



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**EVOLUCIÓN DE LOS SISTEMAS DE REPARACIÓN
DE DNA DEPENDIENTES DE OXÍGENO Y ANÁLISIS
DE SU RELACIÓN EVOLUTIVA CON ENZIMAS
INVOLUCRADAS EN LA FORMACIÓN DE
COLÁGENA Y PENICILINA**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

BIÓLOGO

P R E S E N T A:

WOLFGANG FRANCISCO COTTOM SALAS



**DIRECTOR DE TESIS:
DR. ANTONIO LAZCANO-ARAUJO REYES
2016**

CIUDAD UNIVERSITARIA, CDMX



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Hoja de Datos del Jurado

1. Datos del Alumno

Cottom
Salas
Wolfgang Francisco
55 48 13 49
Universidad Nacional Autónoma de México
Facultad de Ciencias
Biología
309125846

2. Datos del Tutor

Doctor
Antonio
Lazcano-Araujo
Reyes

3. Datos del Sinodal 1

Doctora en Química
Mendoza
Negrón
Alicia

4. Datos del Sinodal 2

Doctora
García
Colín
María

5. Datos del Sinodal 3

Doctor
López
Valdés
Víctor Manuel

6. Datos del Sinodal 4

Doctor
Arredondo
Delaye
Luis José

7. Datos del Trabajo Escrito

Evolución de los sistemas de reparación de DNA dependientes de oxígeno y análisis de su relación evolutiva con enzimas involucradas en la formación de colágena y penicilina
85p
2016

A mis padres, Bolfy y Cynthia. Los pilares de lo que soy.

AGRADECIMIENTOS

Al Doctor Antonio Lazcano y al Doctor Arturo Becerra que me ha abierto las puertas de su conocimiento y su laboratorio. Al laboratorio de Origen de la Vida, Ricardo Hernández, Claudia Álvarez, José Campillo, Carolina Rocha, Alberto Vázquez, Alejandro Álvarez, Ervin Silva, Sara Islas, Israel Muñoz, Coral Luna, Mario Rivas y Rodrigo Jácome; por sus sugerencias y ayuda en la realización de este trabajo.

A los profesores miembros del sínodo por sus muy importantes comentarios y revisiones de este trabajo.

Dra. Alicia Negrón Mendoza

Dra. María Colín García

Dr. Víctor Manuel Valdés López

Dr. Luis José Delaye Arredondo

A las profesoras del Taller “Problemas históricos y filosóficos de la Biología”, Dra. Ana Barahona Echeverría, Dra. Erica Torrens y a la M en C. Alicia Villela por su valiosa instrucción en la carrera.

A todos los profesores de los que tuve la dicha de aprender en mi vida y en especial a los de la Facultad de Ciencias de nuestra Universidad Nacional Autónoma de México.

A mis amigos cercanos y constantes de la universidad y la carrera: José, Julia, Alfonso, Carolina, Yunuén. Gracias por su compañía y sus consejos.

A quienes no tuvieron la oportunidad de estudiar en esta universidad.

Gracias.

*“La ciencia es una parte integral de la cultura. No es cosa extranjera,
realizada por un arcano sacerdocio. Es una de las glorias de la tradición
intelectual humana”*

Stephen Jay Gould

Índice	
1.Resumen/Abstract	8-9
2.Introducción	10
2.1 Evolución atmosférica, el Gran Evento Oxidativo (GOE)	10
2.2 Evolución temprana de la vida: RNA, DNA y proteínas	14
2.3 Daño al DNA por oxígeno	16
2.4 Daño al DNA y sistemas de reparación	18
2.5 Reparación del DNA dependiente de oxígeno. El caso de AlkB	21
2.6 Estructuras cristalográficas de AlkB y sus homólogos	23
3. Objetivos	24
3.1 Objetivo general	24
3.2 Objetivos particulares	24
4.Metodología	25
4.1 Análisis evolutivo con secuencias de residuos de aminoácidos	25
4.2 Análisis evolutivo de AlkB y sus homólogos a partir de sus estructuras terciarias	26

5.Resultados	27
5.1 De análisis evolutivo con secuencias de aminoácidos	
5.2 De análisis evolutivo con estructuras terciarias	30
6. Discusión	32
6.1 Implicaciones del estudio comparativo de la estructura primaria de AlkB y sus homólogos	32
6.2 Implicaciones del uso de estructura terciaria para analizar homología del dominio 2OG-Fe(II)_Oxy	35
6.3 Implicaciones biológicas y evolutivas del dominio 2OG-Fe(II)_Oxy	36
7.Conclusiones	41
8. Referencias	42
9. Apéndice	47
9.1 Bases de datos consultadas	47
9.2 Apéndice de resultados	49
9.3 Identificadores UniProt de las secuencias utilizadas en este trabajo y alineamiento local entre AlkB de <i>Escherichia coli</i> y el dominio del virus del género <i>Lolium</i> . En rojo los residuos característicos del dominio 2OG_Fe(II)_Oxy	81

Resumen

El oxígeno es un elemento químico de importancia central para la vida. Sin embargo, a lo largo de la historia de la Tierra no siempre ha estado disponible como O₂ para su utilización por la biósfera. Debido a su naturaleza altamente oxidante, han surgido evolutivamente una serie de mecanismos para controlar, aprovechar y corregir los productos de las reacciones oxidativas con componentes celulares como membranas, proteínas, ácidos nucleicos y otras moléculas. Las lesiones por adición de grupos alquilo en el DNA y el RNA son reparados mediante la escisión directa de las bases y, en ocasiones, la escisión total del nucleótido. Aravind y Koonin (2001) mediante iteraciones de PSI-BLAST (Altschul y Madden., 1997) identificaron homólogos remotos de la proteína reparadora de DNA AlkB, descrita en *Escherichia coli* y otras bacterias, en el humano y otros eucariontes e incluso, homólogos en virus pero ausentes en el dominio Archaea. Si bien AlkB es una enzima reparadora de DNA, hay homólogos que poseen al menos un dominio similar a AlkB, y que en su mayoría tienen funciones distintas a la reparación de DNA. Sorprendentemente, AlkB también repara RNA, lo cual ha sido interpretado por Logan y Poole (2005) y otros autores como prueba de mecanismos de reparación en etapas evolutivas anteriores al DNA. El estudio evolutivo de este sistema de reparación, que forma una superfamilia proteínica llamada "oxigenasas- α -ketoglutarato(α KG)Fe(II)-dependientes" nos puede dar información evolutiva sobre el desarrollo de estrategias de reparación, así como sobre las implicaciones evolutivas de la utilización de O₂ en enzimas similares. En este trabajo se reporta el resultado de un análisis evolutivo del sistema de reparación de polinucleótidos dependiente de oxígeno (AlkB y sus homólogos) merced a: a) un análisis filogenético con sus homólogos a partir de la estructura primaria, secundaria y terciaria en proteínas; b) una datación relativa en función de su naturaleza dependiente de O₂; y; c) sus implicaciones evolutivas en algunos grupos biológicos.

Palabras clave: GOE (Gran Evento Oxidativo), AlkB, biosíntesis de colágena, biosíntesis de penicilina

Abstract

The oxygen is a chemical element Important for life. Nevertheless, in the Earth history had not been available not like O₂ for their utilization by the biosphere. Because their high oxidant nature, mechanisms several have arrived by evolution to control to take advantage and correct the oxidative products in cellular components like membranes, proteins, nucleic acids and other molecules. The injuries by alkyl groups in the DNA and RNA are corrected by the direct basis-cleavage and once in a while by the total nucleotide-cleavage. Aravind and Koonin (2001) using PSI-BLAST iterations, (Altschul and Madden. 1997) Have Identified remote homologs from the repair protein of DNA AlkB Described in Escherichia coli and other bacteria, in human and other eukaryots, even homologs have been Described homologs in viruses but have not been identified in the domain Archaea. Even though, AlkB is a repair enzyme of DNA, some homologs that have at least the same domain to AlkB and in fact, the great majority have different functions instead of the DNA repair. Surprisingly, AlkB repairs to RNA, this has been interpreted by Logan and Poole (2005) and other authors as a proof of the existence of repair mechanisms in earlier evolutive stages to DNA. The evolutive study of this repair system, part of super protein family called " α -ketoglutarate oxygenases- (α KG) Fe (II) -dependent" can give us information by the development of repair strategies as well as the evolutive consequences of the O₂ utilization in a similar Enzymes. In this study, we report the results of a evolutionary analysis of the polynucleotides repair oxygen system dependent (AlkB and their homologs) in order to obtain: a) a phylogenetic analysis homologs in primary structure, secondary structure and tertiary structure; b) a relative datations in function of the nature of oxygen dependent AlkB an homologs and ; c) a evolutionary implicaciones in some biological groups.

Key words: GOE (Great Oxidant Event), AlkB, biosynthesis of collagen, biosynthesis of penicillin

2.Introducción

2.1 Evolución de la atmósfera terrestre: el Gran Evento Oxidativo (GOE)

El oxígeno en su forma (O_2) actualmente es una molécula de importancia central para muchas especies pero también es el responsable del daño oxidativo al DNA y a otras estructuras celulares. Este daño es inevitable y ha estado presente a lo largo de millones de años y resulta del propio metabolismo celular oxidativo (Aguirre *et al.*, 2006). Un tipo de daño al DNA es el endógeno, inducido por especies reactivas de oxígeno (ROS), eventos hidrolíticos y radiaciones ionizantes que pueden polimerizar, romper, alquilar u oxidar al DNA. Algunas entidades biológicas generan energía mediante la respiración aerobia, proceso que incluye un paso de reducción de oxígeno molecular a agua. Estas especies parcialmente reducidas producen intermediarios como subproductos de la respiración aerobia, incluyendo O_2^- , H_2O , y OH^\bullet . Todos ellos son potentes oxidantes electrofílicos que en el caso de los eucariontes, pueden escapar de las mitocondrias y atacar los componentes celulares vitales (Bruner *et al.*, 2000; Cunningham, 1997; Cooke *et al.*, 2003)

Como lo resumieron Lazcano y Miller (1996), no existe un acuerdo generalizado sobre la composición de la atmósfera primitiva, que bien podría haber sido altamente reductora con combinaciones de compuestos como ($CH_4 + N_2$, $NH_3 + H_2O / CO_2 + H_2 + N_2$), o bien pudo haber sido neutra, con una composición de ($CO_2 + H_2O, N_2$) Lazcano y Miller (1996) A pesar de las pocas evidencias geológicas para definir una composición química precisa de la atmósfera primitiva, al menos se acepta que el O_2 libre estuvo ausente o en muy bajas concentraciones, y distribuido globalmente en forma no uniforme. El requerimiento de una atmósfera reductora fue en principio esencial para la formación de aminoácidos, azúcares, purinas y pirimidinas, las primeras biomoléculas en sintetizarse abióticamente luego de la formación del planeta hace unos 4,500 millones de años. Esta posibilidad se ve apoyada por el análisis de meteoritos como el Murchinson que tiene una edad de 4.6×10^9 años. Se ha propuesto que el impacto constante en la Tierra primitiva de meteoritos, cometas y asteroides también pudo ser fuente de diversos compuestos orgánicos esenciales para la vida.

La teoría Oparin-Haldane (1924 y 1928) sostiene que las primeras formas de vida probablemente fueron heterótrofas y que consumían los compuestos sintetizados abióticamente en el ambiente primitivo. Debido a que un metabolismo heterótrofo es más simple que uno autotrófico, Oparin argumentó que esta forma de vida debió necesariamente evolucionar primero (Lazcano, 2010). Los antecedentes formativos particulares de Alexander I. Oparin, le permitieron basar su teoría en una perspectiva darwiniana que coalesce con las ideas evolutivas actuales, lo que da validez evolutiva a su propuesta. Por ejemplo, los experimentos de síntesis prebiótica como los de Miller (1953) han demostrado la facilidad de síntesis de compuestos orgánicos bajo una atmósfera reductora, aunque experimentos recientes han comprobado también la síntesis de aminoácidos bajo condiciones neutrales (Cleaves *et al.*, 2008).

Como se ha mencionado antes, el O₂ molecular actualmente es para muchos organismos una molécula esencial que permite el intercambio energético en la fotosíntesis oxigénica y la respiración celular. La distribución de la utilización del O₂ está presente en bacterias fotosintéticas, cianobacterias y en todos los eucariontes. Se ha sugerido que el origen del grupo de las cianobacterias tuvo lugar hace unos 2,400 millones de años y marcó el inicio del evento geológico conocido como "*Great Oxidation Event*" (GOE por su acrónimo en inglés), un evento cuyo inicio resulta difícil de marcar con precisión y que debemos como un proceso gradual y no lineal. Hasta hace poco, la evidencia de este *Gran Evento Oxidativo* (GOE), estaba sustentada por la presencia de sedimentos de aspecto rojo (oxidados). Sin embargo, la oxidación de estos estratos ricos en Fe puede producirse aún en ausencia de O₂ producto de la fotosíntesis anoxigénica o por efecto de la radiación UV (Lyons *et al.* 2014). Se debe subrayar que el metabolismo anoxigénico representa en los organismos una alternativa donde el producto final de dicho metabolismo no incluye al O₂. De hecho, Lyons *et al.*, (2014), proponen tres donadores alternativos de electrones que pudieron ser clave en la acumulación de sedimentos ricos en carbono orgánico durante el *Arqueano* sin la necesidad de O₂, el ácido sulfhídrico (H₂S), el hierro en su forma catiónica 2+ (Fe²⁺) y el hidrógeno molecular (H₂). Sin embargo, la fotosíntesis con estos compuestos como donadores de electrones parece ser poco probable por lo que la fotosíntesis oxigénica continúa siendo la mejor explicación para la acumulación de compuestos orgánicos de carbono antes de el Gran Evento Oxidativo.

Todas las rocas del *Arqueano* han sufrido procesos de metamorfismo que han modificado los registros geoquímicos, dificultando la interpretación y reconstrucción de la evolución atmosférica de la Tierra (Lyons *et al.*, 2014). Más aún, fósiles microbianos del mismo eón poseen morfologías simples y es complicado asignarles un metabolismo específico como productores de oxígeno molecular por medio de la fotosíntesis. De la misma manera, los estromatolitos del *Arqueano* y fósiles microbianos en sedimentos han sido objeto de una amplia discusión en su interpretación fosilífera. Se ha sugerido recientemente que la utilización de biomarcadores permite tener una mayor certidumbre para inferencias de tiempo profundo como los eones Arqueano-Proterozoico. De hecho, Brocks (1999) ha sugerido que a partir de datos de un biomarcador, (fósil molecular, un lípido, 2 α - methylphane, encontrado en Pilbara Craton, Australia) se puede inferir la presencia de cianobacterias en rocas de al menos 2,700 millones de años de edad.

No hay duda alguna de que el oxígeno molecular (O₂) que existe en la atmósfera es producto de la fotosíntesis oxigénica, pero hoy en día mucho de este oxígeno es consumido constantemente por organismos aerobios. La acumulación y depósito de carbono orgánico cambia la composición isotópica del carbono inorgánico en el océano debido a que el carbono orgánico tiene proporción menor de C¹³/C¹²; el depósito dramático de carbono orgánico debe entonces manifestarse con una exclusión del isótopo C¹³. Esta aproximación ha sido usada con frecuencia para estimar la edad y naturaleza del depósito de carbono y el contenido atmosférico de O₂ a través del tiempo (Berner, 2004)

Hace más de una década Holland (2006) propuso cinco etapas en la historia geológica para catalogar los ambientes atmosféricos tempranos: los ambientes anoxigénicos, durante el *Gran Evento Oxidativo* y después del *Gran Evento Oxidativo*. Holland analizó esta secuencia temporal tomando como punto de partida la situación hace 3,850 millones de años (3.5 Gyr). Así, durante la etapa 1 (Figura 1, 2 y 3) de hace 3,850 millones de años a 2,450 millones de años, la atmósfera era enteramente anóxica y también los océanos con la posible excepción de depósitos de oxígeno en los océanos superficiales.

Durante la etapa 2 (Figura 1, 2 y 3) de hace 2,450 millones de años hasta 1,850 millones de años, los niveles de oxígeno atmosférico rozan los valores estimados entre 0.02 y 0.04 atm. Los océanos superficiales llegan a estar ligeramente oxidados, pero los océanos profundos continúan con una condición anóxica. Durante la etapa 3 (Figura 1, 2 y 3) hace 1,850 millones de años y hasta 850 millones de años no ocurren grandes cambios y se dan momentos “inapetentes” geológicamente. Los niveles de oxígeno atmosféricos no cambian significativamente. Gran parte de la superficie de los océanos están medianamente oxigenada, y lo mismo ocurre con las profundidades oceánicas. La etapa 4 (Figura 1, 2 y 3) (de hace 850 millones de años hasta 540 millones de años) muestran valores de oxígeno atmosférico no mayores a 0.2 atm. Los océanos poco profundos continuaron igual, pero los océanos profundos fueron anóxicos al menos durante las eras de hielo del Neoproterozoico. Durante la última etapa, la 5 (hace 540 millones de años- presente) el O₂ probablemente alcanzó un máximo valor de 0.3 atm durante el Carbonífero antes de regresar a los valores actuales de... Los océanos superficiales están oxigenados, pero la oxigenación de los océanos profundos fluctúa considerablemente debido a que es una etapa de corta duración geológicamente.

Con estos antecedentes histórico-biológicos en mente, debemos entender que la acumulación del O₂ en la atmósfera fue un evento que marcó y sigue afectando desde entonces la evolución de los metabolismos de los organismos que están en contacto directo con esta molécula. La evolución ha favorecido y diversificado aquellas estrategias que permiten a los organismos sobrevivir, aprovechar y heredar rasgos adaptados como las superóxido dismutasas que catalizan algunas especies reactivas de oxígeno; algunas DNA glicosilasas; y la utilización dependiente del oxígeno en la reacción de la proteína AlkB.

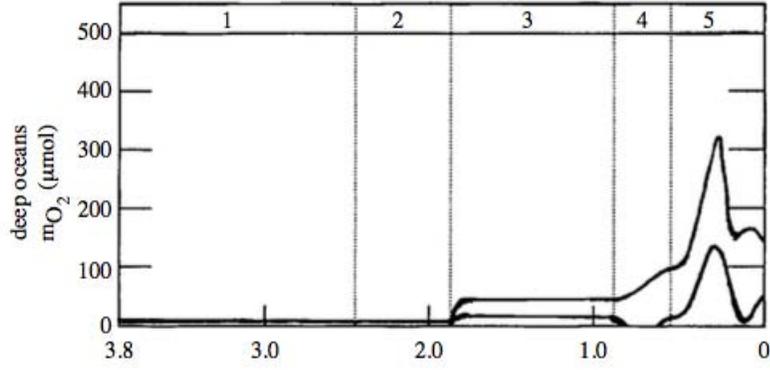


Figura 1. Evolución de pO_2 a través del tiempo. Se muestran las cinco etapas propuestas por Holland (2006) (Tomado de Holland 2006)

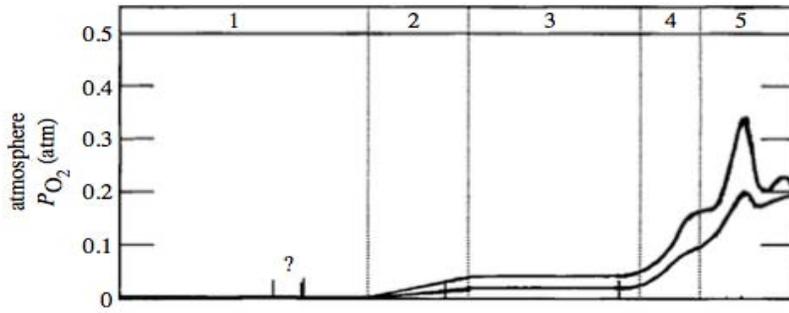


Figura 2. Concentración relativa de O_2 a través de las etapas propuestas por Holland (2006) en los océanos superficiales.

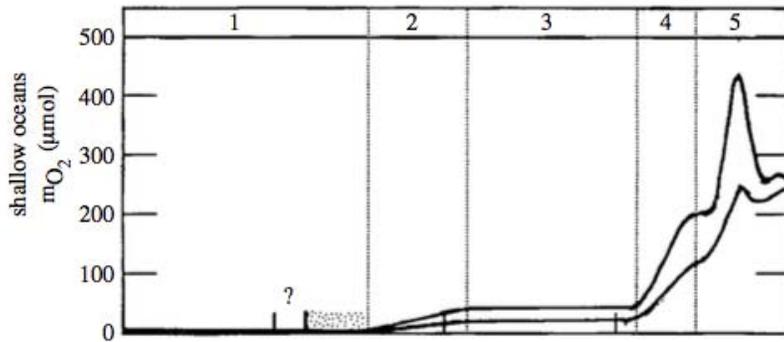


Figura 3. Concentración relativa de O_2 a través de las etapas que Holland (2006) para en los océanos profundos.

Como podemos ver en la Figura 4., tomada de Lyons *et al.* (2014), se debe considerar que el GOE es una etapa heterogénea, bajo el tenor de las concentraciones de O₂ en la atmósfera, y no una etapa lineal en el aumento de la concentración de O₂ libre atmosférico en toda la atmósfera terrestre.

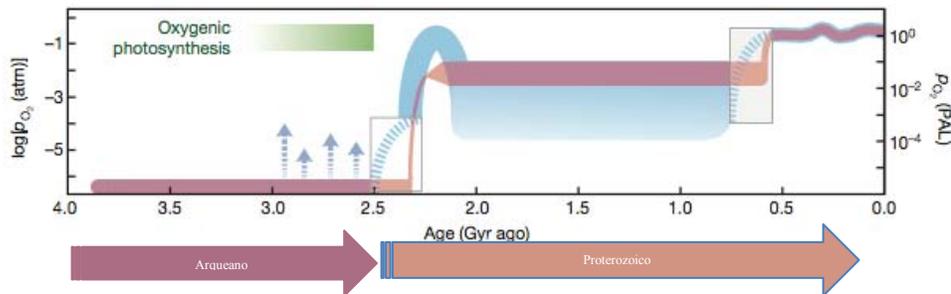


Figura 4. La curva de color rojo degradado nos muestra la visión clásica de dos pasos en la evolución atmosférica. Mientras la curva azul nos muestra el modelo emergente (pO₂, presión parcial de O₂) En el eje de la derecha se muestra la pO₂ relativo al presente y en el eje izquierdo las posibles emergencias de O₂ en el eón Arcaico tardío aunque no es bien entendida esa etapa de la evolución atmosférica. (modificado de Lyons 2014)

2.2 Evolución temprana de la vida: RNA, DNA y proteínas

Como es bien sabido, el DNA es un polímero aniónico presente en todos los sistemas celulares exceptuando a las entidades biológicas como los virus, algunos de los cuales pueden contener genomas compuestos por RNA. Lazcano *et al.* (1988) han propuesto tres fases evolutivas para discernir la evolución del material genético (Figura 5). La utilización del RNA como un polímero genético por formas de vida temprana y su habilidad catalítica abren una discusión evolutiva sobre cuándo arribó la síntesis de proteínas. Una de las cuestiones más importantes está en el estudio del origen del primer material genético, su evolución y la naturaleza del material genético presente en Last Common Ancestor (LCA por sus siglas en inglés), el último ancestro común de los seres vivos actuales. El RNA por si solo pudo ser el resultado evolutivo de diferentes series de polímeros genéticos (Lazcano & Miller, 1996). El mundo del RNA es un periodo hipotético de la biósfera temprana donde tanto la información genética como la actividad catalítica en los procesos biológicos dependían de moléculas de RNA. Esta hipótesis ha sido discutida extensamente, sobre todo el periodo de transición de un mundo de pre-RNA a un mundo de RNA, y la pregunta que surge es si la síntesis de proteínas evolucionó desde un mundo de RNA a un mundo moderno de DNA/RNA/proteína donde el DNA almacena la información y el RNA sirve de auxiliar funcional y catalítico – como ocurre en ribozimas

y el ribosoma – y las proteínas realizan la catálisis (Dworkin *et al.* 2003) Si el RNA evolucionó primero, ¿quién evolucionó después? ¿Proteínas o el DNA?

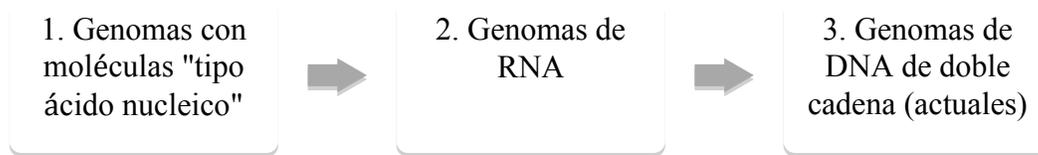


Figura 5. Transición evolutiva del RNA al DNA propuesta por Lazcano *et al.* 1988.

La principal corriente de investigación que plantea una biósfera RNA/proteína hasta ahora está basada en una pequeña pero robusta evidencia. Stephen Freeland y colaboradores han examinado nueva evidencia que sugiere una visión de un ambiente temprano de RNA/proteína con el DNA evolucionando al último, estableciendo una cronología que involucra el contexto evolutivo del código genético permitiéndoles así hacer predicciones acerca de la distribución del RNA como maquinaria fundamental de la vida (Freeland *et al.*, 1999)

La gran diferencia a nivel químico entre el RNA y el DNA es la presencia de ribosa en el polímero, lo que afecta la estabilidad en el esqueleto fosfodiéster del DNA. Las enzimas encargadas de la conversión de la ribosa a una desoxirribosa (ribonucleótido reductasas) juegan un papel central en la transición evolutiva del polímero genético actual (Lazcano *et al.* 1988; Nordlund y Reichard., 2006).

La transición evolutiva del RNA como material genético celular a DNA de doble hebra no ha sido explicado con detalles. Sin embargo, algunos argumentos pueden ser numerados: 1) la mayor estabilidad del esqueleto con enlaces fosfodiéster debido a la presencia de desoxirribosa ; 2) la presencia de mecanismos de reparación; 3) resistencia a degradación química; y 4) menor sensibilidad a la radiación UV del DNA comparada con el RNA. La doble hebra del DNA es más estable a la depurinización mientras que las reacciones hidrolíticas como desaminaciones, depurinizaciones y depirimidizaciones es en un orden de magnitud una a dos veces mayor en una sola hebra de DNA. (Lazcano *et al.*, 1988).

En particular, todo indica que para lograr una fidelidad genética adecuada en un genoma, es indispensable la presencia de ciertos mecanismos de reparación. En un inicio pudo haber sido mucho más complejo el reconocimiento de errores de incorporación de productos de reacciones hidrolíticas, incluso más importante, la necesidad de los ácidos nucleicos de doble hebra y la corrección de errores y reparación del daño hidrolítico (Dworkin *et al.*, 2003).

2.3. Daños causados al DNA por el oxígeno

El estrés oxidativo celular puede ser definido como un desajuste del balance entre oxidantes y antioxidantes lo que resulta en un incremento en el daño celular. Debemos tener en cuenta que el estrés oxidativo debe darse en un contexto celular con al menos dos constituyentes: el estructural y el genético. Por estructural debemos entender aquellos elementos que dan soporte a una célula procarionte o eucarionte (pared, membrana, organelos, proteínas, lípidos, polisacáridos); mientras que el componente genético todo constituyente celular que porte información heredable en forma de un polímero de desoxiribonucleótidos o ribonucleótidos (DNA y RNA). Es en este contexto donde podemos apreciar que el interés de este proyecto está en aquellos daños en el constituyente genético (DNA).

Si bien las bases canónicas del DNA – guanina(G), citosina(C), timina(T) y adenina(A) – pueden sufrir eventos de oxidación que alteran su composición y su naturaleza química, la guanina es en particular, debido a su bajo potencial redox, la base más vulnerable a ser oxidada (Neeley *et al.*, 2006). El producto más abundante debido a esta oxidación es la 7,8-dihidro-8-oxoguanina (Figura 6) (también conocida como 8-oxoguanina; 8-oxoG). De los diversos daños inducidos por las especies reactivas de oxígeno, la más estudiada es la 7,8-dihidroxi-8-oxoguanina. Aunque la guanina tiene una fuerte preferencia de aparearse en forma canónica con citosina, la oxo-G puede aparearse con citosina en el modo Watson-Crick, y en el modo Hoogsteen con adenina, lo que suele ocurrir durante la replicación *in vitro* e *in vivo* elevando la probabilidad de que al darse el cambio G•C -> T•A ocurran transversiones (Bruner *et al.* 2000).

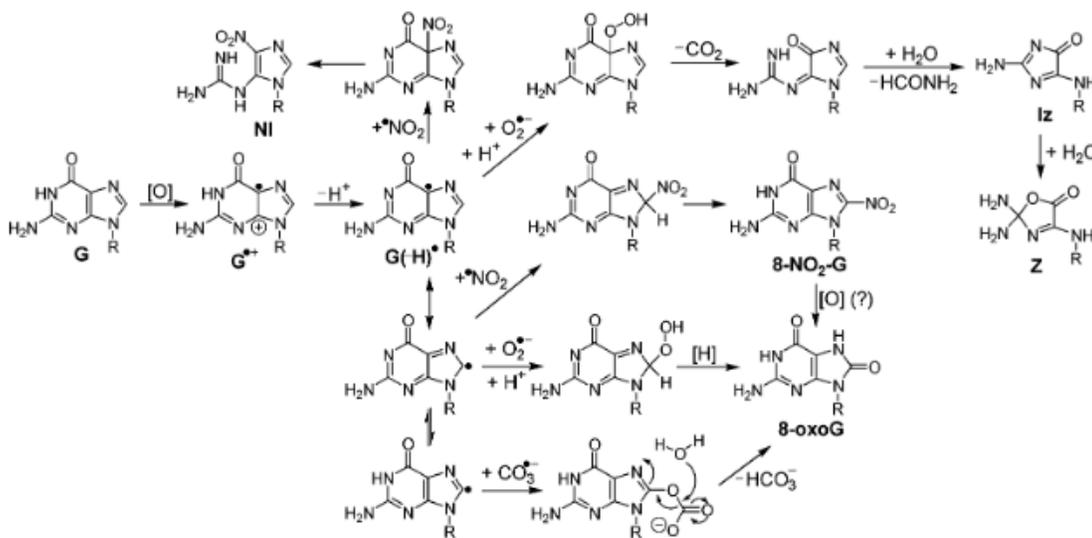


Figura 6. Principales mecanismos de oxidación de la guanina. Tomado de Neeley *et al.* (2006)

Las DNA glicosilasas son proteínas que reconocen de manera específica las bases dañadas y las remueven mediante la escisión el enlace *N*-glicosídico que une a la base con la desoxirribosa, lo que da como resultado un sitio apurínico/apirimídico (AP) que posteriormente es procesado. Específicamente, una AP

endonucleasa escinde el carbono 5' del sitio AP, y una fosfodiesterasa retira el fosfato 5'-desoxirribosa restante para dejar un hueco (gap) en el DNA, que es posteriormente reparado por una polimerasa y sellado por una ligasa. Se han descrito un gran número de glicosilasas con una variedad de sustratos específicos que podrían estar evolutivamente relacionadas (Cunningham.,1997; Jacobs y Schär., 2012)

De acuerdo a Cunningham (1997), las glicosilasas pueden dividirse en dos grandes categorías: las glicosilasas “puras”, que remueven una base inapropiada o alterada sin romper el enlace fosfodiéster adyacente a la base dañada como la uracil-DNA glicosilasa, y las demás glicosilasas, que tienen una actividad asociada que rompe el DNA en el sitio del daño. Desde la perspectiva bioquímica, es la de Drohat y Maiti (2013), propone una clasificación en torno a su función, sea ésta monofuncional o bifuncional. Las glicosilasas monofuncionales tienen una actividad única de escisión de bases, mientras que las bifuncionales también tienen una función de DNA-liasa. El mecanismo de reacción de una glicosilasa y los pasos subsiguientes de la ruta de reparación por escisión de bases (BER) difieren sustancialmente entre estos dos tipos de enzimas (David, S *et al.*, 2007; Van der Veen y Tang., 2015) Las glicosilasas monofuncionales escinden el enlace glicosídico mediante una reacción en donde el agua es el nucleófilo produciendo un producto común, un sitioapurínico y apirimídico (AP).

Otras enzimas involucradas en BER completan el proceso de reparación. Las enzimas bifuncionales escinden las bases dañadas catalizando una transferencia glicosídica, donde el nucleófilo es una amina provista por la enzima. La porción restante del nucleótido dañado es reconocida y removida por una AP endonucleasa que libera al grupo 3'-OH listo para la síntesis de reparación. Algunas glicosilasas bifuncionales suelen catalizar una eliminación- δ que libera el carbono 5' del enlace con O, removiendo así completamente el nucleótido dañado.

Las DNA glicosilasas pueden ser clasificadas de acuerdo a su estructura secundaria, en seis superfamilias que reconocen un amplio espectro de lesiones al DNA. La superfamilia helix-hairpin-helix (hélice-horquilla-hélice, HhH o Endo III) es un conjunto grande y diverso que contiene enzimas monofuncionales, y están representadas por la enzima MutY de *Escherichia coli*, que remueve adenina de un apareamiento A•oxoG, otras enzimas que reparan daños de alquilación y, en eucariontes, la 8-oxoguanin-glicosilasa (OGG1). La superfamilia hélix-two-turn-helix (hélice- dos giros- hélice, H2TH) incluye enzimas bifuncionales que reconocen bases oxidadas, la enzima MutM y las glicosilasas Nei (Figura 6).(Drohat y Maiti, 2013, Cunningham 1997).

Oxo-G es el mutágeno endógeno que más contribuye a la transformación celular espontánea. La enzima MutT es una trifosfatasa que limpia el precursor del pool de bases catalizando una hidrólisis de oxo-GTP a oxo-dGMP + Pi. MutY es una DNA glicosilasa que inicia la reparación de productos erróneos de replicación como oxo-G•A, catalizando una hidrólisis del enlace glicosídico que une la base con su azúcar. El último componente es la Ogg1, una DNA glicosilasa/ β -liasa que reconoce oxo-G•C (Bruner *et al.*, 2000).

2.4 Daños físicos al DNA y sistemas de reparación

El material genético de cualquier entidad biológica es susceptible de sufrir daños físicos. Factores endógenos y exógenos constantemente generan daños celulares, produciendo aductos en el DNA. Como resultado, en el curso de la evolución han surgido diversas estrategias para contrarrestar el daño en el DNA (Friedberg, 1985)

El daño al DNA activa diversas rutas bioquímicas. Primero, enzimas específicas de diversa complejidad reconocen y eliminan el daño. Segundo, los puntos de control de daño (checkpoints) son activados por los daños al DNA, pueden llegar a bloquear el ciclo celular y, en algunos casos, ayudan a recuperar la competencia celular. Por último, los daños al DNA activan la transcripción de ciertos genes. En algunos casos, algunas formas de reparación ocurren independientemente de las tres condiciones a lo que varios autores han llamado “la reparación directa al DNA”. Esta susceptibilidad ha estado presente como consecuencia del daño latente, diversas rutas y mecanismos de reparación se han desarrollado con el fin de contrarrestar estos daños en el material genético.

De acuerdo a Dexheimer (Figura 7) (2012), podemos clasificar los mecanismos de reparación en cuatro categorías; 1.- Reparación por escisión de bases (BER); 2.- Reparación de Mismatch; 3.- Reparación por escisión de nucleótidos (NER); y 4.- Reparación por ruptura de doble Hélice (NHEJ). Esta clasificación, de hecho, excluye otros mecanismos de reparación directa al DNA. Sancar *et al.*, (2004), proponen una clasificación más amplia (Tabla 1)

Tabla 1. Mecanismos de reparación del DNA de acuerdo a la clasificación de Sancar *et al.* 2004

Mecanismo de reparación	Características
Reparación directa	Por inducción de dímeros de pirimidinas por radiación UV y remoción de grupos metilo. (Fotoliasa, metiltransferasa y metiltransferasa oxidativa)
Reparación por escisión de bases	Por oxidación o reducción de bases. (DNA glicosilasas)
Reparación por escisión de nucleótidos	Es el sistema mayor de reparación que remueve lesiones en DNA por exposición a radiación o químicos; incluyendo adiciones protéicas al DNA. (Nucleasa de escisión)
Reparación por ruptura de doble hebra	Por radiación ionizante o especies reactivas de oxígeno. Reparación por recombinación homóloga y recombinación no-homóloga
Reparación por recombinación	Por especies reactivas de oxígeno, radiación ionizante y compuestos que generan especies reactivas de oxígeno. (Recombinación homóloga y no homóloga)
Reparación de cross-link.	Por sustancias y agentes que inducen daño al DNA. (Reparación por escisión de nucleótido o recombinación homóloga)

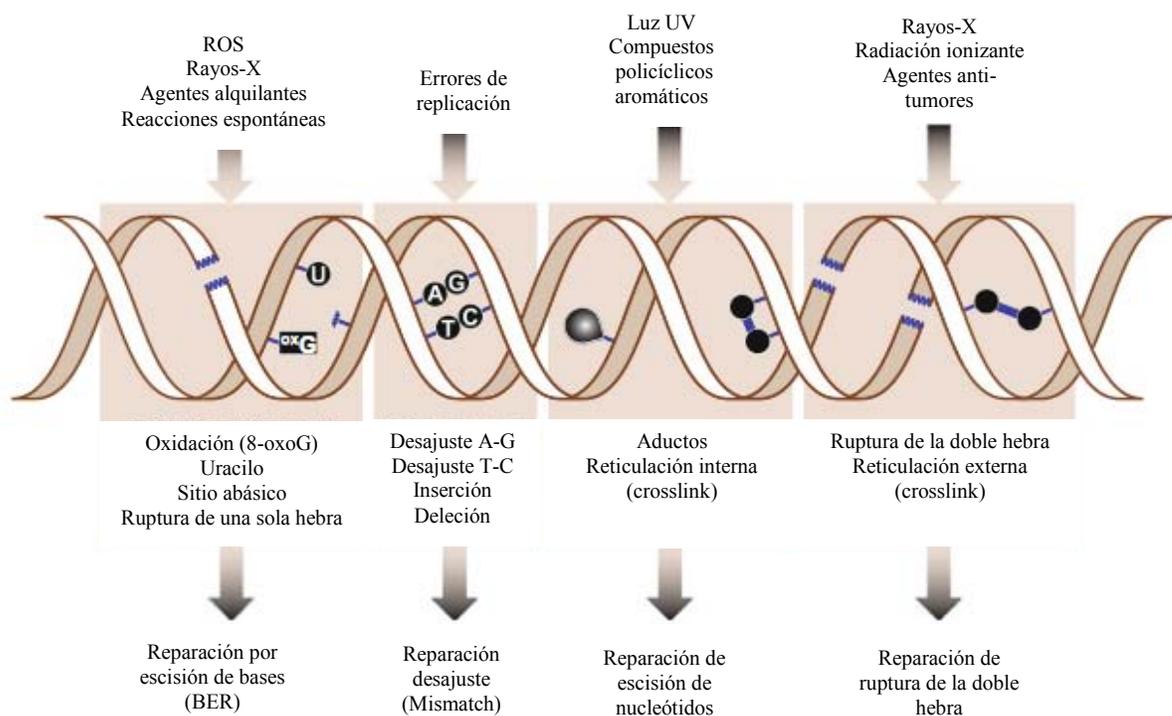


Imagen 7. Daños posibles al DNA y mecanismos de reparación y edición. Tomado y modificado de Dexheimer (2012)

Tomando en cuenta únicamente la reparación directa de DNA, Yi y He (2013) proponen tres clasificaciones para catalogar a los mecanismos de reparación directa al DNA al considerar la función que llevan a cabo: a) fotoliasas, b) alquil-transferasas y c) dioxigenasas (Figura 8)

Estos mecanismos de reparación directa de DNA han sido identificados como: I. fotoliasas que corrigen fotolesiones inducidas por radiación UV; II. alquiltransferasas, que corrigen el daño de alquilaciones en O-guaninas; y III. AlkB dioxigenasas, que corrigen aductos N-aquilados.

I. Se han identificado dos tipos de daños por radiación UV en el DNA: los dímeros de ciclobutano-pirimidinas y los foto-productos 6-4 de pirimidinas. Todas las fotoliasas se unen al cofactor esencial flavin-adenin-dinucleótido (FAD), y sólo el FADH reducido es catalíticamente competente.

II. Agentes alquilantes reaccionan con heteroátomos de las bases del DNA, para generar variedades de aductos citotóxicos y mutagénicos covalentes desde simples grupos metilo hasta adiciones alquilantes grandes. En contraste con la ruta de reparación de escisión de bases que también corrige lesiones N-alquilantes, existen dos rutas de reparación directa de DNA para combatir efectos deletéreos: a) AGTs reversas de guaninas alquiladas, y b) las dioxigenasas de la familia AlkB, que también corrigen las N-alquilaciones que bloquean apareamientos Watson-Crick.

III. Se ha observado que AlkB también repara lesiones en RNA (Aas *et al.*, 2003; Westby *et al.*, y se han identificado hasta nueve homólogos en humanos (ALKBH1-8 y FTO). Es la mayor clase de

desmetilasas DNA/RNA en humanos. En el evento de desmetilación oxidativa la familia de proteínas AlkB utilizan un sitio con Fe(II) para activar la molécula de oxígeno molecular para la oxidación de los grupos alquilo aberrantes. Los grupos alquilo hidroxilados que son unidos por el N1 de una adenina, o por el N3 de una citosina,⁹ para después escindir fácilmente el enlace C-N formando una base sin modificar y formaldehído. Al igual que las fotoliasas y alquiltransferasas, las AlkB y otras dioxigenasas relacionadas también necesitan retirar la base dañada de una doble hebra de DNA. Se ha demostrado que AlkB repara lesiones alquilantes en ssDNA y dsDNA con una preferencia en ssDNA.

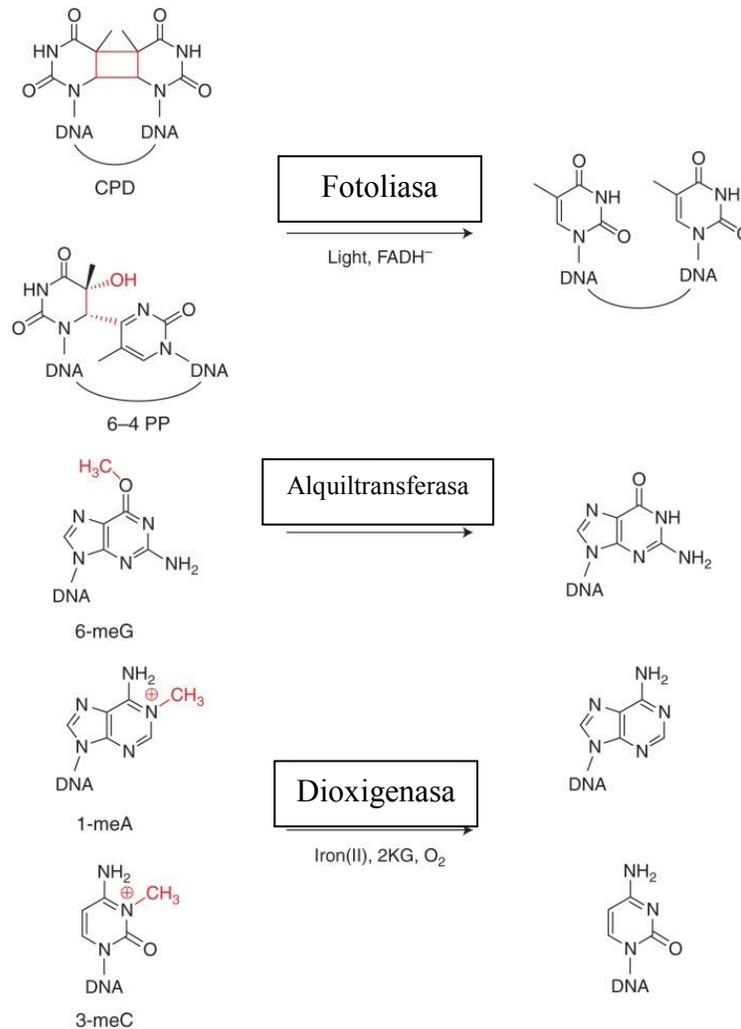


Imagen 8. Reparación directa de DNA. Tomado y modificado de Yi y He *et al.* 2013

2.5 Reparación de DNA dependiente de oxígeno. El caso de AlkB

Es importante subrayar la distinción entre la edición de DNA y la reparación de DNA. Mientras que la edición corrige daños en el ordenamiento de la secuencia nucleotídica, la reparación busca y corrige errores de carácter físico en toda la estructura del DNA e incluso de RNA.

No sólo el oxígeno, también otros factores endógenos y exógenos de naturaleza alquilante pueden producir daños en el DNA o en el RNA. Otras reacciones químicas pueden alterar o provocar la pérdida de la estructura en regiones codificantes del DNA. La importancia de la reparación de material polinucleotídico estriba directamente en la conservación de la información genética (Silvestrov *et al.*, 2014)

AlkB es un gen inducible a agentes alquilantes y se puede ver como una respuesta adaptativa al daño en el material genético. Este gen codifica para una enzima involucrada en la reparación al DNA, que corrige la alquilación inducida por agentes del tipo S_N2. Sin embargo, AlkB repara principalmente lesiones generadas en una sola hebra de DNA, siendo 1-metiladenina (1meA) el sustrato con mayor frecuencia de reconocimiento y catálisis que 3-metilcitosina (3meC) (Sedwick *et al.*, 2007) La razón de que estas lesiones no ocurran en una doble hebra de DNA debido a que los anillos de nitrógeno en la doble hélice son protegidos por la otra hebra complementaria de DNA.

La corrección del sustituyente alquilo (grupos metilo) de 1meA y 3meC en una sola hebra de DNA requiere de manera estricta de Fe²⁺ y 2OG. El grupo metil unido al anillo de nitrógeno es hidroxilado con un intermediario oxi-ferril. El producto de la reacción es reactivo y se descompone fácilmente, y es retirado como formaldehído, y la base dañada es regenerada. Se consume oxígeno molecular y el co-sustrato, 2OG, se convierte en succinato y en CO₂ (Vågbø *et al.* 2013) (Figura 9)

A partir del estudio de AlkB en *E. coli*, se reconoce que los organismos anaerobios no pueden emplear esta estrategia de reparación contra estos daños del tipo alquilante, y que estos deberían usar procesos que una enzima tipo-polimerasa, una reparación por escisión de bases o la remoción directa de 1meA o 3meC por mecanismos desconocidos. No se han detectado homólogos en el dominio celular Archaea (Sedgwick *et al.*, 2007)

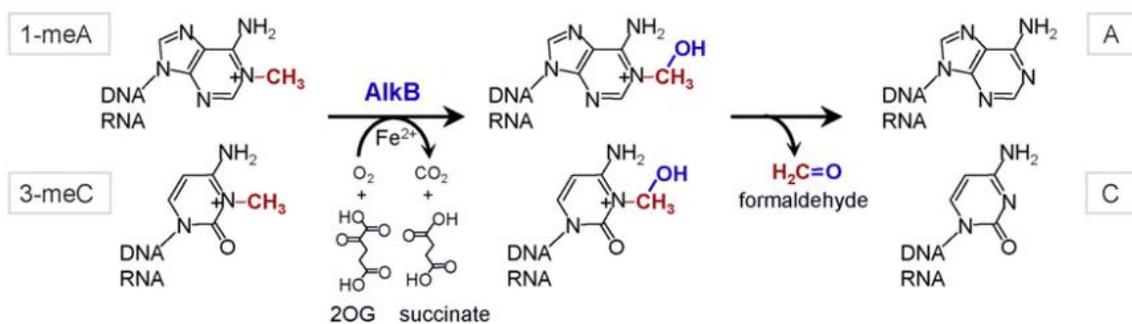


Figura 9. De lado izquierdo se ilustra el anillo de nitrógeno de la adenina con un grupo alquilo, metilo formando la 1-metiladenina (reacción superior) y la citosina con un grupo sustituyente metilo formando 3-meC (reacción inferior). Es ilustrada la dependencia de O₂, Fe²⁺ y 2OG como co-sustrato para después hidroxilar el grupo metilo en ambas bases modificadas. La base es restaurada y se libera CO₂ y el 2OG cambia a succinato y se forma formaldehído. Tomado de Vågbo *et al.* (2013)

Aunque la especificidad de AlkB no es absoluta, el mejor sustrato para esta enzima es DNA de una sola hebra que contenga 1meA y 3meC. El trinucleótido d(1-meA-T) es también un buen sustrato. AlkB no requiere un poli-desoxirribonucleotido para tener actividad. Puede actuar en nucleósidos metilados y en 3'-mononucleótidos, aunque con una eficiencia pobre y mucho menos en un 5'-monucleótido. Otros sustratos, como productos de alquilaciones de DNA son 3-metil timina (3meT) y 1-metilguanina (1meG) también pueden ser reparados pero la de-metilación es menos eficiente que sus sustratos principales 1meA y 3meC.

Al igual que AlkB en *E. coli*, los homólogos en humanos actúan en el mismo sustrato 1meA y 3meC, pero pueden actuar en DNA de una doble hebra y en DNA de una sola hebra. ABH2 tiene una afinidad mayor por una doble hebra de DNA y, al contrario, como AlkB y ABH3 tienen una afinidad definida a una sola hebra de DNA. ABH2 repara con mayor eficiencia 1meA y con una menor eficiencia 3meC, y al contrario ABH3 repara con mayor eficiencia 3meC y no a 1meA, hay evidencia de esta relación en homólogos de AlkB del ratón (Sedgwick *et al.*, 2007:2004)

Los análisis *in vitro* en los que se evaluaban las constantes de Michaelis-Menten de una serie de bases alquiladas muestran una alta actividad de reparación de N1- metiladenina (m1A) y N3-metilcitosina (m3C), comparado con la baja actividad en reparar 1,N⁶-etenoadenina (Frick *et al.*, 2007; Mishina *et al.*, 2005) y con una actividad nula en N1-metilguanina o N3-metil timina (Figura 10) Bomina *et al.* (2009) analizaron y observaron dos conformaciones globales discretas de AlkB donde la diferencia radica en la capacidad de difusión de O₂ mediante un "túnel" que controla la entrada al sitio activo. Las interacciones estéricas entre una serie de loops son las responsables de este control de entrada al sitio activo, lo que sugiere un mecanismo plausible para prevenir el anclaje de O₂ antes de la unión del nucleótido.

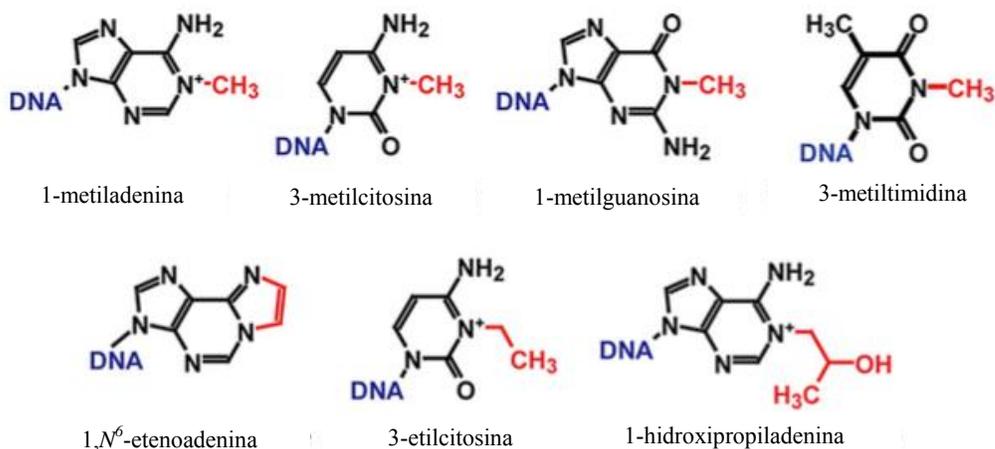


Figura 10. Estructuras de los sustratos de AlkB. Las lesiones que repara AlkB están en color rojo. Tomado de Bomina *et al.* (2009)

2.6 Estructuras cristalográficas de AlkB y sus homólogos

Por interés de este trabajo, se buscó en la literatura trabajos de cristalografía reportados para AlkB y sus homólogos. Existen complejos de enzima-sustrato de AlkB con estructura terciaria resuelta obtenidos por Sedgwick *et al.* (2007) que han ayudado a revelar detalles del proceso de desmetilación del DNA y que también han permitido concluir que los 11 residuos de aminoácidos ubicados en la parte N-terminal no son requeridos para la actividad enzimática. Para llevar a cabo este trabajo se realizó una búsqueda en RSCB PDB y en PDB-EMBL-EBI y se encontraron cristales de AlkB de una resolución $< 2.6 \text{ \AA}$. En estos datos están representados los dos dominios celulares, Eukarya y Bacteria, en donde se ha reportado la presencia de AlkB.

3. Objetivos

3.1 Objetivo general

Analizar desde el punto de vista evolutivo el sistema de reparación de polinucleótidos dependiente de oxígeno *AlkB* y sus homólogos

3.2 Objetivos particulares

- I. Formar una base de datos con las secuencias propuestas por Aravind & Koonin (2001) y contrastar el análisis con secuencias de proteínas y estructuras terciarias
- II. Realizar una filogenia de la familia de las 2OG-Fe(II) oxigenasas.
- III. Tratar de establecer un fechamiento relativo del surgimiento de los homólogos de AlkB
- IV. Discutir el significado biológico y la historia evolutiva de AlkB y sus homólogos

4. Metodología (Figura 10)

4.1 Análisis con secuencias de residuos de aminoácidos

- I. Se coleccionaron las secuencias homólogas de AlkB de acuerdo a Aravind & Koonin (2001), mediante BLAST (*blastp*) con un e-value $\leq 1 \times 10^{-6}$ en los *Reference Proteomes* y se tomaron sólo las secuencias de *Reviewed(Swiss-Prot)* de la base de datos *UniProt*.
- II. Se curó a mano y se recortaron las secuencias de las proteínas multidominio revisando la correspondencia del dominio 2OG Fe(II) en la base de datos *pfam* mediante *hmmscan*, para obtener las secuencias que optimizaran el alineamiento local para cada grupo de homólogos.
- III. Se realizó un alineamiento múltiple *de estructura secundaria* en línea en el servidor PROMALS 3D (Pei y Grishin., 2007) y se envió al servidor ProtTest (http://darwin.uvigo.es/software/prottest2_server.html) para calcular su modelo evolutivo.
- IV. Se obtuvo la filogenia más verosimil mediante máxima verosimilitud con *PhyML 3.1* con el modelo de evolución de proteínas WAG y una prueba estadística de remuestreo de 500 Bootstrapping (Figura 11)

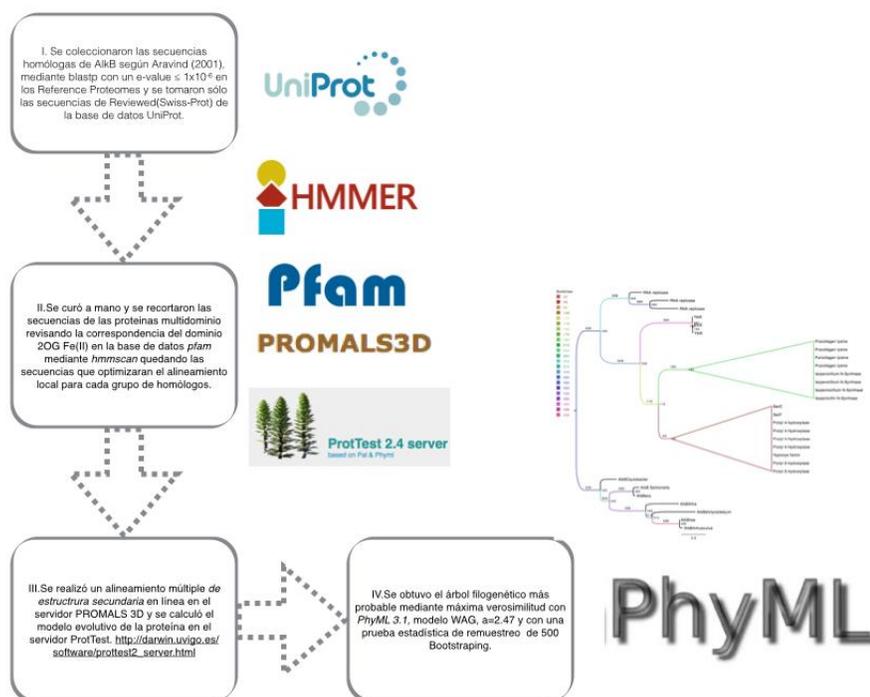


Figura 10. Metodología

4.2 Análisis evolutivo de AlkB y sus homólogos a partir de sus estructuras terciarias

- I. Mediante una búsqueda en la base de datos de cristales de proteínas RCSB-PDB, se coleccionaron los cristales disponibles para AlkB y sus homólogos.
- II. Se realizaron alineamientos estructurales pareados usando el servidor FATCAT-RIGID pairwise alignment (<http://fatcat.sanfordburnham.org/fatcat-cgi/cgi/fatcat.pl?-func=pairwise>) para evaluar el alineamiento estructural del dominio 2OG-Fe(II)_Oxy y de los residuos catalíticos siguiendo una metodología similar a Jácome *et al.*, 2015 y Salazar, A., 2015
- III. Se coleccionaron imágenes de cada alineamiento estructural

5. Resultados

5.1 Del análisis de las secuencias de aminoácidos

Algunas secuencias utilizadas por Aravind y Koonin (2001) para describir los homólogos de AlkB mediante PSI-BLAST, ya no están disponibles en el NCBI por lo que de la base de datos Uniprot, se extrajeron las secuencias ya no disponibles en el NCBI. Lanzando una secuencia por homólogo de AlkB, se obtuvieron más secuencias (al menos tres) mediante un BLAST e-value $\leq 1 \times 10^{-6}$.

La revisión y curación a mano de tres secuencias (por cada OTU), mostró la presencia en todas ellas del dominio 2OG-Fe(II)_Oxy bajo el modelo de *pfam*, sin embargo, hay dos homólogos de AlkB que tienen un dominio proteínico adicional. En la Tabla 2. están las proteínas con dominios *pfam* adicionales al dominio 2OG-Fe(II)_Oxy.

Si bien en la metodología se afirma que se tomaron sólo secuencias *Reviewed*, *Swiss-Prot*, las proteínas SanC, SanF y YbiX, no están aún revisadas en Uniprot (Diciembre 2015). A pesar de ello, se coleccionaron las secuencias. Una vez coleccionadas, se curaron a mano para evitar errores de anotación de la base de datos, obteniendo una proporción de tres secuencias por OTU para alinear las secuencias.

Tabla 2. Dominios adicionales a 2OG-Fe(II)_Oxy en homólogos de AlkB.

PROTEÍNA	Dominio(s) adicionales <i>pfam</i>
Isopenicillin N-Synthase	DIOX_N
RNA replicase	viral_helicase1/RdRP_2/Vmethyltransf

El alineamiento fue luego convertido a formato phylip 3 en http://www.ebi.ac.uk/Tools/sfc/emboss_seqret/ para construir una filogenia con la paquetería PhyML 3.1, el modelo de evolución WAG para proteínas sugerido por ProtTest, a-2.47 y con una prueba de 500 bootstrap (Figura 11)

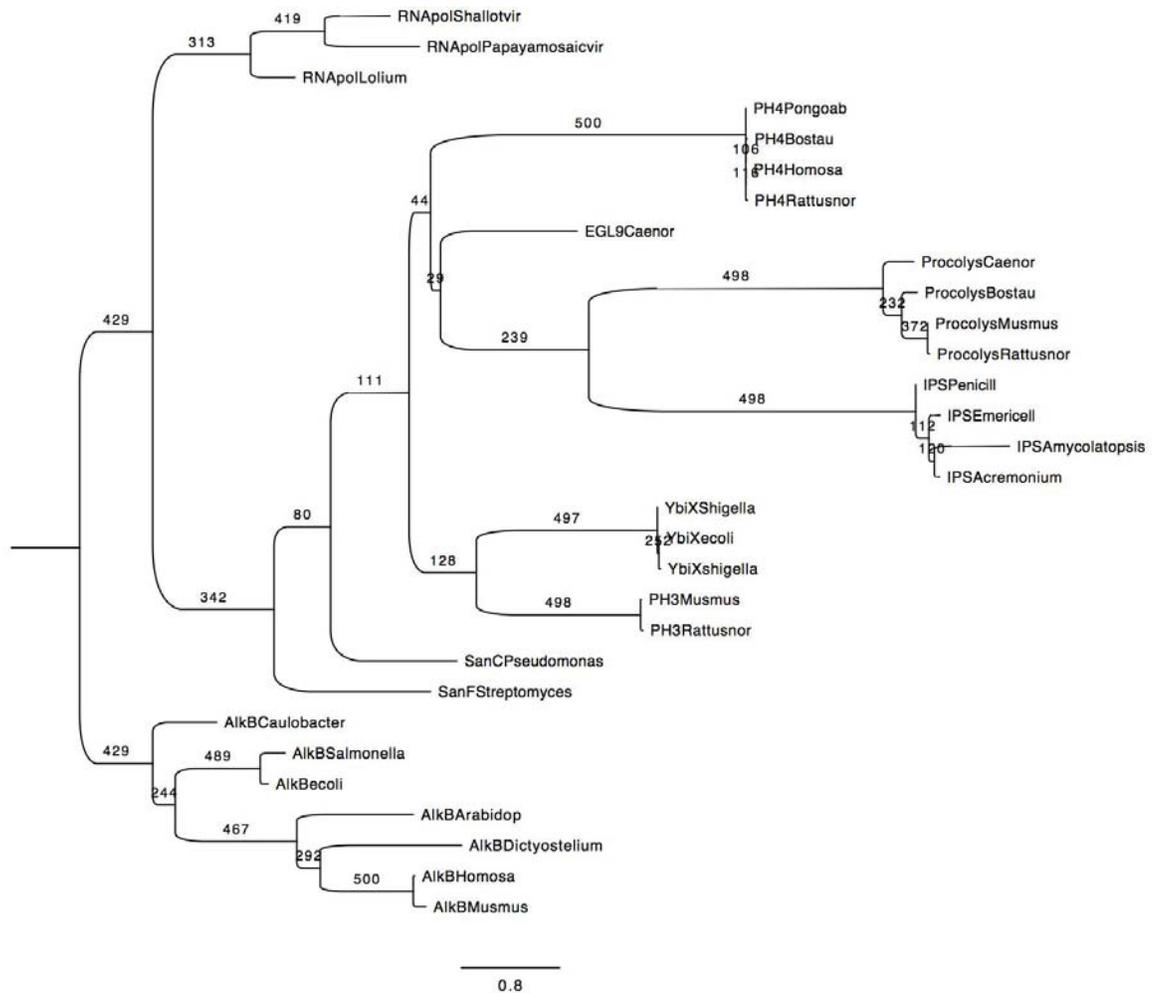


Figura 11. Fenograma de la familia “2OG-Fe(II) oxygenase superfamily” construido con PhyML 3.1,-WAG-, a=2.47, 500 Bootraps.

Al enviar las secuencias de los homólogos de AlkB ya curadas al servidor PROMALS 3D para predecir su estructura secundaria, el servidor PROMALS 3D realizó un PSI-BLAST, iterando tres veces la búsqueda de homólogos e incluyendo nuevos homólogos remotos con un corte de 0.001 y el parámetro para alinear PROMALS, más preciso pero más tardado que MAFFT. El archivo de salida *clustal* fue analizado con JalView Version 2 y editado con EasyDraw . Destacan los residuos H, D y H propios del dominio 2OG-Fe(II)_Oxy (ver Figura 12).

5.2 Del análisis con estructuras terciarias

En la base de datos RCSB-PDB se buscaron cristales tanto de AlkB y sus homólogos. No todos los homólogos están cristalizados hasta la fecha de este trabajo, aun así se coleccionaron los cristales disponibles que a continuación se enlistan tomando en cuenta el organismo de donde fue extraída la proteína, su función enzimática y la resolución del cristal.

Tabla 3. Cristales de AlkB y sus homólogos coleccionados y analizados. Tomada de RCSB-PDB (coleccionado en diciembre 2015 y traducido al español)

PDB ID	Organismo	Función	Resolución
3KHB	<i>Escherichia coli</i>	OXIDOREDUCTASA, AlkB EC 1.14.11.33 DNA demetilasa oxidativa DNA-base-CH ₃ + 2-oxoglutarato + O ₂ = DNA-base + formaldehído + succinato + CO ₂	2.9 Å
1IPS	<i>Aspergillus nidulans</i>	OXIDOREDUCTASA Isopenicilin N-Sintasa EC 1.21.3.1 N-[(5S)-5-amino-5-carboxipentanoil]-L-cysteinil-D-valina + O ₂ = isopenicilin N + 2 H ₂ O	2.5 Å
4QKD	<i>Homo sapiens</i>	OXIDOREDUCTASA, ALKBH7 EC 1.14.11 ¿? Con 2-oxoglutarato como un donador e incorporación de un átomo de oxígeno dentro. Necrosis mitocondrial.	1.35Å

4NJ4	<i>Homo sapiens</i>	<p>OXIDOREDUCTASA</p> <p>ALKH5</p> <p>EC 1.14.11</p> <p>¿?</p> <p>Con 2-oxoglutarato como un donador e incorporación de un átomo de oxígeno dentro RNA demetilasa</p>	2.02 Å
4MG2	<i>Homo sapiens</i>	<p>OXIDOREDUCTASA</p> <p>ALKBH2</p> <p>DNA demetilasa oxidativa</p> <p>EC 1.14.11.33</p> <p>DNA-base-CH₃ + 2-oxoglutarat + O₂ = DNA-base + formaldehído + succinato + CO₂</p>	2.3Å
2IUW	<i>Escherichia coli</i>	<p>OXIDOREDUCTASA</p> <p>ALKBH3</p> <p>EC 1.14.11</p> <p>Con 2-oxoglutarato como un donador e incorporación de un átomo de oxígeno dentro RNA/DNA enzima de reparación(2015)</p>	1.5 Å
2HBT	<i>Homo sapiens</i>	<p>OXIDOREDUCTASA</p> <p>Factor de hipoxia inducible-prolina dioxigenasa.</p> <p>EGLN-1</p> <p>EC 1.14.11.29</p> <p>Factor-hipoxia-inducible -L-prolina + 2-oxoglutarato + O₂ = factor hipoxia-inducible -<i>trans</i>-4-hidroxi-L-prolina + succinato + CO₂</p>	1.6 Å

6. Discusión

6.1 Implicaciones del estudio comparativo de la estructura primaria de *AlkB* y sus homólogos

Existen metodologías y modelos evolutivos adecuados para estudiar la evolución de una proteína a nivel de estructura primaria, calculando la divergencia en función de la probabilidad de cambio de un aminoácido por otro en distintos escenarios posibles. Sin embargo, las proteínas están sujetas a distintas presiones evolutivas y a eventos azarosos que podemos inferir con estos modelos, pero no podemos asegurar que su historia evolutiva haya sido así. Los cambios en secuencias de proteínas son menos probable en comparación con el DNA. Si bien, la evolución proteínica es de naturaleza distinta a la evolución del DNA o RNA, la funcionalidad total de un sistema o familia de proteínas puede verse afectado mediante procesos de mutación en el material genético.

La determinación de la tasa evolutiva de una proteína es hoy la herramienta general para cuantificar la importancia de la selección natural y la deriva génica, así como para identificar fuerzas selectivas en los datos genómicos. El modelo de sustitución más simple supone que todos los tipos de sustitución de aminoácidos son iguales y que estas sustituciones siguen una distribución de Poisson. Sin embargo, las sustituciones ocurren con mayor frecuencia entre aminoácidos con propiedades bioquímicas similares. Matrices empíricas, como BLOSUM y PAM, han analizado grandes colecciones de datos de proteínas mediante alineamientos múltiples y criterios estadísticos, para calcular la probabilidad de cambio de un aminoácido por otro; estas matrices pueden ser elegidas en función de la naturaleza de las secuencias que se estén estudiando (Pál *et al.*, 2006; Henikoff *et al.*, 1992)

¿Qué ocurre con aquellos datos biológicos que sugieren un origen temporal proteínico particular en donde no tenemos certeza si los modelos como las matrices de sustitución empíricas son capaces de reconstruir la historia de la proteína en cuestión?

¿Qué ocurre con aquellas proteínas que han tenido procesos divergentes y que una región conservada ya no está representada en estructura primaria?

Si una estructura primaria de una proteína ha divergido sustancialmente, ¿la homología está representada en gran parte por el plegamiento conservado en estructura terciaria y posiblemente en un mecanismo catalítico similar?

Evidentemente, todos estos modelos de evolución de proteínas tienen limitaciones particulares y ninguno contempla la historia evolutiva total o completa de un grupo de proteínas en particular. Este el caso del dominio 2OG-Fe(II)_Oxy, de interés en este estudio.

De las proteínas descritas por Aravind y Koonin (2001), la Isopenicilina N-Sintasa y la RNA replicasa viral, poseen en su arquitectura proteínica de dominios de acuerdo a la base de datos *pfam* más de un dominio y por lo tanto son proteínas multidominio. Al observar el apartado de resultados, se puede identificar el intervalo que corresponde a cada uno de los dominios en estas proteínas multidominio.

Como se ha descrito en la metodología, estas proteínas fueron lanzadas mediante *hmmscan*, para verificar el contenido del dominio “2OG-Fe(II) oxygenase_Oxy”. Al realizar un primer alineamiento múltiple por metodologías tradicionales como *Muscle*, *Clustal* y *Mafft*, los resultados eran caóticos y sin identidades observables, lo que incluso se reflejaba en los aminoácidos catalíticos como las dos histidinas y el ácido aspártico, que de acuerdo con la literatura son residuos de aminoácidos que coordinan el Fe II catalítico (Silvestrov *et al.*, 2014; Sundheim *et al.*, 2006). He ahí una limitante a los métodos tradicionales que son óptimos para proteínas con una única funcionalidad y un dominio *pfam*. El surgimiento de nuevas propiedades catalíticas y, más aún, la adición de un dominio parecen tener un papel, preponderante al considerar el estudio evolutivo de proteínas con la naturaleza del dominio en cuestión.

Debido a que el material que se ha analizado contiene tanto proteínas multidominio y proteínas monodominio, se alinearon con algoritmos tradicionales como *Muscle*, *Clustal* y *Mafft* para generar un fenograma que no dio resultados óptimos en la reconstrucción de la historia evolutiva de nuestro dominio de interés. Esto es fácil de comprender, ya que los dominios adicionales posteriores son agregados evolutivos que permiten a distinguir unas de otras proteínas, provocando agrupamientos sin un sentido evolutivo *per se*.

A lo anterior, cada una de las secuencias fue enviada a *pfam* para tomar el criterio de identificación de intervalo en la secuencia global de aminoácidos que envuelve a dominio 2OG-Fe(II) oxygenase_Oxy (Apéndice 9.2). Esto se curó a mano, conservando sólo el intervalo en la secuencia que envuelve al dominio en cuestión. Ello generó una mayor certidumbre durante la alineación múltiple, lo que permitió encontrar regiones conservadas y los aminoácidos catalíticos conservados.

Como puede observarse en los resultados, el tamaño de las secuencias conservadas de las proteínas multidominio tiene una longitud variable, en donde la RNA replicasa viral es la proteína más grande con tres dominios. La AlkB procarionte de *Escherichia coli*, posee un solo dominio, al igual que la AlkB eucarionte de *Homo sapiens* y de *Arabidopsis thaliana*, a pesar de ser homólogos funcionales del dominio 2OG-Fe(II)_Oxy. Al finalizar el corte de todas las secuencias bajo el algoritmo de *hmmscan*, el tamaño se redujo, y en promedio, el dominio 2OG-Fe(II)_Oxy resultante posee una longitud de 132 residuos aminoácidos. Como se puede ver, el alineamiento múltiple, a pesar de haber retirado los dominios adicionales en proteínas multidominio y estar bajo el modelo de dominios *hmmr-pfam*, el dominio 2OG-Fe(II)_Oxy no muestra grandes regiones conservadas (identidades). Sin embargo, se alinean las dos histidinas y el ácido aspártico que coordinan el Fe(II) en el plegamiento *jelly roll* característico del dominio 2OG-Fe(II)_Oxy (Silvestrov *et al.* 2014; Sundheim *et al.*, 2006)

El alineamiento múltiple permitió hacer una predicción de estructura secundaria usando PROMALS 3D para el dominio 2OG_Fe(II)_Oxy. Esto es consistente con el resultado de Yu *et al.*, (2006), lo que sugiere que la cristalización de los complejos catalíticos de AlkB asigna a las enzimas que contienen el dominio 2OG-Fe(II)_Oxy, en la región carboxi-terminal, un core de ocho hojas-beta formando un plegamiento “jellyroll”.

Según Sundheim *et al.*, (2006) al cristalizar el homólogo ABH3, la estructura revela que en el dominio 2OG-Fe(II)_Oxy, está directamente involucrado en la coordinación de un centro activo catalítico de Fe(II) que descansa en una regla His¹-X-Asp-X_n-His². Esto es totalmente coherente con los resultados y el alineamiento múltiple de las secuencias de los homólogos de AlkB. Desde el origen del dominio 2OG-Fe(II)_Oxy, posiblemente, diversas fuerzas evolutivas como la deriva génica, han conducido a la variabilidad de la secuencia del dominio, pero siempre manteniendo la regla His¹-X-Asp-X_n-His² (ver página 23 y resultados)

La filogenia (Figura 11 y 13) fue enraizada con AlkB de *Escherichia coli*, mostrando claramente como las enzimas AlkB procariontes se agrupan junto a las AlkBH humanas y de otros eucariontes

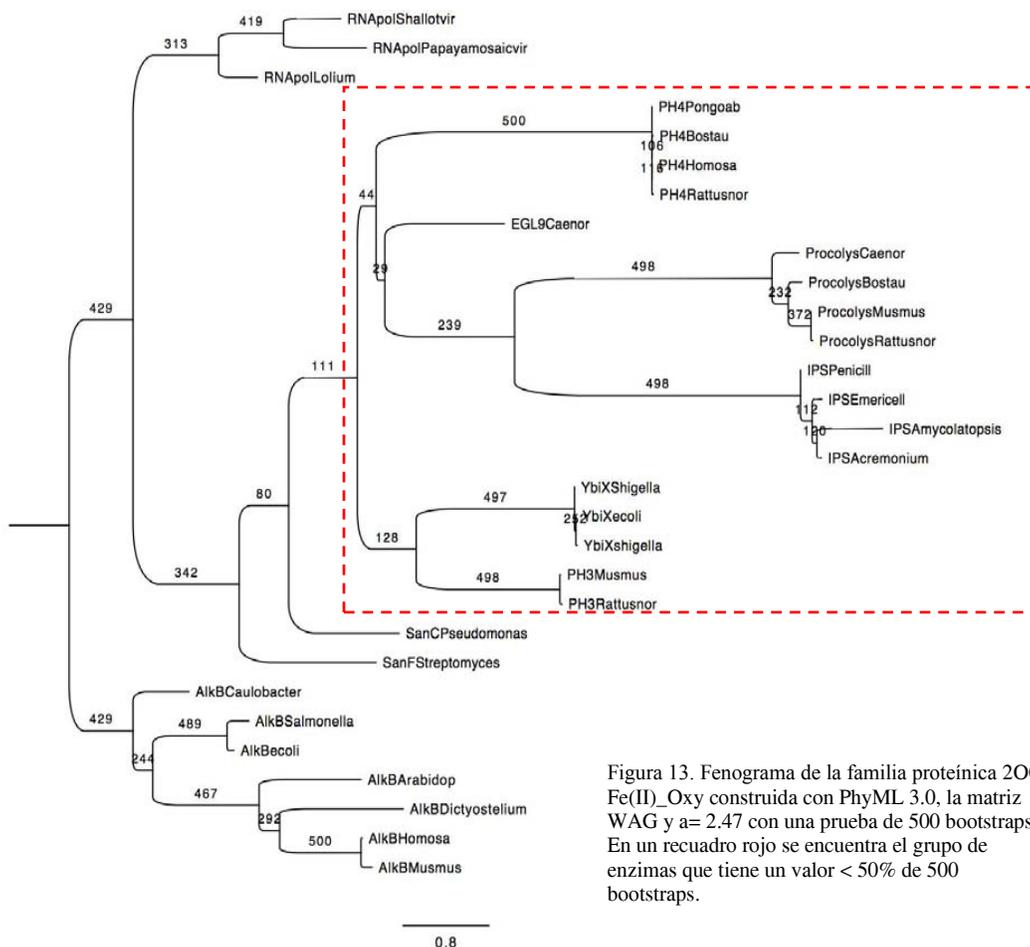


Figura 13. Fenograma de la familia proteínica 2OG-Fe(II)_Oxy construida con PhyML 3.0, la matriz WAG y $\alpha = 2.47$ con una prueba de 500 bootstraps. En un recuadro rojo se encuentra el grupo de enzimas que tiene un valor < 50% de 500 bootstraps.

Esta agrupación implica una posible ancestría común de estas enzimas reparadoras procariontes y una posterior transferencia horizontal de esta enzima a los eucariontes y posteriores divergencias en enzimas homólogas de AlkB como las AlkBH en humanos. Ougland *et al.*, (2015), hicieron una revisión de las funciones de los homólogos de AlkB, estas funciones no son las mismas, y abarcan un espectro funcional que involucra problemas de salud como la obesidad, la sensibilidad a la inflamación, la infertilidad, el crecimiento retardado y múltiples deformaciones en humanos. La ramas contenidas dentro de un recuadro rojo (Figura 13), tienen un valor <50% de bootstraps (500, - 250), por lo que no tenemos certidumbre de la agrupación de estas dos ramas y no podemos con esta metodología proponer relaciones evolutivas con un soporte estadístico robusto de al menos 50% de los bootstraps salvo la ancestría común del dominio contenido en todas las enzimas pertenecientes al grupo de la enzima AlkB.

6.2 Implicaciones del uso de estructura terciaria para analizar la homología del dominio 2OG-Fe(II)_Oxy

Como se ha discutido en el punto anterior, se ha sugerido que el centro catalítico de las enzimas que contienen el dominio 2OG-Fe(II)_Oxy, es una plegamiento *jelly roll* de ocho hojas-beta. Diversos trabajos, han analizado no sólo a la AlkB procarionte, sino que también han cristalizado homólogos, principalmente humanos, cuya función en la mayoría de los casos no ha sido analizada. De hecho Sedwick (2004), ha hecho una revisión muy amplia de las propiedades catalíticas de las enzimas AlkB y sus homólogos, y propone un alineamiento múltiple donde no muestra las ocho hojas-beta que Yu *et al.* (2006) proponen. En contraste, el trabajo de Wang *et al.*, (2014), para los homólogos de AlkBH7 (una enzima descrita en humanos), propone hasta diez hojas-beta en el plegamiento.

El predictor de estructura secundaria de PROMALS 3D discrimina visualmente sólo seis hojas-beta. Contrastando con la bibliografía, es posible que la deficiencia se deba a que el predictor no posee la sensibilidad necesaria para calcular con precisión los plegamientos reales o bien, que en algún intervalo de predicción, no logra discriminar visualmente dos posibles hojas-beta contiguas.

La cristalización de proteínas nos da un acercamiento más preciso a los plegamientos reales de la misma proteína. Los homólogos de AlkB han divergido de tal manera que ha sido necesario apoyarse en predicción de estructura secundaria y cristales de los mismos para proponer un origen en común del dominio 2OG-Fe(II)_Oxy. Una forma de confirmar la conservación y, por lo tanto, robustecer la propuesta de homología de un dominio en cuestión, es la comparación pareada de los cristales disponibles bajo el supuesto de que la divergencia actúa en un sentido que la mutación modifica la secuencia de aminoácidos. Sin embargo, las propiedades físico-químicas de los aminoácidos equivalentes o similares, son seleccionadas dando como resultado una conservación del plegamiento-estructura espacial.

El análisis con estructura terciaria (ver Apéndice de resultados 9.2 y *Análisis con estructuras terciarias 5.2*), ha confirmado la hipótesis de la homología del dominio 2OG-Fe(II). Al observar las gráficas

del alineamiento espacial, sin considerar el alineamiento en secuencias, hay una clara correlación lineal en todos los alineamientos estructurales. Si bien la correlación lineal es interrumpida por gaps, hay una relación clara. Un instrumento estadístico del servidor FATCAT (p-value) fue usado para evaluar la significación (que no haya sido un resultado debido al azar) de una similitud estructural. Los valores < 0.05 son significativamente similares. En consecuencia, al realizar los alineamientos pareados de cada una de las estructuras de cristales de homólogos de AlkB, bajo el parámetro estadístico de FATCAT, todos los alineamientos pareados de los cristales resultaron significativamente similares. Podemos pensar, pues, que la disposición espacial de las proteínas homólogas a AlkB propuestas por Aravind y Koonin (2001) mediante una iteración PSI-BLAST, mantienen una conservación también estructural. En el Apéndice de resultados (9.2), quedan perfectamente espacialmente alineados los residuos de la regla His¹-X-Asp-X_n-His² y se ven marcados en un cuadro.

6.3 Implicaciones biológicas y evolutivas del dominio 2OG-Fe(II)_Oxy

La dependencia funcional del dominio 2OG-Fe(II)_Oxy respecto al oxígeno molecular O₂ tiene implicaciones temporales en el origen del mismo. El Gran Evento Oxidativo (GOE), marca el inicio de uno de los cambios más severos y significativos en la historia geológica y biológica en la Tierra. Es imposible suponer que AlkB jugó un papel en el mantenimiento de la integridad del material genético DNA/RNA antes del Gran Evento Oxidativo, hace aproximadamente 2,500 millones de años, o en la etapa 2 de Holland (2006). Al ser un periodo de acumulación de oxígeno molecular en la atmósfera y en los océanos gradual y heterogéneo, el origen del dominio tendría que tener lugar en un periodo de acumulación homogénea en la atmósfera o mares primitivos o bien, surgir en un entorno con presión parcial de O₂ suficiente para llevar a cabo la función enzimática.

Como lo muestra la Figura 13, el origen del dominio debe haber ocurrido luego del Gran Evento Oxidativo. Después de surgir como un dominio que coordina la catálisis mediante la inclusión de 2-oxoglutarato, que se puede acumular por la desanimación de ácido glutámico o la descarboxilación del isocitrato en el ciclo de Krebs (Meléndez-Hevia *et al.*, 1996)- al igual que la inclusión de O₂ y la coordinación de Fe(II), el reconocimiento del esqueleto polianiónico del DNA o RNA debió estar fuertemente ligado a la evolución de la reacción enzimática.

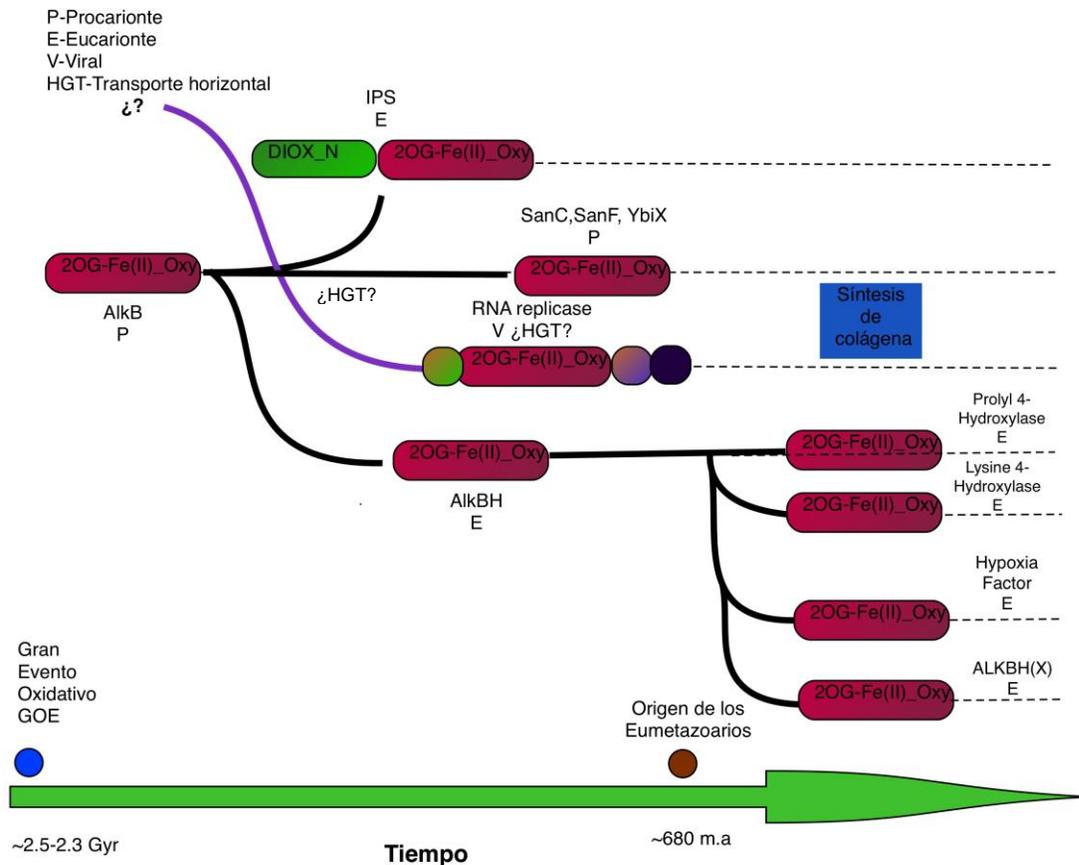


Figura 13. Representación gráfica de la posible evolución del dominio 2OG-Fe(II)_Oxy. El origen del dominio 2OG-Fe(II)_Oxy sólo pudo ocurrir después del Gran Evento Oxidativo (hace 2.5-2.3 Gyr). Por una posible transferencia horizontal se estableció en eucariontes y tuvo eventos divergentes. El origen de los eumetazoarios pudo tener lugar al existir una enzima que hidroxilara a la colágena. Figura realizada con EasyDraw para este trabajo.

Westbye *et al* (2008), han sugerido que el homólogo AlkBH1 es una proteína mitocondrial que principalmente demetila DNA y RNA. Las primeras aproximaciones mediante un BLASTP en este trabajo y la información que Westbye *et al.*, 2008 sugieren que el origen de los demás homólogos AlkBH es mitocondrial. La Figura 11 nos puede estar indicando una posible transferencia horizontal de las bacterias hacia los eucariontes pero excluyendo a las arqueas en este evento pues este grupo carece de homólogos de AlkB.

Al menos para AlkB de *Escherichia coli*, se tiene claro cuáles son los residuos que tienen interacción con el esqueleto de DNA. Se ha propuesto una cavidad exclusiva para la difusión del O₂. La coordinación del Fe por los residuos His, Asp, His no es exclusiva de este ión metálico. Los remplazamientos experimentales con Co(II) o Mn(II) han producido una estructura similar. Se ha sugerido que AlkB posee un diseño molecular “elegante” que es flexible gracias a una estructura segmentada lo que reconoce esqueletos invariantes de ácidos nucleicos, y que produjo gran afinidad y habilidad a gran diversidad de bases modificadas en su núcleo catalítico. Las adaptaciones estructurales contribuyen a la promiscuidad de la enzima en la reparación para mantener la eficiencia catalítica para diversos sustratos. (Yu *et al.*, 2006; Yu y Hunt., 2009).

El reconocimiento del DNA por parte de AlkB ocurre en el complejo AlkB-dsDNA muestra que las bases que flanquean una lesión como 1-meAdenina, son “exprimidas”, de tal manera que se apila una sobre otra a una distancia de 3.4 Å. Posteriormente, un azúcar se invierte y es forzada por los residuos 51-52 (Thr-Pro-Gly) formando finalmente un puente de hidrógeno con el fosfato despejado (Yang *et al.*, 2008).

Probablemente, la promiscuidad del sustrato es la responsable del éxito de esta enzima y la propagación de su dominio en diversas funciones monodominio. SanC y SanF son genes que se han descrito en *Streptomyces ansochromogenes* y están involucrados en la síntesis de nicomicina, un antifúngico que inhibe la síntesis de quinina (Xu y Tan., 2003; Chen *et al.*, 2000). En otro sentido, YbiX fue descrito en una comparación genómica entre *Escherichia coli* K-12 y *Shigella flexneri* donde se analizan ideas en torno a la patogenicidad (Jin *et al.*, 2002). La función de *novo* de Isopenicilina N-Sintasa, es otro claro ejemplo del éxito evolutivo del dominio 2OG-Fe(II)_Oxy. La adición posterior del dominio *pfam* DIOX_N tiene una significancia en una de las fases de síntesis de penicilina (Figura 14)

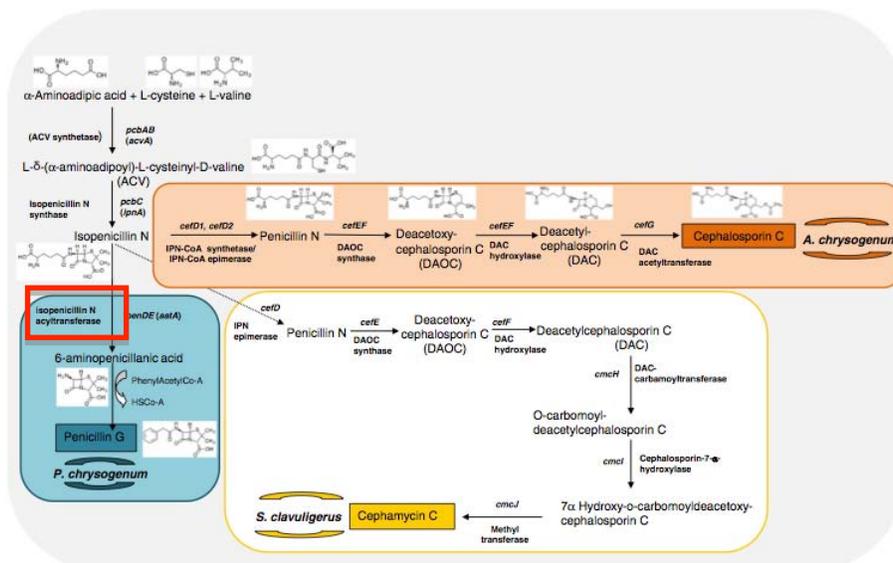


Figura 14. Tomado y modificado de Ozcengiz y Demain (2013) Biosíntesis de penicilina G por *Penicillium chrysogenum*, *Acremonium chrysogenum* y *Streptomyces clavuligerus* respectivamente. Se señala en un cuadro rojo a Isopenicillin N-Synthase.

La reacción catalizada por la isopenicilina N-sintasa, ($N-((5S)-5\text{-amino-5-carboxipentanoil})\text{-L-cisteinil-D-valina} + O_2 = \text{isopenicilina N} + 2 H_2O$), igualmente es dependiente de O_2 lo que restringe también el origen posterior al Gran Evento Oxidativo y en una temporalidad al menos posterior al Cámbrico, periodo geológico donde se ha propuesto el comienzo de la colonización terrestre de los hongos (Brundrett *et al.*, 2002; James *et al.*, 2006; Schoch *et al.*, 2009).

Además de la síntesis de la isopenicilina, existe otro proceso de catálisis donde un homólogo de AlkB juega un papel crítico en el tejido de los animales, la biosíntesis de colágena. La colágena es el componente principal del espacio extracelular en tejidos conectivos propios de animales y probablemente estaba presente en el ancestro común de los metazoarios (Urmetazoa) que ha sido propuesto como un organismo similar al phylum Porifera. Durante los últimos años, cDNA's que codifican a proteínas involucradas en interacciones célula-célula, célula tejido han sido clonados de esponjas, lo que ha permitido la identificación de elementos de la matriz extracelular. Estos datos confirman que todos los metazoarios probablemente descienden de un ancestro común.

En las esponjas, el espacio (mesohilo) entre la capa externa (picanodermo) y la capa interna (coanodermo) no es una mezcla homogénea sino que está compuesta de colágena, moléculas tipo-fibronectina y dermatopontina (Müller., 2003). La existencia de moléculas de adhesión probablemente permitió el surgimiento de un organismo colonial (Figura 15)

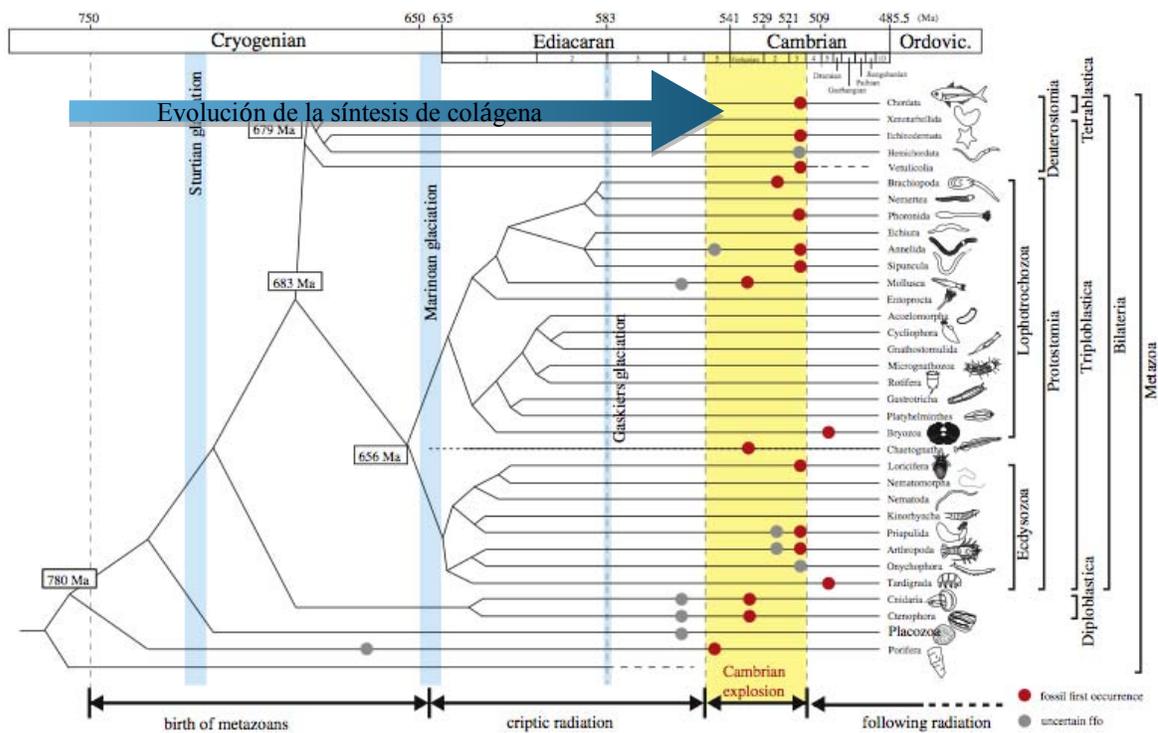


Figura 15. Modificado de Shu *et al.* 2014. Filogenia y registro fósil.

Los primeros fósiles de poríferos y de varios grupos de eumetazoarios no se encuentran en el registro fósil sino hasta el inicio del Cámbrico, cuando tuvo lugar la llamada “Explosión del Cámbrico” hace alrededor de 541 millones de años (Shu *et al.*, 2014). Es posible que el surgimiento de la síntesis de colágena en cuya biosíntesis participan la prolyl y la lysil 4(3)-hidroxilasas, tuvo sus inicios evolutivos incluso mucho antes de la explosión del Cámbrico, hace 583 millones de años pero el registro fósil es incierto para la fauna de Ediacara.

En otro sentido, la propuesta de Logan y Poole (2005) involucran una premisa en la evolución temprana de la reparación de RNA y la presencia directa de un sistema de dioxigenasas en el Last Universal Common Ancestor (LUCA). “...primero, un compendio de argumentos han tomado lugar en la reparación de RNA como una característica de LUCA...”, “... Bajo este escenario, la reparación genómica de RNA debe ser considerada como un paso en la transición RNA a DNA...” Poole y Logan, sin embargo, han ignorado la ausencia de O₂ en etapas evolutivas cercanas a LUCA, lo que hace imposible la existencia de un sistema tipo AlkB dependiente de O₂ para ese momento. La edición de DNA pudo haber surgido incluso antes de un sistema de reparación directa de DNA como AlkB, ya que no hay una dependencia de O₂ para su funcionamiento (Figura 16)

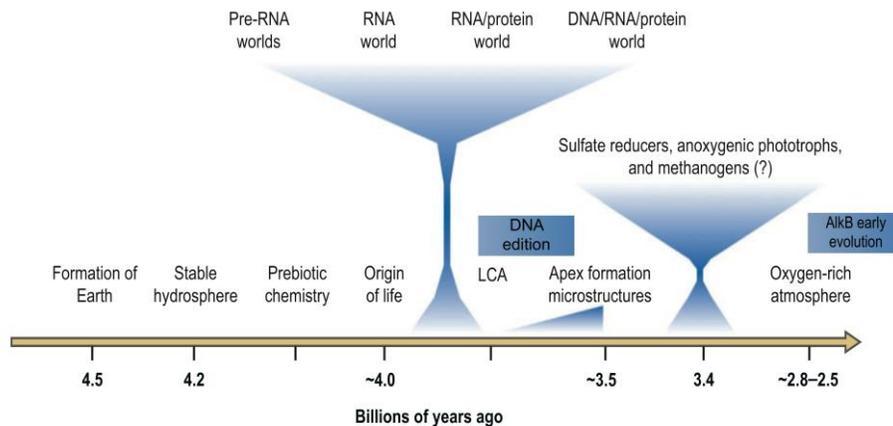


Imagen 16. Modificado de Becerra *et al.* 2007 LCA (LUCA), la edición de DNA y la evolución de AlkB.

Es evidente que el oxígeno abrió grandes posibilidades de utilizar su potencial electronegativo como auxiliar en las reacciones enzimáticas, no sólo de forma directa de este elemento como ocurre en AlkB y sus homólogos, sino también indirectamente como producto blanco en el acarreamiento del elemento por hemoglobinas o hemeritrinas.

7. Conclusiones

1. La gran divergencia en secuencia de aminoácidos plantea un reto en el análisis e identificación de entidades homólogas. Metodologías como la predicción de estructura secundaria, y alineamientos estructurales, pueden abrir una nueva alternativa de análisis donde la premisa es la conservación estructural-espacial.
2. El dominio 2OG-Fe(II) debió surgir después del Gran Evento Oxidativo, cuando las presiones parciales de O₂ permitieran la catálisis enzimática de las proteínas con el dominio 2OG-Fe(II)_Oxy.
3. El surgimiento de la penicilina también está restringido temporalmente por la disposición del O₂ libre, y después de la aparición del ancestro común de hongos y animales, al igual que todos los homólogos de AlkB y los homólogos del dominio 2OG-Fe(II)_Oxy.
4. La síntesis de colágena es un evento medular en el origen de los metazoarios. El que las enzimas responsables de la hidroxilación de los residuos de prolina y lisina contengan el dominio 2OG-Fe(II)_oxy ayuda a comprender porqué los animales surgieron hace no más de 541 millones de años.
5. La ausencia de homólogos del dominio 2OG-Fe(II)_Oxy en el dominio Arquea nos plantea una o varias barreras ecológicas durante la historia evolutiva del dominio en cuestión. Esto debilita la hipótesis de Poole y Logan (2005) ya que si LCA hubiera tenido la capacidad de reparar material genético dependiente de oxígeno debería entonces haber homólogos en el dominio celular Archa.
6. Los sistemas de edición de DNA, al no depender de oxígeno molecular, pueden haberse originado incluso antes del Gran Evento Oxidativo y ser más antigua la edición de DNA que la reparación directa del DNA por AlkB
7. La diversidad de reacciones enzimáticas en donde el oxígeno está involucrado nos demuestra el fascinante aprovechamiento evolutivo de una presión evolutiva tan intensa después del Gran Evento Oxidativo.

8. Bibliografía

- Aas, *et al.* **Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA.** *Nature*, Vol. 40, 2003, p. 859-863
- Aravind, L & Koonin, E. **The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases.** *Genome Biology* 2001
- Aguirre, Jesús *et al.* **Reactive oxygen species and development in microbial *eukaryotes*** *Trends in Microbiology*, Volume 13, Issue 3, p. 111 - 118
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* Vol. 25, 1997
- Becerra, A. *et al.* **The Very Early Stages of Biological Evolution and the Nature of the Last Common Ancestor of the Three Major Cell Domains.** *Annu. Rev. Ecol. Evol. Syst.* 2007. Vol. 38 p. 361-79
- Berner, R. **The Phanerozoic carbon cycle: CO₂ and O₂:** *Oxford University Press*, 2004, p 41
- Brocks, J *et al.*, **Composition and syngeneity of molecular fossils from the 2.78 to 2.45 billion-year-old Mount Bruce Supergroup, Pilbara Craton, Western Australia,** *Geochimica et Cosmochimica Acta*, Vol. 67 N°22, 2003
- Bruner, S. *et al.* **Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA.** *Nature*, Vol. 403, 2000, p. 859-866
- Chen *et al.*, **Cloning, sequencing, and function of sanF: A gene involved in nikkomycin biosynthesis of *Streptomyces ansochromogenes*.** *Curr Microbiol*, 2000 Vol. 41 p. 312-6.
- Cleaves, H. *et al.* **A Reassessment of Prebiotic Organic Synthesis in Neutral Planetary Atmospheres** *Origins of Life and Evolution of Biospheres*, 2008, Volume 38, Issue 2, p. 105-115
- Cooke, M. *et al.* **Oxidative DNA damage: mechanisms, mutation, and disease.** *The FAESB Journal*, Vol. 17, 2003, p. 1195-1211
- Cunningham, R. **DNA glycosylases.** *Mutation Research*, Vol. 383, 1997, p. 189-196
- David, S *et al.* **Base-excision repair of oxidative DNA damage.** *Nature*, Vol. 447, 2007, p. 941-950
- Dexeimer, T. **DNA Repair Pathways and Mechanisms, DNA repair of cancer stem cells.** Cap. 2. 2013, *Springer Netherlands*, p. 19-29

- Drohat, A; Maiti, A. **DNA Glycosylases: Mechanisms.** *Reference Module in Biomedical Sciences, from Encyclopedia of Biological Chemistry*, 2013, p.16-20
- Dworkin, J *et al.* **The roads to and from the RNA world.** *Journal of Theoretical Biology*, Vol. 222, 2003, p. 127-134
- Freeland *et al.* **Do Proteins Predate DNA?** *Science*, Vol. 286, 1999, p 690-692
- Friedberg, E., **DNA repair.** *W.H Freeman & Company*, 1985 p.111.
- Frick LE, *et al.* **Alleviation of 1,N6-ethanoadenine genotoxicity by the Escherichia coli adaptive response protein AlkB.** *PNAS* 2007, Vol 104 p. 755–760.
- Henikoff S, Henikoff JG: **Amino acid substitution matrices from protein blocks.** *Proc Natl Acad Sci USA* 1992, Vol. 89: p.10915-10919.
- Holland, H. **The oxygenation of the atmosphere and oceans.** *Phil. Trans. R. Soc. B*, 2006, Vol. 361, p. 903–915
- Jacobs, A & Schär, P. **DNA glycosylases: in DNA repair and beyond.** *Chromosoma*, Vol. 121, 2012, 1-20
- Jácome *et al.* **Structural Analysis of Monomeric RNA-Dependent Polymerases: Evolutionary and Therapeutic Implications.** *PLoS ONE*, 2015, 10(9): e0139001.
- James *et al.* **Reconstructing the early evolution of Fungi using a six-gene phylogeny.** *Nature*, Vol 443, 2006, 818-882
- Jin *et al.* **Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157.** *Nucleic Acids Research*, 2002, Vol. 30 No. 20
- Lazcano, A. (2010) **Historical development of origins research.** *Cold Spring Harb Perspect Biol.* 2
- Lazcano *et al.* **The Evolutionary Transition from RNA to DNA in Early Cells.** *Journal of Molecular Evolution*, Vol. 27, 1988, 283-290
- Lazcano, A & Miller S. **The Origin and Early Evolution of Life: Prebiotic Chemistry, the Pre-RNA world, and Time.** *Cell*, Vol. 85, 1996, 793-798
- Logan; D Poole, A. **Modern mRNA Proofreading and Repair: Clues that the Last Universal Common Ancestor Possessed an RNA Genome?** *Mol Biol Evol.* 2005 No.6, p. 1444-55

- Lyons, T *et al.* **The rise of oxygen in Earth's early ocean and atmosphere.** *Nature*, Vol. 506, 2014, p. 307-314
- Meléndez-Hevia *et al.* **The Puzzle of the Krebs Citric Acid Cycle: Assembling the Pieces of Chemically Feasible Reactions, and Opportunism in the Design of Metabolic Pathways During Evolution.** *J Mol Evol*, 1996 Vol. 43, p. 293-303
- Mishina Y, *et al.* **Direct repair of the exo-cyclic DNA adduct 1,N⁶-ethenoadenine by the DNA repair AlkB proteins.** *J Am Chem Soc* 2005, Vol 127 p. 14594-14595.
- Miller S. L. (1953). **Production of Amino Acids Under Possible Primitive Earth Conditions.** *Science* Vol. 117 p. 528-529
- Müller, W. **The Origin of Metazoan Complexity: Porifera as Integrated Animals.** *Integr Comp Biol.*, 2003, Vol. 43 p. 3-10
- Nordlund, P., Reichard, P. **Ribonucleotide Reductases,** *Annurev.Biochem*, 2006, Vol.75; 700-701
- Brundett, M. **Coevolution of roots and mycorrhizas of land plants.** *New Phytologist*, 2002, Vol. 154, p. 275-304
- Neeley, W & Essigmann, J. **Mechanisms of Formation, Genotoxicity, and Mutation of Guanine Oxidation Products.** *Chemical Research in Toxicology*, 2006, Vol. 19:4, p.491-501
- Ougland, R. **Non-homologous funtionoes of the AlkB homologs.** *Journal of Molecular Cell Biology*.2015, 2-11
- Ozcengiz, G; Demain, A. **Recent advances in the biosynthesis of penicillins, cephalosporins and clavams and its regulation.** *Biotechnology Advances*, 2013, Vol. 31, p. 287-311
- Pál *et al.* **An integrated view of protein evolution,** *Nature Reviews Genetics*, 2006, Vol 7, p. 337-348.
- Pei J, Grishin NV: **PROMALS: towards accurate multiple sequence alignments of distantly related proteins.** *Bioinformatics* 2007 Vol., 23, p. 802-808.
- Salazar, A. **Evolución Molecular de la Biosíntesis de Histidina y Otros Compuestos Imidazólicos,** UNAM, 2015
- Sancar, A. **Moleular mechanisms of mammalian DNA repair and the DNA checkpoints.** *Annu. Rev. Biochem*, 2004 Vol, 73 p. 39-85

- Schoch *et al.* **The Ascomycota Tree of Life: A Phylum-wide Phylogeny Clarifies the Origin and Evolution of Fundamental Reproductive and Ecological Traits.** *Syst Biol.* 2009 Vol.58:2, p. 224–239
- Sedgwick, B. **Repairing DNA-methylation damage.** *Nature Reviews Molecular Cell Biology*, 2004, Vol. 5, p. 148-157
- Sedgwick *et al.* **Repair of alkylated DNA: Recent advances.** *DNA repair*, 2007, Vol. 6 p. 429–442
- Shu *et al.* **Birth and early evolution of metazoans.** *Gondwana Research*.2014, Vol. 25, p. 884-895
- Silvestrov *et al.* **Homology modeling, molecular dynamics, and site-directed mutagenesis study of AlkB human homolog 1 (ALKBH1).** *Journal of Molecular Graphics and Modelling*, 2014 Vol.54, p. 123-130.
- Sundheim, O *et al.* **Human ABH3 structure and key residues for oxidative demethylation to reverse DNA/RNA damage.** *EMBO*, 2006, Vol. 25, p. 3389-3397.
- Van der Veen, S y Tang C. **The BER necessities: the repair of DNA damage in human-adapted bacterial pathogens.** *Nature Microbiology*, Vol 13, 2015, 83-93
- Yang, C *et al.* **Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA.** *Nature*, 2008, Vol. 452, p. 961-966
- Yi y He, **DNA Repair by Reversal of DNA Damage**, 2013, *CSH*. p. 1-18
- Yu. *et al.* **Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB.** *Nature*, 2006, Vol.439, 2006 p. 879-884
- Yu B y Hunt J. **Enzymological and structural studies of the mechanism of promiscuous substrate recognition by the oxidative DNA repair enzyme AlkB** *PNAS*, Vol. 106:34, p. 14315–14320
- Wang *et al.* **The Atomic Resolution Structure of Human AlkB Homolog (ALKBH7), a Key Protein for Programmed Necrosis and Fat Metabolism.** *Journal of biological chemistry*, 2014, Vol.289: 40, p. 27924-27936
- Xu, J; Tan, H.. **sanC – a novel gene involved in nikkomycin biosynthesis in *Streptomyces ansiochromogenes*.** *Letters in Applied Microbiology*, Vol 36, p. 234-238.
- Vågbø *et al.* **Methylation damage to RNA induced in vivo in Escherichia coli is repaired by endogenous AlkB as part of the adaptive response.** *DNA Repair*, 2013, Vol. 12, p. 188–195

- Westbye *et al.* **Human AlkB Homolog 1 Is a Mitochondrial Protein That Demethylates 3-Methylcytosine in DNA and RNA.** *Journal of biological chemistry.* 2008, Vol. 283: 36, p. 25046–25056
- Wu *et al.* **The Protein Information Resource.** *Nucleic Acids Research.* 2003 Vol. 31:1, p. 345-347
- Wu CH, *et al.*: **The Universal Protein Resource (UniProt): an expanding universe of protein information.** *Nucleic Acids Res,* 2006, Vol.34 p.187-191.

9.Apéndice

9.1 Bases de datos consultadas

Pfam

Dirección web: <http://pfam.xfam.org/>

Los modelos de cadena de Markov (HMM en inglés), son usados para la inferencia estadística de homología, se construyen modelos representativos de secuencias curadas que definen a una familia de proteínas. Una semilla de alta calidad es esencial lo que provee las bases de las frecuencias posición-específica de aminoácidos, los gaps y las longitudes de los parámetros en el perfil de cadena de Markov. Las regiones de la secuencia obtienen un umbral curado para evitar falsos positivos (o cercanos al umbral) . Las entradas curadas están referenciadas como entradas Pfam-A. El perfil HMM es construido usando HMMER software suite (<http://hmmer.janelia.org>).

Pfam está disponible en los servidores del Reino Unido (<http://pfam.sanger.ac.uk/>) y en los Estados Unidos de Norteamérica (<http://pfam.janelia.org/>), es comúnmente usada como una base de datos de familias de proteínas conteniendo (2014) 14831 entradas curadas manualmente en la versión 27.0. Desde su última actualización (2 años antes de 2014), se han generado 1182 nuevas familias y se han mantenido la cobertura de secuencias de UniProtKnowledgebase (UniProtKM) cercano al 80% a pesar del incremento del 50% en secuencias de la mencionada base de datos. En contraste con versiones anteriores, ahora se proveen alineamientos basados en cuatro diferentes bases de datos de proteomas y una nueva interface de búsqueda de DNA. Recientemente, se han agregado estructuras 3D cristalográficas a Pfam.

Hmmer

El sitio web HMMER está disponible en <http://www.ebi.ac.uk/Tools/hmmer/>, provee acceso algoritmos de búsqueda de homólogos protéicos encontrados en el software HMMER. Desde su primera aparición en 2011, el repertorio de búsqueda ha sido expandido al incluir búsquedas iterativas con jackhmmer. El continuo crecimiento de las bases de datos representa que las tabulaciones tradicionales de la significancia de las secuencias hits pueden ser abrumadoras al usuario. En consecuencia, otras vías adicionales de búsqueda de homología han sido desarrolladas permitiendo observar la distribución taxonómica o la arquitectura de dominios. Las representaciones taxonómicas y de arquitectura de dominios pueden ser usadas en combinación con un filtrado de acuerdo a las necesidades del usuario. Las búsquedas pueden incluso estar restringidas antes con un filtro taxonómico, el cual no solamente asegura que los resultados son específicos al grupo taxonómico solicitado sino que también mejoran la búsqueda. El repertorio de perfiles de cadena de Markov que son usados para la anotación de las secuencias query con familias de proteínas y dominios han sido expandidas para incluir librerías de CATH-Gene-D, PIRSF, Superfamily y TIGRFAMs. El servidor está alojado en el Instituto Europeo de Bioinformática EMBL-EBI.

Última modificación del 28 agosto de 2015

El Recurso Universal de Proteína (UniProt), es un recurso comprensivo de secuencias de proteínas y sus anotaciones. Las bases de datos UniProt son: UniProt Knowledgebase (UniProtKB), Clusters de referencia UniProt (UniRef) y el Archivo UniProt (UniParc).

UniProt es el resultado de una colaboración entre el Instituto de Bioinformática Europeo (EMBL-EBI), el Instituto Suizo de Bioinformática (SIB) y el Recurso de Información de Proteínas (PIR). A través de tres instituciones y más de 100 personas, están involucradas en diferentes tareas como la curación de la base de datos, desarrollo de software y apoyo.

EMBL-EBI junto con SIB, han producido Swiss-Prot y TrEMBL mientras PIR ha producido la Base de Datos (PIR-PSD). Estos dos sets de datos coexisten con diferentes coberturas en secuencias de proteínas y prioridades de anotación. TrEMBL (Datos de Secuencias de Nucleótidos Traducidos EMBL) fue originalmente creada porque las secuencias generadas excedieron la capacidad de Swiss-Prot. Mientras tanto, PIR mantuvo el PIR-PSD y las bases de datos relacionadas, incluyendo iProClass, una base de datos de secuencias de proteínas y familias curadas. En 2002, tres institutos decidieron invertir y crear el consorcio UniProt.

El consorcio UniProt es liderado por Alex Bateman y Ioannis Cenarios, apoyados por el equipo clave y reciben valiosos datos por un consejo científico independiente.

9.2 Apéndice de resultados

1.-

AlkB *Escherichia coli* K12

Base de datos: UniProt

Organismo: *Escherichia coli*

Estructura primaria

>AlkB OS=Escherichia coli (strain K12)

```
MLDLFADAEPWQEPLAAGAVILRRFAFNAAEQLIRDINDVASQSPFRQMVTPGGYTMSVA
MTNCGHLGWTTHRQGYLYSPIDPQTNKPWPAMPQSFHNLCQRAATAAGYPDFQPDACLIN
RYAPGAKLSLHQDKDEPDLRAPIVSVSLGLPAIFQFGGLKRNDPLKRLLEHGDVVVWGG
ESRLFYHGIQPLKAGFHPLTIDCRYNLTFRQAGKKE
```

Sequence Matches and Features

Pfam [2OG-FeII_Oxy_2](#) 216

disorder coiled-coil tm & signal peptide

Pfam Matches [Standard](#)

	Family		Clan	Description	Start - End		Alignment		Model			Bit Score	Domain E-values	
	Id	Accession			Start	End	Start	End	Start	End	Length		Ind.	Cond.
>	2OG-FeII_Oxy_2	PF13532.3	CL0029	2OG-Fe(II) oxygenase superfamily	18	210	20	210	3	199	199	189.56	6.5e-56	4.0e-60

Your search took: 0.08 secs

2.-

Factor de hipoxia inducible, prolil-hidroxilasa, EGL9

Base de datos: Uniprot

Organismo: *Caenoradbitis elegans*

Estructura primaria

```
>sp|G5EBV0|EGL9_CAEE Hypoxia-inducible factor prolyl hydroxylase
OS=Caenorhabditis elegans GN=egl-9 PE=1 SV=1
MSSAPND DCEIDK GTPSTASLFTTLM LSQLPSSSTAVLQCTYCGSSCTSSQLQTCLFCGTV
AYCSKEHQQLDWLTHKMICKSLQTSGMVPSNLMPQAAPAVMAPI PPTVSFDDPALTTSL
LSLQNNPILNQ TISNFPPTFSITSKTEPEPSIPIQIPQRISSTSTVPFSSEGSFAFKPYRN
THVFN SISESMSSMCT SHEASLEHMSSASLAMFPTSSTAQSDISR LAQVLSLAGDSPAS
LALVTT SVPSTASTATIPPPATTTSSATSSGKSETITVGKEKI IQTDDPDIQIIETEGGS
KPTVSRTRKRPTPSNSADPKINYKDHKNVYVYSTTLQEHQKHLQNRGLALS IHQAMVLRL
RYIAEHVIRSLNEFGWAVVDNFLGSDHYKFTAKEIERLYERGLFSPGQLMEAKHKDEFHI
KDIRSDHIYWDGYDGRAKDAATVRL LISMIDSVIQHFKKRIDHDIGGRSRAMLAIYPGN
GTRYVKHVDNVPKDGRCITTIYYCNENWDMATDGGTLRLYPETSMT PMDIDPRADRLVFF
WSDRRNPHEVMPVFRHRFAITIWYMDKSERDKALAKGESDAACASKKENDPTSSSLNSL
IGSLLRPRKNPSTHDL SKLDLRLFPSTSSDPALVSAADED RVDISADFQSTSSLAHPEST
DSGVSLSTFNVAHNHMERTTSLQSI SDHFRSERSHERRSSTSSDQDLDEGLPPPSTNPE
YYI
```

Pfam: 2OG-Fe(II) oxygenase superfamily, Clan: Cupin (CL0029)

Sequence Matches and Features

Pfam  723

disorder coiled-coil tm & signal peptide

Family		Clan	Description	Start	End	Alignment		Model		Bit Score	Domain E-values		
Id	Accession					Start	End	Start	End		Length	Ind.	Cond.
> 2OG-FeII_Oxy_3	PF13640.2	CL0029	2OG-Fe(II) oxygenase superfamily	472	565	474	565	3	100	100	61.05	1.3e-16	1.6e-20
> zf-MYND	PF01753.14	CL0175	MYND finger	39	79	39	79	1	40	40	28.36	1.2e-06	1.5e-10

Your search took: 0.05 secs

3.-

Prolyl 3-hydroxylase 1

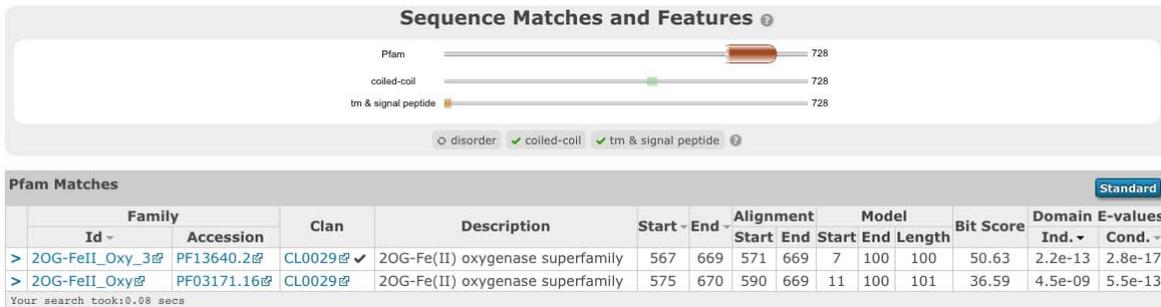
Base de datos: Uniprot

Organismo: *Rattus norvegicus*

Estructura primaria

```
>sp|Q9R1J8|P3H1_RAT Prolyl 3-hydroxylase 1 OS=Rattus norvegicus GN=P3h1
PE=1 SV=1 MVAVAAAAASRATAESEPEWNVVAAPDLLYAEGTAAYARGDWPVVLNMERALRSRAALRA
LRLRCRTRCATELPWAPDLDLGPASSLNHDPGAAALHDLRFFGALLRRAACLRRCLGPPS
AHLLESEELDLEFNKRSPYNYLQVAYFKINKLEKAVAAAHTFFVGNPEHMEMRQNLDDYYQT
MSGVKEEDFKDLEAKPHMHEFRLGVRLYSEEKPLEAVPHLEAALQEYFVADEECRALCEG
PYDYDGYNLYLDYSADLFQAITDHYVQVLSCKQNCVTELASHPSREKPFEDFLPSHYNYLQ
FAYYNIGNYTQAIECAKTYLLFFPNDEVMSQNLAYYTAVLGEEEAASSISPRENAQEYRHR
SLLEKELLFFAYDIFGIPFVDPDSWTPPEEVI PKRLQEKQKSERETAVRISQEI GNLMKEI
ETLVEEKTKEKSLDVSRLTREGGPLLYEGINLTMNSKVLNGSQRVVMDGVI SDDECQELQR
LTNAAATSGDGYRGQTS PHTPNEKFGYVTVL KALKLGQEGKVPLQSAHMYYNVTEKVRV
MESYFRLDTPLYFSYSHLVCRTAIEESQAERKDS SHPVHVDN CILNAESLVCIKEPPAYT
FRDYSAILYLNDFDGGNFYFTELD AKTVTAEVQPQC GRAVGFSSGTENPHGVKAVTRGQ
RCAIALWFTLDP RHSERDRVQADDLVKMLFSP EEVLDLPQEQLPDQQGSPKPGEEESLSDR
ESQPKDEL
```

Pfam: 2OG-Fe(II) oxygenase superfamily, Clan: Cupin (CL0029)



4.-

SanF

Base de datos: Uniprot

Organismo: *Streptomyces ansochromogenes*

Estructura primaria

```
>tr|Q9KIT9|Q9KIT9_9ACTN SanF OS=Streptomyces ansochromogenes GN=sanF PE=4  
SV=1 MVLTLDSALEERTQPFQLFRGRDLLDDGQLDELLATVPTADVAKIAVEDPKHEKQYRMNL  
VDLIVLEKEAPVFPRLPGVWRDLVEDLRGAFTAWLEKSTGIELAGLQRSIGLYTHRNGD  
YLSVHKDKPTKAITVILYLNRDWPVEAGGFQIFASPKEGPTEEISPVGGLLAFFPPTDK  
SWHAVSKIEHPEGTERITVQIEYWLTTGLAGSAYRPQN
```

Pfam: 2OG-Fe(II) oxygenase superfamily, Clan: Cupin (CL0029)

Sequence Matches and Features

Plam  218

disorder coiled-coil tm & signal peptide

Pfam Matches Standard

	Family		Clan	Description	Start - End		Alignment		Model			Bit Score	Domain E-values	
	Id	Accession			Start	End	Start	End	Start	End	Length		Ind.	Cond.
>	2OG-FeII_Oxy_3	PF13640.2	CL0029 ✓	2OG-Fe(II) oxygenase superfamily	111	204	112	199	2	95	100	59.77	3.2e-16	3.9e-20
>	2OG-FeII_Oxy_4	PF13661.2	CL0029	2OG-Fe(II) oxygenase superfamily	100	160	110	153	12	62	70	21.78	0.00011	1.4e-08

Your search took: 0.06 secs

5.-

SanC

Base de datos: Uniprot

Organismo: *Streptomyces ansochromogenes*

Estructura primaria

```
>tr|Q9L732|Q9L732_9ACTN SanC OS=Streptomyces ansochromogenes GN=sanC PE=4  
SV=1 MSLVDIETISRAELEQDPFGHVVIPRSFVTDVAEKLKRGFEFSDDGFEESSERTDLNRSKQ  
YRMFNRLNLTAGAVDEDQVRTLPANWQQLVADVVSRYREALSALTGVDLERCLVEARMT  
RYARGCWIEPHTDRPDKAVTHLFYFNDGWDPEWRGDLRLRLRSADMADCAKRVAPTTGTSV  
VLVRSDRSWHGVPPVADTCPVDRRALLVHFVRP
```

Pfam: 2OG-Fe(II) oxygenase superfamily, Clan: Cupin (CL0029)

Sequence Matches and Features ?

Pfam  213

disorder coiled-coil tm & signal peptide ?

Pfam Matches Standard

	Family		Clan	Description	Start	End	Alignment			Model			Bit Score	Domain E-values	
	Id	Accession					Start	End	Start	End	Length	Ind.		Cond.	
>	2OG-FeII_Oxy_3	PF13640.2	CL0029 ✓	2OG-Fe(II) oxygenase superfamily	117	212	118	210	2	98	100	63.64	2.0e-17	2.5e-21	
>	2OG-FeII_Oxy_4	PF13661.2	CL0029	2OG-Fe(II) oxygenase superfamily	113	162	118	155	14	57	70	28.14	1.2e-06	1.5e-10	

Your search took: 0.06 secs

6.-

Procollagen-lysine

Base de datos: Uniprot

Organismo: *Mus musculus*

Estructura primaria

```
>sp|Q9R0E1|PLOC3_MOUSE Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3
OS=Mus musculus GN=Plod3 PE=1 SV=1
MAAAGPEPRLLLLLLLLLLPPVTSASDRPRGANAVNPDKLLVITVATAETEGYRRFLQ
SAEFFNYTVRTLGLGQEWRRGGDVARTVGGGQKVRWLKKEMEKYADQKDMIIMFVDSYDVI
LASSPTTELLKKKFVQSGSHLLFSAESFCWPEWGLAEQYPEVGMGKRFLNSGGFIGFAPTIH
QIVRQWNYKDDDDDDQLFYTQLYLDPLREKLLKSLDHKSRI FQNLNGALDEVILKFDQNR
VRIRNVAYDTLPVVVHGNPPTKLQLNLYLGNYPNGWTPQGGCGFCNQLRRTLPGGQPPPR
VLLAVFVEQPTPFLPRFLQRLLLLDYPPDRISLFLHNSEVYHEPHIADAWPQLQDHFSAV
KLVGPEEALSAGEARDMAMDCRQNPCECFYFSLDADAVLTNPETLRVLI EQNRKVIAPM
LSRHGKLWSNFWGALSPNEYARSEDYVELVQRKRVGVWVNPYISQAYVIRGETLRTELP
QKEVFSSSDTDPDMAFCKSVRDKGI FLHLSNQHEFGRLLATSRYDTHLHPDLWQIFDNP
VDWREQYIHENYSRALDGEGLVEQPCPDVYWFPLLTEQMCDELVEEMEHYQWSGGRHED
SRLAGGYENVPTVDIHMKQVGYEDQWLQLLRITYVGPMT EYLFPGYHTKTRAVMNFVVRYP
PDEQPSLRPHHDSSTFTLNVALNHKGVDYEGGCRFLRYDCRISSPRKGWALLHPGRLTH
YHEGLPTTRGTRYIMVSFVDP
```

Pfam: 2OG-Fe(II) oxygenase superfamily, Clan: Cupin (CL0029)



Pfam Matches															
	Family		Clan	Description	Start - End		Alignment				Model		Bit Score	Domain E-values	
	Id	Accession			Start	End	Start	End	Start	End	Length	Ind.		Cond.	
>	Anp1	PF03452.10	CL0110	Anp1	286	366	299	341	21	63	264	8.76	0.9	0.00017	
>	Anp1	PF03452.10	CL0110	Anp1	370	456	385	429	135	179	264	9.95	0.39	7.3e-05	
>	Anp1	PF03452.10	n/a	Anp1	594	628	599	621	205	227	264	-0.88	790	0.15	
>	2OG-FeII_Oxy_3	PF13640.2	CL0029	2OG-Fe(II) oxygenase superfamily	655	740	664	738	9	98	100	17.47	0.0048	9.0e-07	
>	Glyco_tranf_2_3	PF13641.2	n/a	Glycosyltransferase like family 2	297	515	300	444	2	147	229	14.75	0.019	3.5e-06	

Your search took:0.08 secs

7.-

Prolyl 4-hydroxylase

Base de datos: Uniprot

Organismo: *Pongo abelii*

```
>sp|Q5RAG8|P4HA1_PONAB Prolyl 4-hydroxylase subunit alpha-1 OS=Pongo
abelii GN=P4HA1 PE=2 SV=1
```

```
MIWYILIIIGILLPQSLAHPGFFTSIGQMTDLIHTEKDLVTSLKDYIKAEDKLEQIKKWA
EKLDRLTSTATKDPEGFVGHVPVNAFKLMKRLNTEWSELENLVLKDMSDGFISNLTIQRQY
FPNDEDQVGAALKALLRLQDTYNLDTDTISKGNLPGVKHKSFLTAEDCFELGKVAYTEADY
YHTELWMEQALRQLDEGEIISTIDKVSVDYLSYAVYQQGDLKALLLTKKLELDPEHQ
RANGNLKYFEYIMAKEKDVNKSASDDQSDQKTPKKGVAVDYLPERQKYEMLCRGEIKM
TPRRQKKLFCRYHGDGNRNPKFILAPAKQEDEWDKPRIIRFHDIISDAEIEIVKDLAKPRL
RRATISNPITGDLETVHYRISKSAWLSGYENPVVSRINMRIQDLTGLDVSTAEELQVANY
GVGGQYEPHFDFARKDEPDAFKELGTGNRIATWLFYMSDVSAGGATVFPVGVASVWPKKG
TAVFWYNLFFASGEGDYSTRHAACPVLVGNKWSNKWLHERGQEFRPCTLSELE
```

Pfam: 2OG-Fe(II) oxygenase superfamily, Clan: Cupin (CL0029)

Sequence Matches and Features

Pfam: P4Ha_N (24-155) | tm & signal peptide (24-534)

disorder
 coiled-coil
 tm & signal peptide

Pfam Matches Standard

Family	Accession	Clan	Description	Start	End	Alignment		Model			Bit Score	Domain E-values	
						Start	End	Start	End	Length		Ind.	Cond.
> P4Ha_N	PF08336.7	n/a	Prolyl 4-Hydroxylase alpha-subunit, N-terminal region	24	155	24	155	1	135	135	137.24	3.0e-40	5.6e-44
> 2OG-FeII_Oxy	PF03171.16	CL0029	2OG-Fe(II) oxygenase superfamily	412	519	413	518	2	100	101	35.20	1.2e-08	2.2e-12
> 2OG-FeII_Oxy_3	PF13640.2	CL0029	2OG-Fe(II) oxygenase superfamily	415	518	416	505	2	86	100	27.15	4.7e-06	8.7e-10

Your search took: 0.08 secs

8.-

RNA polymerase

Base de datos: Uniprot

Organismo: *Lolium latent virus*

```
>sp|B1PS76|RDRP_LOLV RNA-directed RNA polymerase OS=Lolium latent virus (isolate
Lolium/USA/US1/-) GN=ORF1 PE=3 SV=1
MTTLALNLAQYNDPITKGI IAQESLRRVRLKAVVNVNPNYAIPTSAALVLEKLGIGTHP
MSLAVHHPHAPCKAIENQILNTVGHLLPKEQPVTFMAMKKSCLNMLRRHPSRADTFINPIY
HPRDNVRYGLDPPDQDASITASFLEVKTSTAFMQDTLHLYLTPEDLLDIFETSPKLENLI
ASFVLPVEATRNMKSLYDLYSIHYTHGGFQWAPSGHLGDAYFHEPWQLYWLRGSLTRL
IEEETTTTQVPPPGFEGEAYTRVRKEVRELKIYAERVTSIGAHHLFIFSRNAKATPRVRS
YSQNGKWWTLPRIFRPVSHNVQTPKQEVANSLMLTYAVRPSLKDVAKVQRQKDFDEKDL
AEHSPLEITHLINIYIYIDQRAYLTNDDILSDNLLKRWIFTPQAAYKKAKGFLGPD
FQKLLKALEWQPVFVDYAVDHYKSNPWRHASRTGAKMRQLKNFLSRTGLCEAVEDDAL
EGSELLKELEGNMWFRNPSEVEKEVFDIVADLPPDSQRSELILEDPNNSARDYLPIQQD
QPSVTQGPDEAAQQTPTATSMVTNSVAARMPPVPRTRLREHHEQFPTTVRASFEQPSST
RAPETSENTPAESVTPSPRAIYVGDFAIMDANASTSRASSPSRRTVVSQRPEAQNET
LGRQIVSRVSPAHCQMSPYAAQLAQHQLSEQTAYTDVIGQRMVAFYSQHSRSYKYGRHEH
RSQSWLPVIDSLQVALGLDESVDHCLIQRYRKHARVGLHADDEECYEPDSTIVTLNLYGN
ADFLIERNTDKASSETITLQHNDMLFMPSGMQVTHRHAVCSLYEGRVSI TFRNKTKDYLRK
SAPDMNPVEQPGASAGQLTGPLDHRPEELPWEHWI PRLNRLGFTGLQKQTDPEGKLIYPI
TEIRQDMVYVFPNCCPGPLRKDLEAMGRRPVRYTVDTGRLTASDIKNNRVGALLQNA
DLTWKTLLEVEYCRMEPTSVPMTVIHGAGGSGKSKLLQDHLNRAELNVV IIVPTRVLQQDW
RNKMTFEPFSLVQTYEAAMMESAPQMVVDDYGKLPHG YIDLFCQFHP SVEYVILTG DAR
QSTYYEYNSDAGIRNLPTNIEVFKQYCGYYINCTHRNKQDLANMLGVYSEKMGSTHFTFG
NTCTGSLLLVPSGTQKTVMGEGHKTETAGCQGITADKQVIMIDHDTHKSADSHMYTA
LSRATEHIHFYNSIAGLNTARFHAKNLTPYLKTFIQVITERAAAETEPAEYTVQAPTAR
THIPVENCSTFLEKDLLEEQRAKEDREVY TQAGATNVFQ TNSPIVQC FQHQQPKDGALS I I
THAKRLQYASAEANQAEYRAKLQIGAA LWENFKTAMEI PDEPVPFIRDLWEQSEAEVLST
YLSKSEMAIKNGKNRQDPDWQDERMFVYLKAQWVTKASKFNLP TAKAGQTI SAFKQAVVM
KFGAMARYLRRITPKPDNIRINCEMQPQDISK WALGLD SHNRPTKQKWNFERPAFASDFE
AFDQSQDGAMLHFALWARHFNVPSSLI EEYLF LKMHAQAPKGYLTIMRLTGEPTFDAN
TACSIAYNHTRYEIPKSCMQLYAGDDMLLDQVPVEKTGFKNIAAGLKL TAKTEIFEQKRG
KWGEFCSWMMTPYGLVKDPITLYHRILLASEIGDLSKKIDAYAIEAEPAYALQGR LFDCF
NEEQMTAHYGTIRRLIIEGKTNFSLDDQAPRHQDLQGLINCM DAKFMYH
```

Pfam: 2OG-Fe(II) oxygenase superfamily, Clan: Cupin (CL0029)



Pfam Matches														Standard	
Id	Family		Clan	Description	Start	End	Alignment		Model			Bit Score	Domain E-values		
	Accession						Start	End	Start	End	Length		Ind.	Cond.	
> Vmethyltransf	PF01660.13	n/a		Viral methyltransferase	39	364	39	356	1	317	340	191.62	1.5e-56	4.6e-60	
> Viral_helicase1	PF01443.14	CL0023	✓	Viral (Superfamily 1) RNA helicase	982	1212	982	1212	1	234	234	143.05	1.1e-41	3.2e-45	
> 2OG-Fe(II)_Oxy_2	PF13532.2	CL0029	✓	2OG-Fe(II) oxygenase superfamily	660	831	684	831	24	199	199	50.79	2.1e-13	6.6e-17	
> RDRP_2	PF00978.17	CL0027	✓	RNA dependent RNA polymerase	1366	1645	1380	1630	104	350	440	47.43	1.1e-12	3.5e-16	
> 2OG-Fe(II)_Oxy_2	PF03171.16	CL0029	✓	2OG-Fe(II) oxygenase superfamily	742	834	743	832	2	99	101	24.88	2.0e-05	6.1e-09	

Your search took: 0.1 secs

9.-

AlkB1, DNA oxidative demethylase

Base de datos: Uniprot

Organismo: *Mus musculus*

```
>sp|P0CB42|ALKB1_MOUSE DNA oxidative demethylase ALKBH1 OS=Mus musculus
GN=Alkbh1 PE=1 SV=1
MGKMAAAVASLATLAAEPREDAFRKLFRRFYRQSRPGTADLGAVIDFSEAHLARSPKPGVP
QVVRFPPLNVSSVTERDAERVGLPEVSKWRAYGLEGYPGFIFIPNPFPLPGCQRHWVKQCLK
LYSQKPNVCNLDKHMKTKEETQGLWEQSKEVLRKSKEVTKRRPRSLLERLRWVTLGYPHYNWD
SKKYSADHYTPFPSDLAFLEQVATACGFGQGFQAEAGILNYYRLDSTLGIHVDRSELDHS
KPLLSFSFQSAIFLLGGLKRDEAPTAMFMHSGDIMVMSGFSRLLNHAVPRVLPHPDGECL
PHCLETPLPAVLPSNSLVEPCSVEDWQVCATYLRTARVNMTVVRQVLATGQDFPLEPVEE
TKRDIAADGLCHLHDPNSPVKRKRLNPNS
```

Pfam: 2OG-Fe(II) oxygenase superfamily, Clan: Cupin (CL0029)

Sequence Matches and Features

Pfam: 2OG-Fe(II)_Oxy_2 389
tm & signal peptide 389

✓ disorder ✓ coiled-coil ✓ tm & signal peptide

Pfam Matches Advanced

	Family	Accession	Clan	Description	Start	End	Domain Ind.	E-values Cond.
>	2OG-Fe(II)_Oxy_2	PF13532.2	CL0029 ✓	2OG-Fe(II) oxygenase superfamily	98	344	3.0e-35	1.9e-39

Your search took: 0.07 secs

10.-

ALKBH

Base de datos: Uniprot

Organismo: *Arabidopsis thaliana*

```
>sp|Q9SA98|ALKBH_ARATH Alpha-ketoglutarate-dependent dioxygenase alkB
OS=Arabidopsis thaliana GN=At1g11780 PE=2 SV=2
MYESANVSDDADRTAFRRAEKKYKLYEQDSKFSRKKKLPKPIDLSELLDFNLISQNFNN
DGVLDPDGIRVSKVDSSPVFCIDNRPGFYFIPDALSLEQCKWIKESLTSFPQPPNRTNHN
AIYGPIDDLFDSAKENKVLVQDDLTTNNKWKFYEEVDIEKATRSSCKSVSASVLLRKLRS
TLGLQFDWSKRNYDVSLPHNNIPDALCQLAKTHAAIAMPDGEFRPEGAIVNYFGIGDTL
GGHLDDMEADWSKPIVSM SLGCKAIFLLGGKSKDDPPHMYLRS GDVVL MAGEARECFHG
IPRIFTGEENADIGALESELSHESGHFFAEYIKTSRININIRQVF
```

Pfam: 2OG-Fe(II) oxygenase superfamily, Clan: Cupin (CL0029)

Sequence Matches and Features

Pfam 2OG-Fell_Oxy_2 345

✓ disorder ✓ coiled-coil ✓ tm & signal peptide

Pfam Matches Standard

Id	Family	Accession	Clan	Description	Start - End		Alignment		Model			Bit Score	Domain E-values	
					Start	End	Start	End	Start	End	Length		Ind.	Cond.
> 2OG-Fell_Oxy_2		PF13532.2	CL0029 ✓	2OG-Fe(II) oxygenase superfamily	86	342	86	342	1	199	199	133.36	1.1e-38	6.7e-43

Your search took: 0.07 secs

11.-

IPS

Isopenicillin N-Synthase

Base de datos: Uniprot

Organismo: *Emericella nidulans*

>sp|P05326|IPNS_EMENI Isopenicillin N synthase OS=*Emericella nidulans* (strain FGSC A4 / ATCC 38163 / CBS 112.46 / NRRL 194 / M139) GN=*ipnA* PE=1 SV=1

MGSVSKANVPKIDVSPFLFGDDQAAKMRVAQQIDAASRD TGFFYAVNHGINVQRLSQKTKE
FHMSITPEEKWDLAIRAYNKEHQDQVRAGYYLSIPGKKAVESFCYLNPNFTPDHPRIQAK
TPTHEVNVWPDETKHPGFQDFAEQYYWDFGLSSALLKGYALALGKEENFFARHFKPDDT
LASVVLIRYPYLDPYPEAAIKTAADGTKLSFEWHEDVSLITVLYQSNVQNLQVETAAGYQ
DIEADDTGYLINCOSYMAHLTNNYYKAPIHRVKWVNAERQSLPFFVNLGYDSVIDPFDPR
EPNGKSDREPLSYGDYLVSLINKNGQT

Sequence Matches and Features

Pfam DIOX_N 331

✔ disorder ✔ coiled-coil ✔ tm & signal peptide

Pfam Matches Standard

	Family		Clan	Description	Start	End	Alignment		Model		Bit Score	Domain E-values		
	Id	Accession					Start	End	Start	End		Length	Ind.	Cond.
>	DIOX_N	PF14226.3	n/a	non-haem dioxygenase in morphine synthesis N-terminal	9	132	9	132	1	118	118	57.88	1.5e-15	1.8e-19
>	2OG-FeII_Oxy	PF03171.17	CL0029	2OG-Fe(II) oxygenase superfamily	181	288	184	286	4	99	101	46.10	4.9e-12	6.0e-16

Your search took: 0.06 secs

3KHB/4QKD

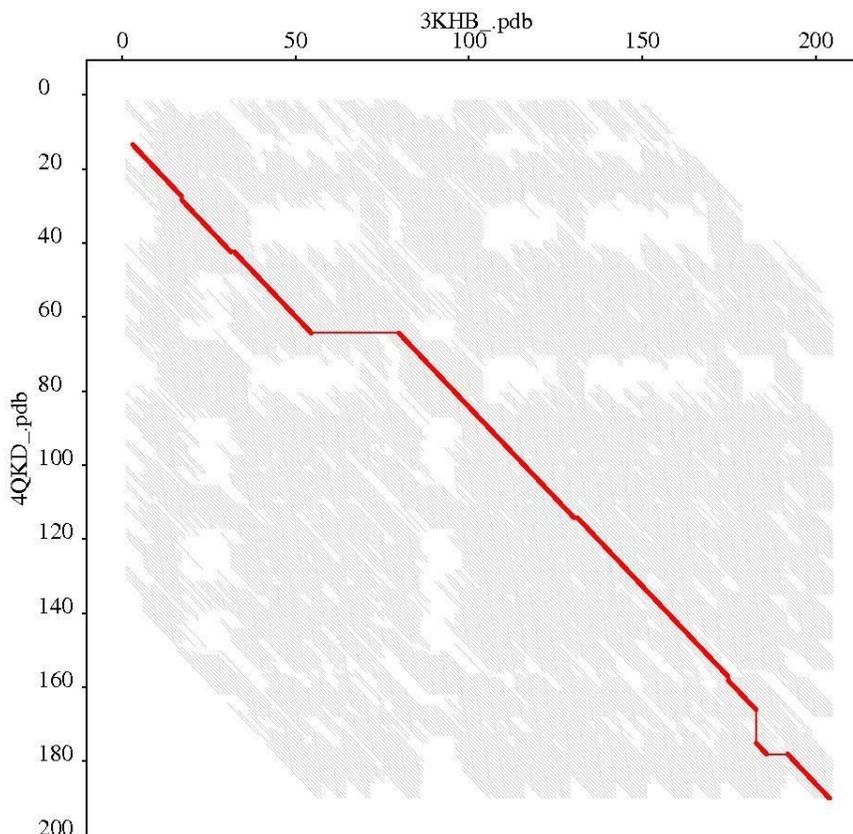
Align 3KHB_.pdb 203 with 4QKD_.pdb 190
Twists 0 ini-len 128 ini-rmsd 2.43 opt-equ 166 opt-rmsd 3.02 chain-rmsd 2.43 Score 259.31
align-len 210 gaps 44 (20.95%)
P-value 3.47e-07 Afp-num 11748 Identity 14.29% Similarity 24.76%
Block 0 afp 16 score 259.31 rmsd 2.43 gap 68 (0.35%)

```
Chain 1: 15 LAAGAVILRRFAFN-AAEQIIRDINDVASQSPFRQMVTGGYTMSVAMTNCGHLGWTTHRQGYLYSPIDP
          11111111111111 11111111111111 11111111111111111111111111
Chain 2: 29 LQDAAVVRPGFLSTAEETLSRELEPELR-RRRYEYDHWDAAIHGFRETEKS-----
```

```
Chain 1: 84 QTNGPWPAMPQSFHNLQRAATAAGYPDFQPDACLINRYAPGAKLSLHQDKDEPDLRAPIVSVSLGLPAI
          1111111111111111111111111111111111111111111111111111111111111111111111111111
Chain 2: 80 -----RWSEASRAILRRVQAAAFGPQTLSSVHVLDEARGYIKPHVDSIKFCGA-TIAGLSLLSPSV
```

```
Chain 1: 154 FQFGGLKRNDPLKRLLEHGDVVVWGGESRL-FYHGIQPL-----KAGFHPLTIDCRYNLTFRQAG
          1111111111111111111111111111111111111111111111111111111111111111111111111111
Chain 2: 143 MRLVHTQEPGEWLELLEPGSLYILRGSARYDFSHEILRDEESFFGERRIPR-----GRRISVICRSLP
```

Note: positions are from PDB; the numbers between alignments are block index



3KHB/4MG2

Align 3KHB_.pdb 203 with 4MG2_.pdb 195
Twists 0 ini-len 136 ini-rmsd 3.71 opt-equ 170 opt-rmsd 3.02 chain-rmsd 3.71 Score 251.63
align-len 218 gaps 48 (22.02%)
P-value 6.08e-07 Afp-num 12764 Identity 17.89% Similarity 29.82%
Block 0 afp 17 score 251.63 rmsd 3.71 gap 63 (0.32%)

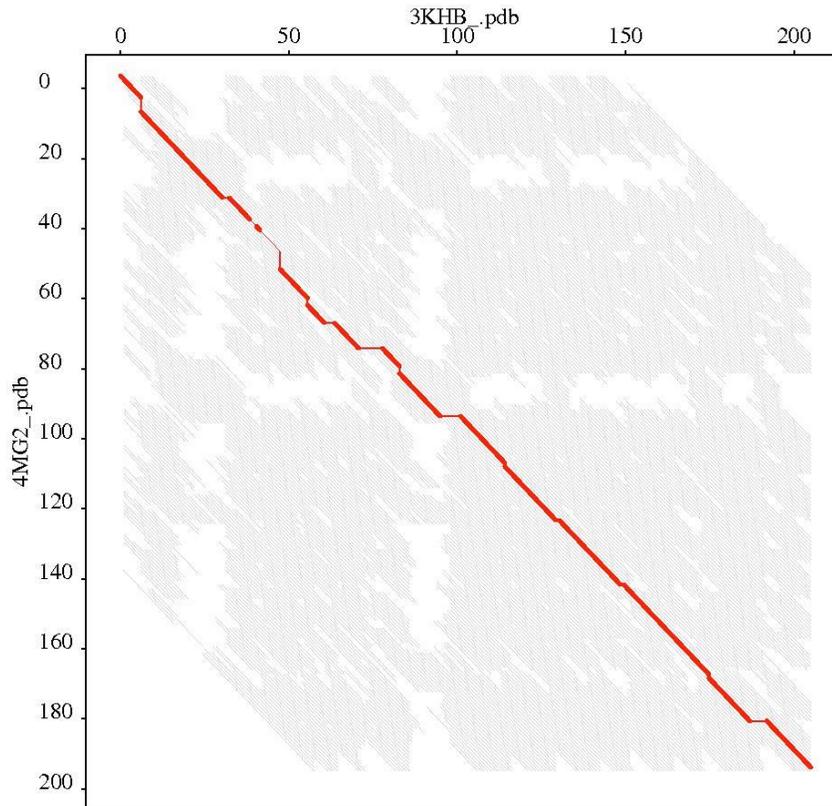
```
Chain 1: 12 QEPLAA---GAVILRRFAFNAAEQLIRDINDVASQSPFRQMVTGGYTMS-----VAMTNCGH--LGWT
111111 11111111111111111111111111111111 111111 1
Chain 2: 56 SWRHIRAEGLDSSYTVLFGKAEADEIFQELEKEV--EYFTGALARVQVAGKWHSVPRKQATYGDAGLTYT

Chain 1: 71 THRQGYLYSPIDPQTNKPWPAMP--QSFHNLQRAATAAGYPDFQPDACLINRYAP-GAKLSLHDKKDEP
1 1111111 11111 1111111111111 1111111111111 1111111111111 1111111111111111
Chain 2: 124 F---SGLTSLP-----KPWIPVLERIRDHVSGVTG-----QTFNFVLINRYKDGSDHICEHRDERE

Chain 1: 138 DLRAPIVSVSLGLPAIFQFGGLKRNPLKRLLEHGDVVVWGGESRL-FYHGIQPLKAGFHPLTIDCRYN
11 11111111111111111111 111111111111111111111111111111 111111111111 111111
Chain 2: 178 LA-SPIASVSFGASRDFVFRH-KDSVAVVRLPLAHGSLMMNHPNTHWYHSLPVRKKVL-----APRVN

Chain 1: 207 LTFRQAGK
11111111
Chain 2: 251 LTFRKILL
```

Note: positions are from PDB; the numbers between alignments are block index



3KHB/2IUW

Align

3KHB_.pdb 203 with 2IUW_.pdb 205

Twists 0 ini-len 136 ini-rmsd 3.47 opt-equ 181 opt-rmsd 3.08 chain-rmsd 3.47 Score 266.24

align-len 220 gaps 39 (17.73%)

P-value 2.66e-07 Afp-num 12929 Identity 12.73% Similarity 26.36%

Block 0 afp 17 score 266.24 rmsd 3.47 gap 64 (0.32%)

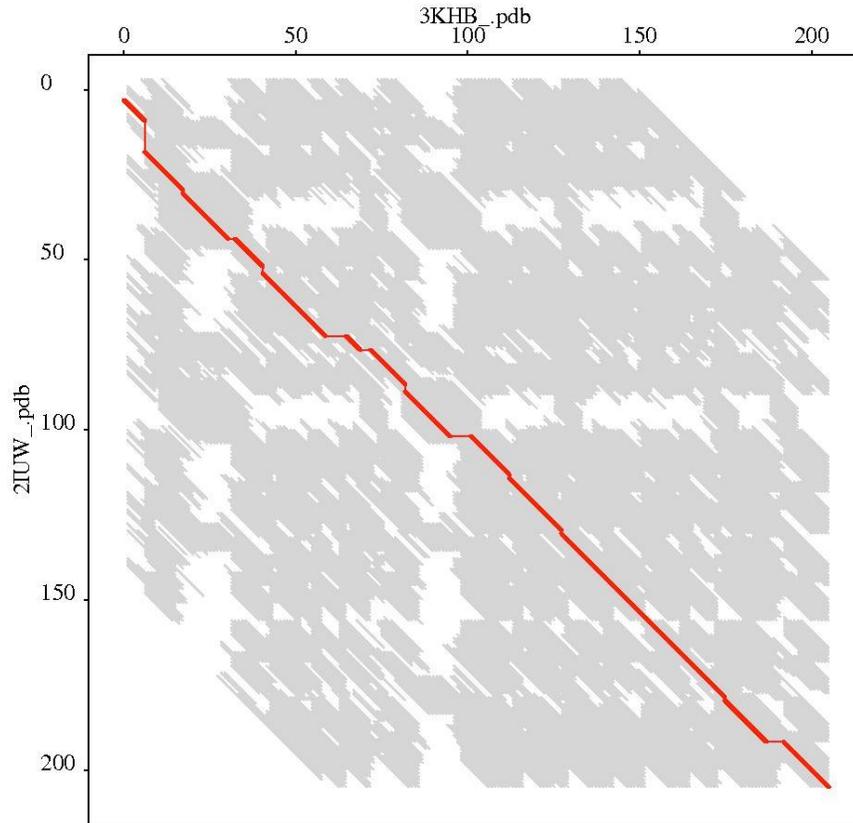
```
Chain 1: 12 QEPLAA-----GAVILRRFAFN-AAEQLIRDINDVASQSPFRQMT--PGGYTMSVAMTNCGHLGW
          111111          111111111111 111111111111 11111111 11111111111111111111
Chain 2: 74 REGVYEISLSPTGVS RVCLYPGFVDVKEADWILEQLCQDV--PWKQRTGIREDITYQQPRLTAWYGELPY
```

```
Chain 1: 70 TTHRQGYLYSPIDPQTNKPWPAM--QSFHNLQRAATAAGYPDFQPDACLINRY-APGAKLSLHODKDE
          1111 1111111111 1111111111111 11111111111 1111111111 1111111111111111
Chain 2: 142 -----TYSR---ITMEPNPHWHPVLR LTKNRIEENTG-----HTFNSLC NLRYNEKDSVDWHSDEP
```

```
Chain 1: 137 P-DLRAPIVSVSLGLPAIFQFGGLKRNDPLKRLLEHG DVVWVGESRL-FYHGIQPLKAGFHPLTIDCR
          1 111111111111111111111111111111111111111111111111111 111111111111 111
Chain 2: 197 SLGRCP IIASLSFGATR TTFEMRKKPPYVERVKIPLDHG TLLIMEGATQADWCHRV PKEYHSR-----EPR
```

```
Chain 1: 205 YNLTFRQAGK
          11111111111
Chain 2: 270 VNLTFRTVYP
```

Note: positions are from PDB; the numbers between alignments are block index



3KHB/2HBT

Align 3KHB_.pdb 203 with 2HBT_.pdb 224
 Twists 0 ini-len 128 ini-rmsd 3.03 opt-equ 149 opt-rmsd 3.06 chain-rmsd 3.03 Score 234.53
 align-len 216 gaps 67 (31.02%)
 P-value 1.46e-05 Afp-num 13754 Identity 9.26% Similarity 19.44%
 Block 0 afp 16 score 234.53 rmsd 3.03 gap 85 (0.40%)

```

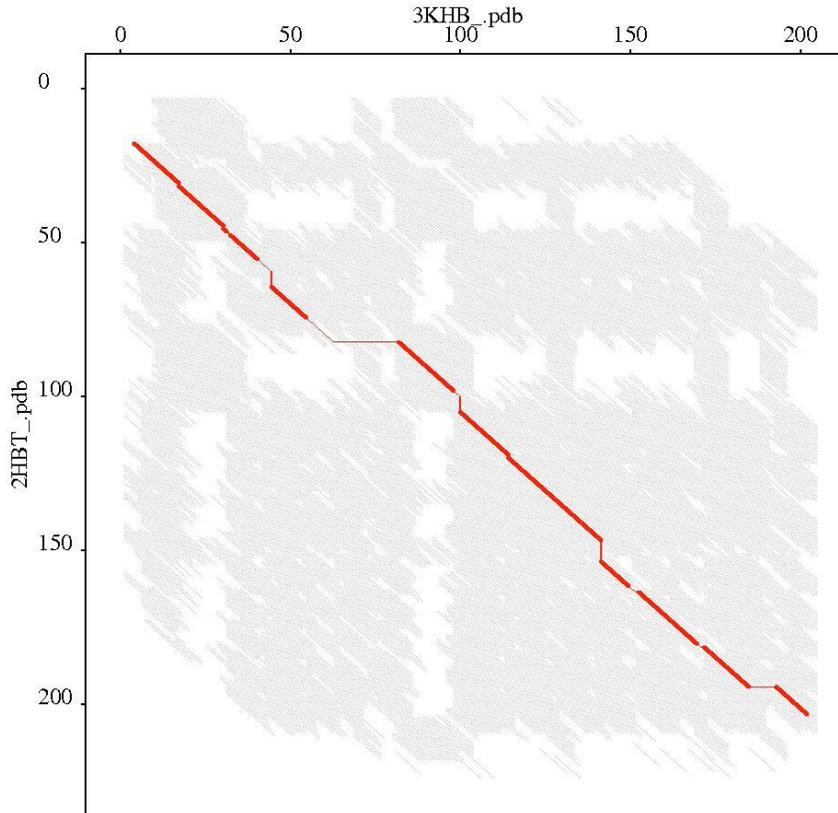
Chain 1: 16 AAGAVILRRFAFN-AAEQLIRDINDVA-SQSPFRQMVTPGGY-----TMSVAMTNCGHLGWTTHRQGYLY
          11111111111111 11111111111111 1 11111111 1111111111
Chain 2: 204 KHGICVVDDFLGKETGQQIGDEVRALHDTGKFTDGLVLSQKSDSSKDIRGDKITWIEGKEPGCET-----

Chain 1: 79 SPIDPQTNKFWPAMPQSFHNLQRAATAAGYP-----DFQPDACLINRYAP-GAKLSIHQKDEPDLRAP
          1111111111111111 11111111111111 1111111111111111111111111111
Chain 2: 269 -----IGLLMSSMDDLIRHCNGKLSYKINGRTKAMVACYPGNGTGYVHVDNPNNGDGRCV

Chain 1: 143 IVSVSLGLP-----AIFQFGGLKRNDPLKRLLEHGDVVVWGGESRIFYHGIQPLKAGFHPLTIDCRY
          1111111111 11111111 11111111111111111111 11111111111111 111
Chain 2: 325 TCIYYLNKDWDAKVSQGILRIFPEGK-AQFADIEPKFDRLFFWS-DRRNEHEVQPAYA-----TRY

Chain 1: 206 NLTFRQ
          111111
Chain 2: 385 AITVWY
  
```

Note: positions are from PDB; the numbers between alignments are block index



1IPS/4MG2

Align 1IPS_.pdb 328 with 4MG2_.pdb 195
 Twists 0 ini-len 120 ini-rmsd 6.34 opt-equ 145 opt-rmsd 3.18 chain-rmsd 6.34 Score 172.27
 align-len 272 gaps 127 (46.69%)
 P-value 3.88e-03 Afp-num 18879 Identity 5.88% Similarity 13.60%
 Block 0 afp 15 score 172.27 rmsd 6.34 gap 147 (0.55%)

```

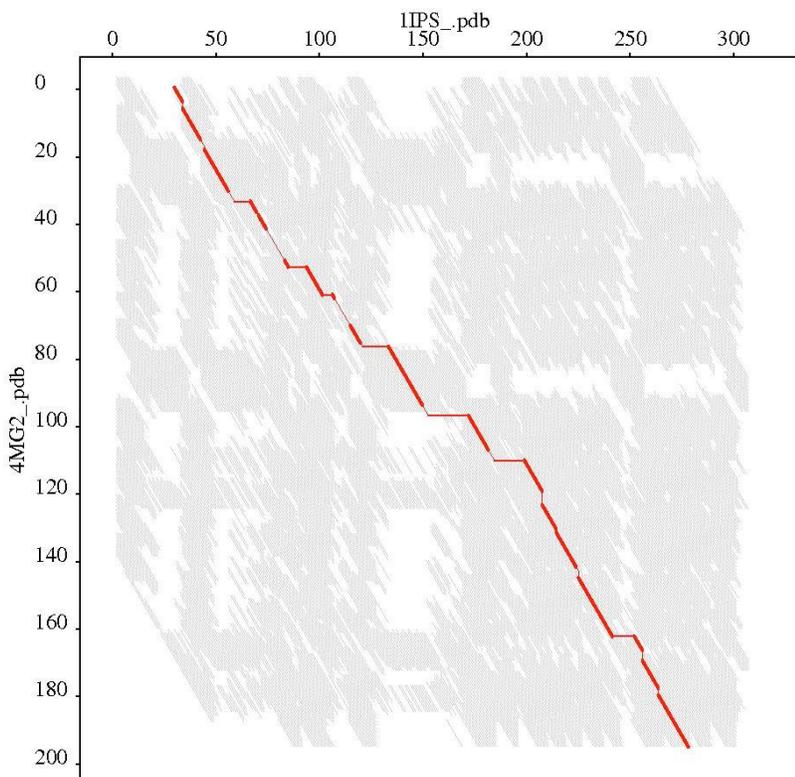
Chain 1: 35 ASRD--TGFFYAVNHGI-NVQRLSQKTKEFHMSITPEEKWDLAIRAYNKEHQDQVRAGYLSIPGKKAVE
          1111 1111111111 111111111111          111 1111          11          1
Chain 2: 59 HIRAEGLDSSYTVLFGKAEADEIFQELEKEVEY-----FTGALARVQVAGKWHVPR-----K

Chain 1: 102 SFCYLNPNFTPDHPRIQAKTPTHEVNVVWDETKHPGFQDFAEQYYWDFGLSSALLKGYALALGKEENFF
          1111111 1 11111          1111111111111111
Chain 2: 112 QATYGDA----GLTYTFSGLTLPKP-----WIPVLERIRDHVSGVTGQTF-----

Chain 1: 172 ARHFKPDDTLASVVLIRYPYLDPYPEAAIKTAADGTKLSFEWHEVVS----LITVLYQ-SNVQNLQVETA
          1111111111          111111111 1111111 11111111111
Chain 2: 154 -----NFVLINRYKDGSD-----HICEHRDDERELASPIASVSVFGASRDFVFRHK

Chain 1: 237 --AGYQDIEADDTGYLINCOSYMAHLTNNYKAP---IHRVKWVN--AERQSLPFFVNLYGD
          111111111111111111          1111 11111111 1111111111111111
Chain 2: 201 DSVAVVRLPLAHGSLMMN-----HPTNTHWYHSLPVRKKVLAPRVNLTFRKILLM
  
```

Note: positions are from PDB; the numbers between alignments are block index



1IPS/2IUW

Align 1IPS_.pdb 328 with 2IUW_.pdb 205
 Twists 0 ini-len 120 ini-rmsd 3.55 opt-equ 152 opt-rmsd 3.23 chain-rmsd 3.55 Score 174.60
 align-len 315 gaps 163 (51.75%)
 P-value 3.60e-03 Afp-num 19819 Identity 6.98% Similarity 14.29%
 Block 0 afp 15 score 174.60 rmsd 3.55 gap 140 (0.54%)

```

Chain 1:   5 SKANVPKIDVSPFLFGDDQAAKMRVAQQIDAASRD-----TGFFYAVNHGI--NVQRLSQKTKEFHMSIT
          11111111          11111          11111111111          11111111111111
Chain 2:  69 MRVIDREG-----VYEISLSPTGVSRVCLYPGFVDVKEADWILEQLCQDV-----

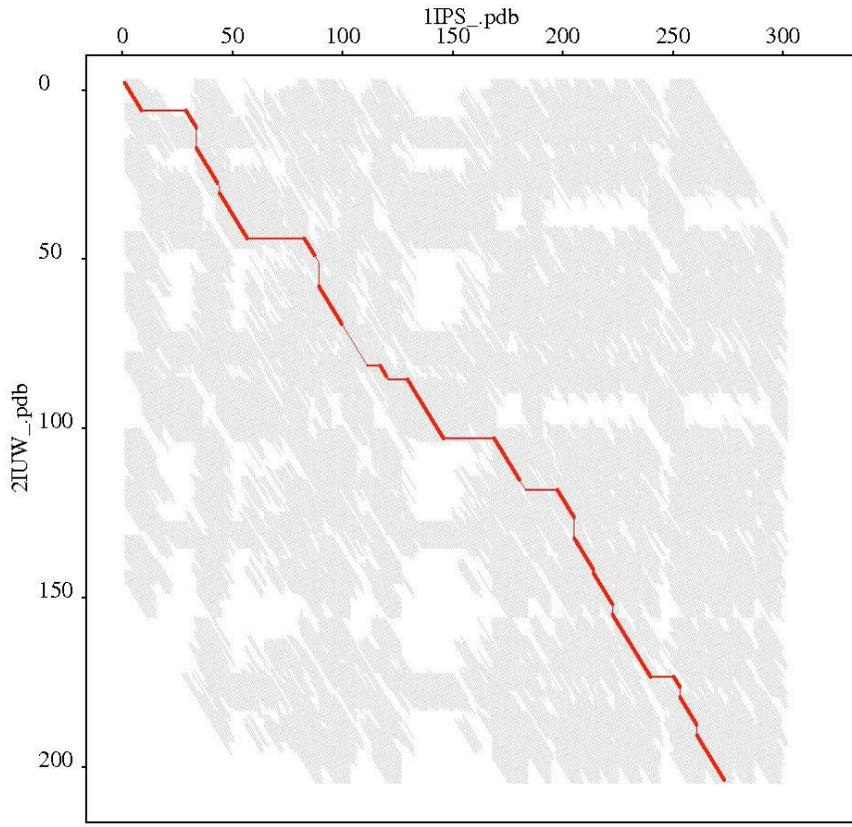
Chain 1:  67 PEEKWDLAIRAYNKEHQDQVRAGYYLSIPG-----KKAVESFCYLNPNFTPDHPRIQAKTPTHEVNVW
          11111          11111111111          111
Chain 2: 114 -----PWKQRTGIREDITYQQPRLTAWYGELPYTYSRITMEP-----NPHW

Chain 1: 130 PDETKHGPQDFAEQYYWDFGLSSALLKGYALALGKEENFFARHFKPDDTLASVVLIRYPYLDPYPEAA
          111111111111111111          1111111111111111
Chain 2: 155 -----HPVLRTLKNRIEENTGH-----TFNSLLCNLYRNEKD-----

Chain 1: 200 IKTAADGTKLSFEWHEDV-----SLITVLYQS-NVQNLQVET---AAGYQDIEADDTGYLINCOSYMAH
          11111111          1111111111          1111111111111111111111111111
Chain 2: 187 -----SVDWHSDEPSLGRCPFIASLSFGATRTFEMRKKPPYVERVKIPLDHGHTLLIME-----

Chain 1: 260 LTNNYYKA---PIHRVKWV---NAERQSLPFFVNL
          111 111111111          1111111111111111
Chain 2: 249 -----GATQADWQHRVPKEYHSREPRVNLTFRTVY
  
```

Note: positions are from PDB; the numbers between alignments are block index



4QKD/4NJ4

Align 4QKD_.pdb 190 with 4NJ4_.pdb 218
 Twists 0 ini-len 136 ini-rmsd 2.66 opt-equi 167 opt-rmsd 3.05 chain-rmsd 2.66 Score 308.91
 align-len 221 gaps 54 (24.43%)
 P-value 3.41e-08 Afp-num 12173 Identity 13.12% Similarity 26.70%
 Block 0 afp 17 score 308.91 rmsd 2.66 gap 72 (0.35%)

```

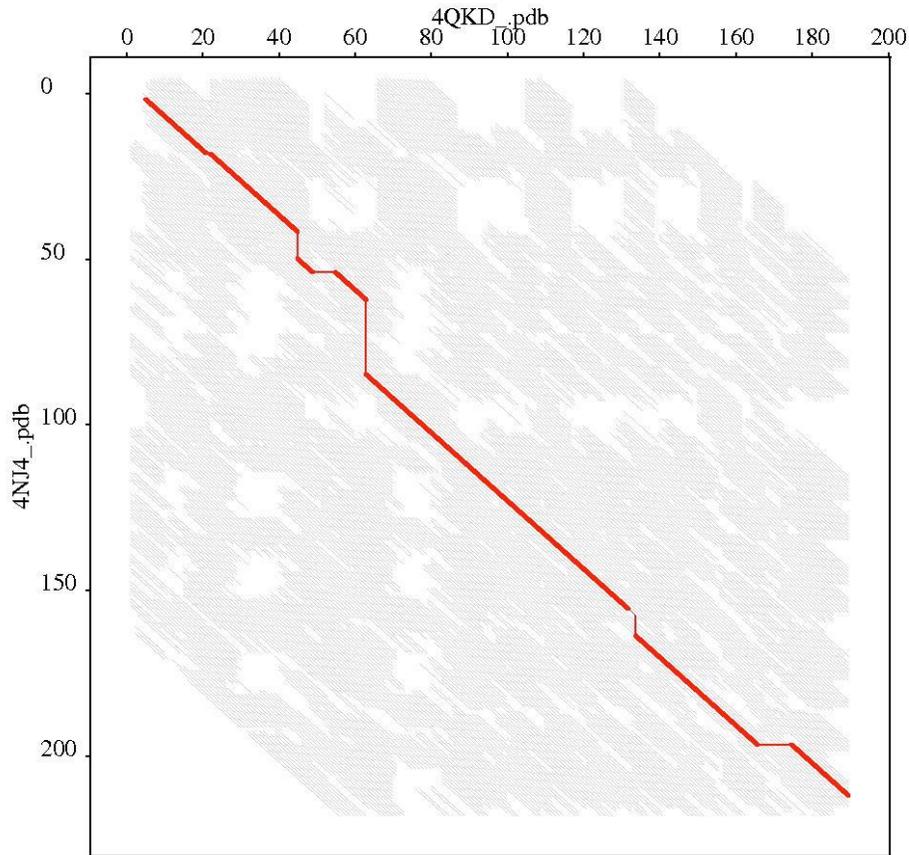
Chain 1: 21  . . . . . : . . . . . : . . . . . : . . . . . :
      SGPSVLSRLQDAAVVRPGFLSTAEETLSRELEPELRRRR-----YEYDHWDAAIHGFRETEK----
      1111111111111111 111111111111111111111111111111111111 1111 11111111
Chain 2: 77  QKEEEARKVKSGIRQM-RLFSQDECAKIEARIDEVVSRAEKGLYNEHTVDRA-----PLRNKYFFGEGY

Chain 1: 79  . . . . . : . . . . . : . . . . . : . . . . . : . . . . . :
      -----SRWSEASRAILRRVQAAAFGPQTLLSSVHVLDLEARGYIKH HVDSTKFCGAT
      11111111111111111111111111111111111111111111111111111111111111111111
Chain 2: 140 TYGAQPGQERLYPPGDVDEIPEWVHQLVIQKLVHRVIPEGFVNSAVINDYQPGGCIVSHVDPIHIFERP

Chain 1: 132  . . . . . : . . . . . : . . . . . : . . . . . : . . . . . :
      IAGLSLLSPSVMRLVHTQE-----PGEWLELLEPGSLYILRGSARYDFSHEILRDEESFFGERRIPRG
      11111111111111111111 1111111111111111111111111111111111111111111111111
Chain 2: 215  IVSVSFFSDSALCFGCKFQFKPIRVSEPLSLPVRRGSVTVLSGYAADEITHCIRPQ-----DIKE

Chain 1: 196  . . . . . :
      RRISVICRSLP
      111111111111
Chain 2: 276  RRAVIILRKTR
  
```

Note: positions are from PDB; the numbers between alignments are block index



4NJ4/2IUW

Align 4NJ4_.pdb 218 with 2IUW_.pdb 205
 Twists 0 ini-len 120 ini-rmsd 1.64 opt-equ 171 opt-rmsd 3.04 chain-rmsd 1.64 Score 287.42
 align-len 222 gaps 51 (22.97%)
 P-value 1.91e-07 Afp-num 12960 Identity 12.16% Similarity 27.48%
 Block 0 afp 15 score 287.42 rmsd 1.64 gap 75 (0.38%)

```

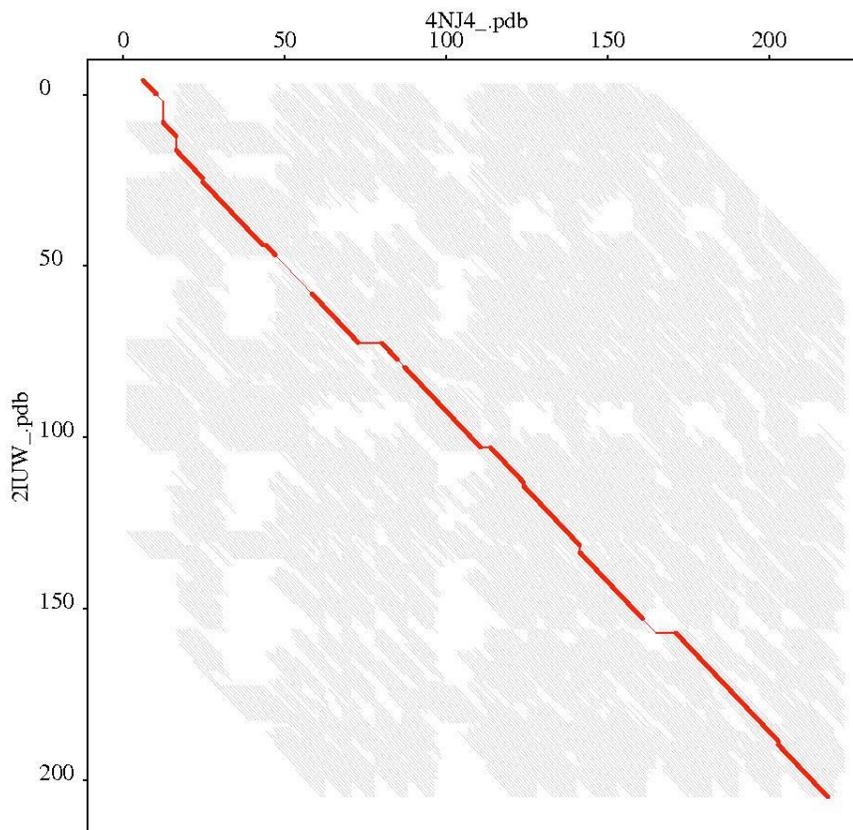
Chain 1: 76 LQKEEE-----ARKV----KSGIRQMR-LFSQDECAKIEARIDEVVSRAEKGLYNEHTVDRAPLRNKYF
      1111      1111      111111111 111111111111111111 111      111111111
Chain 2: 67 SHMRVIDREGVYEISLSPTGVSrvCLYPGFVDVKEADWILEQLCQDV-PWKQRTGIREDDITYQQPRLTAW

Chain 1: 135 FGEGYTYGAQPGQERLYPPGDVDEIPEWVHQLVIQKLVEHRVIPEGFVNSAVINDY-QPGGCIVSHVDPI
      111111      11111 111111111111111111111111 1111111111 11111111111111
Chain 2: 136 YGELPY-----TYSRITMEPNPHWHPVLRTLKNRIEENTGH---TFNSLLCNLYRNEKDSVDWHSDD

Chain 1: 209 HIFE--RPIVSVSFFSDSALCFGCKFQFKPIRVSEPVLSLPVRRGSVTVLSGYAADEITHCIRPQD-IKE
      1111 11111111111111111111      11111111111111111111111111111111 111
Chain 2: 196 PSLGRCPPIIASLSFGATRTEFEMRKKPPYV-----ERVKIPLDHGTLIMEGATQADWQHRVPKEYHSRE

Chain 1: 276 RRAVILRKTRL
      11111111111111
Chain 2: 268 PRVNLTFRTVYP
  
```

Note: positions are from PDB; the numbers between alignments are block index



4MG2/2IUW

Align 4MG2_.pdb 195 with 2IUW_.pdb 205 Twists 0 ini-len 160 ini-rmsd 2.11 opt-equ 186 opt-rmsd 3.04 chain-rmsd 2.11 Score 372.32 align-len 203 gaps 17 (8.37%) P-value 1.05e-10 Afp-num 12789 Identity 30.05% Similarity 43.84% Block 0 afp 20 score 372.32 rmsd 2.11 gap 40 (0.20%)

```

      . : . : . : . : . : . : . : . :
Chain 1: 56 SWRHIRAEG-----LDSSYTV-LFGKAEAEIFQELEKEVEYFTGALARVQVAGKWSVPRKQATYGDAG
      111111111 1111111 1111111111111111111111111111 111 111111111111111 1
Chain 2: 74 REGVYEISLSPTGVS RVCLYPGFVDVKEADWILEQLCQDVPWKQR--TGIRE--DITYQQPRLTAWYG-E

```

```

      . : . : . : . : . : . : . : . :
Chain 1: 120 LTYTFSGLTLSPK-PWIPVLERIRDHVSGVTGQTFNFVNLINRYKDGSDHICEHRDDERELA--SPIASVS
      11111111111111 111111111111111111111111111111111111111111111111111111111 11111111
Chain 2: 139 LPYYSRITMEPNPHWHPVLR LTKNRIEENTGHTFNSLLCNLYRNEKDSVDWHSDEPSLGRCPIIASLS

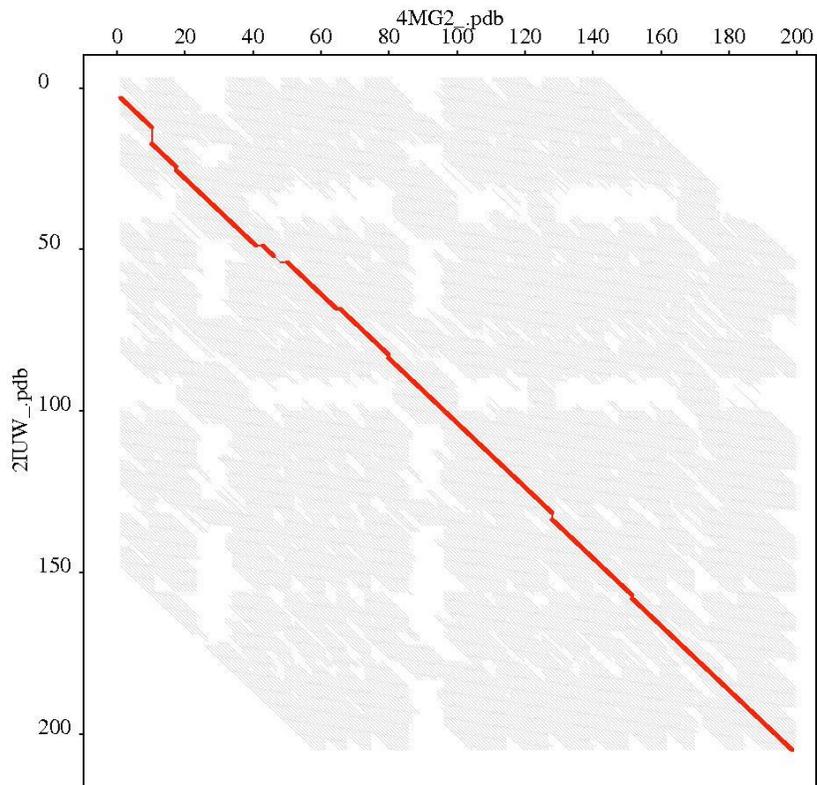
```

```

      . : . : . : . : . : . : . : . :
Chain 1: 189 FGASRDFVFRHKDSVA-VVRLPLAHGSLMMNHPTNTHWYHSLPVRKKVLAPRVNLTFRKILL
      1111111111111111 1111111111111111111111111111111111111111111111111111111
Chain 2: 209 FGATRTFEMRKKPPYVERVKIPLDHTLLIMEGATQADWQHRVPKEYHSREPRVNLTFRTVYP

```

Note: positions are from PDB; the numbers between alignments are block index



4MG2/2HBT

Align 4MG2_.pdb 195 with 2HBT_.pdb 224

Twists 0 ini-len 128 ini-rmsd 4.75 opt-equ 152 opt-rmsd 3.02 chain-rmsd 4.75 Score 247.49

align-len 221 gaps 69 (31.22%)

P-value 4.19e-06 Afp-num 13383 Identity 7.24% Similarity 16.29%

Block 0 afp 16 score 247.49 rmsd 4.75 gap 77 (0.38%)

```

Chain 1: 56 . . . . . : . . . . . : . . . . . : . . . . . :
      SWRHRAEGLDSSYTV-LFGKAEADEIFQELEKEV---EYFTGALARVQVAGKWHVSPRKQATYGDAGLT
      111 1 11111111 111111111111111111 11111 11111111
Chain 2: 196 EYIVPCMNKHGICVVDNFLGKETGQQIGDEVRALHDTGKFTDGGQLVVSQKSDSSKDIRGDKITWIEGKEPG

```

```

Chain 1: 122 . . . . . : . . . . . : . . . . . : . . . . . :
      YTFSGLTLSPKPWIPVLERIRDHVSQVGTGQ-----TFNFVLINRYKDGSDHICEHRDDE-RELA
      111111111111111111 111111111111111111111111111111 1111
Chain 2: 266 -----CETIGLLMSSMDDLIRHCNGKLGSYKINGRTRKAMVACYPGNGTGYVRHVDNPNNGDGR

```

```

Chain 1: 182 . . . . . : . . . . . : . . . . . : . . . . . :
      SPIASVSFGAS-----RDFVFRHKDSVAVVRLPLAHGSLMNMNHPNTNHWYHSLPVRKKVLAPRVNLT
      111111111111 11111111111111111111111111111111 111111111111 11111111
Chain 2: 323 CVTCIYYLNKDWDAKVSGGILRIFPEGKAQFADIEPKFDRLLEFFW--SDRRNEHEVQPAY---ATRYAIT

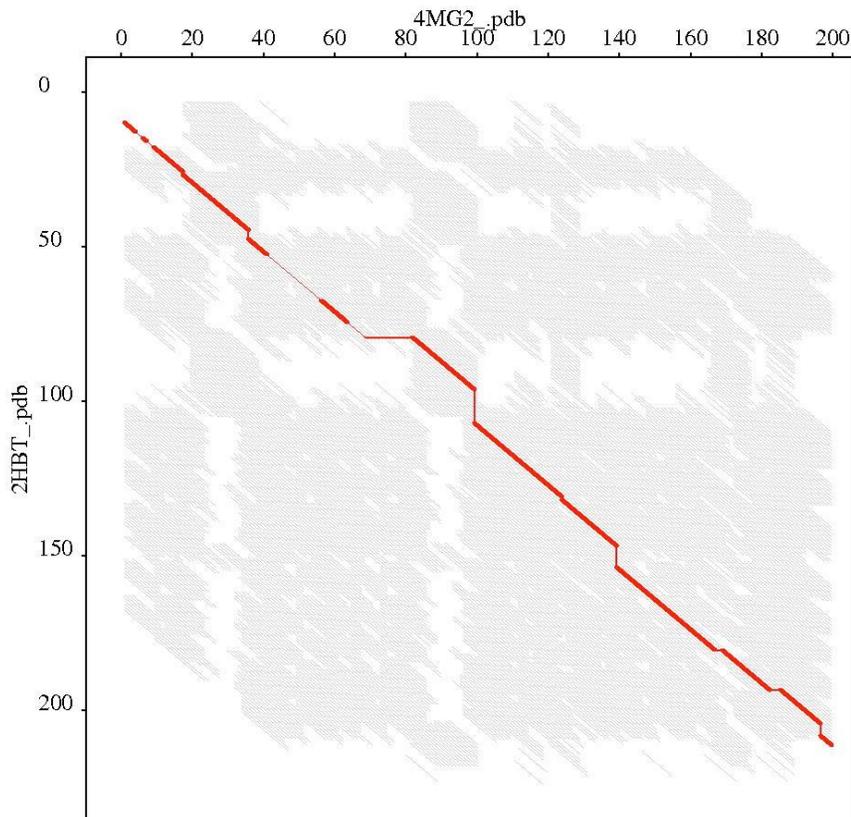
```

```

Chain 1: 253 . . . . . :
      FRKI----LLM
      1111 111
Chain 2: 388 VWYFDADERAR

```

Note: positions are from PDB; the numbers between alignments are block index



9.3 Identificadores UniProt de las secuencias utilizadas en este trabajo

G5EBV0(Hypoxia-inducible factor prolyl hydroxylase)

Q9R1J8(Prolyl 3-hydroxylase 1)

Q3V1T4(Prolyl 3-hydroxylase 1)

Q32P28(Prolyl 3-hydroxylase 1)

Q9KIT9(San F)

Q20679(Procollagen-lysine,2-oxoglutarate 5-dioxygenase)

Q9R0E1(Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3)

Q5U367(Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3)

O77588(Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1)

P13674(Prolyl 4-hydroxylase subunit alpha-1)

Q5RAG8(Prolyl 4-hydroxylase subunit alpha-1)

Q1RMU3(Prolyl 4-hydroxylase subunit alpha-1)

P54001(Prolyl 4-hydroxylase subunit alpha-1)

B1PS76(RNA-directed RNA polymerase)

P20951(RNA replication protein)

Q04575(RNA replication protein)

Q9SA98(Alpha-ketoglutarate-dependent dioxygenase alkB)

Q54N08(Alpha-ketoglutarate-dependent dioxygenase alkB)

P0CB42(Alkylated DNA repair protein alkB homolog 1)

Q13686(Alkylated DNA repair protein alkB homolog 1)

P05326(Isopenicillin N synthase)

P08703(Isopenicillin N synthase)

P05189(Isopenicillin N synthase)

P27744(Isopenicillin N synthase)

P75779(PKHD-type hydroxylase YbiX)

B2TVC6(PKHD-type hydroxylase YbiX)

Q32I80(PKHD-type hydroxylase YbiX)

P05050(EcoAlpha-ketoglutarate-dependent dioxygenase AlkB)

B8GWW6(CaulobacterAlpha-ketoglutarate-dependent dioxygenase AlkB homolog)

P37462(SalmonellaAlpha-ketoglutarate-dependent dioxygenase AlkB)

Secuencias nuevas re-extraídas de UniProt

>AlkB OS=Escherichia coli (strain K12)

MLDLFADAEPWQEPLAAGAVILRRFAFNAAEQLIRDINDVASQSPFRQMVTGGYTMSVA
MTNCGHLGWTTHRQGYLYSPIDPQTNKPWPAMPQSFHNLQRAATAAGYPDFQPDACLIN
RYAPGAKLSLHQDKDEPDLRAPIVSVSLGLPAIFQFGGLKRNDPLKRLLEHGDVVVWGG
ESRLFYHGIQPLKAGFHPLTIDCRYNLTFRQAGKKE

>alkB homolog 2 Mus musculus

MDKFLVRPDLRDLQGGGEEPAPTGGASDLKSPDWRHLRAEGLSCDYTVLFGKAEADKIF
RELEQEVEYFTGALAKVQVFGKWHVPRKQATYGDAGLTYTFSGLTTPKPWVPVLERVR
DRVCEVTGQTFNFVLVNRKYDGC DHIGEHRDDERELAPGSPIASVSFGACRDFIFRHKDS
RGKRPRRTVEVVRLQLAHGSLMMNPPTNTHWYHSLPIRKRVLAPRVNLTFRKILPTKK

>alkB homolog 3Homo sapiens

MEEKRRRARVQGAWAAPVKSQAIAQPATTAKSHLHQKPGQTWKNKEHHLSDREFVFKPEQ
QVVRRAPEPRVIDREGVYEISLSPTGVS RVCLYPGFVDVKEADWILEQLCQDVPWKQRTG
IRE DITYQQPRLTAWYGELPYTYSRITMEPNPHWHPVLR TLKNRIEENTGHTFN SLLCNL
YRNEKDSVDWHS DDEPSLGRCPIIASLSFGATR TFEMRKKPPPEENG DYTYVERVKIPLD

HGTLIMEGATQADWQHRVPKEYHSREPRVNLTFRFTVYPDPRGAPW

>alkB homolog 2 *Arabidopsis thaliana*

MTNPLNSTAANRSNQSSDGISDGQITNEEAESLINKKNCSGHKLKEVTDSDTFSDNGKD
DSDTKKRFBYHQDQRRMSLTSIVAVESPSSSNAPSRKTIDLGHGSDLIYIQRFLPFQSW
TFFDYLDKHIPWTRPTIRVFGRSCLQPRDTCYVASSGLTALVYSGYRPTSYSWDDFPPLK
EILDAIKVLPGSRFNLLLLNRYKGASDYVAWHADDEKIYGPTPEIASVSFGCERDFVLK
KKKDEESSQGKTGDSGPAKKRLKRSSREDQQSLTLKHGSLVMRGYTQRDWIHSVPKRAK
AEGTRINLTFRLVL

>Isopenicillin N synthase *Streptomyces cattleya*

MPVLMPSADVPTIDISPQLFGTDPTPRRTSRGRSTRPARGSGFFYASHHGIDVRRRLQTWS
NESTTMTDQRSTTWRSTRYNENNSHVRNGYYMARPGRETVESWCYLNPSFGEDHPMMKAG
TPMHEVNVWPDEERHPDFGSFGEQYHREVSASRRCRCCGASRWRQAGESSSNEVTEEDTL

77

SAVSMIRYPYLDPYPEAAIKTGPDGTRLSFEDHLDVSMITVLSKTEVQNLQVETVDGWQS
LPTSGENFLINCGTYLGYLTDYFPAPNHRVKYVNAERLSLPPFLHAGQNSVMKPFTRRT
GDRKLNPAVTYGEYLQEGFTR

>Isopenicillin N synthase *Penicillium chrysogenum*

MASTPKANVPKIDVSPLFGDNMEEKMKVARAIDAASRDTGFFYAVNHGVDVKRLSNKTRE
FHFSITDEEKWDLAIRAYNKEHQDQIRAGYYLSIPEKKAVESFCYLNPNFKPDHPLIQSK
TPTHEVNVWPDEKKHPGFREFAEQYYWDVFGLSALLRGYALALGKEEDFFSRHFKKEDA
LSSVVLIRYPYLNPIPPAAIKTAEDGTKLSFEWHEDVSLITVLYQSDVANLQVEMPQGYL
DIEADDNAYLVNCGSYMAHITNNYYPAPIHRVKWVNEERQSLPFFVNLFNDTVQPWDPS
KEDGKTDQRPISYGDYLNGLVSLINKNGQT

>Isopenicillin N synthase *Streptomyces griseus*

MPIPMLPAHVPTIDISPLSGGDADDKKRVAQEINKACRESGFFYASHHGIDVQLLKDVVN
EFHRTMTDEEKYDLAINAYNKNNPRTRNGYYMAVKGKKAVERWCYLNPSFSEDHPQIRSG
TPMHEGNIWPDEKRHRQFRPFCEQYYRDVFSLSKVLMRGFALALGKPEDFFDASLSLADT
LSAVTLIHYPYLEDYPPVKTGPDGKLSFEDHLDVSMITVLFQTEVQNLQVETADGWQDL
PTSGENFLVNCGYMGYLTNDYFPAPNHRVKFINAERLSLPFFLHAGHTTVMEPFSPEDT
RGKELNPPVRYGDYLLQASNALIAKNGQT

>Isopenicillin N synthase *Amycolatopsis lactamdurans*

MKMPSAEVPTIDVSPLFGDDAQEKVRVGQEINKACRSGGFFYAANHGVDVQRLQDVVNEF
HRTMSPQEKYDLAIHAYNKNNSHVRNGYYMAIEGKKAVERFCYLNPSFSEDHPEIKAGTP
MHEVNSWPDEEKHPSFRPFCEEYWTMHRLSKVLMRGFALALGKDERFFEPELKEADTLS
SVSLIRYPYLEDYPPVKTGPDGKLSFEDHFDVSMITVLYQTQVQNLQVETVDGWRDLPT
SDTDFLVNAGTYLGHLTNDYFPSPLHRVKFVNAERLSLPFFFHAGQHTLIEPFFPDGAPE
GKQGNEAVRYGDYLNHGLHSLIVKNGQT

>YbiX *Shigella flexneri*

MMYHIPGVLSPOQDVAHFREQLEQAEWVDGRVTTGAQGAQVKNNQQVDTRSALYAALQNEV
LNAVNHQHALFFAAALPRTLSTPLFNRYQNNETYGFHVDGAVRSHQPNGWMRTDLSATLFL
SDPQSYDGGELVVNDTFGQHRVKLPAGDLVLYPSSSLHCVTPVTRGVRVASFMWIQSMIR
DDKKRAMLFELDKNIQSLKSRYGENEEILSLLNLYHNLLREWSEI

>YbiX OS=*Escherichia coli*

MMYHIPGVLSPOQDVARFREQLEQAEWVDGRVTTGAQGAQVKNNQQVDTRSTLYAALQNEV
LNAVNHQHALFFAAALPRTLSTPLFNRYQNNETYGFHVDGAVRSHQPNGWMRTDLSATLFL
SDPQSYDGGELVVNDTFGQHRVKLPAGDLVLYPSSSLHCVTPVTRGVRVASFIWIQSMIR
DDKKRAMLFELDKNIQNIQSLKSRYGENEEILSLLNLYHNLLREWSEI

>SanF *Streptomyces ansochromogenes*

