lesiones en las bases nitrogenadas lo que puede ocasionar rompimiento de cadena sencilla y de cadena doble ocasionando mutaciones, bloqueo en los procesos de replicación y transcripción, afectando la expresión genética [3], de este modo la reparación del ADN es fundamental para mantener la estabilidad genómica en las células [4]. Las células responden al daño del ADN a través de una serie de vías de señalización para detectar lesiones y promover la reparación del ADN, los principales mecanismos de reparación son: escisión de bases (BER), escisión de nucleótidos (NER), errores de apareamiento (MMR), recombinación homóloga (HR) y unión de extremos no homólogos (NHEJ) [5], [6].

BER es un mecanismo de reparación del ADN en respuesta al daño ocasionado por procesos como oxidación, hidrólisis, desaminación y alquilación de las bases nitrogenadas. La reparación de bases del ADN es un proceso altamente conservado a través de la evolución e involucra la participación de múltiples enzimas, en humanos participan alrededor de 30 proteínas [7]. El primer paso en la BER es remover bases dañadas o modificadas mediante ADN glicosilasas, para cortar el enlace N-glucosídico, dejando un sitio apurínico o apirimidínico (AP), como se puede ver en la figura 1, el cual es el sustrato para la endonucleasa apurínica/apirimidínica (APE1) y una APLiase encargada de cortar el lado 3' del sitio AP para eliminar el azúcar, en el espacio libre dejado por la base dañada se incorporan nucleótidos por la ADN polimerasa y finalmente sellado por una ADN ligasa [8],[9]. Este mecanismo previene la acumulación de daño oxidativo y otras



modificaciones en las bases nitrogenadas tanto en el genoma nuclear y mitocondrial. Las ADN glicosilasas son las primeras enzimas reclutadas para reparar las lesiones del ADN por medio de BER, la función que llevan a cabo es crucial para identificar y remover las bases alteradas en el ADN [10], además de que tienen una amplia distribución tisular, en el ser humano se han identificado 11 ADN glicosilasas que pueden subdividirse en tres grupos; enzimas monofuncionales que escinden la base dañada dejando un sitio AP y un esqueleto de fosfodiéster intacto; glicosilasas bifuncionales que eliminan la base y escinden el enlace fosfodiéster en el lado 3' de la base dañada generando un aldehído 3 - α,β -insaturado (eliminación β) y glicosilasas similares a Nei (NEIL) las cuales pueden catalizar una reacción de eliminación β/δ en el que el enlace fosfodiéster es escindido [11]. Las glicosilasas analizadas en este estudio realizan funciones muy específicas; MUTYH se encarga de escindir la adenina insertada frente a la 8-oxoguanina (8-oxoG), además está involucrada en la reparación posreplicación [12], NEIL 1 tiene actividad dual de ADN glicosilasa y beta/delta liasa, NTHL1 es una enzima bifuncional, pertenece a la familia de las endonucleasas III [13], MPG tiene la función de una alquiladenina ADN glicosilasa en las bases alquiladas del ADN [14] y OGG1 es responsable de escindir a la 8-Oxoguanina (8-oxoG), dejando un sustrato para la endonucleasa APE1 [15].

Diferentes estudios han demostrado que la deficiencia, sobreexpresión o mutaciones de las proteínas que participan en la vía de BER particularmente de las glicosilasas están relacionadas con algunos procesos patológicos [16], [17]. Además, se ha reportado que la acumulación de lesiones y mutaciones en el ADN puede tener como consecuencia la inestabilidad genómica, acumulación de mutaciones, senescencia celular [18], lo cual tiene relación principalmente con trastornos neurodegenerativos y carcinogénesis [19], también se ha relacionado que las deficiencias en los procesos por BER son un factor importante en la inflamación y

enfermedades metabólicas [20]. Una de las características en las células tumorales es la deficiencia en la reparación del ADN por lo tanto las mutaciones en genes involucrados en la detección del daño al ADN y reparación ocasionaría una predisposición al cáncer [21].

Por otro lado, el estudio de genomas, transcriptomas y proteomas mediante la secuenciación de nueva generación (NGS) y espectrometría de masas ha generado una gran cantidad de información lo cual ha motivado a desarrollar nuevas investigaciones [22]. Actualmente en el GeneBank se han reportado 25 billones de pares de bases de más de 3 700 millones de secuencias de nucleótidos para 557 000 especies [23]. El uso de metodologías computacionales en los trabajos de investigación ha sido cada vez más frecuente en las ciencias biológicas [24]. En los últimos años se han desarrollado diferentes programas y softwares para el análisis de genes y proteínas como Expasy (Sistema experto de análisis de proteínas) Genome tools, por mencionar algunos, los cuales tienen la ventaja de que se pueden realizar en poco tiempo y a bajo costo, estas herramientas pueden predecir las características de diferentes moléculas, en algunos casos también se puede realizar la predicción de mutaciones, sobreexpresión, inhibición, regulación de la expresión genética, señalización celular, en diferentes procesos celulares y patologías [25], [26].

Las glicosilasas tienen un papel importante en la reparación del ADN y su deficiencia está asociada con la carcinogénesis, inflamación, envejecimiento y trastornos neurodegenerativos, por esta razón nos propusimos caracterizar los genes y proteínas de las glicosilasas debido que existen pocos estudios *in silico* en los procesos de reparación del ADN, por lo que los resultados obtenidos en este análisis pueden ser utilizados como una alternativa en el desarrollo de nuevas estrategias aplicadas en el área médica.



2. Metodología

En el presente estudio, se utilizaron servidores en línea y aplicaciones de softwares disponibles de manera gratuita, en la figura 2 se muestra la metodología análisis que se realizaron en este trabajo. Las secuencias de nucleótidos y aminoácidos de las glicosilasas implicadas en la

reparación del ADN por BER en *Homo sapiens* fueron recuperadas del GenBank en formato FASTA del Centro Nacional de Base de datos de información biotecnológica (<https://www.ncbi.nlm.nih.gov/gene/>). Para este estudio las glicosilasas seleccionadas fueron: OGG1, MPG, MUTYH, NEIL1 y NTHL1.

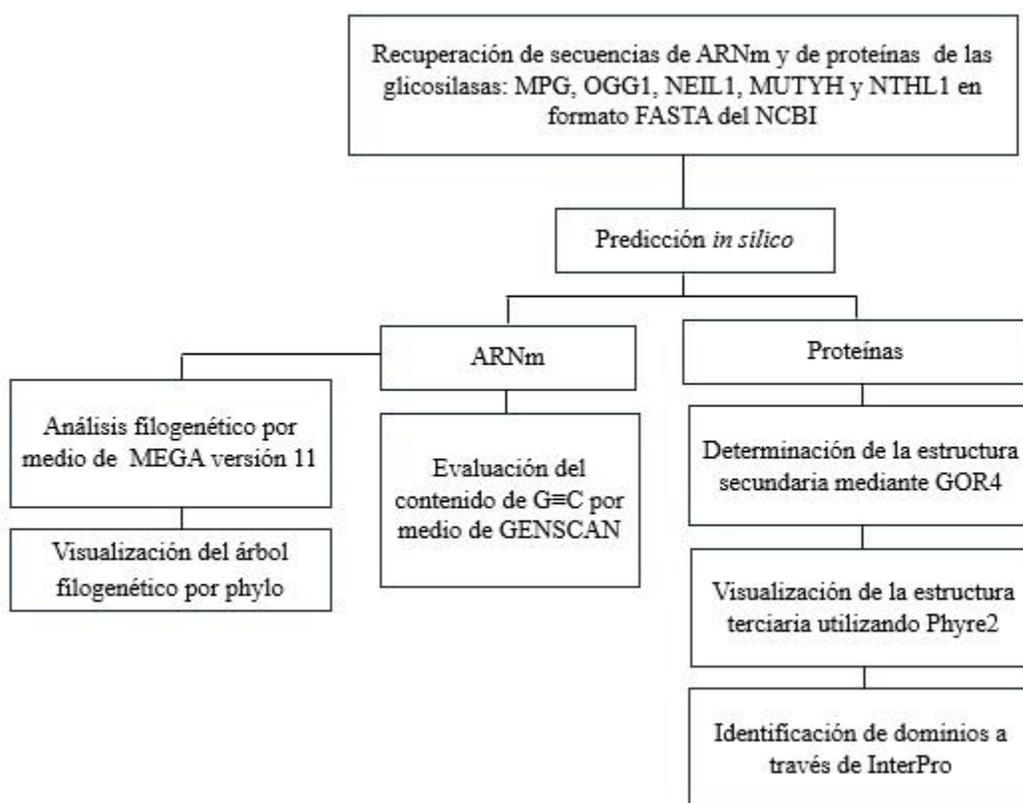


Figura 2. Diagrama de flujo de la metodología empleada en la predicción *in silico* de los ARN y proteínas de las glicosilasas.

Se descargaron las secuencias de los ARNm, en el caso de los transcritos con más de dos isoformas se seleccionó la isoforma más representativa de cada uno de los genes (MANE SELECT), además tomamos en cuenta los números de acceso, nombre de los genes, del mismo modo se descargaron las secuencias de aminoácidos de las proteínas, así como a longitud de cada secuencia

de nucleótidos y de aminoácidos. Una vez que se realizó la descarga de las secuencias se procedió a realizar el análisis del porcentaje del contenido de G=C de acuerdo con las secuencias de nucleótidos de los ARNm de cada gen se determinó con el software GENSCAN (<http://hollywood.mit.edu/GENSCAN.html>), esta herramienta es fácil de usar y tiene la ventaja de



predecir variaciones en regiones codificantes. La predicción de las estructuras secundaria de las proteínas se realizó con el software en línea GOR4 secondary structure prediction (https://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html), el cual permitió obtener el porcentaje de las estructuras correspondientes a alfa hélice, estructura extendida y plegamiento aleatorio. Para el análisis de la estructura terciaria se utilizaron las mismas secuencias de los aminoácidos para este objetivo se utilizó el software en línea Phyre2 Protein Fold Recognition Server (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>), este software es de fácil acceso, los algoritmos comparan la homología/analogía de estructuras de proteínas conocidas, lo cual permite predecir de la estructura terciaria de forma más precisa.

La predicción de los dominios presentes en las proteínas se realizó en el software InterPro (<https://www.ebi.ac.uk/interpro/>) con las secuencias de los aminoácidos previamente descargados. Finalmente, la relación filogenética entre los genes de las glicosilasas que participan en la reparación del ADN por el mecanismo de BER, se determinó mediante la construcción de un árbol filogenético utilizando las secuencias de nucleótidos de los ARNm recuperados del NCBI, mediante el software Molecular Evolutionary Genetics Analysis (MEGA), versión 11. Para este objetivo se realizó un alineamiento de las secuencias en formato FASTA y posteriormente se utilizó el método de pares no ponderado utilizando la media aritmética (UPGMA) con 1000 repeticiones bootstrap, la visualización del árbol filogenético se llevó a cabo a través del software phylo (<https://phylo.io/>).

Tabla 1. Secuencias de nucleótidos y aminoácidos de las glicosilasas que participan en el mecanismo de BER, recuperados del GenBank y contenido de G≡C.

Gen	Número de acceso al GenBank	Proteína	Número de acceso al GenBank	Nucleótidos (bp)	Aminoácidos (aa)	Contenido G≡C (%)
OGG1	NM_002542.6	N-glicosilasa/ADN liasa isoforma 1a	NP_002533.1	1630	345	59.20
MPG	NM_001015052.3	ADN-3-metiladenina glicosilasa isoforma b	NP_001015052.1	1036	293	63.80
MUTYH	NM_001048174.2	Adenina ADN glicosilasa isoforma 4	NP_001041639.1	1708	521	60.48
NEIL1	NM_024608.4	Endonucleasa 8-de tipo 1 isoforma 2	NP_078884.2	3711	390	61.33
NTHL1	NM_002528.7	Isoforma 1 de la proteína 1 de tipo endonucleasa III	NP_002519.2	1030	304	63.50

3. Resultados y Discusiones

En este trabajo analizamos a las principales glicosilasas involucradas en esta vía de reparación. En la tabla 1 se describen; el nombre del gen (ARNm), proteína, número de acceso del GeneBank, longitud de las secuencias tanto de

nucleótidos y aminoácidos, así como el contenido de G≡C. Se determinó el porcentaje de G≡C en las secuencias de cada transcrito, el gen MPG posee el mayor porcentaje de GC (63.80%), seguido por el gen NTHL1 (63.50%), NEIL1 (61.33%), MUTYH (60.48%), el valor más bajo en el contenido de C≡G se observó en el gen



OGG1 (59.20%). Se ha evidenciado que el contenido de G=C en el ARNm tiene la función de promover la exportación nuclear [27], [28].

Por otro lado, se ha sugerido que la eficiencia de la expresión genética depende del contenido de G=C en el ARNm en células de mamíferos por lo que las proteínas resultantes muestran una estructura más estable [29]. De acuerdo con los resultados mostrados anteriormente, los genes que codifican a las glicosilasas poseen porcentajes igual o mayor a 60%, lo cual sugiere que la estabilidad y función enzimática podría estar regulada desde la estructura del ARNm, para realizar eficientemente el proceso de remoción de bases dañadas. Interesantemente en un estudio de secuenciación se identificaron los sitios AP como marcadores de daño oxidativo al ADN. Se observó que, en regiones con alto contenido de GC, el daño al ADN se redujo, lo que sugiere que los genes en estas regiones pueden sufrir menos mutaciones [30].

La estructura secundaria de las proteínas se puede ver afectada principalmente por puentes de hidrógeno e interacciones hidrofóbicas en la cadena polipeptídica, por lo que la predicción de la estructura secundaria de las proteínas proporciona información relevante sobre la función, plegamiento, así como interacciones proteína-proteína para realizar funciones

específicas en diferentes procesos celulares [31-33].

Nuestro siguiente objetivo en este trabajo fue realizar el análisis de la estructura secundaria de las glicosilasas a través del software GOR4 en función de la secuencia de aminoácidos de las proteínas, esto permitió comparar el contenido de alfa hélice, hebra extendida y plegamiento aleatorio, las cuales se pueden observar en la tabla 2, el cálculo de los porcentajes de las estructuras secundarias se realizó de acuerdo con el número de aminoácidos (aa) de cada estructura secundaria sobre el total de aminoácidos de cada glicosilasa, la cantidad de aminoácidos se observa en la tabla 1. La proteína NTHL1 presenta el mayor porcentaje de alfa hélice con 43.42%, seguido de MUTYH con 40.31%, OGG1 con 35.36%, MPG con 32.42% y NEIL1 30.26% el cual representa el porcentaje más bajo de alfa hélice. La proteína OGG1 presenta el mayor porcentaje de hebras extendidas con 16.25%, seguido por MUTYH con 13.44%, MPG con 12.63%, NEIL1 con 12.05%, el valor más bajo fue para la proteína NTHL1 con 8.55%. El plegamiento aleatorio de las secuencias de aminoácidos de las glicosilasas se observó de la siguiente manera: la proteína NEIL1 tiene el mayor porcentaje con 57.69 % seguido de MPG con 54.95%, OGG1 con 48.41%, NTHL1 con 48.03% y MUTYH con el menor porcentaje para plegamiento aleatorio con 46.26%.

Tabla 2. Estructuras secundarias de las proteínas de las glicosilasas que participan en el proceso de reparación del ADN por BER

Gen	Alfa hélice	Estructura extendida	Plegamiento aleatorio	Longitud de la proteína (aa)	Alfa hélice (%)	Hebra extendida (%)	Plegamiento aleatorio (%)
OGG1	122	56	167	345	35.36	16.23	48.41
MPG	95	37	161	293	32.42	12.63	54.95
MUTYH	210	70	241	521	40.31	13.44	46.26
NEIL1	118	47	225	390	30.26	12.05	57.69
NTHL1	132	26	146	304	43.42	8.55	48.03



De acuerdo con lo anterior, el plegamiento aleatorio representa la estructura secundaria más abundante de las glicosilasas estudiadas, la hélice alfa fue la segunda estructura secundaria predominante y la hebra extendida representa el porcentaje más bajo, los porcentajes observados se relacionan con la función que realiza cada proteína. El plegamiento aleatorio tiene conformaciones extendidas y compactas, por lo que esta estructura es muy flexible para cambiar la conformación de la proteína, este cambio permite regular las interacciones en la exposición de sitios activos con otras proteínas o moléculas como ADN y ARN [30]. La estructura secundaria alfa hélice es una de las más abundantes en las proteínas, determinan en gran parte la estructura global e interaccionan con secuencias específicas del ADN a través de diferentes motivos en las proteínas [32], en este trabajo se observó que la alfa hélice es la segunda estructura abundante en las glicosilasas, la menos abundante la estructura extendida la cual se caracteriza por tener una función

estabilizadora en las proteínas, además de que pueden formar regiones cohesivas entre las proteínas y definen regiones donde las fluctuaciones estructurales son más o menos probables [33].

La estructura terciaria de macromoléculas como las proteínas se han determinado principalmente por los métodos biofísicos como la cristalografía de rayos X y espectroscopia de resonancia magnética nuclear (RMN) [34]. La mayoría de las estructuras terciarias de las proteínas registradas en las bases de datos como UniProt, el banco de datos de proteínas (PDB, por sus siglas en inglés), han sido determinadas por alguna de estas técnicas [35]. En el caso de las glicosilasas, las estructuras terciarias reportadas en el PDB han sido determinadas por difracción de rayos X, no obstante, la principal ventaja de los algoritmos *in silico* es la apreciación de las estructuras tridimensionales en diferentes ángulos.

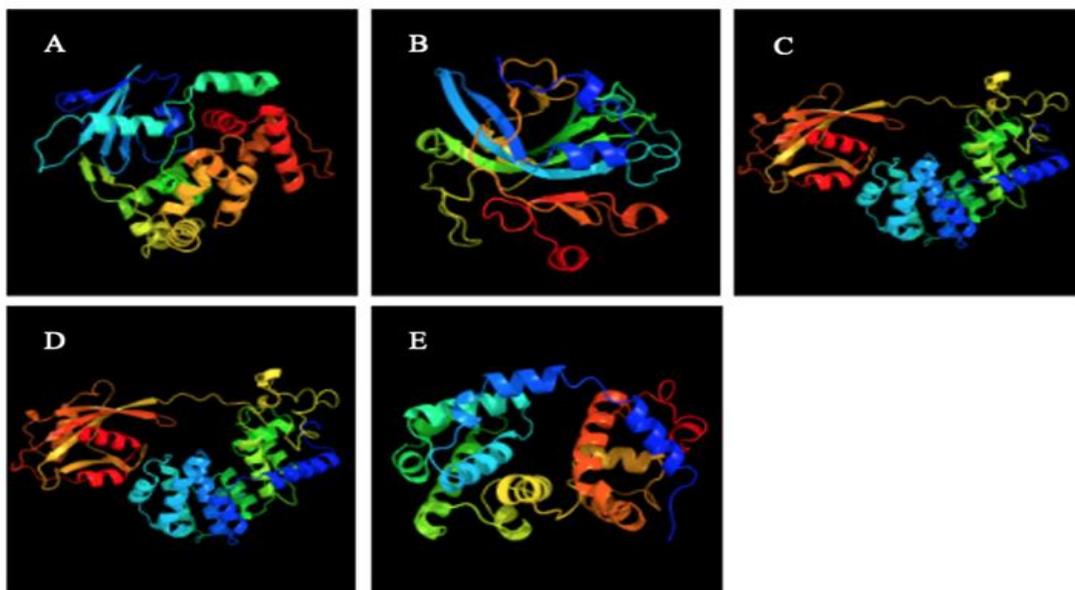


Figura 3. Estructura terciaria predicha de las glicosilasas en humanos. A) OGG1, B) MPG, C) MUTYH, D) NEIL1, E) NTHL1.

Las estructuras terciarias de las glicosilasas fueron obtenidas con el servidor Phyre2, estas se

pueden observar en la figura 3, (A) OGG1 principalmente se compone de alfa hélices y



plegamiento aleatorio, MPG (B) de plegamientos aleatorios y giros beta, MUTYH (C) se compone en gran parte de alfa hélice, las hebras extendidas y el plegamiento aleatorio en menor cantidad, (D) NEIL y NTHL1 (E) contienen elevado contenido de alfa hélices. Tomando en cuenta las características de las estructuras secundaria y terciaria, estos son importantes debido a que determinan la función biológica, en este caso en la función catalítica de remover nucleótidos dañados. Las predicciones *in silico* de proteínas tienen la limitante de que puede haber un margen de error en los algoritmos empleados, de este modo las estructuras generadas no sustituyen a los estudios experimentales y estos deben utilizarse como una herramienta en el estudio de proteínas [36], además se debe tomar en cuenta que las proteínas son moléculas dinámicas y esta es una desventaja que los algoritmos no pueden predecir, así como las combinaciones de mutaciones [37].

Las estructuras proteicas obtenidas mediante diferentes softwares deben ser validadas por técnicas experimentales como la difracción de rayos X de cristales, la RMN y la microscopía electrónica. A pesar de las limitaciones, este tipo de estudios ha sido utilizado en la ingeniería de proteínas, en el diseño de experimentos como la mutagénesis dirigida y también para caracterizar los sitios activos y catalíticos de las proteínas, predicción de plegamientos [38], interacciones moleculares como el acoplamiento molecular (molecular docking), este último tiene la utilidad en el diseño de fármacos [39].

Tabla 3. Predicción de los dominios en las secuencias de las glicosilasas realizada en el programa InterPro.

Glicosilasa	Dominios	Posición en la secuencia de aminoácidos
OGG1	HhH-GPD	139 – 319
	endo3end	146 – 316
	HhH-GPD	143 – 285
	ENDO3c	139 – 319
	OGG N	26 – 141
MPG	No reportado	No reportado
MUTYH	HhH-GPD	100 – 260
	endo3end	108 – 260
	HhH-GPD	199 – 238
	ENDO3c	100 – 258
	MutY_C	340 – 469
	NUDIX	357 – 466
	DNA_Glycosylase_C	340 – 469
NUDIX hydrolase	339 – 470	
NEIL1	Endonuclease VIII-like 1	252 – 290
	FPG_cat	1 – 124
	FPG_CAT	2 – 121
	Fapy_DNA_glyco_2	2 – 124
	Fapy_DNA_glyco	1 – 123
	DNA glycosylase/AP lyase, DNA-bd	142 – 224
NTHL1	H2TH_2	142 – 224
	HhH-GPD	122 – 280
	ENDO3c	122 – 278
	endo3end	130 – 280
	HhH-GPD	127 – 191, 216 - 263



Las glicosilasas se han identificado en, bacterias, arqueas y eucariontes, lo cual sugiere que hubo una diversificación para alterar la especificidad del sustrato durante la reparación del ADN [40]. Además del plegamiento, un factor que influye en la función de una proteína es la presencia de dominios, los cuales están conformados por secuencias de aminoácidos que pueden formar una unidad estructural y funcionalmente independiente con características que se conservan en ciertos grupos de familias de proteínas [41], [42].

En nuestro estudio realizamos la predicción de los dominios presentes en las glicosilasas a través de la plataforma InterPro, como se puede ver en la tabla 3, las enzimas OGG1, MUTYH y NTHL1 presentan el dominio HhH, mientras que las glicosilasas NEIL1 y MPG carecen de este dominio, pero presentan otros dominios con funciones específicas. La familia con el dominio

HhH-GPD (hélice- hairpin-hélice) y su bucle rico en glicina/prolina forma parte de una amplia gama de proteínas reparadoras de ADN estructuralmente relacionadas [43]. Las glicosilasas que presentan el dominio HhH, son unas de las familias más versátiles para reconocer diversas lesiones en el ADN principalmente en la reparación de purinas y pirimidinas oxidadas, además este grupo de glicosilasas participan en diversos procesos, como la regulación genética y el control del ciclo celular [44]. La importancia de este dominio y la eficiencia de las glicosilasas se debe tomar en cuenta debido a que la alteración o inactivación de estos dominios puede afectar los procesos de reparación del ADN [45]. La vía de reparación por medio de BER es un proceso evolutivamente conservado, diversos estudios han identificado alrededor de 80 tipos diferentes de lesiones de bases nitrogenadas [46] y azúcares de manera directa o indirecta en el ADN [47], [48].

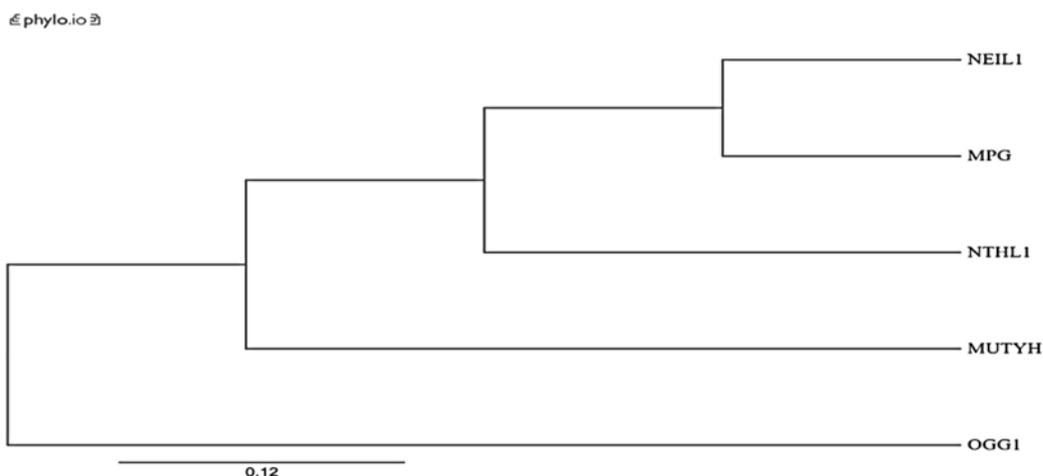


Figura 4. Árbol filogenético generado por el método UPGMA con base en la comparación de secuencias de genes de las glicosilasas que participan en la reparación del ADN por BER.

El último objetivo fue analizar la relación filogenética de los transcritos de los genes que codifican a las glicosilasas, el árbol filogenético muestra la relación entre las enzimas, así como su historia evolutiva enraizado a partir de un

ancestro común, se utilizó el software MEGA y en este se observan cuatro clústeres (figura 4), cada rama o grupo tiene secuencias de nucleótidos similares, los genes NEIL1 y MPG están representados en el grupo uno, el grupo dos



por el gen NTHL1, el grupo tres por el gen MUTYH, y el grupo cuatro es el gen OGG1 el cual representa el ancestro común de todos los genes analizados en este estudio. Las descripciones de las estructuras secundarias, el modelado de la estructura terciaria de las proteínas, la presencia de dominios, así como el análisis filogenético de los genes, muestran que cada grupo (clado) posee una función similar. En la figura 5 se puede apreciar el mecanismo de reparación por medio de BER, observamos la importancia de las glicosilasas funcionales y no funcionales o defectuosas. Las predicciones de la estructura de los genes y proteínas de las glicosilasas reportadas en este estudio pueden tener aplicaciones en diversas disciplinas biológicas como la epidemiología, genética de poblaciones y en la biomedicina [49].

Además, esta información puede ser utilizada en diferentes análisis como, por ejemplo, en la agrupación de genes con una función determinada, cálculos de similitud de estructuras proteicas, en el caso de los dominios identificados en mecanismos de actividad, genómica funcional, proteómica, así como en el diseño de experimentos para validar las predicciones. Interesantemente la caracterización tanto de genes y proteínas, en la identificación de nuevos marcadores moleculares de ARN o proteínas aplicadas en el diagnóstico o incluso en el tratamiento de enfermedades como cáncer, enfermedades neurodegenerativas, inflamatorias y autoinmunes [50].

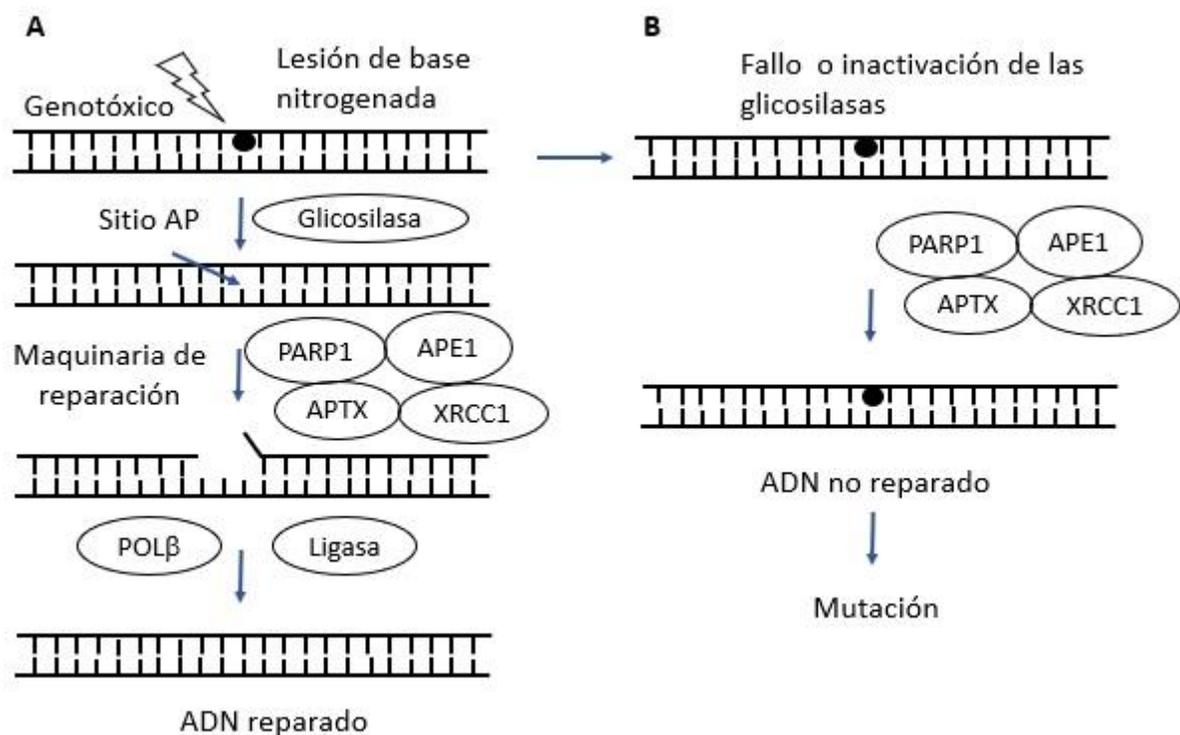


Figura 5. Mecanismo de reparación por BER. A, las lesiones en las bases nitrogenadas del ADN son eliminadas por las glicosilasas, posteriormente un conjunto de enzimas se encarga de sintetizar una nueva hebra de ADN. B, el fallo en las glicosilasas ocasionaría que el ADN no se repare, generando acumulación de mutaciones.



4. Conclusiones

Los resultados muestran que las glicosilasas analizadas en este estudio tienen un contenido de G≡C superior al 60%, lo que sugiere una alta estabilidad del ARNm y un posible impacto positivo en la eficiencia de la expresión génica en células de mamíferos. Además, las estructuras secundarias predominantes en estas proteínas son la alfa hélice y el plegamiento aleatorio, lo que puede influir en sus interacciones y funciones durante la reparación del ADN, además la presencia del dominio HhH en las glicosilasas permite identificar y actuar específicamente sobre las bases dañadas del ADN, facilitando su interacción con el ADN para reconocer y eliminar las modificaciones en las bases,

iniciando así el proceso de reparación por BER. En conclusión, este estudio demuestra que las herramientas bioinformáticas son una metodología valiosa para la caracterización inicial de proteínas involucradas en diversos procesos celulares y mecanismos esenciales, como la reparación del ADN. Su aplicación puede extenderse a numerosos campos, desde la biomedicina hasta la genética evolutiva.

5. Agradecimientos

Los autores expresan su agradecimiento a la Facultad de Ciencias Químico Biológicas de la Universidad Autónoma de Campeche.

6. Agradecimientos de autores

Marco Antonio Popoca Cuaya: Análisis de datos, metodología, discusión, escritura, versión final y edición; *Estrella Alexandra Pinkney Rivas:* Investigación, metodología, software y borrador.



Referencias

- [1] D. Cucchi, A. Gibson, and S. a Martin, "The emerging relationship between metabolism and DNA repair," *Cell Cycle*, vol. 20, no. 10, pp. 943–959, May 2021, <https://doi:10.1080/15384101.2021.1912889>.
- [2] U. S. Srinivas, B. W. Q. Tan, B. A. Vellayappan, and A. D. Jeyasekharan, "ROS and the DNA damage response in cancer," *Redox Biol*, vol. 25, p. 101084, Jul. 2019, <https://doi:10.1016/j.redox.2018.101084>.
- [3] N. Chatterjee and G. C. Walker, "Mechanisms of DNA damage, repair, and mutagenesis," *Environ Mol Mutagen*, vol. 58, no. 5, pp. 235–263, Jun. 2017, <https://doi:10.1002/em.22087>.
- [4] T.-H. Lee and T.-H. Kang, "DNA Oxidation and Excision Repair Pathways," *Int J Mol Sci*, vol. 20, no. 23, p. 6092, Dec. 2019, <https://doi:10.3390/ijms20236092>.
- [5] M. B. S. Mota, M. A. Carvalho, A. N. A. Monteiro, and R. D. Mesquita, "DNA damage response and repair in perspective: *Aedes aegypti*, *Drosophila melanogaster* and *Homo sapiens*," *Parasit Vectors*, vol. 12, no. 1, p. 533, Dec. 2019, <https://doi:10.1186/s13071-019-3792-1>.
- [6] Z. Nikitaki, C. E. Hellweg, A. G. Georgakilas, and J.-L. Ravanat, "Stress-induced DNA damage biomarkers: applications and limitations," *Front Chem*, vol. 3, Jun. 2015, <https://doi:10.3389/fchem.2015.00035>.
- [7] Y. Baiken *et al.*, "Role of Base Excision Repair Pathway in the Processing of Complex DNA Damage Generated by Oxidative Stress and Anticancer Drugs," *Front Cell Dev Biol*, vol. 8, Jan. 2021, <https://doi:10.3389/fcell.2020.617884>.
- [8] R. J. Carter and J. L. Parsons, "Base Excision Repair, a Pathway Regulated by Posttranslational Modifications," *Mol Cell Biol*, vol. 36, no. 10, pp. 1426–1437, May 2016, <https://doi:10.1128/MCB.00030-16>.
- [9] D. Gohil, A. H. Sarker, and R. Roy, "Base Excision Repair: Mechanisms and Impact in Biology, Disease, and Medicine," *Int J Mol Sci*, vol. 24, no. 18, p. 14186, Sep. 2023, <https://doi:10.3390/ijms241814186>.
- [10] J. Woodrick *et al.*, "A new sub-pathway of long-patch base excision repair involving 5' gap formation," *EMBO J*, vol. 36, no. 11, pp. 1605–1622, Jun. 2017, <https://doi:10.15252/embj.201694920>.
- [11] F. Hans, M. Senarisoy, C. Bhaskar Naidu, and J. Timmins, "Focus on DNA Glycosylases—A Set of Tightly Regulated Enzymes with a High Potential as Anticancer Drug Targets," *Int J Mol Sci*, vol. 21, no. 23, p. 9226, Dec. 2020, <https://doi:10.3390/ijms21239226>.
- [12] A. Konopka and J. D. Atkin, "The Role of DNA Damage in Neural Plasticity in Physiology and Neurodegeneration," *Front Cell Neurosci*, vol. 16, Jun. 2022, <https://doi:10.3389/fncel.2022.836885>.
- [13] B. Van Houten, G. A. Santa-Gonzalez, and M. Camargo, "DNA repair after oxidative stress: Current challenges," *Curr Opin Toxicol*, vol. 7, pp. 9–16, Feb. 2018, <https://doi:10.1016/j.cotox.2017.10.009>.
- [14] Gh. R. Bhat, I. Sethi, B. Rah, R. Kumar, and D. Afroze, "Innovative in Silico Approaches for Characterization of Genes and Proteins," *Front Genet*, vol. 13, May 2022, doi:10.3389/fgene.2022.865182.
- [15] X. Zhang *et al.*, "In silico Methods for Identification of Potential Therapeutic Targets," *Interdiscip Sci*, vol. 14, no. 2, pp. 285–310, Jun. 2022, <https://doi:10.1007/s12539-021-00491-y>.
- [16] D. Mitra *et al.*, "Evolution of bioinformatics and its impact on modern bio-science in the twenty-first century: Special attention to pharmacology, plant science and drug discovery," *Computational Toxicology*, vol. 24, p. 100248, Nov. 2022, <https://doi:10.1016/j.comtox.2022.100248>.
- [17] B. D. Freudenthal, "Base excision repair of oxidative DNA damage from mechanism to disease," *Frontiers in Bioscience*, vol. 22, no. 9, p. 4555, 2017, <https://doi:10.2741/4555>.
- [18] G. J. Grundy and J. L. Parsons, "Base excision repair and its implications to cancer therapy," *Essays Biochem*, vol. 64, no. 5, pp. 831–843, Oct. 2020, <https://doi:10.1042/EBC20200013>.
- [19] G. S. Leandro, P. Sykora, and V. A. Bohr, "The impact of base excision DNA repair in age-related neurodegenerative diseases," *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, vol. 776, pp. 31–39, Jun. 2015, <https://doi:10.1016/j.mrfmmm.2014.12.011>.
- [20] D. Hanahan, "Hallmarks of Cancer: New Dimensions," *Cancer Discov*, vol. 12, no. 1, pp. 31–46, Jan. 2022, <https://doi:10.1158/2159-8290.CD-21-1059>.
- [21] E. J. Duncavage *et al.*, "Recommendations for the Use of in Silico



- Approaches for Next-Generation Sequencing Bioinformatic Pipeline Validation,” *The Journal of Molecular Diagnostics*, vol. 25, no. 1, pp. 3–16, Jan. 2023, <https://doi:10.1016/j.jmoldx.2022.09.007>.
- [22] E. W. Sayers *et al.*, “GenBank 2024 Update,” *Nucleic Acids Res*, vol. 52, no. D1, pp. D134–D137, Jan. 2024, <https://doi:10.1093/nar/gkad903>.
- [23] A. J. Lee and S. S. Wallace, “Hide and seek: How do DNA glycosylases locate oxidatively damaged DNA bases amidst a sea of undamaged bases?,” *Free Radic Biol Med*, vol. 107, pp. 170–178, Jun. 2017, <https://doi:10.1016/j.freeradbiomed.2016.11.024>.
- [24] D. Jurkovicova, C. M. Neophytou, A. Č. Gašparović, and A. C. Gonçalves, “DNA Damage Response in Cancer Therapy and Resistance: Challenges and Opportunities,” *Int J Mol Sci*, vol. 23, no. 23, p. 14672, Nov. 2022, <https://doi:10.3390/ijms232314672>.
- [25] E. A. Mullins, A. A. Rodriguez, N. P. Bradley, and B. F. Eichman, “Emerging Roles of DNA Glycosylases and the Base Excision Repair Pathway,” *Trends Biochem Sci*, vol. 44, no. 9, pp. 765–781, Sep. 2019, doi: <https://10.1016/j.tibs.2019.04.006>.
- [26] Y. Ouyang *et al.*, “Recent advances in biosensor for DNA glycosylase activity detection,” *Talanta*, vol. 239, p. 123144, Mar. 2022, <https://doi:10.1016/j.talanta.2021.123144>.
- [27] A. F. Palazzo and Y. M. Kang, “GC-content biases in protein-coding genes act as an ‘mRNA identity’ feature for nuclear export,” *BioEssays*, vol. 43, no. 2, Feb. 2021, <https://doi:10.1002/bies.202000197>.
- [28] Y. S. Rao, X. W. Chai, Z. F. Wang, Q. H. Nie, and X. Q. Zhang, “Impact of GC content on gene expression pattern in chicken,” *Genetics Selection Evolution*, vol. 45, no. 1, p. 9, Dec. 2013, <https://doi:10.1186/1297-9686-45-9>.
- [29] M. Courel *et al.*, “GC content shapes mRNA storage and decay in human cells,” *Elife*, vol. 8, Dec. 2019, <https://doi:10.7554/eLife.49708>.
- [30] A. R. Poetsch, S. J. Boulton, and N. M. Luscombe, “Genomic landscape of oxidative DNA damage and repair reveals regioselective protection from mutagenesis,” *Genome Biol*, vol. 19, no. 1, pp. 215–237, Dec. 2018, <https://doi:10.1186/s13059-018-1582-2>.
- [31] L. Degrève, C. A. Fuzo, and A. Caliri, “Extended secondary structures in proteins,” *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1844, no. 2, pp. 384–388, Feb. 2014, <https://doi:10.1016/j.bbapap.2013.10.005>.
- [32] P. Craveur *et al.*, “Protein flexibility in the light of structural alphabets,” *Front Mol Biosci*, vol. 2, May 2015, <https://doi:10.3389/fmolb.2015.00020>.
- [33] W. Yang, Y. Liu, and C. Xiao, “Deep metric learning for accurate protein secondary structure prediction,” *Knowl Based Syst*, vol. 242, p. 108356, Apr. 2022, <https://doi:10.1016/j.knosys.2022.108356>.
- [34] S. Arce-Solano and E. Hernández-Carvajal, “Implementación de las técnicas de RMN y cristalografía de macromoléculas para la caracterización estructural de proteínas de interés biomédico,” *Revista Tecnología en Marcha*, Sep. 2019, <https://doi:10.18845/tm.v32i9.4627>.
- [35] P. Choudhary, S. Anyango, J. Berrisford, M. Varadi, J. Tolchard, and S. Velankar, “Unified access to up-to-date residue-level annotations from UniProt and other biological databases for PDB data via PDBx/mmCIF files,” *bioRxiv*, 2022, <https://doi.org/10.1038/s41597-023-02101-6>.
- [36] P. Katsonis, K. Wilhelm, A. Williams, and O. Lichtarge, “Genome interpretation using in silico predictors of variant impact,” *Hum Genet*, vol. 141, no. 10, pp. 1549–1577, Oct. 2022, <https://doi:10.1007/s00439-022-02457-6>.
- [37] O. Carugo and K. Djinović-Carugo, “Structural biology: A golden era,” *PLoS Biol*, vol. 21, no. 6, p. e3002187, Jun. 2023, <https://doi:10.1371/journal.pbio.3002187>.
- [38] V. Cicaloni, A. Trezza, F. Pettini, and O. Spiga, “Applications of in Silico Methods for Design and Development of Drugs Targeting Protein-Protein Interactions,” *Curr Top Med Chem*, vol. 19, no. 7, pp. 534–554, May 2019, <https://doi:10.2174/1568026619666190304153901>.
- [39] Q. Jiang, X. Jin, S.-J. Lee, and S. Yao, “Protein secondary structure prediction: A survey of the state of the art,” *J Mol Graph Model*, vol. 76, pp. 379–402, Sep. 2017, <https://doi:10.1016/j.jmgm.2017.07.015>.
- [40] P. Prorok *et al.*, “Evolutionary Origins of DNA Repair Pathways: Role of Oxygen Catastrophe in the Emergence of DNA Glycosylases,” *Cells*, vol. 10, no. 7, p. 1591, Jun. 2021, <https://doi:10.3390/cells10071591>.
- [41] M. F. Aziz and G. Caetano-Anollés, “Evolution of networks of protein domain organization,” *Sci Rep*, vol. 11, no. 1, p. 12075, Jun. 2021, <https://doi:10.1038/s41598-021-90498-8>.
- [42] S. H. Wilson, “The dark side of DNA repair,” *Elife*, vol. 3, May 2014, <https://doi:10.7554/eLife.03068>.
- [43] C. H. Trasviña-Arenas, M. Demir, W.-J. Lin, and S. S. David, “Structure, function and



evolution of the Helix-hairpin-Helix DNA glycosylase superfamily: Piecing together the evolutionary puzzle of DNA base damage repair mechanisms,” *DNA Repair (Amst)*, vol. 108, p. 103231, Dec. 2021,

<https://doi:10.1016/j.dnarep.2021.103231>.

[44] M. De Rosa, S. A. Johnson, and P. L. Opresko, “Roles for the 8-Oxoguanine DNA Repair System in Protecting Telomeres From Oxidative Stress,” *Front Cell Dev Biol*, vol. 9, Nov. 2021, <https://doi:10.3389/fcell.2021.758402>.

[45] H. Sampath, “Oxidative DNA damage in disease—Insights gained from base excision repair glycosylase-deficient mouse models,” *Environ Mol Mutagen*, vol. 55, no. 9, pp. 689–703, Dec. 2014, <https://doi:10.1002/em.21886>.

[46] M. Stratigopoulou, T. P. van Dam, and J. E. J. Guikema, “Base Excision Repair in the Immune System: Small DNA Lesions With Big

Consequences,” *Front Immunol*, vol. 11, May 2020, <https://doi:10.3389/fimmu.2020.01084>.

[47] H. E. Krokan and M. Bjoras, “Base Excision Repair,” *Cold Spring Harb Perspect Biol*, vol. 5, no. 4, pp. a012583–a012583, Apr. 2013, <https://doi:10.1101/cshperspect.a012583>.

[48] E. Nischwitz *et al.*, “DNA damage repair proteins across the Tree of Life,” *iScience*, vol. 26, no. 6, p. 106778, Jun. 2023, <https://doi:10.1016/j.isci.2023.106778>.

[49] G. Munjal, M. Hanmandlu, and S. Srivastava, “Phylogenetics Algorithms and Applications,” 2019, pp. 187–194. https://doi:10.1007/978-981-13-5934-7_17.

[50] J. Wang, “Editorial: Methods and Applications in Molecular Phylogenetics,” *Front Genet*, vol. 13, Jul. 2022, <https://doi:10.3389/fgene.2022.923409>.

Derechos de Autor (c) 2024 Estrella Alejandra Pinkney Rivas, Marco Antonio Popoca Cuaya



Este texto está protegido por una licencia [Creative Commons 4.0](https://creativecommons.org/licenses/by/4.0/).

Usted es libre para compartir —copiar y redistribuir el material en cualquier medio o formato— y adaptar el documento —remezclar, transformar y crear a partir del material— para cualquier propósito, incluso para fines comerciales, siempre que cumpla la condición de:

Atribución: Usted debe dar crédito a la obra original de manera adecuada, proporcionar un enlace a la licencia, e indicar si se han realizado cambios. Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que tiene el apoyo del licenciante o lo recibe por el uso que hace de la obra.

[Resumen de licencia](#) - [Texto completo de la licencia](#)