



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

DETERMINACIÓN DE LA ENERGÍA DE UNIÓN INTERMOLECULAR COMO UN MÉTODO
PARA LA BÚSQUEDA DE GENES REGULADOS POR EL FACTOR DE TRANSCRIPCIÓN
MEF2A

TESIS

QUE PARA OPTAR POR EL GRADO DE:
Maestro en Ciencias

PRESENTA:
Biol. QUETZALCOATL ESCALANTE COVARRUNBIAS

TUTOR PRINCIPAL

Dr. León P. Martínez Castilla, Facultad de Química - UNAM

MIEMBROS DEL COMITÉ TUTOR

Dra. Elena Alvarez-Buylla Roces, Instituto de Ecología - UNAM
Dr. Gabriel del Río Guerra, Instituto de Fisiología Celular - UNAM

MÉXICO, D. F. Noviembre, 2015



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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.

ÍNDICE

Nota al lector	1
Lista de abreviaturas	2
RESUMEN	3
JUSTIFICACIÓN	4
OBJETIVOS E HIPÓTESIS GENERALES	5
1. CAPÍTULO I - Consideraciones previas del sistema de estudio	6
1.1. Planteamiento del problema del capítulo I	6
1.2. Objetivos particulares del capítulo I	7
1.3. Hipótesis del capítulo I	7
1.4. Sistema experimental del capítulo I	7
1.5. Software del capítulo I	7
1.6. Introducción al capítulo I	8
1.6.1. Factores de transcripción	8
1.6.2. Factores de transcripción con dominio MADS	8
1.6.2.a. Subestructura	9
1.6.2.b. Reconocimiento de un motivo específico de DNA	9
1.6.2.c. Formación de dímeros	10
1.6.2.d. Roles biológicos	10
1.6.2.e. Secuencias CArGbox	11
1.6.2.f. Relaciones filogenéticas entre proteínas MADS	11
1.6.3. La proteína MEF2A	12
1.6.3.a. Estructura y función	12
1.6.3.b. Actividades de regulación de la expresión genética	13
1.6.3.c. Estructura de su CArGbox	13
1.7. METODOLOGÍA DEL CAPÍTULO I	15
1.7.1. Determinación de la RMSD	15
1.8. RESULTADOS DEL CAPÍTULO I	16
1.8.1. Comparación de las proteínas y ácidos nucleicos presentes en los distintos complejos MEF2A-CARbox alternativos	16
1.9. DISCUSION DEL CAPITULO I	18
2. CAPÍTULO II – Estudio de Dinámica molecular	21
2.1. Planteamiento del problema del capítulo II	21
2.2. Objetivos particulares del capítulo II	22
2.3. Hipótesis del capítulo II	22
2.4. Sistema experimental del capítulo II	22
2.5. Software del capítulo II	22
2.6. Introducción al capítulo II	23
2.6.1. Factores que intervienen en la unión de un TF y su secuencia blanco	23
2.6.1.a. El llamado a distancia	23
2.6.1.b. Interacciones de cercanía física	24
2.6.2. Dinámicas moleculares	24
2.6.2.a. Interacciones atractivas y repulsivas	26
2.6.2.b. Representación de las interacciones moleculares	26
2.6.2.c. Campo de fuerza (FF)	27
2.6.2.d. Campo de fuerzas AMBER	29
2.6.2.e. Funcionamiento de un programa de dinámica molecular	29
2.6.2.f. Límites del sistema simulado	30
2.6.2.g. Métodos de integración	30
2.7. METODOLOGÍA DEL CAPÍTULO II	32

2.7.1. Especificación de la caja donde se aloja el sistema	32
2.7.2. Solvatación del sistema	33
2.7.3. Creación del archivo <i>.tpr</i>	33
2.7.4. Neutralización de la carga total del sistema	34
2.7.5. Flujo de ejecución de la preparación del sistema	34
2.7.6. Minimización de la energía total del sistema	34
2.7.7. Dinámica de posiciones restringidas	35
2.7.8. Dinámica molecular de todos los átomos	36
2.7.9. Herramientas de análisis posterior a las dinámicas experimentadas por el sistema	36
2.7.9.a. Obtención de la gráfica de energía potencial	36
2.7.9.b. RMSD	37
2.7.9.c. Obtención de las gráficas de temperatura, presión y densidad	37
2.7.9.d. Creación del grupo <i>Protein_DNA</i> y el aislamiento de subgrupos atómicos	37
2.7.9.e. Centrado del complejo dentro del sistema	37
2.7.9.f. Determinación de los puentes de hidrógeno existentes entre MEF2A y su CArGbox	38
2.7.9.g. Visualización de estructura secundaria	39
2.8. RESULTADOS DEL CAPITULO II	39
2.8.1. Inspección de los puentes de hidrógeno presentes entre MEF2A y la CArGbox	39
2.8.2. Mapa de interacciones por puentes de hidrógeno en el complejo MEF2A-CARgbox	49
2.8.3. Registro del cambio en la estructura secundaria del factor de transcripción MEF2A durante la dinámica molecular	51
2.9. DISCUSION DEL CAPITULO II	54
2.9.1. Minimización energética del sistema	54
2.9.2. Fijación de las posiciones del soluto durante la relajación del disolvente	54
2.9.3. Control de parámetros físicos a lo largo de la dinámica molecular de todos los átomos	56
2.9.4. Evolución en el acomodo espacial del complejo MEF2A-CARgbox	57
2.9.5. Estudio sobre la evolución de la estructura secundaria de MEF2A a lo largo de la dinámica	60
2.9.6. Distribución de los puentes de hidrógeno en las interacciones aminoácido-ácido nucleico	64
2.9.7. Interacciones entre la CArGbox con residuos de Lys y Arg	67
3. CAPÍTULO III – Identificación de nuevas CArGbox a las que se une el factor de transcripción MEF2A en el genoma de <i>Homo sapiens</i>	68
3.1. Planteamiento del problema del capítulo III	68
3.2. Objetivos particulares del capítulo III	68
3.3. Hipótesis del capítulo III	68
3.4. Sistema experimental del capítulo III	68
3.5. Recursos informáticos del capítulo III	69
3.6. INTRODUCCIÓN AL CAPITULO III	70
3.6.1 Calculo de la ΔG_{bind}	70
3.7. METODOLOGIA DEL CAPITULO III	73
3.7.1. El programa <i>MultiCARG</i>	73
3.7.1.a Creación del Universo 4 ¹⁰	73
3.7.1.b. Anexión de las bases complementarias a la secuencia lineal de nucleótidos	73
3.7.1.c. Construcción del archivo " <i>bp_step2.par</i> "	76
3.7.1.d. Construcción del archivo <i>.pdb</i> que contiene la estructura	

tridimensional de la nueva CArGbox	75
3.7.1.d.a. Mutación de A por T (Hebra líder)	76
3.7.1.d.b. Mutación de T por A (Hebra complementaria)	76
3.7.1.d.c. Mutación de G por A (Hebra líder)	76
3.7.1.d.d. Mutación de C por T (Hebra complementaria)	76
3.7.1.d.e. Mutación de una C por A (Hebra líder)	77
3.7.1.d.f. Mutación de una G por T (Hebra complementaria)	77
3.7.1.e. Adición de los átomos de H	77
3.7.1.f. Formación del complejo MEF2A-CArGbox	78
3.7.1.g. Computo de la energía libre de unión	78
3.7.2. Búsqueda de nuevas CArGbox	81
3.7.3. Arreglo de datos	82
3.7.4. Búsqueda de genes utilizando motivos descubiertos	82
3.7.5. Construcción de matrices	83
3.8. RESULTADOS DEL CAPITULO III	83
3.8.1. El programa MultiCArG	83
3.8.2. Archivo “ <i>Universooligos.txt</i> ”	83
3.8.3. Librerías de átomos de hidrógenos	84
3.8.4. Obtención de estructuras tridimensionales de DNA en formato <i>.pdb</i>	84
3.8.5. Complejos MEF2A-CArGbox	84
3.8.6. Complejos MEF2A-CArGbox.. Determinación de la energía de unión	85
3.8.7. Determinación de la Energía libre de Unión generada en diversos complejos MEF2A unidos a diversas secuencias de DNA	85
3.8.8. El conjunto CArGrango	87
3.8.9. Matrices de puntuación por posición específica (PSSMs)	92
3.8.10. Nuevos elementos-cis descubiertos en el genoma de <i>Homo sapiens</i> .	93
3.9. DISCUSION DEL CAPITULO III	108
3.9.1. Construcción de complejos MEF2A-CArGbox.	108
3.9.2. Determinación de la ΔG_{bind} .	109
3.9.3. Matrices de puntuación por posición específica obtenidas bajo los parámetros de éste estudio.	109
3.9.4. Identificación de regiones reguladas por MEF2A.	110
3.9.5. Variación del motivo consenso en posiciones 1 y 10 de la CArGbox.	111
3.9.6. Búsqueda de elementos cis con PSSMs contruidas por el método presentado en éste estudio.	112
3.9.7. ¿Por qué buscar nuevos métodos para el estudio de los factores de transcripción?	114
4. CONCLUSIONES	116
5. PERSPECTIVAS	118
6. BIBLIOGRAFÍA	119
APÉNDICE A	123
APÉNDICE B	126
APÉNDICE C	132
APÉNDICE D	136
APÉNDICE E	137
APÉNDICE F	138



NOTA AL LECTOR

El cuerpo de esta tesis se halla dividido en tres capítulos, cada uno posee su propio planteamiento del problema, objetivo e hipótesis particulares, junto con su respectiva sección de introducción y metodologías. Recuérdese que cada capítulo se halla acotado en todo momento por los objetivos e hipótesis generales, por lo que se recomienda acudir a ellos cuando se requiera contextualizar cada sección. El acomodo antes mencionado tiene dos razones: 1) Que el lector advierta la secuencia temporal en la que se ejecutó cada uno de los tópicos para comprobar la hipótesis y cumplir con el objetivo general; ya que aunque cada sección responde a éste y guarda una estrecha relación con el resto de los capítulos, posee características teóricas y metodológicas específicas que es mejor tratar por separado para evitar confundir al lector al mezclarlas entre sí; y 2) una razón didáctica, que tiene por objeto permitir el uso de cada capítulo como una guía para aquellos que deseen incorporar las técnicas aquí expuestas al estudio de los factores de transcripción, y de este modo facilitarles la comprensión de la teoría y la aplicación de las metodologías, ya sea siguiendo un flujo de elaboración como el que se plantea o incorporándose a la estación metodológica que se requiera.

Cada capítulo también posee su sección particular de resultados y discusión, con lo que se le facilitará al lector comprender los alcances teóricos y prácticos obtenidos en cada sección; aunado a ello al final de la tesis se cuenta con un apartado de conclusiones generales en las que se engloban los aportes del presente trabajo.

AGRADECIMIENTOS

A la beca entregada por CONACyT y al Programa de Apoyo a los Estudios de Posgrado (PAEP).



LISTA DE ABREVIATURAS

DNA:	Ácido desoxirribonucleico
AMBER	Campo de fuerzas cuyas siglas provienen del inglés <i>Assisted Model Building and Energy Refinemed</i>
CArGbox	Motivo compuesto por una secuencia de DNA localizada en la región upstream del sitio de inicio de la transcripción, en aquellos genes regulados por proteínas con dominio MADS.
DM	Dinámica molecular.
Elemento-cis	Sitio de reconocimiento en el DNA ubicado en la región <i>upstream</i> de los genes regulados por factores de transcripción.
FF	Sus siglas provienen del inglés <i>force field</i> , que se traduce como campo de fuerzas y hace referencia a la descripción de la interacción atómica dentro de un sistema.
L-J	Hace referencia al potencial Lennard-Jones.
MADS	Acrónimo conformado por los los genes en cuya secuencia fue identificada la caja MADS por primera vez: <i>MCM1</i> , <i>AGAMOUS</i> , <i>DEFICIENS</i> y <i>SRF</i>
MEF2A:	Factor de transcripción que potencia la diferenciación de las células que forman el tejido muscular (provienen sus siglas del inglés <i>Myocyte Specific Enhancer Factor 2A</i>).
PBC	Hace refrencia a las condiciones periódicas del contorno en un sistema de dinámica molecular.
PDB	Banco de dato de proteínas (provienen sus siglas del inglés <i>Protein Data Bank</i>), con esta abreviatura también se alude al tipo de archivo que se resguarda en este banco de datos.
pK _a	Logaritmo negativo de la constante de disociación ácida de un ácido débil.
RMSD	Sus siglas provienen del inglés <i>root-mean-square deviation</i>
TF	Factor de transcripción (provienen sus siglas del inglés <i>Transcription Factor</i>).
VMD	Sus siglas provienen del inglés <i>Visual Molecular Dynamics</i> , programa que tiene por finalidad la visualización 3D de proteínas.
VW	Fuerzas de atracción o repulsión entre moléculas distintas al enlace covalente y fuerzas iónicas, nombradas en honor a Johannes Diderik van der Waals
ΔG_{bind}	Energía libre de unión



RESUMEN

En la regulación de la expresión genética que deriva hacia la generación o represión de los productos funcionales del genoma, participan proteínas especializadas denominadas *factores de transcripción*, que al interactuar con secuencias cortas de DNA que reconocen de forma específica, conforman un importante sistema en la regulación del metabolismo de los seres vivos.

En el presente trabajo se propone una metodología que permite abordar el análisis de los factores de transcripción y sus sitios de unión al DNA a través de estrategias bioinformáticas, ejemplificadas en el estudio de la proteína MEF2A y sus secuencias blanco, conocidas como “CArGbox”.

En esta metodología se aspira a que el estudio de los sistemas moleculares no provistos de movimiento sea sólo un punto y no el único de los que se abordan en la investigación, ya que para lograr comprender el modo en que el factor de transcripción estudiado interactúa con su secuencia blanco de DNA, se recurrió a dinámicas moleculares con el fin de proveer al sistema de cambio y movimiento, logrando revelar algunos de los factores fundamentales para que se de la formación del complejo formado entre MEF2A y su CArGbox.

Concomitante a la metodología teórica, se creó el programa MultiCArG con fines prácticos, que aunado a otras paqueterías es capaz de calcular la energía libre de unión generada entre la proteína MEF2A y un grupo de secuencias CArGbox mutantes; el objetivo: descubrir por medio de este parámetro secuencias de DNA viables de ser blanco del factor de transcripción MEF2A.

La evaluación de la energía libre de unión (ΔG_{bind}), derivó en la construcción de matrices de puntuación por posición específica, con las que se inició la búsqueda y asignación de nuevos *elementos-cis* en el genoma de *Homo sapiens*, dando como resultado final, la detección de nuevos genes regulados por el factor de transcripción MEF2A.



JUSTIFICACIÓN

El desarrollo de métodos bioinformáticos como auxiliares en la investigación de los factores de transcripción y sus secuencias blanco, contribuye a profundizar el conocimiento biológico existente en este campo por medio de la creación de aplicaciones prácticas, que en el presente caso, derivan en la determinación de potenciales secuencias blanco a las que se une MEF2A dentro del genoma de *Homo sapiens*, cuestión altamente valiosa en el estudio del control de la expresión genética.

Las metodologías que se presentan a continuación, fueron desarrolladas con la intención de ser aplicables para el estudio de otros factores de transcripción y la detección de sus sitios de unión, con lo que se espera contribuir al desarrollo de éste campo del conocimiento.



OBJETIVO

Proponer una metodología para el estudio de los factores de transcripción basada en herramientas bioinformáticas que derive en la detección de nuevos genes potencialmente regulados por el factor de transcripción MEF2A en el genoma de *Homo sapiens*.

HIPÓTESIS

Los complejos formados entre el factor de transcripción MEF2A y las secuencias de nucleótidos a las que se une, denominadas como CArGbox; poseen características termodinámicas exclusivas y manifiestas en su valor de energía libre de unión (ΔG_{bind}), tales que permiten identificar como CArGbox potenciales aquellas secuencias que posean un valor de ΔG_{bind} semejantes entre sí.



CAPÍTULO I

Consideraciones previas del sistema de estudio

1.1. Planteamiento del problema del capítulo I.

Para realizar la determinación de la energía libre de unión (ΔG_{bind}) fue necesario obtener la estructura tridimensional de los decámeros de DNA correspondientes al resultado de la combinatoria 4^{10} , que representa todas las secuencias potenciales a las que puede unirse MEF2A. Dado que el cómputo de los ensayos de *docking* necesarios para aparear dichas secuencias con el factor de transcripción (TF por sus siglas en inglés) serían temporalmente costosos, se optó por tomar como molde la estructura tridimensional de la CArGbox (denominada como elemento-*cis* dentro de la terminología de los factores de transcripción) descrita en el reporte de Huang *et al.*, 2000 (No. idPDB:1c7u), con el fin de sustituir los nucleótidos presentes en su secuencia sin alterar los valores del perfil estructural de la CArGbox, tales como: ángulo de curvatura, tamaño de los surcos y los parámetros de rotación y traslación de las bases presentes en el decámero de DNA. Se respetaron las coordenadas espaciales de la CArGbox dentro del complejo TF-elemento-*cis* determinado por Huang, bajo la suposición de que la arquitectura del complejo no cambia de forma drástica, incluso si la secuencia de nucleótidos de la CArGbox varía. Atendiendo a esta hecho, es necesario comprobar la veracidad de dicha suposición, dada la posibilidad de que sí existan diferencias significativas en la arquitectura de los complejos MEF2A-CArGbox al modificar la secuencia de nucleótidos, ya que esto predispondría un grado de error significativo durante la determinación de la ΔG_{bind} por estar basado en parámetros iniciales intrínsecamente erróneos.

Afortunadamente, se han realizado determinaciones espectroscópicas de la estructura cuaternaria de MEF2A adicionales a la referida con anterioridad, las cuales no fueron tomadas en cuenta para los experimentos realizados dentro de esta tesis por poseer un mayor grado de complejidad, lo que implica menor eficiencia en el cómputo de los datos. Ejemplos representativos de éstas son las resueltas por el grupo de Wu (Wu *et al.*, 2010) y el de He (He *et al.*, 2011); la primera de ellas corresponde a la estructura cuaternaria de MEF2A que incluye al motivo MEF2S (residuos 59-95) unida a la CArGbox 5'CTATTTATAA3' (No. id:3kov) y la segunda corresponde a este mismo complejo unido al dominio TAZ2 de la proteína p300 (No. id:3p57). Tomando en cuenta los datos recabados en estos estudios, fue posible realizar una comparación estructural de la arquitectura de MEF2A y la CArGbox, lo que permitió considerar si las diferentes condiciones experimentales y biológicas a las que se sometieron los complejos antes referidos, como el cambio de la secuencia en la CArGbox y la unión a la proteína p300, tienen un efecto sustancial en la arquitectura del complejo.

1.2. Objetivos particulares del capítulo I.

Realizar un estudio comparativo de las proteínas y los ácidos nucleicos presentes en los distintos complejos MEF2A-CArGbox reportados en la literatura.

1.3. Hipótesis del capítulo I.

La arquitectura de MEF2A y la CArGbox dentro del complejo FT-elemento-*cis* reportada en el archivo PDB con No.id: 1C7U no varía de forma significativa, incluso si se modifica la secuencia de nucleótidos en la CArGbox.

1.4. Sistema experimental del capítulo I.

- Estructura tridimensional del complejo MEF2A-CArGbox (No. de id:1C7U).
- Estructura tridimensional del complejo MEF2A-CArGbox (No. de id:3KOV).
- Estructura tridimensional del complejo p300-MEF2A-CArGbox (No. de id:3P57).

1.5. Software del capítulo I.

Visual Molecular Dynamics (VMD)

1.6. INTRODUCCIÓN.

Dentro del Universo conviven diversas formas de materia cuya naturaleza intrínseca, está determinada por particularidades que rigen la manera en la que interactúan con el medio que los rodea. Una de las manifestaciones que pertenece a un conjunto bien definido dentro de estas formas de existencia es la vida.

Una de las propiedades más significativas que poseen los seres vivos es la capacidad de auto-organización, la cual ha permitido un aumento en la complejidad de las redes de regulación interna, algunas de las cuales están destinadas a la asimilación de nutrientes, la respuesta a estímulos, al desarrollo y a la reproducción. Bajo estas expresiones de vida subyacen las redes de regulación transcripcional en la que participan los denominados factores de transcripción.

1.6.1. Factores de transcripción.

Los factores de transcripción (TF's por sus siglas en inglés) constituyen el grupo de proteínas encargadas de la represión o activación de genes dentro de un genoma. Su síntesis tiene lugar en el citoplasma donde (en la mayor parte de los casos) tienden a la multimerización formando unidades repetitivas o conjuntos de diferentes proteínas (Pabo & Sauer, 1992); así sometidos a una regulación de su activación y desactivación basada en la unión a otros polipéptidos y cofactores.

Una vez que adquieren la forma idónea para llevar a cabo su actividad biológica, pueden diferenciarse como mínimo dos dominios funcionales: el de unión al DNA y el de trans-activación. El primero de ellos reconoce secuencias específicas de material genético ubicadas corriente arriba del sitio de iniciación de la transcripción, siendo los motivos hélice-giro-hélice, hélice lazo hélice, dedo de zinc y cremallera de Leucina las conformaciones estructurales típicas observadas en este sitio. El dominio de trans-activación está encargado de proporcionarle actividad biológica a la TF por medio del reconocimiento de proteínas y cofactores, su estructura tridimensional es muy variada ya que depende de las particularidades propias de cada familia de TF's (Taleisnik, 2006).

1.6.2. Factores de transcripción con dominio MADS.

Entre todas las familias de TF's encargadas de dirigir una compleja red de regulación genética estrictamente organizada, la familia de proteínas con dominio MADS es una de las más relevantes. El término MADS es un acrónimo derivado de los cuatro miembros fundadores de éste grupo: la proteína MCM1 de levadura, AGAMOUS y DEFICIENS de plantas y la proteína SRF presente en mamíferos. Desde que la proteína MCM1 fue reportada en 1988 (Steven *et al.*, 1988) la familia de TF's MADS adquirió un rol fundamental para entender el control en la transducción de señales durante el desarrollo de animales, plantas y hongos; ya que al activarse determinan la identidad y posición de cada célula con el fin de diferenciar y organizar los tejidos y segmentos en un organismo (Deutsch, 2010).

El dominio MADS está codificado en una secuencia de 180 pares de bases denominada caja homeótica, altamente conservada entre este grupo de genes y que se traduce a una región de aproximadamente 60

aminoácidos que da lugar al dominio MADS ubicado en la región amino terminal, siendo éste el encargado de la unión entre el TF y el DNA (Heffer *et al.*, 2010). El resto de la proteína es altamente variable y es responsable de la dimerización y la transactivación, las cuales son necesarias para que estos TF's adquieran actividad biológica.

1.6.2.a. Subestructura.

Esencialmente, los miembros de la familia MADS poseen dos regiones principales: el dominio mínimo de unión a DNA encargado de la dimerización (compuesto de aproximadamente 56 residuos de aminoácidos conservados entre todos los miembros de la familia), y la extensión del carboxilo terminal altamente variable en cuanto a tamaño y estructura, debido a su heterogeneidad dentro la familia (Fig. 1).

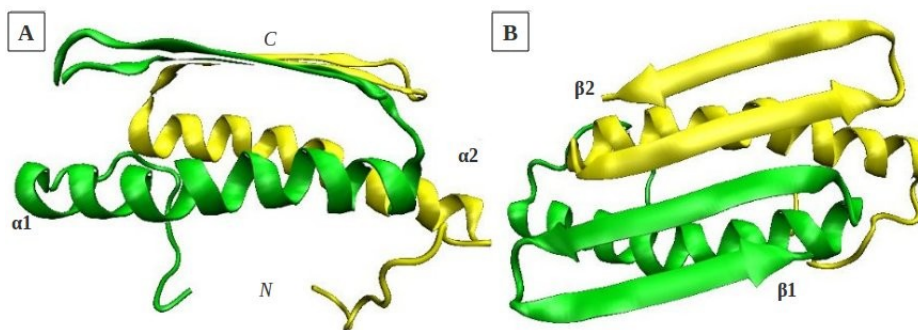


Fig. 1. Estructura cuaternaria del dominio MADS de MEF2A. A) Vista frontal. B) Vista superior. Los primeros 11 residuos de aminoácidos corresponden a una región desordenada, del residuo 12 al al 38 se forma una alfa hélice, del residuo 39 al 41 se observa un loop y del residuo 42 al 57 se forman dos hebras beta que forman una hoja. Las imágenes fueron obtenidas a partir del archivo PDB con número de registro 1C7U.

1.6.2.b. Reconocimiento de un motivo específico de DNA.

Al ser resuelta la estructura tridimensional de MEF2A, estudios deletéreos han identificado los residuos necesarios para que se lleve a cabo la dimerización y la unión con el DNA, en el caso de MEF2A son los primeros 72 aminoácidos que se ubican a partir del N-terminal (Huang *et al.*, 2000), en este caso se les ha asociado a los primeros 11 residuos que forman una región intrínsecamente desordenada y del residuo 12 al 38 que forman una α -hélice que es la que lleva a a cabo la interacción con el DNA. Se reprodujo el mismo caso en análisis similares realizados en la proteína SRF (Norman *et al.*, 1988), MCM1 y AG (Mueller *et al.*, 1991) identificando de este modo una sección mínima del TF que permite la interacción con el DNA. Un gran número de quimeras de MADS han sido construidas con el propósito de identificar y mapear por función y región la subestructura de estas proteínas; una de ellas consistió en una quimera creada a partir de la fusión de 41 aminoácidos de la región amino terminal de MCM1, que fueron transferidos a la región equivalente y del mismo tamaño de SRF, logrando de este modo que éste último factor de transcripción, identificara secuencias de DNA equivalentes a las que identifica MCM1 (Sharrocks *et al.*, 1993), de este modo se concluyó que esta región es la encargada de dotar de especificidad a las proteínas MADS en el reconocimiento de una secuencia de DNA. Esta compatibilidad entre el amino terminal de las MADS y el DNA se debe a que esta región se encuentra enriquecida en aminoácidos básicos, siendo muy comunes la Arginina o la Serina en estas regiones terminales,

comprobado este hecho por estudios de delección en los cuales se eliminaron estos aminoácidos del amino terminal, lo que provocó una reducción en la unión a secuencias de DNA, y además fue causa de ambigüedad en cuanto a la detección de su elemento-cis, perdiendo así su selectividad. (Sharroks *et al.*, 1993). De esta manera, se confirmó que la función del amino terminal es identificar la secuencia del DNA dotando de especificidad al factor de transcripción.

1.6.2.c. Formación de dímeros.

Un factor determinante para comprender la formación de dímeros en las proteínas con dominio MADS, está dada por la simetría propia de las cadenas de DNA a las que se une, estas secuencias en su mayoría son palindrómicas, lo que significa que cada una de las proteínas que forman el dímero se encuentran unidas a la misma secuencia. Este hecho ha sido reportado tanto para homodímeros como para heterodímeros (McDermott *et al.*, 1993).

En el caso de la proteína MEF2 es teóricamente posible crear más de 100 heterodímeros a causa de las cuatro diferentes isoformas obtenidas por *splicing* alternativo de los cuatro genes que codifican a MEF2 (Martin *et al.*, 1994). Esta gran variedad de dímeros se ve enriquecida por la conservación del dominio MADS, el cual permite la dimerización con otros miembros de la familia, como se ha demostrado durante la identificación de un heterodímero formado *in vitro* por MCM1 y SRF (Mueller *et al.*, 1991).

Todas las isoformas de MEF2 poseen un conjunto de aminoácidos altamente conservados en la región C-terminal, conocida como el dominio MEF2 (Gramzow *et al.*, 2010), el cual forma un parche hidrofóbico compuesto por una región de hojas beta que estabilizan su estructura cuaternaria. En estudios de mutagénesis realizados sobre la región C-terminal próxima al dominio MADS, se ha demostrado una reducción en el número de heterodímeros que pueden formarse entre los diferentes miembros de la familia MEF2, sin embargo; los heterodímeros que sí se forman, no ven alterada su unión al DNA ni su especificidad, por lo que se propone que las extensiones C-terminal juegan un rol fundamental durante la dimerización de estos TF's. (Bachinski *et al.*, 2010). Estudios similares en el dominio SAM que corresponde a la región C-terminal de las proteínas MADS de tipo I, corroboran esta hipótesis (Wu *et al.*, 2010).

1.6.2.d. Roles biológicos.

Las proteínas MADS participan en múltiples actividades biológicas, en levadura por ejemplo, la proteína MCM1 está encargada de la respuesta a feromonas, mientras que AGRI está encargada de la transcripción de un gran número de genes que codifican a las enzimas encargadas del metabolismo de la Arginina. Homólogos de SRF ha sido identificada en un gran número de organismos como *Homo sapiens*, donde lleva a cabo la regulación de genes encargados de la diferenciación de la epidermis; en *Drosophila* regula la diferenciación temprana de la tráquea, en mamíferos y en las ranas del género *Xenopus*, este TF está encargado de la regulación de un gran número de genes involucrados en la diferenciación de varios tipos de tejidos musculares; mientras

que en plantas una gran variedad de proteínas MADS son necesarias durante el desarrollo de órganos florales, fruto, tejidos de raíz y hoja (van Dijk & van Ham, 2010).

1.6.2.e. Secuencias CArGbox

El sitio de reconocimiento en el DNA ubicado en la región *upstream* de los genes regulados por factores de transcripción (elemento-cis) de la familia MADS, es denominado CArGbox. Correspondiendo al alto grado de conservación que encontramos en el dominio MADS de los TF's, los sitios a los que se unen dentro del DNA también están relacionados entre sí en cuanto a su composición y secuencia de bases; sin embargo, dentro de esta homogeneidad son fácilmente reconocibles las variantes que determinan la especificidad de cada uno de los integrantes de esta familia para con sus secuencia blanco. El tamaño promedio de la CArGbox es de 10 pb (de Folter & Angenent, 2006), aunque en el caso del heterodímero formado por las proteínas AGL15 y DTA1 de plantas, se han reportado CArGbox de hasta 12 pb de largo (Wang *et al.*, 2004).

Se han realizado un gran número de estudios para la caracterización de los motivos consenso a los que se unen estas proteínas, como es el caso de MEF2A (Andrés *et al.*, 1995). Existe una sección rica en Adeninas y Timinas en la parte media de la CArGbox que comprende por lo general los sitios 4, 5, 6 y 7 de dirección 5' → 3'; en estos, muy rara vez se localizan Citosinas o Guaninas, como lo indica el motivo general de CArGbox CC (A/T)₆GG (Shore & Sharrocks, 1995).

1.6.2.f. Relaciones filogenéticas entre proteínas MADS.

Todos los genes MADSbox poseen una secuencia altamente conservada de 180 pb que corresponde al dominio MADS en éstos factores de transcripción.

Análisis filogenéticos han demostrado que se puede ubicar al dominio MADS en dos linajes llamados tipo I y tipo II (Alvarez-Buylla. *et al.* 2000), ambos grupos se encuentran presentes en animales, plantas y hongos, lo cual sugiere que ocurrió un evento de duplicación en el linaje de las MADS anterior a la divergencia entre animales, plantas y hongos.

La primera de las ramas agrupa a las proteínas MADS de Tipo I, compuesta por aquellas que comparten una filogenia cercana con la proteína SRF (propia de animales y hongos), y las proteínas Tipo I de plantas. Estos dos grupos se diferencian debido a la secuencia de aminoácidos que da origen a la extensión del C-terminal, ya que las proteínas tipo SRF poseen el dominio SAM y las de tipo I de plantas un dominio que aún no se ha logrado identificar y que no tiene relación filogenética con el dominio SAM (De Bond *et al.*, 2003).

La segunda rama esta compuesto por las proteínas MADS de tipo II, que engloba a las proteínas tipo MEF que se encuentran en animales y hongos; y por otro lado, a las proteínas del tipo II de plantas que poseen una extensión hacia el carboxilo terminal compuesto por el dominio de intervención (I), keratin-like (K) y C-terminal (C) (Kaufman *et al.*, 2005), que no comparten relaciones filogenéticas con el dominio MEF, el cual posee características particulares que se mostrarán a continuación.

1.6.3. La proteína MEF2A

Al igual que todos los factores de transcripción MADS el factor de transcripción *myocyte enhancer factor 2* (MEF2) realiza un rol fundamental durante la morfogénesis, en este caso durante la miogénesis de los tipos musculares cardiaco, esquelético y liso (Olson *et al.*, 1995). Estudios bioquímicos y genéticos realizados en vertebrados e invertebrados, muestran que el factor de transcripción MEF2A regula la transcripción genética a través de la formación de homodímeros y heterodímeros que surgen de la unión con múltiples factores de transcripción, que lo ha inmiscuido en variados circuitos de control positivo o negativo .

1.6.3.a. Estructura y función.

Originalmente se asoció a MEF2A con la diferenciación de músculo esquelético a través del reconocimiento de una secuencia de DNA rica en A/T (Gossett *et al.*, 1989), subsecuentemente fue identificado su papel durante la morfogénesis temprana del músculo cardiaco, y en la última década se ha demostrado que bajo la represión de MEF2A se producen severas anomalías en todos los tejidos musculares y en el sistema nervioso central, por lo que se le asocia con la regulación transcripcional de numerosas proteínas estructurales.

En vertebrados existen cuatro genes que codifican diferentes isoformas de la proteína MEF2 referidos como *mef2a*, *-b*, *-c* y *-d* localizados en diferentes cromosomas (Morisaki *et al.*, 1997), que en forma de proteína suelen interactuar entre ellos para formar homodímeros o heterodímeros, lo que conlleva al reconocimiento de una gran gama de secuencias de DNA.

Los primeros 57 aminoácidos de MEF2 que conforman la caja MADS son los encargados del reconocimiento de su CArGbox, pero se ha demostrado que se requiere la interacción de los primeros 29 aminoácidos del dominio MEF2 para aumentar su afinidad al DNA y a la dimerización.

Como miembro de la familia MADS, el TF MEF2A posee una serie de residuos invariantes dentro del dominio MADS, además se ha comprobado que el dominio MEF2 durante el ensamblaje de alta afinidad al DNA, es mediador de la interacción entre proteínas durante la dimerización, y que también interactúa con ligandos accesorios que regulan su actividad (Huang *et al.*, 2000).

Entre los miembros pertenecientes al clado de las proteínas tipo MEF2, existe una gran identidad de secuencia, tanto en el dominio MADS como en el dominio MEF2, e inclusive esta similitud también está presente en las CArGbox a las que se unen estos factores de transcripción; se sabe que las secuencias de DNA a las que se une el TF de *Drosophila* (referido como D-MEF2A), son las mismas a las que se une MEF2A en vertebrados (Lilly *et al.*, 1994). Además dos factores homólogos de MEF2 identificados en levadura, también poseen secuencias blanco de DNA que se corresponden con las de vertebrados (Dodou & Treisman, 1997).

Se conoce la estructura cuaternaria de la proteína MEF2A, ha sido determinada tanto por cristalografía (Santinelli & Richmond, 2000) como por resonancia magnética nuclear (NMR) (Huang *et al.*, 2000). Esta última estructura fue acompañada de ensayos de unión a DNA utilizando diversas estructuras de MEF2A truncadas en el C-terminal, que fueron realizados con el propósito de determinar la porción mínima del dominio MEF2 a conservar para no perder su afinidad por el DNA. Para este experimento se realizaron 5 tratamientos

con fragmentos que van del primer aminoácido del N-terminal hasta el residuo: 85, 80, 76, 73, 70 y un control en el que se digirió completamente a la proteína por lo cual no se espera obtener algún registro de unión al DNA. Cada uno de los tratamientos fue expuesto a la misma cantidad de segmentos de DNA, posterior a la reacción en la que se formó el complejo DNA-MEF2A fueron purificados todos los complejos y corridos en un gel de agarosa con la finalidad de observar la densidad de banda formada por cada tratamiento para con ello deducir el número de complejos formados. Los resultados indicaron que el largo mínimo necesario para que MEF2A no pierda su afinidad por el DNA es de 72 aa partiendo del aminoácido 1 del N-terminal, ya que el ancho de banda que corresponde a éste tratamiento es equivalente al que se obtiene con los 2 primeros tratamientos y en el caso del tratamiento 4 se observa claramente una reducción de complejos formados, no así el caso del tratamiento 5 en el que no se observa ningún complejo formado y se asemeja al control negativo. De éste modo logró obtenerse un modelo confiable que garantiza el estudio adecuado de MEF2A con su CArGbox (Sharrocks *et al.*, 1993)

La estructura del homodímero de MEF2A unida al veinteámero de DNA 5'CTCGGCTATTAATAGCCGAG3' que contiene a la CArGbox del sitio 6 al 15, revela una región desordenada de los primeros 12 aminoácidos partiendo del amino terminal, una α -hélice formada por los residuos 13 al 37, dos hebras beta identificadas como I (residuos 41 al 48) y II (residuos del 52-58), los residuos que van de 61 al 73 se detecta una hélice-alfa que corresponde a parte del dominio MADS y del residuo 75 al 85 no pudieron ser detectados en la resonancia por ser insolubles aunque si participaron durante la unión al DNA. Con respecto a este se observó que está en conformación B con 9.9 pares de bases por vuelta con una distancia de entre 2.7-3.1 Å entre ellas, que adopta una curvatura de 17 grados al estar unido a el homodímeros de MEF2 y que solo el surco menor interactúa con el amino terminal del TF (Santinelli & Richmond, 2000).

1.6.3.b. Actividades de regulación de la expresión genética.

Numerosos estudios han descrito que la actividad de MEF2A como regulador está altamente enriquecida en aquellos genes que contribuyen a la diferenciación de los tejidos musculares y diferenciación del tejido neuronal (Andrés *et al.*, 1995), mientras que también se han hallado inmiscuidos en la regulación transcripcional de otras proteínas con funciones más diversas (Gong *et al.*, 2011), lo que pone en evidencia la compleja red de regulaciones en las que participa MEF2A producto de la regulación propia de ésta proteína.

1.6.3.c. Estructura de su CArGbox.

La secuencia de DNA a la que se une MEF2A es denominada como CArGbox, está formada por 10 pb y desde que se realizó un exhaustivo estudio para identificar los sitios de unión de MEF2A (Andrés *et al.*, 1995; Fickett 1996; Ornatsky & McDermott 1996) se ha aceptado por consenso el motivo canónico CTA(A/T)₄TAG; y aunque han habido nuevos genes identificados en la última década (Feng *et al.*, 2009; Flore *et al.*, 2009; Gong *et al.*, 2011;) el motivo sigue siendo el mismo ya que no se ha realizado un estudio que incorpore las características de estas nuevas secuencias al motivo canónico antes mencionado, del cual se establece que no incluye Citosinas

o Guaninas en las posiciones 4, 5, 6 y 7 y solo se han reportado algunas variaciones en la posición 1 en la que puede existir Adenina y en la posición 10 que puede presentarse Timina, pero en muy raros casos (Ornatsky & McDermott, 1996).

1.7. MÉT ODOLOGÍA DEL CAPÍTULO I.

1.7.1. Determinación de la RMSD.

Con el propósito de comparar las posibles variantes estructurales existentes en el arreglo espacial de la CArGbox y MEF2A dependientes de la composición de los ácidos nucleicos del elemento -cis y las secuencias de DNA en la que esta inmersa la CArGbox, se obtuvo la RMSD proveniente de los diferentes modelos del complejo MEF2A-CArGbox que hasta el momento se han determinado tanto por cristalografía como por RMN. Atendiendo a la compatibilidad estructural entre aminoácidos y nucleótidos, el estudio se dividió en dos grupos: el primero corresponde a los polipéptidos y el segundo constituido por sus secuencias blanco de oligonucleótidos; ambos conjuntos fueron extraídos del Protein Data Bank (www.rcsb.org) respetando las coordenadas originales. La comparación estructural se realizó por medio de la determinación de la distancia media cuadrática mínima o RMSD computada por el programa de modelado molecular y visor de estructuras Visual Molecular Dynamics (VMD), quien obtiene dicho valor al resolver la ecuación:

$$RMSD(N; x, y) = \left[\frac{\sum_{i=1}^N w_i \|x_i - y_i\|^2}{N \sum_{i=1}^N w_i} \right]^{\frac{1}{2}}$$

Donde cada una de las coordenadas atómicas de las proteínas \underline{x} y \underline{y} corresponden al número de carbonos- α (N) ponderados por w_i .

Únicamente se consideró el *backbone* y por lo tanto los carbonos- α en el caso de las proteínas y en el de los nucleótidos fueron tomados en cuenta los esqueletos de fosfatos y azúcares. La RMSD fue medida por los dos métodos presentes en VMD *top* y *average*; en el primer caso la estructura de MEF2A utilizada en la determinación de la energía libre de unión (ΔG_{bind}) que se explica en el Capítulo III (id:1C7U) y la estructura determinada por He en el 2010 (id:3P57) fueron comparadas con el dímero reportado por Wu en el 2011 (id:3KOV). En la determinación por el método *average*, VMD computa el promedio de todas las coordenadas \underline{x} , \underline{y} , y \underline{z} correspondientes a cada uno de los puntos comparables de las tres estructuras, como consecuencia obtiene un modelo que representa el promedio de los valores espaciales de las moléculas comparadas, este modelo virtual posteriormente se compara con cada uno de los modelos reales incluidos en el análisis.

Para la alineación estructural de los dímeros se tomó en cuenta la sección de aminoácidos que va del 2 al 70. En el caso de las CArGbox se dejaron fuera las secuencias de nucleótidos que flanquean el sitio de unión al factor de transcripción, por lo que se incluyó únicamente la región correspondiente a la CArGbox.

1.8. RESULTADOS DEL CAPÍTULO I.

1.8.1. Comparación de las proteínas y ácidos nucleicos presentes en los distintos complejos MEF2A-CArGbox alternativos.

La RMSD obtenida por los métodos *average* y *top* en la comparación estructural de los péptidos y oligonucleótidos se muestra en la tabla 1.

Tabla 1. RMSD de tres estructuras de MEF2A y sus correspondientes CArGbox. Se muestran los diferentes valores de RMSD individuales y totales computados por el método *top* y *average* para polipéptidos y oligonucleótidos.

	id	MEF2A	CArGbox
<i>top</i>	3KOV	0 Å	0 Å
	3P57	0.45 Å	0.69 Å
	1C7U	2.39 Å	0.87 Å
RMSD Total		1.72 Å	0.79 Å
<i>average</i>	3KOV	0.86 Å	0.37 Å
	3P57	0.75 Å	0.46 Å
	1C7U	1.55 Å	0.48 Å
RMSD Total		1.11 Å	0.44 Å

En los resultados individuales de RMSD computados por el método *top*, se obtuvo un mayor grado de similitud entre las estructuras con el No. de id:3KOV y 3P57, en comparación al observado en la alineación de cualquiera de estos dos con el dímero reportado en Huang *et al.*, 2000 (1C7u); ésto sucedió tanto en el grupo de polipéptidos como en el de ácidos nucleicos; lo cual indica una mayor disparidad estructural en la conformación espacial de este último complejo para con el resto (Fig. 2); siendo evidente la misma cuestión al observar los valores obtenidos con el método *average*. En ambos criterios de medición, el valor total de RMSD es bajo, sobre todo en el caso de la comparación realizada a las CArGbox en donde es cercano al 0.

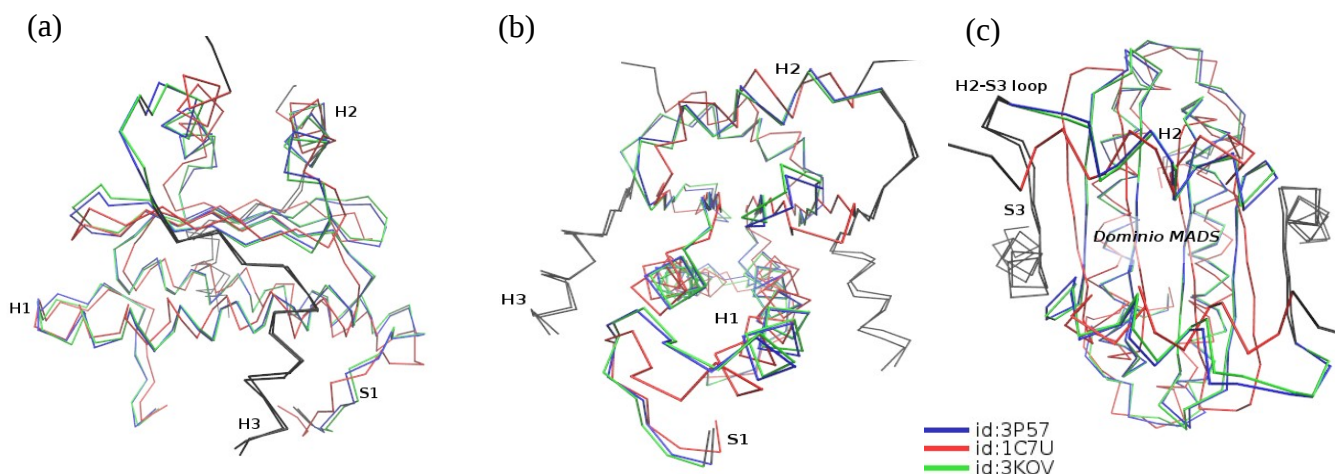


Fig.2. Superposición del esqueleto de las proteínas MEF2A extraídas del complejo TF-elementocis. (a)Vista frontal en diagrama de delineado. En rojo el complejo que no incluye el dominio MEF2 (residuos 2-74), en verde el complejo que incluye el dominio MEF2A (residuos 2-95), en azul el dímero que interactúa con el dominio TAZ2 de la proteína p300 (residuos 2-95) y en negro los residuos que no fueron incluidos en el alineamiento correspondientes a la sección H2-S3loop y H3. (b) Vista lateral en la que se observa estructuralmente conservada la sección inmediata al amino terminal (S1) y la hélice H1 en los tres dímeros. (c) Vista superior del alineamiento en la que se observa el empalme de H2 correspondiente al comienzo del dominio MEF2 y la pérdida de similitud en esta sección debido a la exclusión de S3 y H3 en la estructura reportada por Huang En la parte inferior del diagrama se muestra el No. de id. correspondiente al archivo pdb del cual se extrajeron las estructuras.

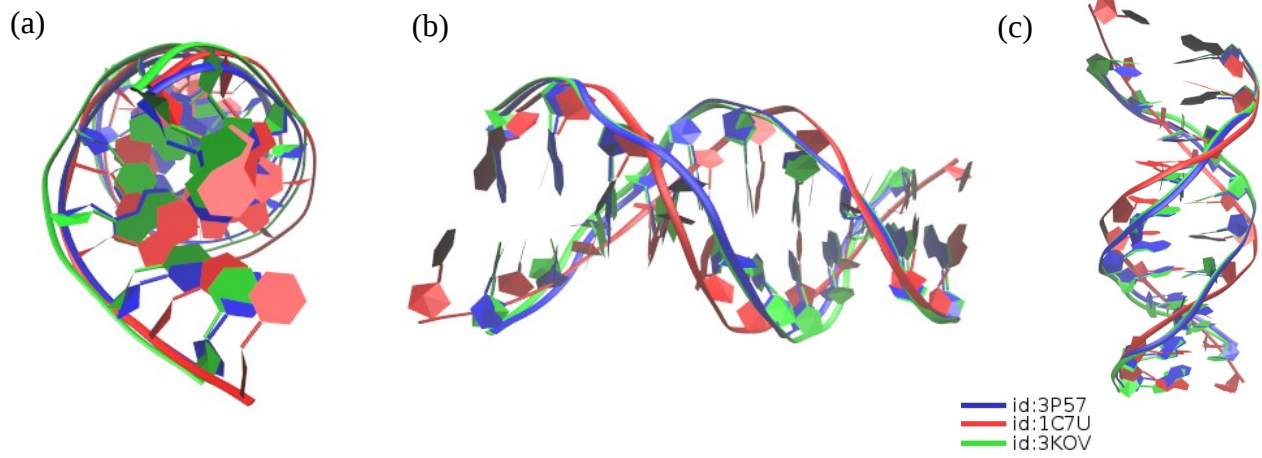


Fig. 3. Alineamiento estructural del esqueleto de fosfatos y azúcares de tres muestras de CARGbox. (a) Vista superior representada en diagrama de cintas; en rojo CARGbox (5'CTATTAATAG3') extraída del complejo formado con MEF2A (residuos 2-74), en verde CARGbox (5'CTATTTATAA3') extraída del complejo formado con MEF2A (residuos 2-95), en azul CARGbox (5'CTATTTATAA3') extraída del complejo formado con TAZ2-MEF2A (residuos 2-95). En la parte inferior de la figura se muestra el No. de id. Del archivo del cual se extrajeron las estructuras. (b) Vista lateral en la que es apreciable la similitud estructural del esqueleto de fosfatos en las tres estructuras. (c) Vista frontal de los oligonucleótidos en la que se observa la complementariedad en la orientación de azúcares y bases.

1.9. DISCUSIÓN DEL CAPITULO I

Con el objetivo de hacer eficiente el tiempo de cómputo durante la determinación de la ΔG_{bind} , se optó por utilizar el complejo MEF2A-CARgbox con No. de id:1C7U debido a que dentro de los reportes en los que se encuentra presente dicha estructura (publicados en el *Protein Data Bank* a la fecha); es la que menos átomos incluye y por lo tanto, *a priori* la más factible de integrar a los experimentos. La proteína MEF2A que conforma dicho complejo posee trunco el dominio MEF2S (residuos 59-74) a diferencia del complejo reportado en Wu *et al.*, 2010 (No. de id:3kov) que incluye el dominio MEF2S constituido por las α -hélices H2 (residuos 61 al 72) y H3 (residuos 80 al 90), al igual que en el dímero reportada en He *et al.*, 2011 (No. de id:3P57), con la variante que ésta fue determinada durante su interacción con el dominio TAZ2 de la proteína p300.

Existió en un inicio la posibilidad de que durante la unión del complejo MEF2A-CARgbox, la proteína y el DNA mostrasen una conformación diferente a la observada en el reporte de Huang, debida a la alteración de la secuencia de nucleótidos en la CARgbox y/o la adición del dominio MEF2S; en ese caso los resultados obtenidos en la medición de la ΔG_{bind} , no serían biológicamente relevantes, ya que la sustitución de los oligonucleótidos en el complejo se realizó sin modificar las coordenadas espaciales de los fosfatos y desoxirribosas presentes en la CARgbox descrita en este trabajo, por lo que si estos parámetros no son cercanos a los que se observan *in vivo*, como podrían sugerir variaciones observadas en otras determinaciones cristalográficas o de RMN, los cálculos de la ΔG_{bind} conllevarían un grado intrínseco de error.

Debido a estas percepciones fue de suma importancia realizar un estudio comparativo con respecto a algunas de las estructuras alternas reportadas, para que se determinase si existe o no, dicha variabilidad. Los resultados de los valores de RMSD totales mostraron mínima disparidad, ya que obteniéndolos tanto por el método *top* y *average* (que en el caso de las proteínas fueron de 1.72 Å y 1.11 Å respectivamente), fueron menores a 3 Å por lo que se les considera con alto grado de similitud (Reva *et al.*, 1998). Aún así, analicemos las diferencias estructurales que existen para ponderar de forma individual su relevancia. En los resultados arrojados bajo los parámetros del método *top*, se observó un mayor grado de similitud entre las proteínas reportadas en los pdb con No. de id:3KOV y 3P57 en comparación con la estructura descrita por Huang; tal hecho resulta evidente si tomamos en cuenta los siguientes factores: la sección H2 de la proteína MEF2A con el dominio MEF2S trunco, posee un mayor grado de desorden en dicha hélice, el cual se acentúa mientras más se aproxima al carboxilo terminal, a diferencia de las proteínas en las que el dominio MEF2S está completo; también se observa un mayor grado de disparidad en la sección H1 del dímero que carece de dicho dominio con respecto a los que lo poseen, esto debido a los posibles choques estéricos que surgen de la interacción entre las α -hélices H3 del dominio MEF2S y las α -hélices H1 presentes en el dominio MADS de MEF2A(Wu *et al.*, 2010).

Con base en estas observaciones, podemos afirmar que la estructura cuaternaria de MEF2A está conservada dentro del complejo que forma con la CArGbox, aún cuando el dominio MEF2S no está presente y por lo tanto si las mediciones de ΔG_{bind} fueran hechas en base a la estructura que sí posee tal dominio, estas serían equivalentes a las descritas en esta tesis.

Los oligonucleótidos poseen un valor de RMSD total de 0.79 Å y 0.44 Å obtenidos por el método *top* y *average* respectivamente, lo cual sugiere que la arquitectura de la CArGbox cuando se encuentra unida a MEF2A no varía de forma radical, aún cuando existan diferencias en su secuencia, como fue el caso de la comparación realizada en los oligonucleótidos 5'CTATTAATAG3' y 5'CTATTTATAA3' presentes en los complejos con No. de id: 1C7U y 3KOV respectivamente. Los resultados anteriores sugieren que la sustitución de bases que se realizó en la estructura tridimensional de la CArGbox reportada en Huang *et al.*, 2000, no poseen un alto grado de error en cuanto a las coordenadas espaciales de las estructuras tridimensionales de los oligonucleótidos construidos *in silico*.

Por su relevancia biológica, cabe recalcar el alto grado de similitud en la arquitectura general del complejo bajo las tres condiciones reportadas, es decir, cuando el motivo MEF2S (que va del aminoácido 59 al 95) está presente en la proteína, cuando está truncado e incluso cuando MEF2A se encuentra unido al dominio TAZ2 de la proteína p300, ya que podemos inferir que durante la evolución de la familia de proteínas MEF2, la estructura y función del dominio MADS fue altamente conservada de forma independiente al dominio MEF2A, hecho que resulta evidente al observar la filogenia de las proteínas con dominio MADS, en la que se observa altamente conservada su secuencia a diferencia de los dominios particulares que dividen a las subfamilias (Shore & Sharocks, 1995). En el análisis realizado por He *et al.*, 2011, observamos que aún sometiendo nuestro dímero a las interacciones electrostáticas provenientes de la unión con la proteína p300, poco tienen que ver estas con el arreglo estructural de MEF2A en la zona de interacción con el DNA, por lo que tomándolo en cuenta podemos sugerir fuertemente que el modelo con el cual estamos trabajando es biológicamente representativo y por lo tanto promoverá que los resultados en el cálculo de la ΔG_{bind} lo sean también.

Los enunciados anteriores están respaldados por los ensayos de unión al DNA realizados por medio de la técnica del análisis de retardo en gel y reportados por Huang *et al.*, 2000. Dichas pruebas consistieron en el apareamiento de la CArGbox con secuencia 5'CTATTTATAG3' y diferentes dímeros de MEF2A, cuya peculiaridad consistía en truncar su secuencia en el extremo carboxilo, obteniendo de este modo polipéptidos con residuos que abarcaban del 1 al 85, del 1 al 80, del 1 al 76, del 1 al 73 y del 1 al 70. El objetivo de estos experimentos consistió en saber cuál era la longitud mínima que debía poseer MEF2A sin que su afinidad por la CArGbox se viera afectada, la conclusión a la que se llegó fue que la proteína con 73 residuos de aminoácidos era el mínimo suficiente para que MEF2A realizara dicha unión, ya que este polipéptido conservaba por

completo el dominio MADS responsable de la afinidad por el DNA y los residuos 60 a 73 del dominio MEF2S, que resultaron ser los mínimos e indispensables para estabilizar al resto de la proteína.



CAPÍTULO II

Estudios de dinámica molecular

2.1. Planteamiento del problema del capítulo II.

La estructura resuelta por NMR del complejo MEF2A-CArGbox reportada en Huang 2000 *et al.*, es de gran valor en el estudio de los factores de transcripción MADS, específicamente porque proporciona un modelo realista y detallado del sistema, en el que tenemos oportunidad de observar sus características con el excepcional detalle del nivel atómico. En este estudio, dicha descripción será utilizada para ampliar la comprensión que hasta ahora se ha mostrado del complejo antes referido al superar las limitantes propias de un modelo estático (aunque hasta el momento nos halla llenado de valiosas observaciones), debido a que es una perspectiva fragmentaria, es decir, una realidad provisional a razón de claudicar ante los principios de movimiento y cambio esenciales para entender a nuestro arquetipo de partida (pdb 1C7U) como una transición entre sus estados pasados y los que le habrán de ocurrir, importante cuestión si recordamos que en la naturaleza nuestro objeto de estudio no permanece estático y existe provisto de cambios de estado y movimiento. En ésta sección nos enfocaremos a los procesos que han de sobrevenir en ambas moléculas una vez que su unión se establece, como se ilustra en el archivo .pdb 1C7U (Huang 2000 *et al.*) y que será tomado como semilla para cumplir el propósito antes expuesto.

La importancia de obtener más información a cerca de las transiciones posteriores a la formación del complejo es fundamental, porque aunque no interfieren en el cálculo de ΔG_{bind} , tienen repercusiones teóricas sobre éste, ya que llevándolas a su ejecución, estaremos en posibilidad de dar un vistazo a los estados futuros del sistema bajo la luz de las fuerzas físicas y químicas que rigen sus propiedades y su comportamiento en un lapso de tiempo determinado, enriqueciendo también el conocimiento que tenemos acerca de los nuevos elementos-cis inferidos en la presente tesis, al tener la posibilidad de comprender de forma general su interacción con MEF2A tras estudiar el comportamiento observado en nuestro arquetipo. Para lograr comprender el sistema en movimiento se echará mano de una dinámica molecular.

Así, el presente trabajo se convierte no sólo en un método alternativo en la determinación de secuencias blanco para un factor de transcripción dado, sino en una metodología completa para el estudio de los factores de transcripción y sus secuencias blanco, al integrar el estudio del sistema en reposo (necesario durante la medición de la ΔG_{bind}) y el estudio del complejo en plena transformación, en el que podrán observarse las consecuencias de su formación.

2.2. Objetivo particular del capítulo II.

Estudiar durante un lapso de tiempo de 25 ns las interacciones moleculares observadas dentro del complejo MEF2A-CArGbox reportado en el archivo pdb 1C7U.

2.3. Hipótesis del capítulo II.

Si el complejo MEF2A-CArGbox es estudiado bajo el fundamento del movimiento, podremos tener una aproximación de las interacciones moleculares y de los residuos específicos que hacen posible la unión del complejo durante un lapso de tiempo determinado.

2.4. Sistema experimental del capítulo II.

- Archivo PDB 1C7U

2.5. Software del capítulo II.

- GROMACS v4.5.5.
- Grace v5.1.23
- VMD
- Poliview 2D

2.6. INTRODUCCIÓN AL CAPÍTULO II

2.6.1. Factores que intervienen en la unión de un TF y su secuencia blanco.

La unión entre el factor de transcripción y su secuencia blanco está mediada por un amplio espectro de características físicas y químicas que determinan su afinidad y su interacción durante la formación del complejo TF-DNA.

2.6.1.a. El llamado a distancia

Generalmente se restringe la interacción entre dos moléculas a una cercanía física representada por la unión de un complejo, por ejemplo la unión de un TF con su secuencia blanco de DNA, sin embargo; la interacción entre moléculas antecede por mucho a una mínima cercanía física. La fuerza eléctrica (como la gravitacional) actúa a distancia, lo que significa que existe una fuerza entre dos objetos, incluso cuando éstos no están en contacto; por ello es de suma importancia conocer los aspectos que refieren al conocimiento del campo eléctrico en las moléculas. Éstas poseen de manera inherente una naturaleza eléctrica que recae en primera instancia en las partículas subatómicas, manifestada en las fuerzas que rigen la interacción entre cada uno de los átomos. Cada molécula posee una carga eléctrica particular que puede permanecer constante o verse alterada a causa de la interacción con su medio o con otras moléculas, surgiendo de este modo fuerzas electrostáticas que influyen en su comportamiento y su interacción con otros cuerpos, debido a la distribución de cargas que emerge alrededor de la molécula y que es descrita por el campo eléctrico, utilizado para describir la interacción entre dos sistemas de naturaleza eléctrica, siendo ésta una razón vital para comprender las interacciones de este tipo entre moléculas orgánicas.

Una forma útil de comprender estas interacciones moleculares es con la idea de campo desarrollada por Michael Faraday. De acuerdo con él y con respecto a la electricidad, un campo eléctrico (E) se extiende hacia afuera desde toda carga, e impregna el espacio, si esta carga es perturbada por una segunda, se le puede denominar carga q_2 . La primera resentirá una fuerza ejercida por el campo eléctrico. En principio es posible determinar el campo eléctrico que rodea a una molécula midiendo la fuerza sobre una carga de prueba positiva, ya que el campo eléctrico se define en términos de la fuerza sobre tal carga de prueba positiva. De modo simplista el campo eléctrico, en cualquier punto en el espacio se define como la fuerza F ejercida sobre una pequeña carga de prueba positiva, dividida por la magnitud de la carga de prueba q. De esta manera E se define como el límite de F/q , con q tendiendo a cero cuando es tan pequeña que no manifiesta fuerza sobre las otras cargas que crean el campo. A partir de esta definición se observa claramente que el campo eléctrico es un vector que indica la dirección como un punto en el espacio al que se desplaza una carga.

$$\vec{E} = \frac{\vec{F}}{q}$$

Si bien estas son las bases, claro está que el anterior es un modelo simplista a comparación de las consideraciones que se deben tener para calcular el campo eléctrico de una proteína. En el caso de estos sistemas

se comienza calculando el radio de cada átomo y su carga, posterior a esto se realiza un cálculo de las interacciones interatómicas de la molécula para obtener una dirección de carga conjunta y también una extensión de la carga, los vectores obtenidos son utilizados para formar la distribución de carga biomoléculas calculada por la ecuación Poisson-Boltzmann (PB) la cuál será explicada dentro de esta misma sección mas adelante.

Como ya se mencionó, el campo eléctrico es una magnitud vectorial, lo que significa que puede deducirse de ella una dirección y una distancia de emisión. Una isosuperficie es una representación gráfica de la distribución de carga biomoléculas que por medio de una superficie volumétrica tridimensional, representa estos valores constantes, así mismo otorga la oportunidad de analizar los parámetros ya mencionados, así como la regionalización de cargas positivas y negativas.

2.6.1.b. Interacciones de cercanía física

En caso de que las fuerzas a larga distancia favorezcan la unión entre dos moléculas, entonces entrarán en acción aquellas fuerzas responsables de mantener estable, aveces por tiempo indefinido, la asociación molecular.

Actualmente el puente de hidrógenos es una estructura energética aceptada en la química, resulta difícil creer que cuando fue propuesto por su autor en la década de los 60's, muchos lo viesen con escepticismo hasta su irrefutable descripción cristalográfica (Taylor & Kennard, 1982). El puente de hidrógeno ha ocupado un elemento muy importante para entender las interacciones entre moléculas biológicas, ya que la gran mayoría de éstas se sostienen por este tipo de fuerzas. El puente de hidrógeno se forma cuando el H está unido a un átomo altamente electronegativo y establece una relación dipolo-dipolo con un átomo donador de electrones.

Desde el punto de vista biológico, resulta muy interesante entender por qué el puente de hidrógeno es el principal mediador de las interacciones intermoleculares. La fortaleza de este enlace varía mucho con respecto a la naturaleza del átomo electronegativo y el donador de electrones, pero siempre será mucho más débil que la de un enlace covalente, su magnitud oscila entre los 5 y 30 kJ/mol, sin embargo en interacciones cooperativas, todas estas fuerzas son sumadas, y en conjunto establecen interacciones cercanas a las de un enlace covalente.

Esta naturaleza ambigua le proporciona a las moléculas la capacidad de asociación manteniendo la fuerza necesaria para quedar unidos y llevar a cabo su actividad biológica, pero al mismo tiempo son los suficientemente débiles para ceder y liberar al complejo antes formado. Esta parecería ser una labor poco relevante, pero tómese en cuenta que si se uniera al TF con el DNA utilizando fuerzas mayores que impidieran la disociación del complejo, la regulación genética se vería alterada temporalmente y lo mismo sucedería si se unieran con fuerzas muy bajas.

2.6.2. Dinámicas moleculares.

Los avances cultivados hace más de un siglo por la química teórica, con la que ha sido posible crear representaciones formales de los sistemas moleculares, han permitido aplicar algoritmos con los que es posible representar la interacción entre moléculas; esta metodología se ha apoyado fundamentalmente en los avances

tecnológicos de la computación por la amplia capacidad que poseen los ordenadores para resolver un conjunto muy elevado de cálculos, que hacen posible el estudio de sistemas moleculares complejos. De esta manera la química computacional tiene un alto impacto para la biología, cuyo objetivo principal es proporcionar modelos realistas que ayuden a comprender el funcionamiento de los sistemas biológicos a través de los principios básicos de la física y la química.

Estudios de este tipo pueden ser realizados en dos niveles de aproximación: el primero utiliza la mecánica cuántica que conlleva un nivel de escrutinio en aproximaciones lo más detalladas posibles, su uso requiere la resolución de la ecuación de Schrödinger, que hasta el momento es computacionalmente imposible de resolver para grandes sistemas biológicos (como asociaciones multimoleculares). Por el contrario, utilizando los métodos de mecánica clásica, se simplifica el cálculo de estos sistemas, economizando los factores de intervención de muchos efectos que pueden ser descartados y que aún así no demeritan el realismo de la simulación. En concreto, los métodos de mecánica clásica abandonan el movimiento electrónico y calculan la energía como una función de las coordenadas nucleares; esto conlleva a que no sea posible estudiar los estados de transición de un enlace, su reactividad o complejas estructuras electrónicas, pero siguen siendo muy útiles para resolver una gran cantidad de planteamientos bioquímicos y biofísicos que explican algún evento biológico en particular (puede contener dentro de su sistema más de 80,000 átomos que solo compongan el solvente, como el caso que se presenta en esta tesis) .

Los métodos de dinámica molecular son los responsables de poner en práctica la teoría antes descrita, se definen simplemente como la descripción general de un sistema químico complejo en términos de un modelo atómico realista que sirve para entender y predecir propiedades macroscópicas basadas en detalles conocidos a una escala atómica. Todo ello observado sobre un modelo dinámico bajo las leyes de la física y la química.

De este modo a partir de un conjunto de coordenadas, tomadas por ejemplo de las moléculas determinadas por RMN o cristalografía, podemos asignar las velocidades y energía propias de cada átomo a partir de los supuestos no cuánticos de la distribución de Boltzmann; dada la temperatura, las sucesivas coordenadas y las velocidades se obtienen de la integración de la ecuación de Newton para el movimiento en cada dirección como una coordenada. En una dimensión, la ecuación se puede escribir como:

$$\frac{d^2 x_i}{dt^2} = \frac{F_{x_i}}{m_i}$$

donde m_i y x_i representan la masa y la posición de cada átomo respectivamente, y F_{x_i} es la derivada del potencial de acuerdo a una ecuación de campo de fuerza (cuyo concepto será explicado en la sección 2.6.2.c.). El resultado es una trayectoria que muestra cómo las posiciones atómicas y velocidades, evolucionan con el tiempo de acuerdo a la influencia de los átomos presentes en el sistema.

De este modo, la dinámica molecular (MD) ha proporcionado muchos conocimientos sobre el comportamiento de macromoléculas desde las primeras proteínas que se estudiaron bajo esta metodología a

finales de la década de los 80's del siglo pasado 25 años (McCammon & Harvey; 1987). Su poder radica en el alcance que tiene para poder investigar movimientos e interacciones moleculares que a menudo no son accesibles a la experimentación por su velocidad y su escala microscópica. Esta herramienta junto a las teorías biológicas, resulta de gran poder en la investigación.

2.6.2.a. Interacciones atractivas y repulsivas.

Para lograr entender la forma en la que la dinámica molecular modela sistemas sujetos a fuerzas atractivas (interacciones de largo alcance) y repulsivas (interacciones de corto alcance), se explicarán brevemente.

Las interacciones de largo alcance pueden dividirse en energía electrostática, energía de inducción y energía de dispersión. Las cuales serán descritas mediante la mecánica clásica.

La energía electrostática es una interacción que se da en especies cargadas como en aquellas en carga neta, pero que presentan un momento bipolar permanente. Son interacciones coulombicas que dependen de la separación y la orientación relativa de las moléculas. La distribución de carga se da en función de los momentos multipolares, por este motivo se requiere una suma que incluya la presencia de los principales momentos.

En el caso de las interacciones dipolo-dipolo globales se promedian las orientaciones relativas, considerando que no todas son igualmente probables.

La energía de inducción surge como consecuencia de los momentos dipolares que dependen de la distribución de carga que surge cuando una molécula se ve afectada por campos eléctricos externos. Para las moléculas isotrópicas este momento dipolar inducido puede cuantificarse en la dinámica molecular a partir del campo eléctrico y la polarizabilidad.

La energía de dispersión descubierta por London en 1930, no puede ser descrita completamente en términos clásicos, ya que su origen es mecanocuántico, sin embargo, una descripción clásica puede ayudar a comprenderla si se considera que en una molécula, debido al continuo movimiento electrónico, puede generarse en un momento dipolar instantáneo, el cual induce un dipolo en un átomo vecino, como en el caso de las fuerzas de inducción clásicas.

Las interacciones de corto alcance describen la repulsión entre moléculas que surge del solapamiento de núcleos y nubes electrónicas de dos átomos que están muy próximos. De acuerdo a esta descripción se toma el modelo de enlace valencia de Heitler-London que permite tener una descripción de dos fuerzas repulsivas en el caso de dos átomos de hidrógeno, que a pesar de no acercarse completamente a la realidad, tiene la ventaja de establecer claramente los orígenes de las fuerzas repulsivas y es sencillo de computar.

2.6.2.b. Representación de las interacciones moleculares.

Es complejo en la actualidad el cómputo de las interacciones moleculares de cualquier sistema mediante estudios mecanocuánticos. Para resolver el problema se obtiene el cálculo de las propiedades efectuado mediante un potencial propuesto, cuyos parámetros se ajustaron a manera de que estas propiedades coincidan con los

resultados experimentales. La función potencial más conocida es la de Lennard-Jones (LJ) en la que el potencial depende de dos parámetros ajustables: el diámetro de colisión y un término repulsivo de corto alcance, resultado del solapamiento de las nubes electrónicas, además de un tratamiento atractivo debido a las fuerzas de van de Waals. El hecho de que no se consideren los términos atractivos superiores y que la energía repulsiva no varíe exponencialmente, como predice la teoría, da lugar a que este exponencial sólo pueda ser utilizado bajo ciertas parametrizaciones de un conjunto de moléculas, por lo que su uso corresponde a una cuestión de simplicidad.

2.6.2.c. Campo de fuerza (FF)

La descripción de la interacción atómica dentro de un sistema puede ser descrito de forma clásica mediante un campo de fuerza. Un campo de fuerza (o FF por sus siglas en inglés) es un conjunto de parametrizaciones de física clásica acerca del movimiento y la energía que contienen los cuerpos. Tiene por objetivo explicar la interacción de moléculas asignándoles propiedades medibles que permiten calcular la magnitud de la interacción de todos los componentes del sistema. El campo de fuerza está compuesto de un conjunto de ecuaciones que se utilizan para generar la energía potencial y los parámetros empleados en estas ecuaciones.

Hoy en día existen cuatro campos de fuerza de uso común para la simulación de sistemas moleculares: AMBER (Pearlman, 1995), CHARMM (Brooks, 1983; MacKerell 1998; Mackerell 2004), GROMOS (Van Gunsteren 1987) y OPLS (Jorgensen 1996).

La energía potencial intramolecular de los enlaces, suele ser representada por osciladores armónicos, durante los momentos angulares y el estiramiento de éstos, por lo que se desarrolla una serie de Fourier para cada ángulo de torsión, la tensión del enlace, y del conteo de las fuerzas coulombicas, además del potencial Lennard-Jones (L-J). De este modo se representan las interacciones entre átomos separados por tres o más enlaces. Los dos últimos términos a que se refiere juntos como términos no-consolidado-se evalúan entre todos los átomos pares en el sistema para producir la energía intermolecular. Estos campos de fuerza calculan la energía como la suma de los términos que representan la elongación del enlace, sus momentos angulares, la deformación dihedral y las interacciones de no enlace de la siguiente manera:

$$E = \sum_{bonds} V^{str} + \sum_{angles} V^{bend} + \sum_{torsion} V^{tors} + \sum_{LJ} V^{LJ} + \sum_{Coulomb} V^{Coul}$$

Las primeras tres sumatorias de la ecuación corresponden a energías de enlace (incluyen átomos conectados hasta en tres enlaces consecutivos) y las dos últimas se refieren a las energías no condicionadas por un enlace. Las sumatorias pueden ser fácilmente calculadas a partir de las coordenadas del sistema en un momento dado. Para cada par de átomos enlazados (i y j), la tensión en el enlace se calcula como:

$$V_{ij}^{str}(r_{ij}) = k_{ij}^r (r_{ij} - r_{ij}^0)^2$$

Donde k_{ij}^t es la fuerza de tensión constante y r_{ij} y r_{ij}^0 son la distancia entre átomos y el equilibrio de la longitud del enlace, respectivamente. Para cada grupo de tres átomos enlazados (i, j, k), el término angular es descrito como:

$$V_{ijk}^{bend}(\theta_{ijk}) = k_{ijk}^\theta (\theta_{ijk} - \theta_{ijk}^0)^2$$

donde k_{ijk}^q representa la fuerza constante de flexión, q_{ijk} y q_{ijk}^0 son los ángulos entre los átomos y su valor de equilibrio respectivamente. Del mismo modo para un grupo de cuatro átomos entrelazados (i, j, k, l) el término dihedral se representa como una expansión del coseno:

$$V_{ijkl}^{tors}(\Phi_{ijkl}) = k_{ijkl}^\Phi [1 + \cos(n\Phi_{ijkl} - \Phi_{ijkl}^0)]^2$$

donde k_{ijkl}^f es un diedro constante que afecta la distancia entre los átomos enlazados, n es el número de mínimos en una rotación de 360 °, y Φ_{ijkl} y Φ_{ijkl}^0 son el ángulo diedro y el valor de equilibrio según la convención bioquímica ($\varphi = 180^\circ$ trans, cis $\varphi = 0^\circ$ y $\varphi = 60^\circ/300^\circ$), respectivamente. El primer término de “no enlace” es a menudo representado por un potencial L-J de 12.6, un modelo matemático sencillo (Lennard-Jones 1931) que representa dos fuerzas distintas (un atractivo y un repulsivo) como se muestra a continuación:

$$V_{ij}^{LJ}(r_{ij}) = \left[\left(\frac{A_{ij}}{r_{ij}} \right)^{12} - \left(\frac{B_{ij}}{r_{ij}} \right)^6 \right]$$

donde A_{ij} y B_{ij} son parámetros que dependen de cada par de átomos. El primer término refleja las fuerzas de atracción de largo alcance (van der Waals o dispersión) y el otro para las fuerzas repulsivas de corto alcance, como resultado de la superposición entre los orbitales electrónicos. Finalmente, el último término en la ecuación de campo de fuerza es un Potencial de Coulomb que describe las interacciones electrostáticas:

$$V_{ij}^{Coul}(r_{ij}) = \frac{q_i q_j}{4\pi \epsilon_0 r_{ij}}$$

en donde q_i y q_j son las cargas de los átomos i y j . r_{ij} es la distancia relativa entre el par de átomos y ϵ la permitividad del vacío.

En resumen, los campos de fuerza están compuestos de la sumatoria de las energías provenientes de la tensión, ángulos y torsiones de los enlaces, así como de las interacciones electrostáticas de largo alcance y parámetros particulares derivados de diferentes métodos teóricos o experimentales, tales como: la espectroscopía de migración para calcular las corrientes de fuerza; rayos X y microondas para calcular las geometrías, contactos

de gases raros para dilucidar las fuerzas de VW, etc. La forma que adquieran estos parámetros será particular de cada campo de fuerza. De éste modo es como los FF permiten una descripción de la energía potencial de las moléculas empleando la descripción de las magnitudes anteriores; de modo que pueden describirse en la siguiente ecuación general:

$$E_{total} = E_{enlaces} + E_{noenlace} + E_{otras}$$

Para lograr esta parametrización, cada campo de fuerzas contiene ciertos bloques de información necesarios y elementales que son: una lista del tipo de átomos que se tomará en cuenta junto con una nomenclatura particular, una lista de cargas atómicas, formas funcionales para los componentes de la expresión energética final y una serie de parámetros fisicoquímicos que lo harán más o menos adecuado para el uso de un sistema biológico en particular. En el presente trabajo se utilizó el campo de fuerzas AMBER.

2.6.2.d. Campo de fuerzas AMBER (*Assisted Model Building and Energy Refinemed*).

En 1984 se desarrolló la primera versión del campo de fuerzas AMBER (Weiner et al., 1984). Introdujo importantes mejoras en los parámetros del término electrostático y las interacciones de van der Waals del campo de fuerzas desarrollado por Karplus anteriormente.

Su función de energía está constituida por seis términos: tensión, flexión, torsión, interacciones de van de Waals, electrostáticas e interacciones electrostáticas por puente de hidrógeno. Los términos de tensión y flexión están descritos mediante funciones armónicas y no se consideran influencias por átomos o grupos químicos vecinos, es decir, el termino cruzado no se maneja para este campo de fuerza.

Las cargas atómicas del término electrostático se derivan a partir de cálculos mecanocuánticos, por lo cual este método permite el uso de una constante dieléctrica que depende de la distancia, cuyo objetivo es mimetizar los efectos de la polarización durante las interacciones atractivas, poniendo el mayor peso a las interacciones más próximas.

Las interacciones de van der Waals y las interacciones electrostáticas, se calculan por átomos separados al menos por tres enlaces, y se incluye un término explícito para representar el enlace por puente de hidrógeno, de esta manera se logra una mejor descripción energética como estructural de los sistemas biológicos.

Otra característica importante que se debe tomar en cuenta para la selección de este campo de fuerzas en la dinámica molecular, es su capacidad en el modelaje de ácidos nucleicos.

2.6.2.e. Funcionamiento de un programa de dinámica molecular.

El proceso general que sigue un estudio de dinámica molecular comienza con la determinación de las distancias interatómicas, ángulos de enlace, ángulos diedros y conectividad de la energía inicial, éste es importante ya que el computo de las expresiones energéticas se extiende sobre todos los términos que conforman

la energía total del sistema, y dado que el modelo clásico no considera los electrones, carece de este elemento para establecer la existencia o no de un enlace.

Posteriormente se continúa con el proceso de optimización, el cual es iterativo, en éste se procesa la energía en función de la geometría. La topología finalmente obtenida, será utilizada para el cálculo de momentos de inercia, momentos dipolares, etc., calculados por el campo de fuerzas elegido durante el lapso total de la dinámica.

2.6.2.f. Límites del sistema simulado.

Los efectos de contorno y los límites del sistema son fundamentales en los métodos de simulación. El enfoque más eficiente para establecer éstas características en el sistema son las condiciones periódicas de contorno (PBC), con las cuales el sistema se replica en todas las direcciones del espacio, generando una matriz periódica. En estas condiciones, cuando una partícula en movimiento abandona los límites de la caja central, es inmediatamente reemplazada por otra “partícula imagen” que se incorpora desde la cara opuesta, por lo que el número de partículas en el sistema permanece constante.

Además de la celda cúbica, que es el sistema más sencillo de visualizar y programar, existen otras formas apropiadas de construir una red periódica mediante operaciones de translación de la caja central en tres dimensiones: prisma hexagonal, paralelepípedo, dodecaedro, octaedro, etc... Su elección depende de valorar las dimensiones de las moléculas a tratar en la simulación y sobre todo, el espacio que habrá entre la molécula de interés y los límites del sistema, para con estas consideraciones conseguir una disminución notable en el número de partículas a simular, lo que en definitiva redundará en una mayor eficiencia computacional.

2.6.2.g. Métodos de integración.

Los cambios conformacionales del sistema son explorados bajo el desarrollo de un algoritmo que sigue la evolución a lo largo de una trayectoria temporal utilizando las coordenadas nucleares x y la energía del sistema U descritas por las ecuaciones de la mecánica clásica Newtoniana.

$$-\frac{dU}{dx} = m \frac{d^2 x}{dt^2}$$

Existe un inconveniente: estas ecuaciones no pueden ser resueltas analíticamente, ya que la fuerza ejercida sobre un átomo depende de un movimiento acoplado de partículas que resulta difícil de computar, porque se toma en cuenta la posición de los átomos circundantes; debido a esto se utilizan métodos de diferencias finitas basados en la fuerza generada en cada átomo, que junto con sus coordenadas y velocidades calcularán las posiciones que habrán de tomar, asumiendo que en este intervalo temporal la fuerza será constante. Uno de los algoritmos utilizados para integrar las diferencias finitas, está dentro del campo de fuerzas AMBER que resulta de una modificación del algoritmo de Verlet denominado *leap-frog*.

El algoritmo de Verlet produce trayectorias estables tras largo tiempo de iteración y considera las expansiones de Taylor alrededor de las configuraciones del paso posterior ($t+Dt$) y el anterior ($t-Dt$):

$$x(t+\Delta t) = xt + \frac{dx}{dt} \Delta t + \frac{1}{2} \frac{d^2 x}{dt^2} \Delta t^2 + \dots$$

$$x(t-\Delta t) = xt - \frac{dx}{dt} \Delta t + \frac{1}{2} \frac{d^2 x}{dt^2} \Delta t^2 + \dots$$

Si se trunca el segundo término se obtiene la siguiente ecuación en donde x y a son los vectores de posición y aceleración respectivamente.

$$x(t+\Delta t) = 2x(t) - x(t-\Delta t) + a(t)\Delta t^2$$

Como se observa el vector de velocidad no se determina directamente, lo que dificulta la simulación a temperatura constante, con ello la primera iteración resulta problemática porque generalmente no se tienen dos conformaciones al inicio de ésta.

El algoritmo *leap frog*, posterior al cálculo del algoritmo de Verlet, calcula las coordenadas a partir de las anteriores junto con las velocidades evaluadas a medio paso de la integración en donde v es el vector de velocidades:

$$x(t+\Delta t) = x(t) + v(t + \frac{1}{2} \Delta t) \Delta t$$

$$v(t + \frac{1}{2} \Delta t) = v(t - \frac{1}{2} \Delta t) + a(t) \Delta t$$

Este algoritmo se utiliza de manera frecuente por ser bastante robusto pese a que no pueda calcularse la energía cinética al mismo tiempo que la potencial, por lo que las coordenadas y velocidades se resuelven de forma no sincronizada asignándolas durante la primera iteración por medio de una distribución de Boltzman para una temperatura determinada.

Unidos todos estos paradigmas en múltiples algoritmos de los que se alimenta una dinámica molecular, ahora podemos observar con claridad la poderosa técnica que está en nuestras manos para estudiar de cerca los átomos y los grupos funcionales de los aminoácidos que han de interactuar con las bases y el esqueleto de fosfato del DNA, así como la mecánica de movimiento general que se observa durante la formación del complejo y sobre todo de las conclusiones e inferencias que se desentrañan de la interacción entre un factor de transcripción y su secuencia blanco de DNA.

2.7. METODOLOGÍA DEL CAPÍTULO II.

Se tomó como semilla la estructura del factor de transcripción MEF2A unida a la CArGbox consenso 5' CTATTAATAG 3' reportado en el archivo *.pdb 1C7U* (Huang *et al.* 2000) resuelto por NMR y constituido por 3772 átomos, de los cuales 2504 corresponden al TF MEF2A y 1268 a la CArGbox con una masa total de 8728.270 a.m.u para cada monómero del TF y de 6097.946 a.m.u para cada cadena del ácido nucleico. La carga total del sistema es de -28.0 e. No se realizaron modificaciones a las coordenadas originales, pero debido a cuestiones “gramaticales” fue necesario corregir la escritura de algunos átomos para que el archivo original contara con la nomenclatura utilizada en las bibliotecas del campo de fuerzas seleccionado y así pudiera leerse. (*cfr.* Apéndice A)

Se utilizó el programa GROMACS en su versión 4.5.5 (Apol *et al.* 2010) para realizar la dinámica molecular con el campo de fuerzas AMBER99 *protein, nucleic.* (Wang *et al.*, J. Comp. Chem. 21, 1049-1074, 2000), que está diseñado para el tratamiento de proteínas, ácidos nucleicos y otros sistemas macromoleculares; su función de energía está compuesta por los términos de flexión, tensión, interacciones de van der Waals electrostáticas, interacciones por puentes de hidrógeno y torsión (*Cfr.* apéndice B), este último sólo depende de los átomos centrales. Las interacciones de largo alcance se calculan entre átomos separados por al menos tres enlaces. Para el cálculo de los puentes de hidrógeno se incluye un término explícito que permite una mejor descripción energética y estructural tan importante en las interacciones proteína-ácido nucleico. Las interacciones disolvente-disolvente y soluto-disolvente están contempladas, y la redistribución sobre los átomos permite la generación de dipolos inducidos en cada átomo.

El modelo de agua seleccionado fue TIP3P que es un prototipo simplista de H₂O que considera a la molécula como un ente rígido (logra ahorrar tiempo de cómputo) que toma en cuenta las interacciones de no valencia entre las moléculas de agua y el sistema MEF2A-CArGbox. La fuerza electrostática está basada en la ley de Coulomb en tanto que la repulsión y dispersión utiliza el potencial Lennard-Jones, que considera tres puntos de carga repartidos de forma exacta en cada átomo de la molécula cuyo ángulo es de 104.52°. (Su formulación puede observarse en el apéndice C).

En cuanto a los parámetros generales presentes en todas las simulaciones, las condiciones de frontera fueron periódicas (Figura 4), la temperatura fue de 36.85°C controlada por el método *v-rescale* como opción al termostato de Berendsen con constantes de acoplamiento de 0.1 ps, finalmente se mantuvo la presión constante de 1 atm controlada por el algoritmo Parrinello-Rahman, por lo tanto el ensamble elegido en esta dinámica fue NPT.

2.7.1. Especificación de la caja donde se aloja el sistema.

Se utilizó como semilla el archivo *1C7U.pdb*, que se convirtió al formato nativo de GROMACS con el comando “*pdb2gmx*”, que tiene por objeto generar dos archivos: el *.gro* que contiene las coordenadas tridimensionales de los átomos del sistema y el *.top* que alberga información acerca de sus masas atómicas, cargas y enlaces.

Una vez realizado el cambio al formato nativo de GROMACS, se configuraron las dimensiones de la caja que albergaron a los átomos, ejecutando el comando “*editconf*”. El largo mayor del complejo es de ~ 72.24 Å, su altura mayor es de ~ 49.07 Å y de ancho tiene ~ 30 Å, se especificó una distancia de 1.5 nm (15 Å), a partir del punto más superficial en los ejes *x*, *y* y *z* el complejo TF-elementocis. La forma de la caja es dodecahédrica que ocupa un volumen 871.24 nm^3 , cuyo centro se halla en las coordenadas 10.721 10.721 10.721 (nm) para los valores de *x*, *y* y *z* respectivamente.

2.7.2. Solvatación del sistema.

Posterior al desarrollo de las especificaciones de la caja se procedió a rellenarla con moléculas H_2O , ya que la simulación de éstas se hizo de manera explícita. Esta acción se realizó con “*genbox*”, programa que modificó el fichero *.gro* y *.top*, para agregar 27687 moléculas de agua basadas en el modelo de coordenadas *spc216.gro* que guarda una distancia mínima de 1.57 Å entre moléculas en su estado inicial y es compatible con otros modelos de tres puntos como el TIP3P.

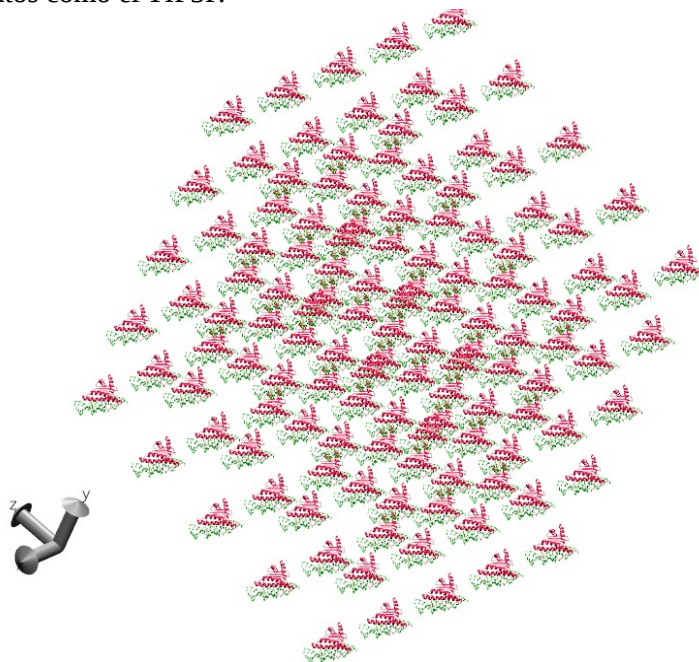


Fig. 4. Condiciones Periódicas del sistema. Proyección ortogonal de las primeras cuatro capas del sistema periódico, se observa al complejo MEF2A-DNA en múltiples repeticiones y en la esquina inferior izquierda un registro del cuadrante retratado en el visualizador molecular VMD.

2.7.3. Creación del archivo *.tpr*

Para continuar con el proceso de la dinámica molecular fue necesario realizar la compilación de los ficheros *.gro*, *.top* y *.mdp* utilizando el comando “*grompp*” que al conjuntarlos genera un archivo *.tpr*, (archivo en código máquina que contiene las coordenadas del sistema, así como la información fisicoquímica de cada átomo y la especificación del campo de fuerzas utilizado durante la simulación). Cabe mencionar que la generación de este archivo es necesaria para llevar a cabo la adición de iones en el mismo, por lo que se realizó una minimización ínfima de 300 ps.

2.7.4. Neutralización de la carga total del sistema.

En un inicio se registró una carga total del sistema de -28.0 e, que fue neutralizada con el fin de aumentar su estabilidad. Se adicionaron 28 iones monoatómicos de Na^+ al remplazar moléculas de agua en lugares energéticamente favorables para su inclusión en el sistema. La adición de iones se lleva a cabo utilizando el programa “*genion*”.

2.7.5. Flujo de ejecución de la preparación del sistema.

Confróntese el apéndice “A” para revisar a detalle las indicaciones otorgadas al programa GROMACS 4.5.5. Cada paso incluye los archivos de entrada y salida con el fin de facilitar la repetición de la dinámica, así como observaciones y notas importantes generados durante las distintas etapas de confección del sistema.

2.7.6. Minimización de la energía total del sistema.

Esta primera parte de la simulación tiene por objetivo recomodar las coordenadas de los átomos presentes en las moléculas disueltas para que las longitudes de sus enlaces y sus ángulos se encuentren en configuraciones de mínima energía potencial. El procedimiento se realizó con el fin de evitar rupturas en la molécula debido a los altos niveles energéticos que puedan afectar la estabilidad de la simulación; este paso acelera las simulaciones posteriores.

Se realizó la compilación de los archivos *MINI.mdp*, *1C7U.top* y *1C7Uneutralizado.gro* con el fin de generar los archivos *1C7Ucompilacion.tpr* y *mdout.mdp*. A continuación se resaltarán las principales especificaciones con las que se corrió la minimización.

En la línea *integrator=steep* se ordenó utilizar el algoritmo *steepest-descent* para la minimización de energía potencial. Este algoritmo se considera una buena opción para acercarse al mínimo energético que depende de tres variables independientes $3N$ que corresponden a 3 componentes inmersos en las coordenadas de los átomos en nuestro sistema, éstos son:

X_k : Vector que contiene la configuración del sistema iteración k .

λ_k : Escalar que indica la magnitud del salto en la dirección de búsqueda de mínima energía.

S_k : Corresponde al vector de búsqueda que nos lleva a la dirección negativa del gradiente.

de este modo el camino de la minimización es la que persigue el gradiente del potencial.

El valor de *nsteps* corresponde al número de pasos de integración al que se someterá al sistema, en éste caso se indicaron 30000 como máximo número de iteraciones.

Con *nlist* se indicó que la frecuencia de actualización de la lista de vecinos atómicos que mantienen una interacción no covalente por debajo de un radio de corte R_c , será actualizada y retenida cada 10 pasos durante la simulación. Se asignó el valor de 1 nm a la distancia de corte R_c , denotada como *rlist* dentro del archivo *.mdp*.

Las interacciones coulombicas entre las partículas cargadas fueron calculadas bajo el esquema de la

suma de Ewald, indicado como *coulombtype=pme*.

El radio utilizado para la búsqueda de vecinos inmersos en interacciones coulombicas se especificó de 1 nm asignado en la línea *rcoulomb=1.0* al igual que el radio considerado en las interacciones de van der Waals especificado en la línea *rvdw=1.0*.

El rango de puntos de corte se especificó con la relación $rvdw \geq rlist$ en la línea *vdw-type=cut-off*.

La minimización de energía se hizo converger cuando la fuerza máxima fuera menor que $100 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ (*emtol=100*) y finalmente, los datos de cálculo de energía fueron guardados cada 10 pasos.

Confróntese el apéndice “A” para revisar a detalle el flujo de ejecución de los comandos durante la minimización.

2.7.7. Dinámica de posiciones restringidas.

Partiendo de la fijación del soluto dentro de nuestro sistema, fue necesario permitir que el disolvente y los iones se "relajasen" a su alrededor en posiciones de mínima energía potencial, se consiguió que el disolvente llenase la caja de manera uniforme e ingresase a los poros presentes en el complejo MEF2A-CARgbox, ya que el programa *genbox* no lo hace. En este paso se impide la formación de vacíos durante la simulación de todos los átomos.

El sistema se encuentra separado en los grupos DNA, PROTEIN, SOL y el que integran los iones. Se fijaron todos los átomos de los grupos que corresponden a los polipéptidos y al oligonucleótido, cuya posición restringida se encuentra definida por una fuerza constante, asociada a un potencial hamiltoniano adicional. Las moléculas de agua no fueron fijadas con el fin de mantener al solvente sin restricción dentro del sistema con los fines antes descritos.

Para activar posiciones restringidas de átomos que se especificaron en ficheros con extensión *.itp* (creados tras la conversión de archivos *.pdn* al archivo nativo de GROMACS) se indicó en el archivo de entrada *PR.mdp: define=-DPOSRES*. La dinámica se ejecutó bajo el algoritmo *leap-frog*, responsable de la integración de las ecuaciones de Newton (especificado en la línea *integrator=md*). El tiempo de simulación fue de 125000 pasos, que equivalen a 250 ps (*dt=0.002*), bajo las siguientes condiciones: se restringieron las distancias de los enlaces con la opción *constraints=all-bonds*, la temperatura se mantuvo constante a (36.85 °C) y se utilizó el algoritmo *v-rescale* como variante del algoritmo de Berendsen, mismo que fue utilizado para mantener la presión constante a 1 atm.

El esquema de vecindad para regular las interacciones electrostáticas fue el mismo que en la minimización, se utilizó el algoritmo *PME* para el cálculo de las fuerzas coulombicas. Con las opciones *nstxout=250* y *nstenergy=100* se indicó el número con el que las coordenadas del sistema y los cálculos de energía se almacenaron para construir los ficheros de trayectoria y energía.

Finalmente con la línea *gen_vel=yes* indicamos que las velocidades aleatorias se generaran desde el inicio de corrida del sistema. Confróntese el apéndice “A” para revisar a detalle el flujo de ejecución de los

comandos utilizados durante la dinámica molecular de posiciones restringidas.

2.7.8. Dinámica molecular de todos los átomos.

En la tercera simulación o dinámica molecular completa se realizaron los cálculos de movimiento que incluyen todos los grupos del sistema, no se restringieron posiciones para ningún átomo (se retiró la opción *-DPOSRES*), con el objetivo de hacer de esta dinámica un reflejo de las interacciones existentes entre el DNA y el factor de transcripción. El archivo *MD.md* contiene todas las especificaciones de corrimiento del sistema, a continuación se indican sus principales determinantes: Los cálculos de la ecuación de movimiento de Newton se resolvieron con el algoritmo *leap-frog* (*integrator=md*), todos los enlaces atómicos fueron considerados como rígidos (indicado en la línea *constraints=all-bonds*). Tanto la temperatura como la presión fueron constantes, la primera fue controlada por el algoritmo *v-rescale* a 36.85 °C y se mantuvo el sistema a 1 atm controlándola con el algoritmo Parrinelo-Rahman.

La lista de vecinos atómicos con interacción no covalente según el radio R_c (de 1 nm) se actualizó cada 10 pasos (*nlist=10*). El esquema de Ewald fue el seleccionado para el cálculo de las interacciones coulombicas (indicado en la línea *coulombtype=pme* del archivo *.mdp*, asociado al valor de *rcoulomb=1.0* que indica un radio de corte de 1 nm). Para esta simulación el radio de van der Waals se acrecentó a 1.4 nm (en las anteriores simulaciones fue de 1 nm) con el fin de considerarlas ampliamente, debido al papel fundamental que tienen durante la interacción entre factores de transcripción y sus secuencias blanco. Cabe mencionar que la relación $r_{vdw} \geq r_{list}$ se mantuvo (*vdw-type=cut-off*).

El tiempo total de la simulación fue de 125000 pasos que equivalen a 25000 ps. ($dt=0.002$ ps.), los cuales fueron registrados cada 2500 pasos para los archivos de trayectoria y velocidades, (*nstxout*, *nstvout* y *nstfout*)- Los archivos de energía fueron construidos cada 100 pasos (indicados en la línea *nstenergy=100*). Confróntese el apéndice “A” para revisar a detalle el flujo de ejecución de los comandos utilizados durante la dinámica molecular de todos los átomos.

2.7.9. Herramientas de análisis posterior a las dinámicas experimentadas por el sistema.

Debido al número de datos obtenidos -tras el desarrollo de la dinámica- fue necesario crear un conjunto de metodologías adecuadas de análisis, para comprender el sistema a través de sus diferentes fases.

2.7.9.a. Obtención de la gráfica de energía potencial.

A través del estudio de la energía potencial podemos inferir la energía almacenada en nuestro sistema, dato importante que permite realizar un esbozo de interacciones sufridas por átomos, tales como las fuerzas de atracción y repulsión.

Durante el período de minimización, el análisis de la energía potencial es un dato fundamental; al observar su variación, podremos observar cómo ha fluctuado en un nivel inferior al inicial, de ello dependerá la optimización de nuestro sistema durante el proceso de minimización.

El proceso de obtención de la energía potencial se realiza a través de la ejecución del programa *g_energy* incluido dentro de la paquetería de GROMACS; su función es extraer los componentes energéticos del archivo con extensión *.edr*, creado como archivo de salida durante los tres procesos de dinámica realizados. La línea de código a ejecutar será: *g_energy -f X.edr -o X*; n donde *-f* representa el archivo de entrada, *-o* el archivo de salida y *X* el nombre del archivo. Este programa es capaz de calcular otros parámetros, se debe seleccionar la opción 10 que corresponde a la energía potencial (aplicable en la versión 4.5.5).

2.7.9.b. RMSD.

La concatenación de los puntos comunes entre dos estructuras dentro de un espacio tridimensional, suele ser muy útil en los análisis posteriores a una dinámica, para comparar los cambios estructurales de las moléculas a través del tiempo. El análisis de RMSD (*Root-Mean-Square Deviation*) correspondiente a la minimización se realizó con el método “*top*” de VMD, se alinearon en su totalidad tanto el *backbone* de la proteína como el esqueleto de fosfatos en el DNA de las estructuras que corresponden al *frame 0* y el último de la simulación que en este caso fue el 74.

2.7.9.c. Obtención de las gráficas de temperatura, presión y densidad.

Examinar los cambios de temperatura, presión y densidad del sistema es fundamental para obtener una idea de las condiciones a las que se sometieron las moléculas durante la simulación, por ejemplo, el control de estos parámetros en la dinámica de posiciones restringidas, nos brinda la certeza de que el sistema total ha llegado a un equilibrio o no, en cuyo caso las oscilaciones de dichas fuerzas físicas aún se considerarían inadecuadas como para progresar con la dinámica de todos los átomos. Al igual que en la obtención de la gráfica de presión, a través del programa *g energy*, podremos construir las gráficas de densidad y temperatura, utilizando la misma línea de comandos y valiéndonos de los datos reportados en el fichero con extensión *.edr* proporcionado después de aplicar el programa *mdrun* de la paquetería de GROMACS.

2.7.9.d. Creación del grupo *Protein_DNA* y el aislamiento de subgrupos atómicos.

Debido a que ciertos análisis sólo tomarán en cuenta al complejo TF-CARGbox, es necesario crear un archivo que contenga tanto sus registros energéticos como los de trayectoria, este fichero será leído por GROMACS como un grupo. Su creación es muy simple, basta con otorgarle al programa *make_ndx* el archivo de configuración *.gro*, y seleccionar de entre los grupos ya existentes o por medio de la selección de átomos específicos valiéndose de su numeración en el archivo de coordenadas; ejemplo: teclee *a 3297-3326*, después de la ejecución del comando *make_ndx* para seleccionar la C 226 de la cadena B. De ser el caso, la concatenación de los grupos se lleva por medio de operadores booleanos.

2.7.9.e. Centrado del complejo dentro del sistema.

Durante la dinámica molecular es probable que el complejo molecular estudiado se desplace fuera de los

límites de la caja de agua, si bien las condiciones periódicas previenen la pérdida de átomos, pueden causar problemas en la visualización y provocar errores durante el análisis de RMSD. En consecuencia, las coordenadas del sistema se distorsionan considerablemente, por ello antes de realizar dicho análisis, se recomienda centrar al complejo dentro del sistema. El programa *trjconv* ejecuta un compendio de tareas que incluyen la conversión de un formato a otro, el cambio en la periodicidad de representación, realiza cambios de reducción, crea particiones en la trayectoria y también sirve para reubicar grupos de átomos dentro del espacio tridimensional del sistema. Esta última tarea se divide en dos secciones: en la creación de un archivo de trayectoria con las especificaciones de la reubicación de átomos, y en la creación de un archivo en el que pueda visualizarse dichos cambios. El primer paso a seguir es ejecutar el programa *trjconv* como se indica en la siguiente línea de comando:

```
trjconv -f X.trr -s X.tpr -n indx.ndx -o archivo.xtc -center -pbc mol
```

en donde *-f* es el archivo de trayectoria, *-s* el fichero compilado con el que se ejecuta el programa *mdrun* y *-o* es el archivo de salida que contiene la trayectoria corregida. La forma de ejecución del programa en cuestión, depende de las opciones con las que se acompañe, en éste caso *-center* coloca al grupo seleccionado en el centro de la caja, la opción *-pbc mol* considera que el centro de masa de la molécula estará al centro de todo el sistema observado y sus contrapartes periódicas.

2.7.9.f. Determinación de los puentes de hidrógeno existentes entre MEF2A y su CArGbox.

Debido a que los puentes de hidrógeno son fundamentales en la formación y reconocimiento de los factores de transcripción y su secuencia blanco, es necesario su estudio para localizar de forma puntual los sitios de interacción; en consecuencia se utilizó el programa *g_hbond* - incluido en la paquetería de GROMACS - para determinar el número de puentes de hidrógeno formado entre dos grupos de átomos (en este caso un aminoácido y un nucleótido) en función del tiempo total de la dinámica. *h_bond* calcula la existencia de puentes de hidrógeno basado en parámetros espaciales y energéticos, por lo que atendiendo a estos fines se designaron los siguientes parámetros: radio de corte del aceptor 0.35 nm (recomendado en estudios de sondeo de HB), 30° para el ángulo del donador, el aceptor y el átomo de hidrógeno, y finalmente se especificó una temperatura de 310 K para computar la energía de Gibbs correspondiente al rompimiento y formación de los puentes. Los registros fueron realizados cada 5 ns.

El sondeo se realizó formando grupos entre cada uno de los nucleótidos dentro de la CArGbox de las cadenas A y B del DNA contra cada uno de los aminoácidos iniciando desde Gly1 hasta Ala38, trecho que corresponde a la zona de unión. Las relaciones fueron conformándose por grupos entre los átomos del aminoácido X y todos los átomos del nucleótido Y, es decir 1 a 1. Los puentes de hidrógeno se rastrearon desde el inicio de la dinámica hasta el nanosegundo 25, es decir a su término y fueron visualizados con *Grace* 5.1.23 (1991-1995 Paul J Turner, Portland, OR; 1996-2012 Grace Development Team).

2.7.9.g. Visualización de estructura secundaria.

Con la finalidad de examinar los cambios de la estructura secundaria a las que están sujetas las cadenas del TF MEF2A, se realizó un registro de éstas cada 500 *frames* (2.5 ns) hasta el final de la simulación. El programa POLYVIEW-2D (Lincoln D. Stein's perl GD graphics Library that can be found at Boutell.Com, Inc. Web server or in any CPAN modules collection) se eligió para realizar un estudio de la estructura secundaria de fácil visualización.

2.8. RESULTADOS DEL CAPITULO II.

Se obtuvieron las siguientes simulaciones de dinámica molecular para el sistema CArG-MEF2A reportado en el archivo .pdb 1C7U:

- 9318 ps de minimización energética bajo el algoritmo *steepest-descent*, con un mínimo de -1.45×10^6 kJ/mol.
- RMSD con valor de 1.13 Å en la comparación de la semilla y el sistema al final de la dinámica.
- 250 ps de dinámica de posiciones restringidas.
- 25 ns de dinámica de todos los átomos.
- Mapeo de los puentes de hidrógeno involucrados en la unión de MEF2A y su CArGbox a lo largo de la simulación.

2.8.1. Inspección de los puentes de hidrógeno presentes entre MEF2A y la CArGbox.

Por medio del programa *g_hbond* se examinaron los residuos de aminoácido del 1 al 38 presentes en el factor de transcripción, ya que es la zona de interacción del complejo. En la tabla 2 se presentan los aminoácidos que resultaron positivos para este rastreo junto a los nucleótidos con los que interactúan; en la cual se observa que 21 de los aminoácidos presentes en la cadena A y B del dímero MEF2A, sostienen uno o más puentes de hidrógeno con 17 de los nucleótidos presentes en la CArGbox. Cabe resaltar que la distribución de las interacciones no es simétrica en el caso de la Arg2, la Ile5, la Thr19, la Lys22 la Gly26 y la Lys29, ya que aunque se puede establecer un eje de simetría bien marcado en el complejo debido a que trabajamos con un homodímero y a una secuencia de DNA palíndroma, cabría esperar que las interacciones fuesen similares para ambos cadenas de aminoácidos, sin embargo no se cumple dicho supuesto.

Tabla 2. Registro de los puentes de hidrógenos formados entre los aminoácidos de MEF2A y los nucleótidos de la caja CArG. Por medio del programa *g_hbond* se realizó la determinación de puentes de hidrógeno existentes entre el factor de transcripción y el DNA, para lo cual se relacionaron de forma individual cada uno de los primeros 38 residuos de aminoácido presentes en la proteína, con cada uno de los 20 nucleótidos presentes en la CArGbox. Se reportan únicamente aquellos ensayos que dieron positivo en la ocurrencia de puentes de hidrógeno. En la primera columna se indica el aminoácido en el que se detectó el puente de hidrógeno, la segunda columna indica la cadena a la cual pertenece el residuo dentro del dímero MEF2A; en la tercera columna se reporta el nucleótido con el que se establece la interacción, en forma de súper índice aparece reportado con una *a* y/o una *b* si el nucleótido se encuentra dirección 5' → 3' o 3' → 5' respectivamente. En forma de subíndice se reporta el número máximo de puentes de hidrógeno que se establecen entre el aminoácido reportado y el nucleótido a lo largo de 25 ns. de dinámica.

Aminoácido	Cadena	Nucleótidos				
Gly1	A	T9 ₁ ^a	T10 ₁ ^a		T13 ₁ ^b	
	B	T9 ₁ ^b	T10 ₁ ^b		T13 ₁ ^a	
Arg2	A	C6 ₁ ^a	T7 ₃ ^a	A8 ₁ ^a	A14 ₁ ^b	G15 ₁ ^b
	B	-	-	-	A14 ₁ ^a	G15 ₂ ^a
Lys3	A	A14 ₁ ^b			G15 ₁ ^b	
	B	A14 ₁ ^a			G15 ₁ ^a	
Lys4	A	T10 ₂ ^a			A11 ₂ ^a	
	B	T10 ₂ ^b			A11 ₂ ^b	
Ile5	A	A11 ₁ ^a				
Asn15	A	G15 ₁ ^b				
	B	G15 ₁ ^a				
Thr19	B	A14 ₁ ^b				
Lys22	A	C6 ₁ ^a	-	A14 ₂ ^b	G15 ₁ ^b	
	B	-	T13 ₁ ^a	A14 ₂ ^a	G15 ₁ ^a	
Arg23	A	T13 ₂ ^b			A14 ₃ ^b	
	B	T13 ₁ ^a			A14 ₃ ^a	
Gly26	A	A14 ₁ ^a				
Lys29	A	A11 ₁ ^b	A12 ₂ ^b	T13 ₁ ^b		
	B	-	A12 ₂ ^a	T13 ₁ ^a		
Lys30	A	A11 ₁ ^b	A12 ₂ ^{a,b}	T13 ₂ ^{a,b}		
	B	A11 ₁ ^a	A12 ₃ ^{a,b}	T13 ₁ ^b		

En total se hallaron 69 interacciones de este tipo, las cuales pueden dividirse claramente en tres grupos con respecto a su tiempo de presencia dentro de la duración total de la dinámica: aquellas que estuvieron presentes más del 80% total del tiempo de duración (grupo I); las que estuvieron presentes arriba de un 30% hasta un 79% de la dinámica (grupo II); y las que se sostuvieron por debajo del 30% de la dinámica (grupo III). Esta forma de clasificación contribuye a la comprensión del papel que cada una de las interacciones cumple en la unión del TF con el DNA. Esta clasificación se muestra en la tabla 3.

Los aminoácidos del grupo I contribuyen considerablemente dentro del tiempo total de la dinámica en el

establecimiento de la interacción entre el dímero y el ácido nucleico, pueden ser considerados como la maquinaria principal de anclaje entre las dos moléculas. El segundo grupo posee estadios relativamente medianos, que es posible contribuyan en menor medida a la unión de las moléculas. El último grupo parece no poseer importancia relevante en la formación del complejo durante el tiempo total de la dinámica, ya que la aparición de estas interacciones está limitada a tan sólo un momento del tiempo total de la simulación. Será interesante analizar estos datos para obtener nociones sobre la formación del complejo, ya que se observan recambios de formación en puentes de hidrógeno a nivel de un mismo aminoácido que posee dos o más interacciones y también entre diferentes aminoácidos que establecen interacciones cuando algún otro residuo ha dejado de participar en una.

Tabla 3. Clasificación de los aminoácidos que sostienen interacción por puente de hidrógeno según su porcentaje de estadía a lo largo de una dinámica de 25 ns. Los porcentajes de estadía fueron obtenidos del registro realizado cada 5 ps por el programa *g_hbond*, estos fueron tomados para calcular el tiempo de estadía de los puentes de hidrógeno conjuntando los registros individuales por cada aminoácido. En la primera fila se registra el aminoácido acompañado de su posición en la cadena polipeptídica partiendo desde el extremo amino. En la columna dos se especifica la cadena a la que pertenece dentro del dímero de MEF2A y finalmente en la columna de la derecha se especifica el grupo al cual pertenece.

Aminoácido	Cadena	Grupo
Gly1	A	I
Gly1	B	I
Arg2	A	I
Arg2	B	I
Lys4	A	I
Lys4	B	I
Lys22	A	I
Arg23	A	I
Arg23	B	I
Lys29	A	I
Lys29	B	I
Lys30	A	I
Lys30	B	I
Lys22	B	II
Lys3	B	II
Lys3	A	III
Ile5	A	III
Asn15	A	III
Asn15	B	III
Thr19	B	III
Gly26	A	III

Los tiempo de participación por interacción se observan claramente en los diagramas A al G que conforman la figura 5 y que fueron registrados cada 5 ps.

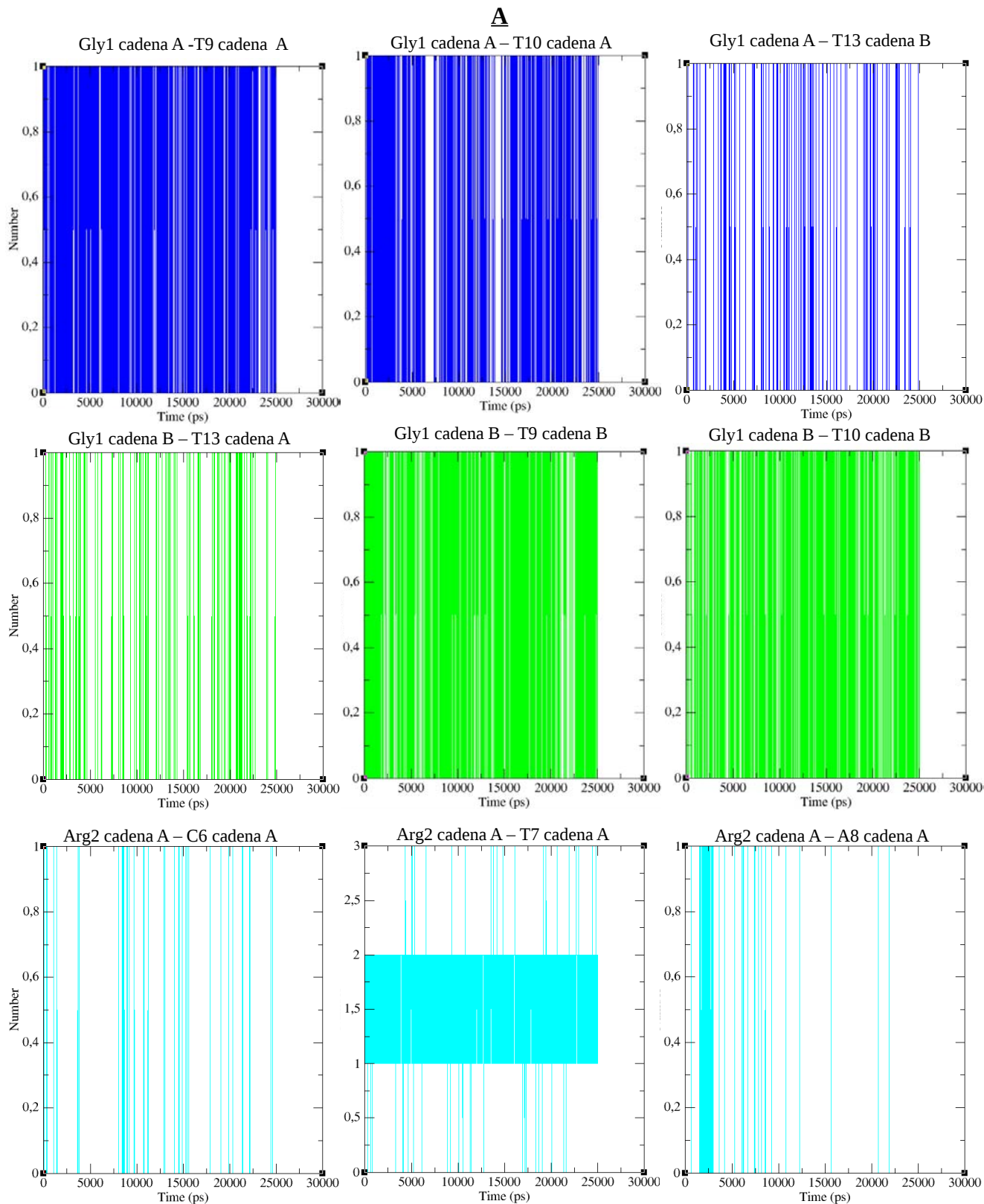


Fig 5. Sección A. Formación de puentes de hidrógeno a lo largo de 25 ns de dinámica molecular en los residuos de Gly1 cadenas A y B y Arg2 de la cadena A. En el eje de las x se muestra el tiempo total de duración de la dinámica y en el eje de las y observamos el número de puentes de hidrógeno formado a lo largo de la dinámica. En la parte superior, en azul, se muestran las interacciones que Gly de la cadena A sostiene con T9 y T10 de la cadena A y T13 de la cadena B. En el medio de color verde se reportan las interacciones de la Gly1 de la cadena B sostenidas con T13 de la cadena A y T9 y T10 de la cadena B. Finalmente en la parte inferior en color azul se observan los puentes de hidrógenos establecidos por Arg2 de la cadena A con C6, T7 y A8 de la cadena A. Los registros de puentes de hidrógeno se tomaron cada 5 ps.

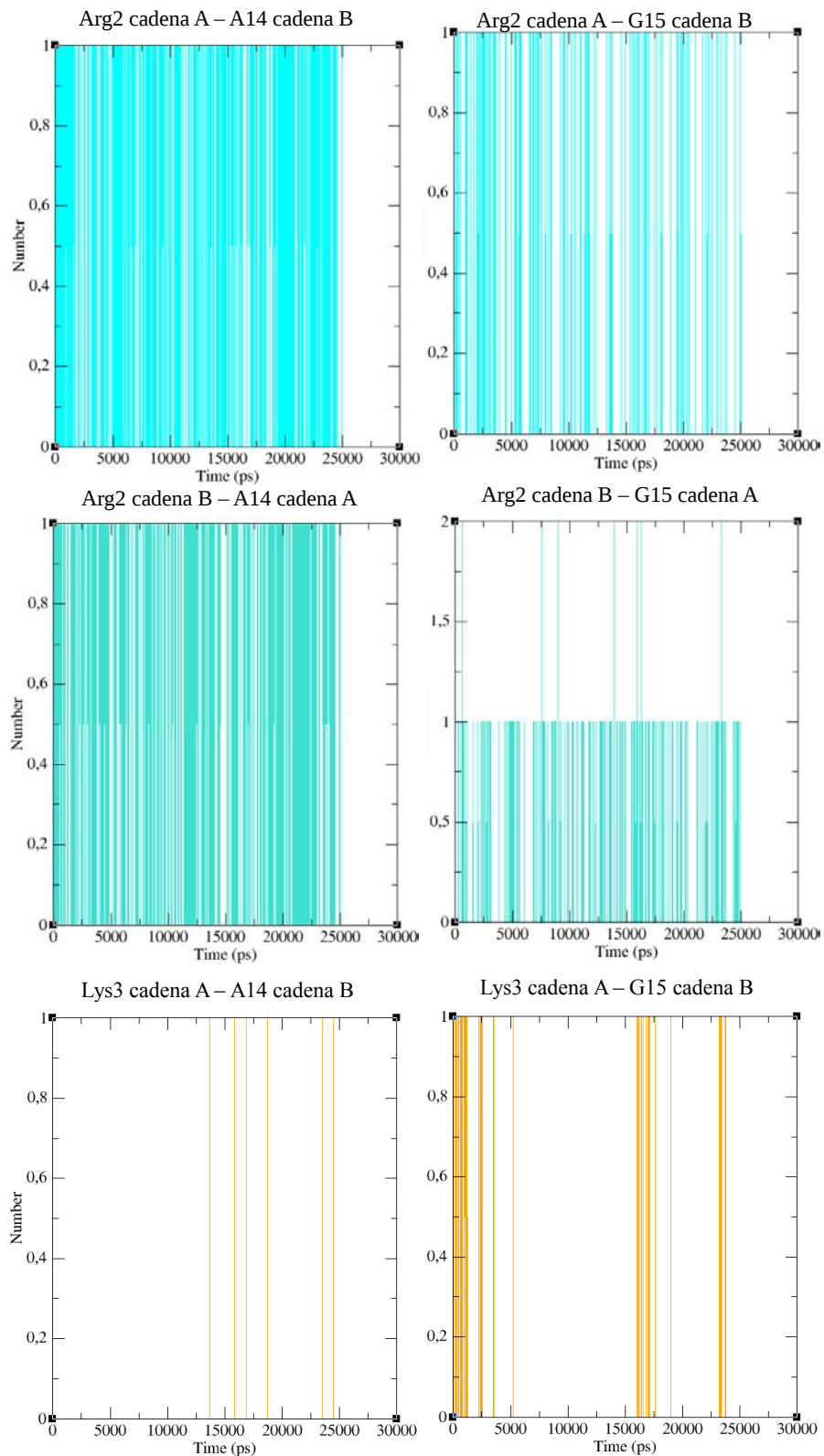
B

Fig 5. Sección B. Formación de puentes de hidrógeno a lo largo de 25 ns de dinámica molecular en los residuos de Arg2 cadenas A y B y Lys3 de la cadena B. En el eje de las x se muestra el tiempo total de duración de la dinámica y en el eje de las y observamos el número de puentes de hidrógeno formado a lo largo de éste periodo. En la parte superior, en azul, se muestran las interacciones que Arg de la cadena A sostiene con A14 y G15 de la cadena B. En el medio de color verde agua se reportan las interacciones de la Arg2 de la cadena B sostenidas con A15 y G15 de la cadena A; finalmente en la parte inferior en color naranja obsérvese los puentes de hidrógenos establecidos por Lys3 de la cadena A con A14 y G15 de la cadena B. Los registros de puentes de hidrógeno se tomaron cada 5 ps.

C

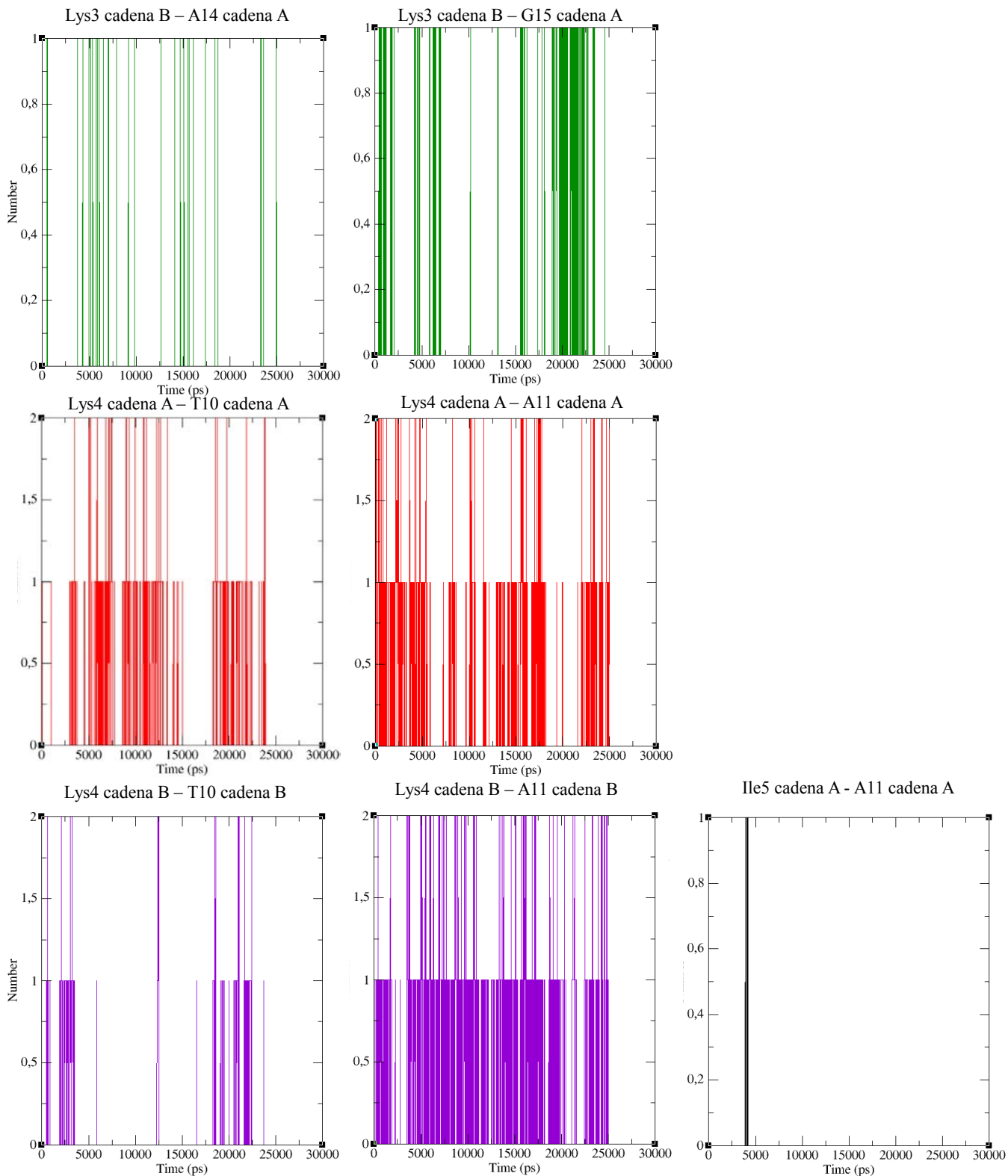


Fig 5. Sección C. Formación de puentes de hidrógeno a lo largo de 25 ns de dinámica molecular en el residuo de Lys3 de la cadena, la Lys4 de la cadena A y B i Ile5 de la cadena A. En el eje de las x se muestra el tiempo total de duración de la dinámica y en el eje de las y observamos el número de puentes de hidrógeno formado a lo largo de éste periodo. En la parte superior, en verde, se muestran las interacciones que Lys3 de la cadena A sostiene con A14 y G15 de la cadena A . En el medio de color rojo se reportan las interacciones de la Lys4 de la cadena A sostenidas con T10 y A11 de la cadena A; finalmente en la parte inferior en color morado obsérvense los puentes de hidrógenos establecidos por Lys4 de la cadena B con T10 y A15 de la hebra B junto con la única interacción reportada para Ile5 de la cadena A en color negro con A11 de la hebra A. Las dataciones de puentes de hidrógeno se tomaron cada 5 ps.

D

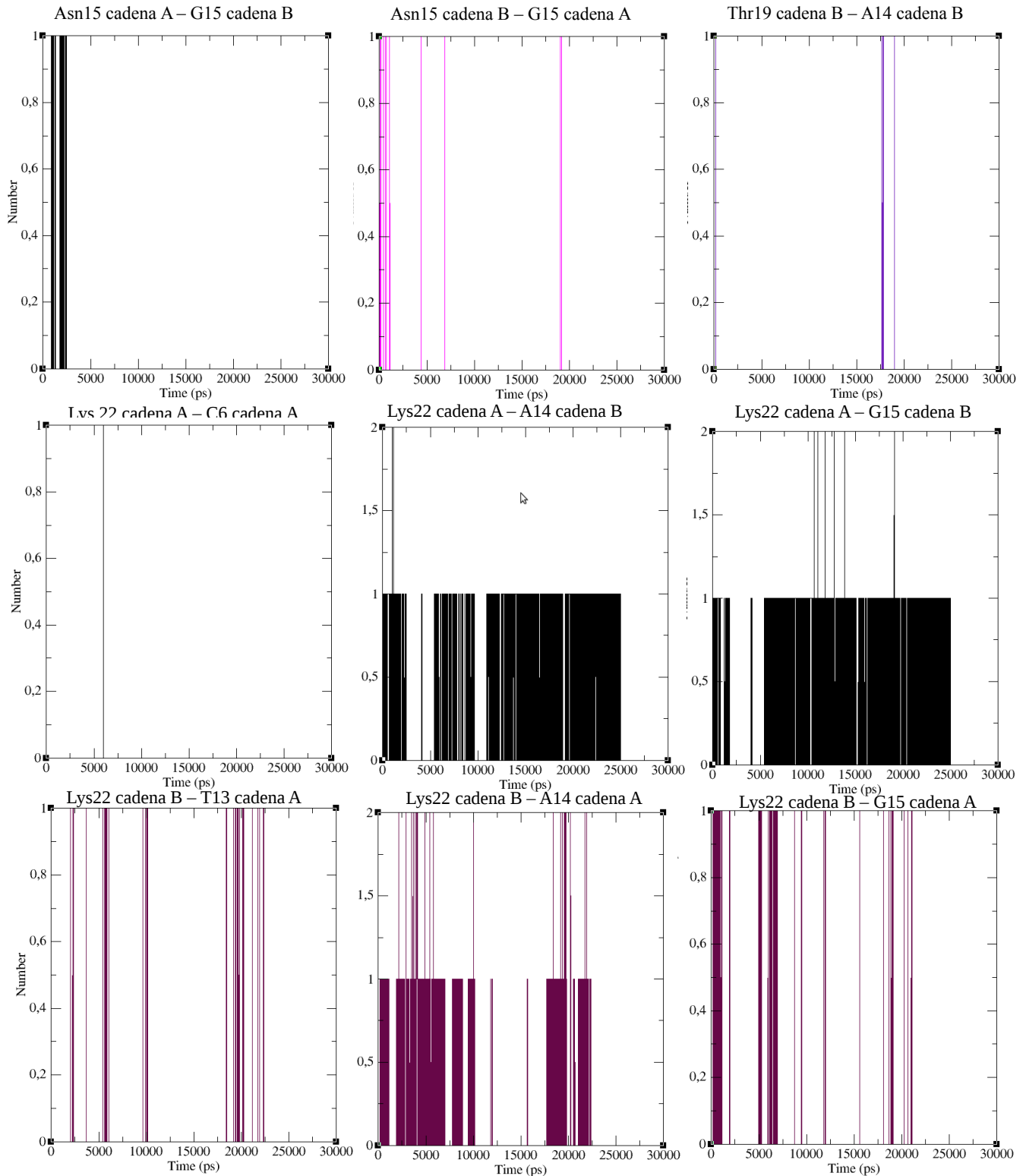


Fig 5. Sección D. Formación de puentes de hidrógeno a lo largo de 25 ns de dinámica molecular en el residuo de Asn15 de las cadenas A y B, Thr19 de la cadena B y Lys 22 de las cadenas A y B. En el eje de las x se muestra el tiempo total de duración de la dinámica y en el eje de las y observamos el número de puentes de hidrógeno formado a lo largo de ese periodo. En la parte superior, en negro y fuxia, se muestran las interacciones que Asn15 de las cadenas A y B sostiene con G15 de la cadena B y A respectivamente; del lado izquierdo en la parte superior de color morado se reporta la interacción de Thr19 de la cadena B con A14 de la hebra B. En el medio de color negro se reportan las interacciones de la Lys22 de la cadena A sostenidas con C6 de la hebra A y A14 y G15 de la cadena B; finalmente en la parte inferior en color uva obsérvense los puentes de hidrógenos establecidos por Lys22 de la cadena B con T13 de la hebra A y A14 y G15 de la hebra A. Las dataciones de puentes de hidrógeno se tomaron cada 5 ps.

E

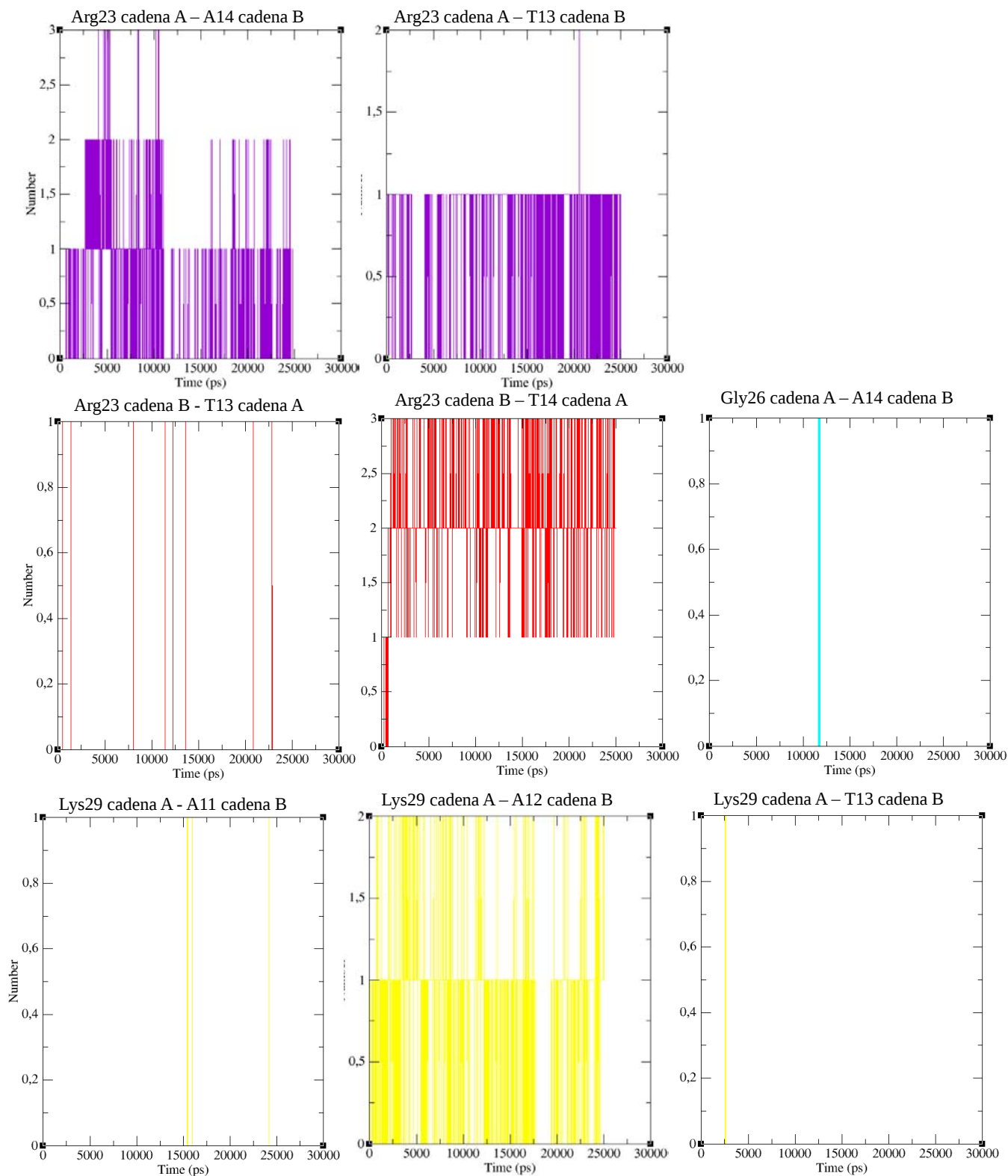


Fig 5. Sección E. Formación de puentes de hidrógeno a lo largo de 25 ns de dinámica molecular en los residuos de Arg23 cadenas A y B, Gly26 de la cadena A y Lys29 de la cadena A. En el eje de las x se muestra el tiempo total de duración de la dinámica y en el eje de las y observamos el número de puentes de hidrógeno formado a lo largo de ese periodo. En la parte superior, morado, se muestran las interacciones que Arg23 de la cadena A sostiene con A14 y T13 de la cadena B. En el medio de color rojo se reportan las interacciones de la Arg23 de la cadena B sostenidas con T13 y T14 de la cadena A, a un lado en color azul se observa la interacción de Gly26 de la cadena A con A14 de la cadena B. En la parte inferior en color amarillo obsérvese los puentes de hidrógenos establecidos por Lys29 de la cadena A con A11, A12 y T13 de la cadena B. Las dataciones de puentes de hidrógeno se tomaron cada 5 ps.

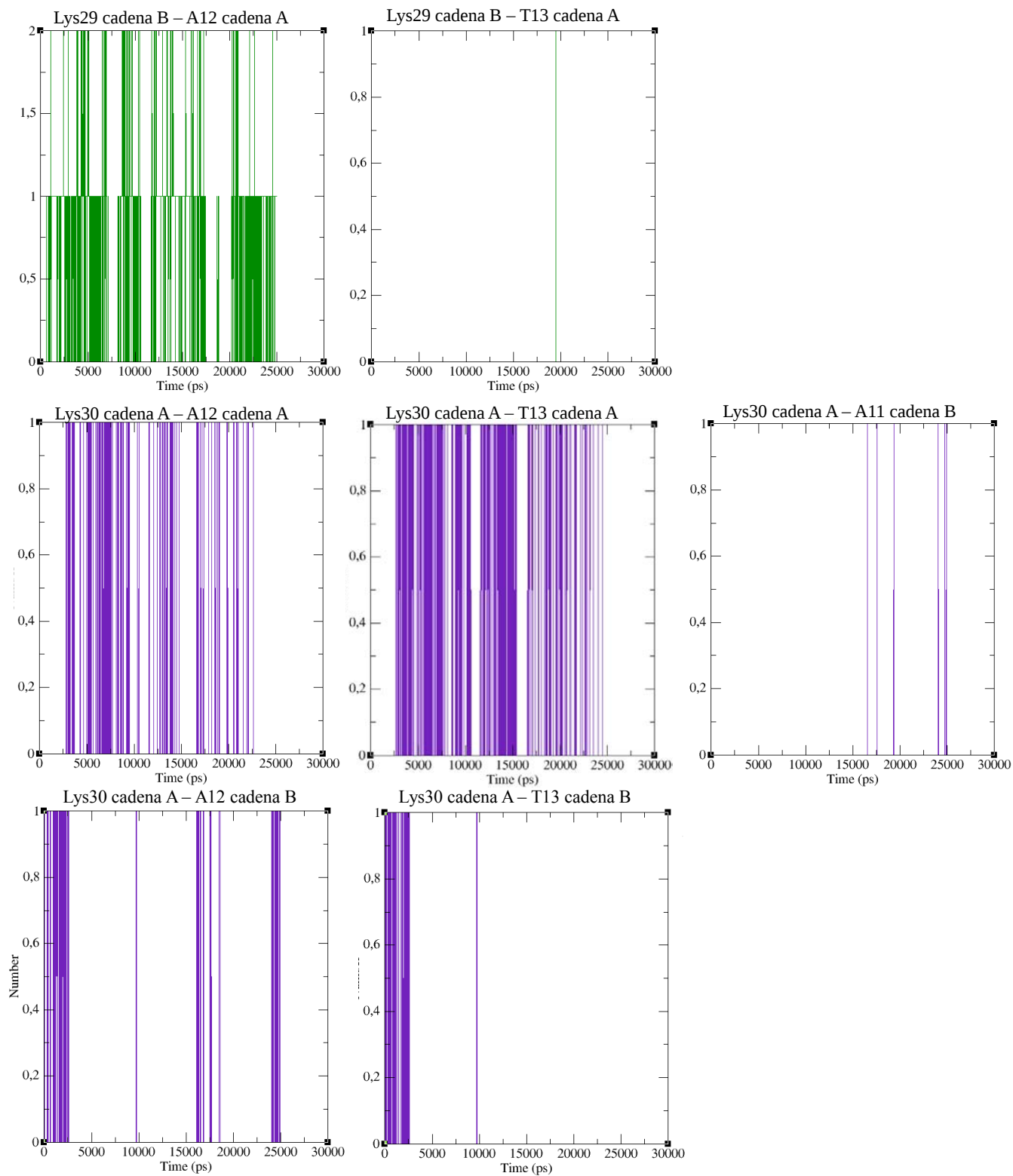
F

Fig 5. Sección F. Formación de puentes de hidrógeno a lo largo de 25 ns de dinámica molecular en los residuos de Lys29 y Lys30 presentes en la cadena A. En el eje de las x se muestra el tiempo total de duración de la dinámica y en el eje de las y observamos el número de puentes de hidrógeno formado a lo largo de ese periodo. En la parte superior, en verde, se muestran las interacciones que Lys29 de la cadena B sostiene con A12 y T13 de la hebra A. En morado se reportan las interacciones que Lys30 de la cadena A sostenidas con A12, T13 de la cadena A y, A11, A12 y A13 de la cadena B. Las dataciones de puentes de hidrógeno se tomaron cada 5 ps.

G

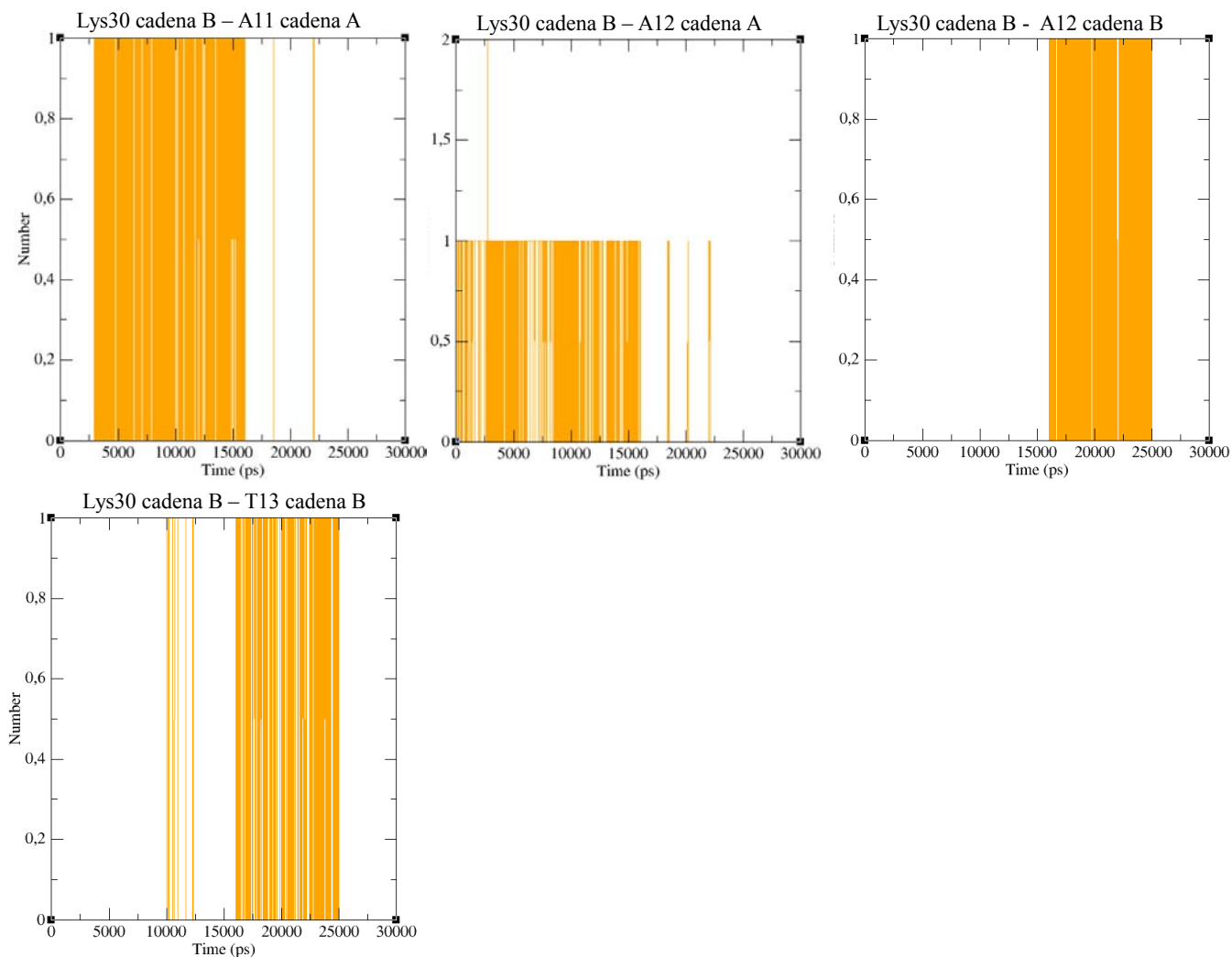


Fig 5. Sección G. Formación de puentes de hidrógeno a lo largo de 25 ns de dinámica molecular en los residuos de Lys30 presente en la cadena B. En el eje de las x se muestra el tiempo total de duración de la dinámica y en el eje de las y observamos el número de puentes de hidrógeno formado a lo largo de ésta. En naranja se muestran las interacciones que Lys30 de la cadena A sostiene con A11 y A12 de la cadena A junto con A12 y T13 de la cadena B. Las dataciones de puentes de hidrógeno se tomaron cada 5 ps.

Fue posible identificar algunos de los aminoácidos que participan de forma importante en la prevalencia del complejo a través de la dinámica. Los aminoácidos de la cadena A que resultaron positivos en el establecimiento de puentes de hidrógeno, muestran una marcada diferencia en el tiempo de duración de estas interacciones en los aminoácidos del grupo I y III, observándose puentes de hidrógeno que permanecen a lo largo de la dinámica como es el caso del registrado entre la Arg2 y la T7 de la hebra A y también interacciones menos consistentes como las hay entre la Gly26 y la A14 de la cadena B. En la figura 6 se reportan los porcentajes de presencia de puentes de hidrógeno a lo largo de 25 ns de dinámica molecular obtenidos para cada uno de los aminoácidos representados: Gly1-96%, Arg2-100%, Lys4-89%, Lys22-81%, Arg23-87%, Lys29-85%, Lys30-92%, Lys3-28%, Ile5-3%, Asn15-8% y Gly 26-3%.

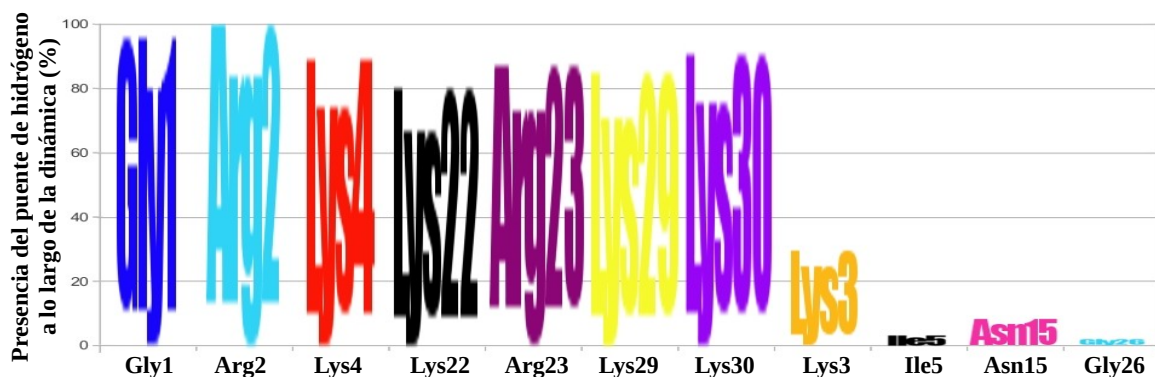


Fig. 6. Porcentaje de duración de los puentes de hidrógeno formados por los aminoácidos de la cadena A del dímero de MEF2A. Se utilizaron los registros de la determinación de puentes de hidrógeno realizados cada 5 ps con el programa *g_hbond* para obtener el porcentaje de presencia a lo largo de la dinámica por lo que 25ns corresponden al 100%. El registro fue culitativo, ya que a éste porcentaje contribuyen igualmente la presencia de uno o más interacciones, en el caso de los aminoácidos que forman dos o más puentes en el mismo lapso de tiempo. Para éste gráfico solo se tomaron en cuenta los aminoácidos presentes en la cadena A de MEF2A y que dieron positivo para el establecimientos de puentes de hidrógeno con la caja CArG y que corresponden a Gly1, Arg2, Lys4, Lys22, Arg23, Lys29, Lys30, Lys3, Ile5, Asn15 y Gly26 diferenciados por colores. Los datos incluidos en la figura descienden directamente de los recabados en la dinámica molecular tratada en éste capítulo.

En el caso de los aminoácidos de la cadena B (como se observa en la figura 7), se hallaron los tres grupos de clasificación con respecto al tiempo de estadía, siendo visibles los siguientes porcentajes de permanencia de los puentes de hidrógeno en los distintos aminoácidos que sostienen este tipo de interacciones con la CArGbox: Gly1-93%, Arg2-91%, Lys4-81%, Arg23-96%, Lys29-78%, Lys30-98%, Lys22-67%, Lys3-45%, Asn15-7% y Thr19-2%. Cabe mencionar con respecto a los porcentajes de ambas cadenas, que los aminoácidos reportados poseen porcentajes similares.

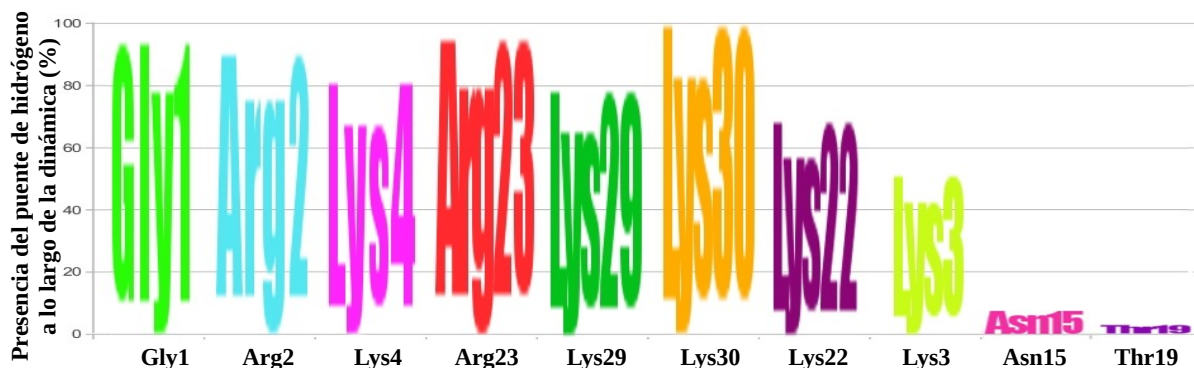


Fig. 7. Porcentaje de duración de los puentes de hidrógeno formados por los aminoácidos de la cadena B del dímero de MEF2A. Se utilizaron los registros de la determinación de puentes de hidrógeno realizados cada 5 ps con el programa *g_hbond* para obtener el porcentaje de presencia a lo largo de la dinámica por lo que 25 ns corresponden al 100%. El registro fue culitativo, ya que a éste porcentaje contribuyen igualmente la presencia de uno o más interacciones, en el caso de los aminoácidos que forman dos o más puentes en el mismo lapso de tiempo. Para éste gráfico solo se tomaron en cuenta los aminoácidos presentes en la cadena B de MEF2A y que dieron positivo para el establecimientos de puentes de hidrógeno con la caja CArG los cuales son: Gly1, Arg2, Lys4, Lys22, Arg23, Lys29, Lys30, Lys3, Asn15 y Thr19 diferenciados por una gama de colores. Los datos incluidos en la figura descienden directamente de los recabados en la dinámica molecular tratada en éste capítulo.

2.8.2. Mapa de interacciones por puentes de hidrógeno en el complejo MEF2A-CArGbox.

Se realizó el mapa de interacciones entre la CArGbox y MEF2A, que consiste en la detección precisa de los puentes de hidrógeno entre los aminoácidos de la cadena A y B de MEF2A, con los nucleótidos presentes en ambas hebras de DNA como se observa a en la figura 8.

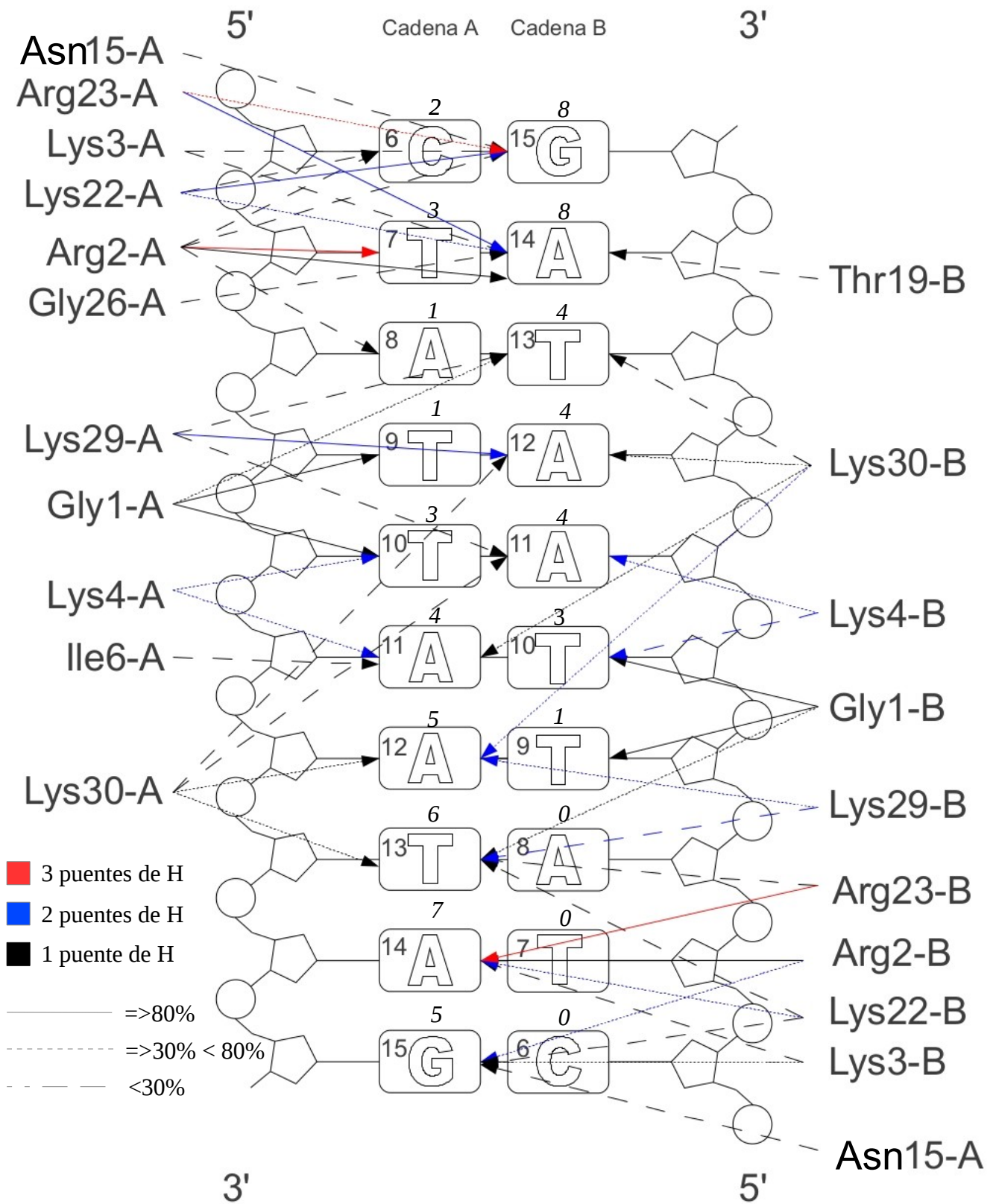


Fig. 8. Mapa de interacciones por puentes de hidrógeno entre MEF2A y su CARGbox. Se utilizaron los registros de puentes de hidrógeno derivados de la dinámica molecular de 25 ns, para representar puntualmente las interacciones sostenidas entre los aminoácidos y los nucleótidos presentes en el complejo MEF2A-CARGbox. A los puentes de hidrógeno se les representa de distintas maneras: aquellos que poseen una duración menor al 30% del total de la dinámica, los que están presentes arriba del 30% y por debajo del 80% y aquellos que permanecen durante 80% o más de la dinámica, también se incluye un esquema de colores como se indica en la parte inferior izquierda de la figura, para puntualizar el número de interacciones que se generan en la

relaciones aminoácido-ácido nucleico, ya que existen casos en las que un aminoácido establece más de un puente de hidrógeno con un solo nucleótido., En la parte superior se especifican el nombre de cada una de las cadenas de DNA, la numeración de los nucleótidos se corresponde con el veinteamero descrito en el archivo *.pdb* con id:1C7U, aunque por fines prácticos solo se reporta la sección que corresponde a la CArGbox, los números en *itálicas* que se encuentran sobre cada nucleótido corresponden al número de interacciones que cada uno presenta.

En el mapa de la figura 8, se observa con mayor especificidad la temporalidad de cada una de las interacciones entre aminoácidos y nucleótidos, como en el caso de la Lys22 de la cadena A que posee interacciones desarrolladas en los tres periodos de tiempo descritos anteriormente. Se observa una concentración relativamente mayor de puentes de hidrógeno en ciertos residuos, como el caso de G15 y A14 de la hebra B, con 8 interacciones cada una de ellas (el mayor número registrado dentro de la CArGbox), de las cuales se distinguen las establecidas con la Lys22A en el caso de G15; Lys23A y Arg2A en el caso de A14 por permanecer más del 80% durante el tiempo total de la dinámica molecular. C6 y T7 de la cadena A, establecen 2 y 3 puentes de hidrógeno respectivamente, por lo que la suma total que aportan los pares C6-G15 y T7-A14 es de 10 en el caso del primer par y de 11 en el segundo, registrando en conjunto la mayor concentración de puentes de hidrógeno.

El número mínimo de interacciones se registra en los nucleótidos C6, T7 y A8 de la hebra B. En los que el programa *g_hbond* no localizó ninguna interacción, sin embargo, las bases con las que se aparean de la hebra A sí establecen interacciones. La base G15 establece 5 puentes de hidrógeno, A14 posee 7 y T13 posee 6. Estas bases se distinguen por ser algunos de los sitios con más interacciones dentro de la CArGbox. Los pares T9-A12, T10-A11, A11-T10 y A12-T9 (nombrando en primer lugar el nucleótido de la hebra A) que conforman la sección más variable de la CArGbox, poseen 5, 7, 7 y 6 puentes de hidrógeno respectivamente, siendo la repartición casi equitativa en las bases que componen los pares T10-A11 y A11-T10, lo que distingue esta región de la CArGbox; no es este el caso de los pares T9-A12 y A12-T9 en los que una de las bases posee casi en su totalidad las interacciones presentes en el par, hecho que parece acentuarse según nos acerquemos a los extremos 3' de cada hebra.

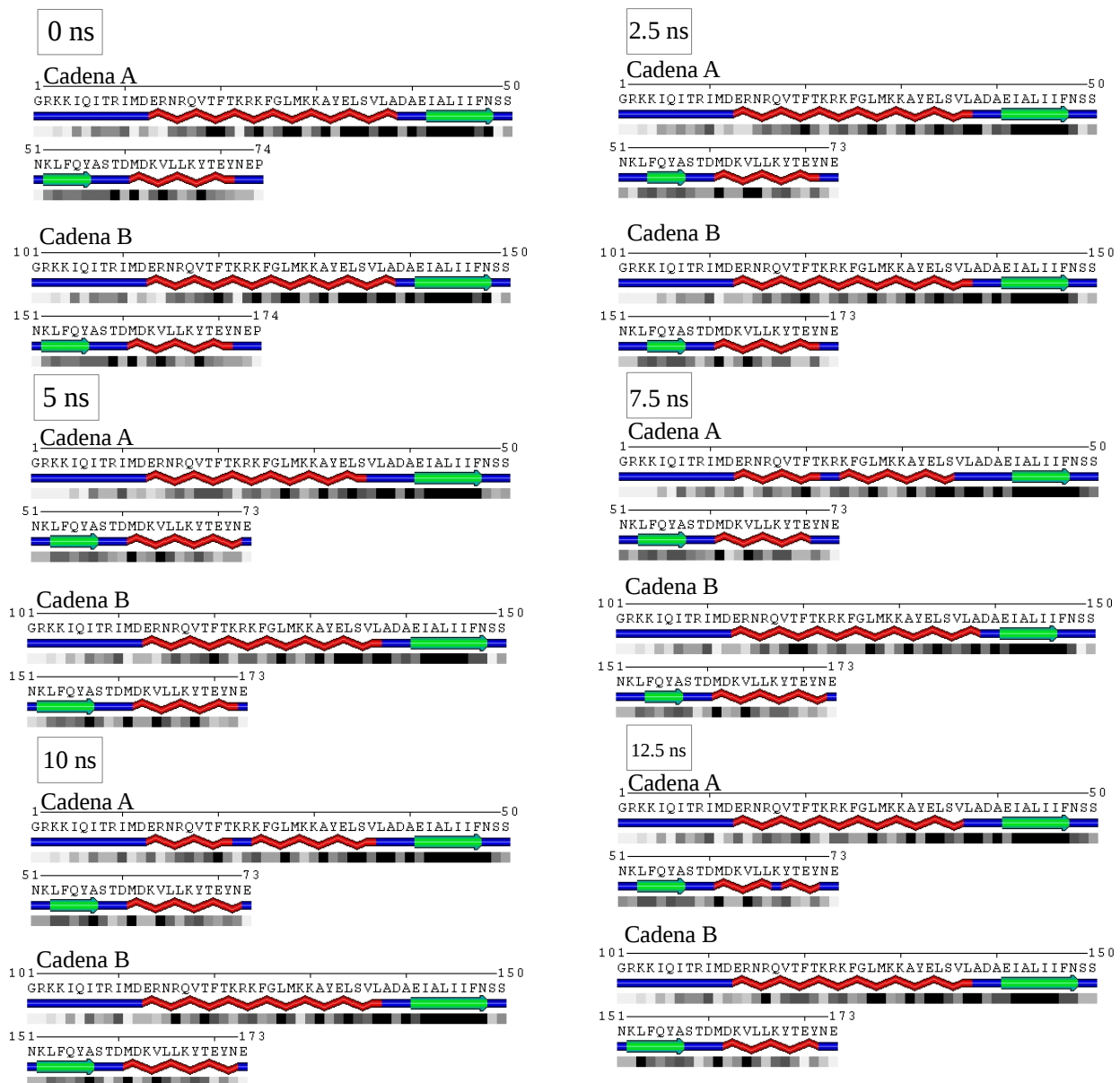
2.8.3. Registro del cambio en la estructura secundaria del factor de transcripción MEF2A durante la dinámica molecular.

Con el fin de obtener un control de los cambios ocurridos en la estructura secundaria de las dos cadenas polipeptídicas que conforman a MEF2A, se realizó un registro de la estructura secundaria cada 2.5 ns, comenzando por el registro de la semilla en el nanosegundo 0 hasta finalizar con la estructura registrada al término de la dinámica utilizando el programa POLIVIEW-2D (Lincoln D. Stein's perl GD graphics Library that can be found at Boutell.Com, Inc. Web server or in any CPAN modules collection). Véase figura 9.

El factor de transcripción estudiado está compuesto de dos cadenas polipeptídicas casi idénticas en cuanto a estructura primaria y secundaria (excepto por el comienzo de la lámina β de la cadena B en Glu 41, es decir una posición antes que en la cadena A), como lo muestran las determinaciones por cristalografía y RMN (Santinelli & Richmond, 2000; Huang *et al.* 2000). Cada una de ellas cuenta con una estructura desordenada que abarca desde el aminoácido 1 hasta el 12, como se observa a los 0ns en la determinación de estructura secundaria realizada con POLIVIEW-2D, tanto en la cadena A como en la B, al paso de 2.5 ns no se observa cambio a lo largo del tiempo de simulación, y se representa como una zona accesible al solvente, sobre todo en los primeros

cuatro aminoácidos. En el α hélice II de ambas cadenas, que abarcan del Ácido glutámico en la posición 13 hasta la Ala38, ocurren cambios de longitud, tanto en la cadena A como en la B, cuya terminación varía entre la Ser35 y la Ala38 reduciendo así la estructura desordenada que se registró en un inicio de la simulación del Asp39 al Glu41 en la cadena A y en la cadena B de Asp39 a Ala40. Aunado a los cambios antes mencionados en la α hélice I, se observa una interrupción en la estructura de la misma, según el rastreo evaluado al tiempo de 7.5 ns en los residuos de Lys22 y Arg23 que se detectaron a los 10, 15, 17.5 y 22.5 ns sólo en la cadena A.

Tanto en la cadena A como en la B, la composición de la lámina β fluctúa entre el Glu41 al Asn48, viéndose constituida por un mínimo de 6 a.a a los 7.5 ns en la cadena A sin mostrarse accesible al solvente a lo largo de la dinámica. Tanto la lámina β II como el α hélice II también poseen cambios en sus sitios de inicio y término en ambas cadenas, la primera fluctúa desde la Lys52 hasta la Ala57, y a la segunda puede encontrarse desde la Met61 hasta la Asn72. Ambas estructuras muestran accesibilidad media la solvente.



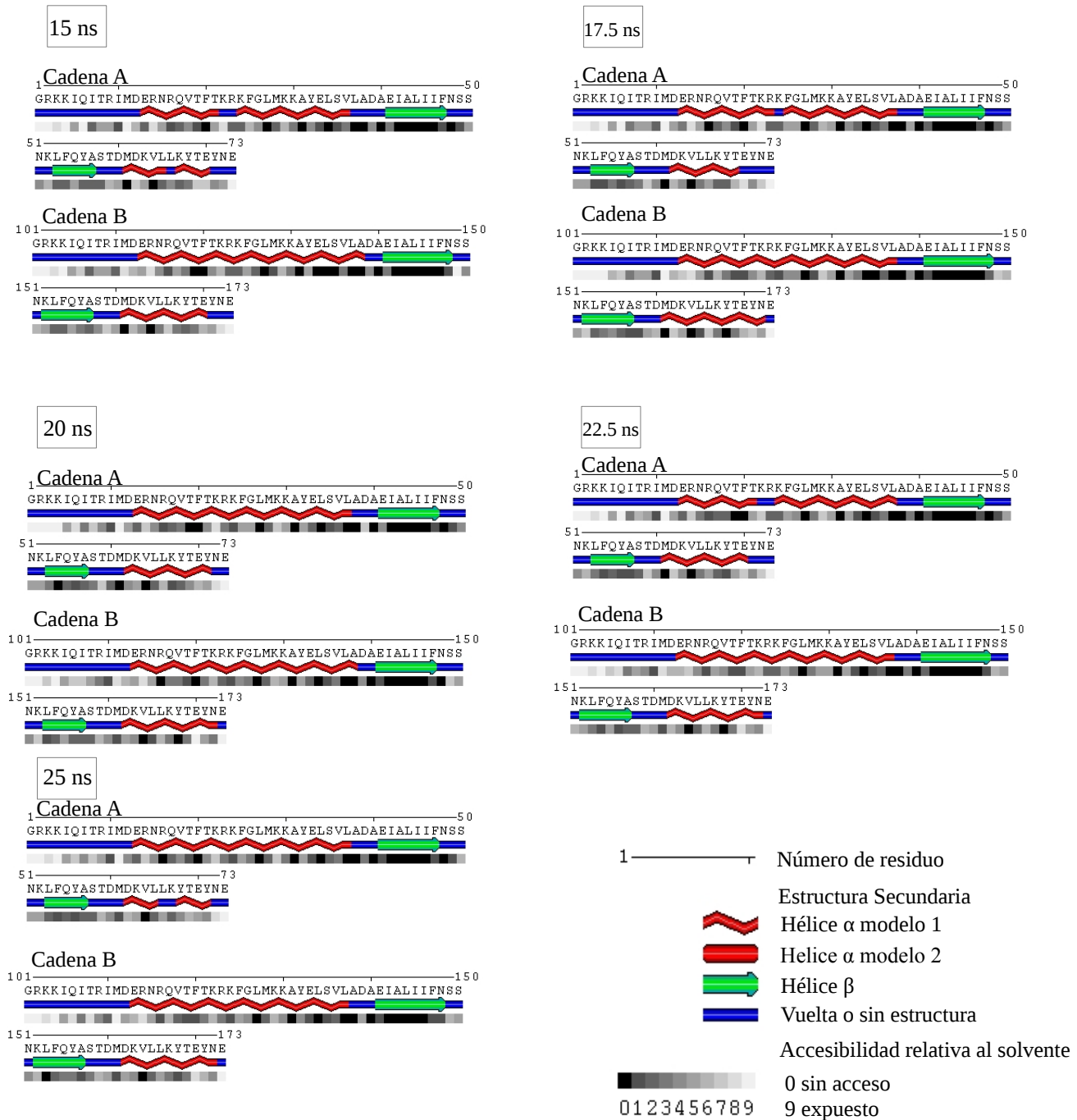


Fig. 9. Evolución de la estructura secundaria de las cadenas A y B de MEF2A a lo largo de 25 ns. Los archivos *PDB* fueron extraídos cada 2.5 ns y fueron procesados en línea con la paquetería de POLYVIEW-2D para obtener la visualización de la estructura secundaria. Cada muestra posee un encabezado que especifica el tiempo de sondeo e indicado como “Cadena A ó cadena B” podrá observarse la estructura secundaria de cada polipéptido; cuya representación está indicada en la parte inferior derecha de la figura. La numeración de los aminoácidos se encuentra en la parte superior de su notación a una letra, en el caso de la cadena B, se comienza desde el 101 para facilitar su diferenciación de la cadena A. En la parte inferior de las representaciones de estructura secundaria de halla indicado la accesibilidad al solvente que corresponde a un gradiente de 0 a 9 expuesto así en la parte inferior derecha de la figura. Dentro de las variaciones observadas vale la pena resaltar el cambio de estructura experimentado en el α helice I de la cadena A ubicado en los residuos de Lys 22 y Arg 23 a los 10, 15 y 22.5, al igual que en la Lys 24 en el ns 17.5.

2.9. DISCUSION DEL CAPÍTULO II.

2.9.1. Minimización energética del sistema.

Con el fin de evitar una ruptura de los enlaces moleculares durante la dinámica y de acelerar las simulaciones posteriores a la minimización, se ajustaron los enlaces e interacciones estéricas de los átomos a conformaciones de mínima energía. Conforme al diseño bajo el cual se realizó la minimización, se esperó su término en los siguientes casos:

- Si el número de iteraciones excedía el rango de los 30000 pasos ó
- Sí las fuerzas presentes en el sistema fueran inferiores al valor máximo de $100 \text{ kJ mol}^{-1} \text{ nm}^{-1}$.

Debido a la ocurrencia del segundo caso, la minimización del sistema cesó a los 9318 pasos que corresponden a 9318 ps al registrarse una fuerza máxima de $90.57 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ en el átomo 54593 que corresponde a un hidrógeno de una molécula de H_2O . El registro de la energía potencial durante este proceso fue de gran ayuda, ya que como se observa en la Fig. 10, el sistema experimentó un descenso súbito de la U durante los primeros 2000 ps para estabilizarse durante un descenso de menor grado hasta el final de la minimización en la que se registró una U_{total} de $-1.45 \times 10^6 \text{ kJ/mol}$, resultado conveniente, ya que se considera como exitosa la minimización con valores de *emtol* iguales a 1000 Kj/mol utilizando el algoritmo *Steepest Descent* y con sistemas de tamaño similar al manipulado (A Strasser *et al.*, 2013).

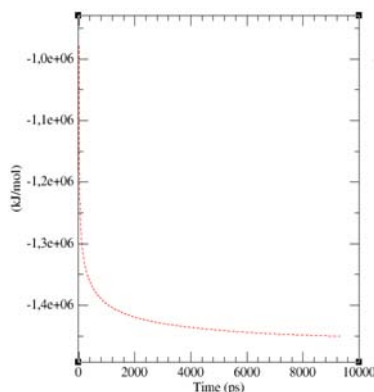


Fig. 10. Registro de la U_{Total} del sistema a lo largo del proceso de minimización durante 9318 ps. Se incluyeron a todos los grupos del sistema (proteína, ácido nucleico, iones y agua) para éste cálculo que fue realizado con los registros basados en un valor de *nstenergy* = 10. La línea roja punteada corresponde al valor de U_{total} en el que se observa un rápido decaimiento de la misma durante los primeros 2000 ps y una meseta de estabilización bien definida de los 8200 a los 9318 ps. El gráfico fue obtenido utilizando el programa *g_energy* cuyo resultado es visualizado en el procesador de gráficos *xmgrace*.

2.9.2. Fijación de las posiciones del soluto durante la relajación del disolvente.

El sistema se sometió a una dinámica de posiciones restringidas durante 250 ps, lapso reportado como adecuado para sistemas simulados con el ensamble NTP (Santucci *et al.*, 2009; Oostenbrink & Van Gunsteren, 2003). Tras éste periodo, se observaron lapsos constantes de estabilidad durante toda la simulación de energía total, densidad, temperatura y presión (Fig, 11), de los primeros tres parámetros se reportan oscilaciones mínimas

alrededor del valor deseado; la presión varia posee variaciones mayores, sin embargo estos cambios se consideran como típicos para el tipo de ensamble ocupado en ésta metodología (Hess B *et al.*, 2000-GROMACS demo for GROMACS version 4.5). Debido a estas observaciones se juzgó como equilibrado al sistema y por lo tanto adecuado para conducir a una dinámica en la que no hubiese restricciones dentro del sistema.

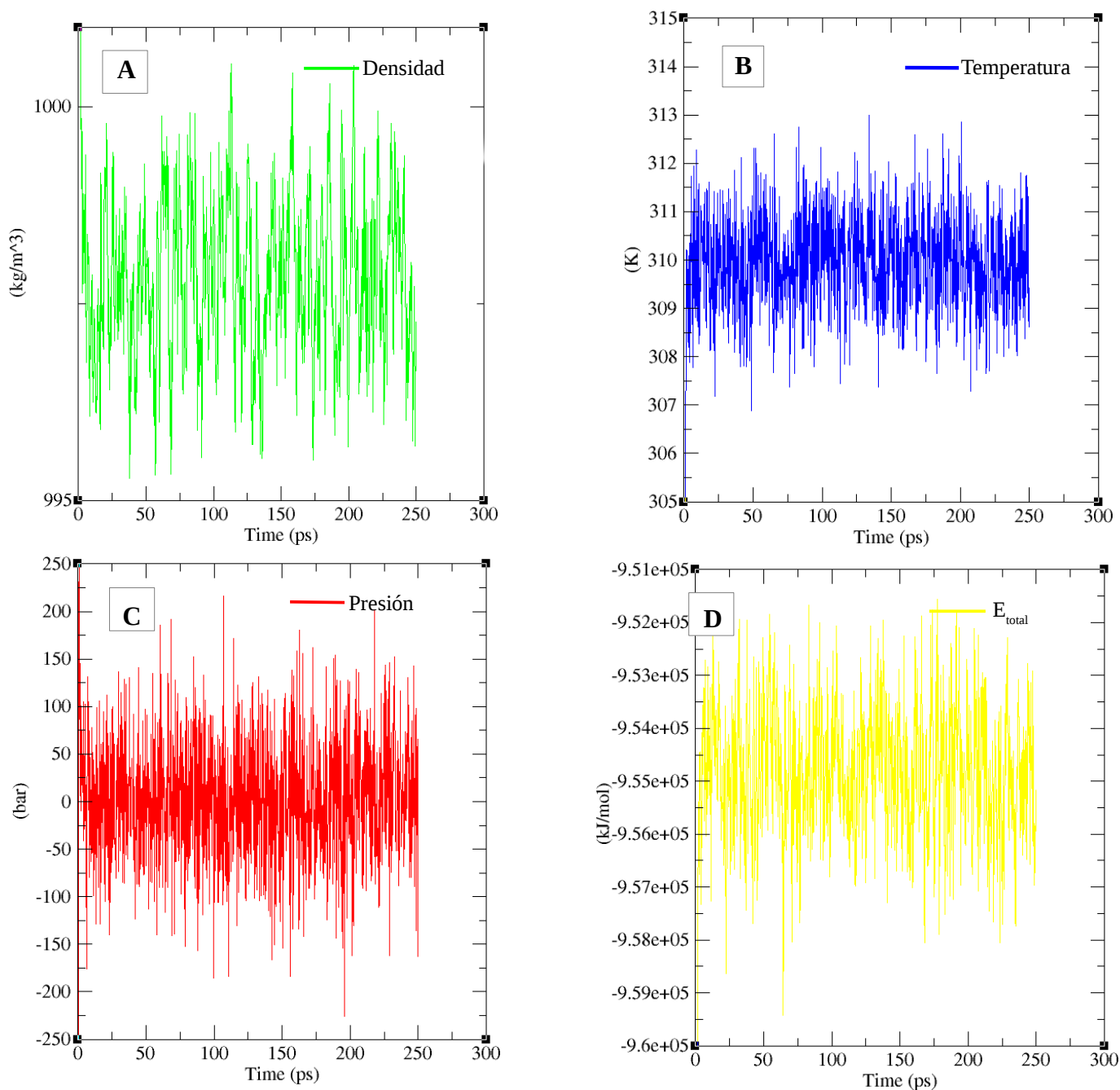


Fig. 11. Registro de densidad, temperatura, presión y E_{TOTAL} a lo largo de 250 ps de dinámica con posiciones restringidas. Todos los parámetros fueron registrados desde el inicio hasta el final de la dinámica sin interrupción procesándolos con el programa *g_energy* y visualizados con *xmgrace*. En el panel A en color verde se observan los registros de densidad, cuya variación se mantiene aprox. entre $997.5 \pm 2.5 \text{ kg/m}^3$ con un perfil estable al igual que lo observado en el panel B en azul que corresponden a la temperatura total del sistema cuyos valores oscilan levemente al rededor de 310 K; el caso de la E_{TOTAL} observada en amarillo en el panel D, ya que su variación es mínima a lo largo de la simulación. Finalmente en el panel C con color rojo se observa una oscilación mucho más contundente en la presión, registrándose valores con amplitud de hasta 200 bar.

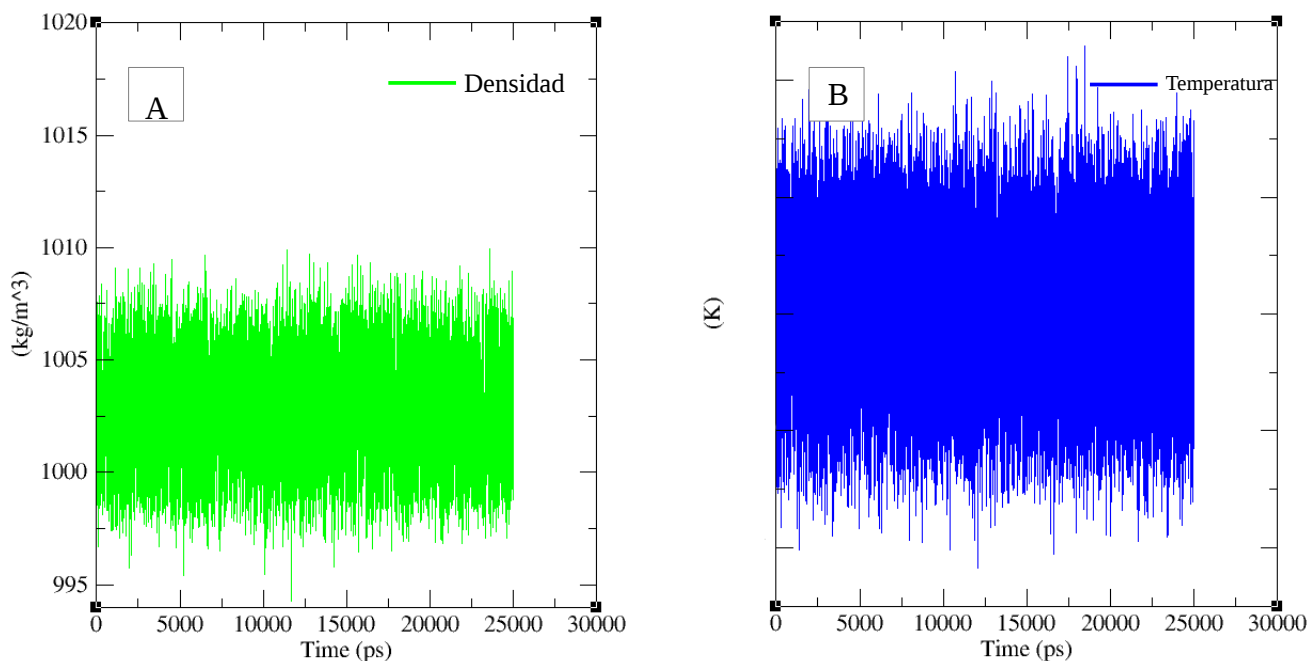
2.9.3. Control de parámetros físicos a lo largo de la dinámica molecular de todos los átomos.

Con el fin de tener un control sobre el comportamiento del sistema durante la dinámica de todos los átomos, se tomó registro de la densidad, temperatura, presión y la E_{TOTAL} a lo largo de los 25 ns de la simulación.

La oscilación de la densidad (ρ) es mucho mayor que en el caso de la dinámica con posiciones restringidas, esto debido a que se corresponde con el aumento y disminución de la presión en el sistema, ya que la relación inversamente proporcional entre p y V provoca aumentos en la densidad, que corresponden a los aumentos de magnitud en p .

Aún tomando en cuenta éstos cambios mayores en ρ y p consideraremos al sistema como estable a lo largo de la dinámica, ya que el rango de variación en el transcurso de ésta es constante y presenta un perfil bien definido como en el caso de la E_{TOTAL} , que no posee oscilaciones abruptas a lo largo de los 25 ns, que podrían ser observados durante la ruptura de un enlace o incluso la disolución del complejo MEF2A-CARbox, hecho no observado durante la visualización de la dinámica en el programa VMD.

En este análisis cabe resaltar la poca variación de temperatura durante la simulación que puede ser debida a la implementación del algoritmo *v-rescale*, que aunque utiliza la misma constante de acoplamiento para el tiempo (τ) que el termostato de Berendsen, introduce un término estocástico adicional que fuerza al sistema hacia una correcta distribución de la energía cinética, por lo que cualquier exceso o defecto de temperatura con respecto al termostato será corregido en poco tiempo. Los parámetros antes mencionados pueden visualizarse en la Fig. 12.



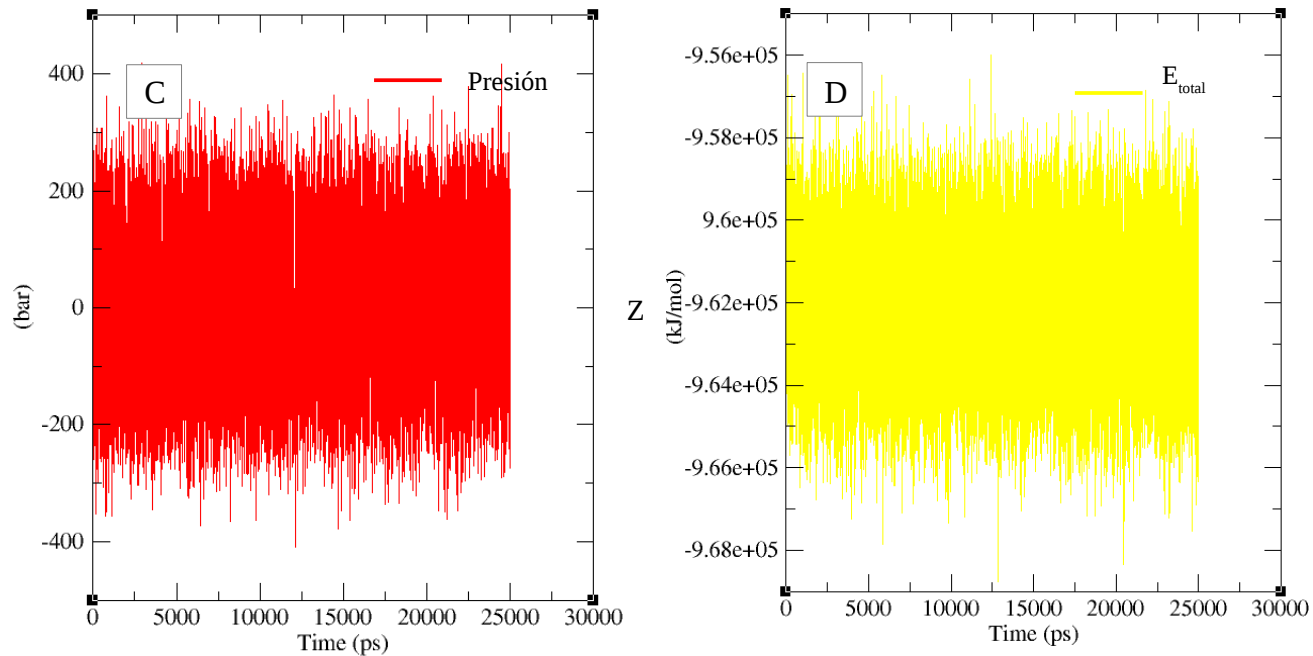


Fig.12. Registro de densidad, temperatura, presión y E_{total} durante 25 ns de dinámica de todos los átomos. Cada una de las gráficas se obtuvieron procesando los datos arrojados desde el inicio hasta el final de la dinámica con el programa *g_energy* y fueron visualizados con *Grace*. En el panel A con color verde se encuentra la gráfica de densidad que presenta valores promedio de $1002.5 \pm 7 \text{ Kg/m}^3$, en el panel B en azul se observa la oscilación moderada de la temperatura al rededor de los 310 K, el mismo patrón de oscilación moderada es seguido por la E_{total} del sistema mostrado con color amarillo en el panel D. En el panel C con color rojo se observa la gráfica de presión.

2.9.4. Evolución en el acomodo espacial del complejo MEF2A-CArGbox.

Debido a la metodología empleada en la medición de la Energía Libre de Unión, es conveniente analizar los cambios acontecidos en la estructura del complejo, ya que la determinación de la ΔG_{bind} está basada en una estructura inerte, que de presentar cambios drásticos a lo largo de la dinámica, invitarían a tratar estos datos con mayor cautela, ya que acaecería un valor intrínseco de error sobre nuestros cálculos, debido a la consideración de posibles estructuras alternas las cuales pudiesen arrojar ΔG_{bind} diferentes a las observadas en la mediciones en el arquetipo.

Se calculó la RMSD entre las estructuras iniciales y finales del complejo, es decir, al inicio y término de la dinámica, con el propósito de cuantificar el cambio en la estructura del mismo. El resultado de este estudio (realizado con la paquetería de GROMACS), arrojó un valor de RMSD de 1.13 \AA , por lo que los cambios en la distribución espacial de la estructura del DNA y la proteína fueron mínimos, como se observa en la Fig. 13.



Fig13. Alineamiento estructural del complejo MEF2A-CARgbox al inicio y al final de la Dinámica Molecular. Se aisló el último cuadro de los 25 ns totales de dinámica y se alineó con la estructura reportada en el archivo .pdb 1C7U que fue tomado como semilla para la simulación. Para la alineación se tomaron en cuenta únicamente el esqueleto de la proteína y el esqueleto de fosfatos de la CARgbox. Representado en diagrama de delineado, en color verde se muestra el complejo que corresponde a la semilla y en rojo el complejo al término de la simulación. Los cambios en el acomodo espacial son mínimos lo que se ve determinado por un RMSD de 1.13 Å, éstas diferencias se localizan principalmente en el extremo 3' de la hebra A del DNA en las regiones del amino terminal en la proteína y sobre todo en el carboxilo terminal. El gráfico fue obtenido con el visualizador de moléculas VMD.

A la par de esta comparación se realizó un estudio de RMSD más completo a lo largo de toda la dinámica, en el que la gama de variaciones estructurales presentes en el complejo se determinaron utilizando el programa *g_rms* cada 5 ps. Se realizaron tres estudios de RMSD, el primero para el complejo, otro para la proteína y el último para el ácido nucleico. En todos los estudios se tomó en cuenta el esqueleto de aminoácidos y el de fosfatos respectivamente, además de que el RMSD para el tiempo t fue obtenido de la comparación con el tiempo $t-0$, de este modo los estudios realizados fueron una comparación de estructuras que se obtuvieron a lo largo de la dinámica con respecto a la estructura inicial, es decir, reflejan que tanto se parece el complejo a lo largo de la dinámica con respecto a la estructura inicial, es decir, reflejan que tanto se parece el complejo a lo largo de la dinámica con respecto a la estructura con la que se inició.

El valor más alto de RMSD proveniente del complejo, corresponde a 0.41nm registrado a los 25000 ps de iniciada la dinámica, el valor promedio es de 0.38 nm con una desviación estándar de 0.63359, lo que refleja valores de RMSD que no rebasan los 3 Å y por lo tanto, están considerados como cambios mínimos dentro de la estructura (Reva *et al.*, 1998). La desviación estándar obtenida de los datos refleja que los valores de RMSD a lo largo de la simulación se mantienen cercanos al promedio, por lo que en la Fig. 14, no se muestran cambios considerables en el acomodo espacial del complejo. Tomando en cuenta que el último valor de RMSD obtenido al final de la dinámica es el más alto se considera que sería prudente llevar a cabo una simulación más para observar en qué punto se estabiliza la estructura. Sin embargo cabe resaltar que a partir del picosegundo 15000 se puede apreciar claramente una meseta, lo que indicaría que probablemente la estructura se estabilizó.

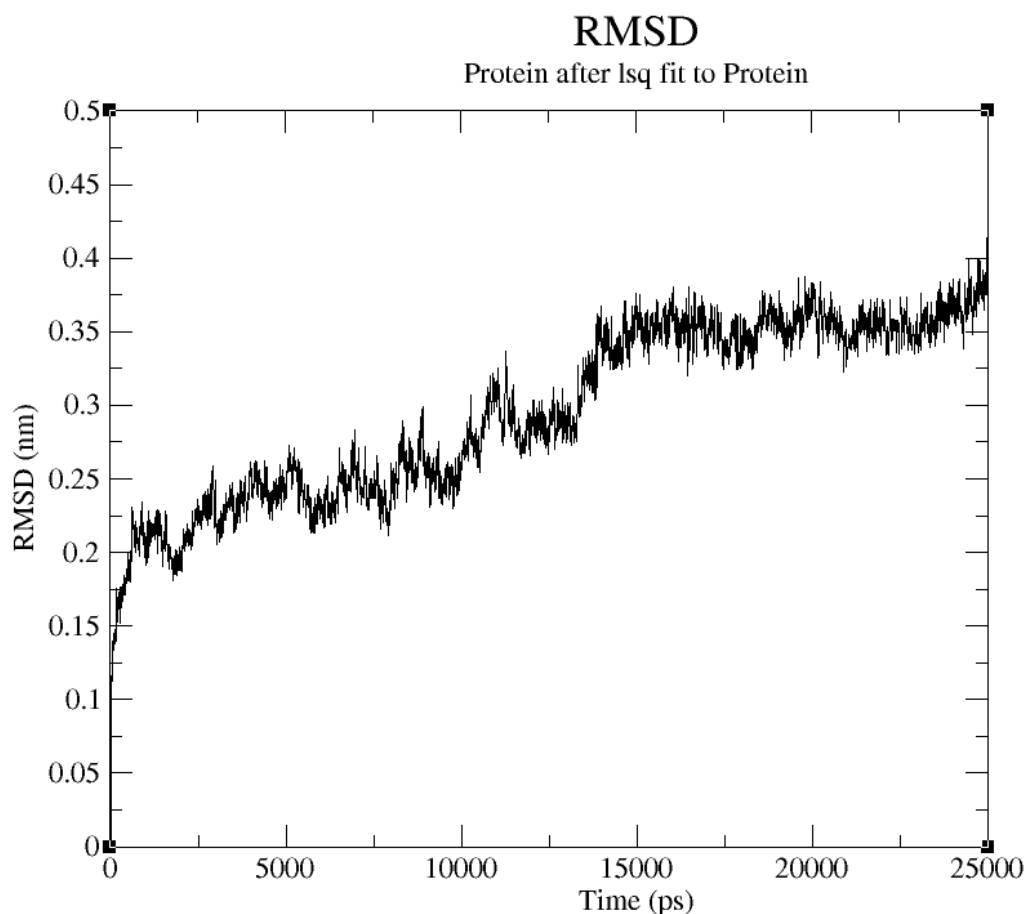


Fig.14. RMSD del complejo MEF2A-CARgbox a lo largo de una dinámica de 25ns. Los registros de RMSD fueron obtenidos cada 5ns, para este estudio se tomó en cuenta únicamente el esqueleto de aminoácidos y el de fosfatos presentes en la proteína y el ácido nucleico. En el gráfico se observa una tendencia a alejarse de la estructura inicial del complejo de la cual se partió, ya que progresivamente el cálculo de RMSD aumenta su magnitud conforme avanza la dinámica, lo que puede deberse al tiempo de corte de la misma en el que no se logró estabilizar por completo al complejo, sin embargo tómesese en cuenta que el valor de RMSD no excede de 1 nm. El estudio fue realizado con el programa *g_rms* incluido en la paquetería de GROMACS y graficado con *xmgrace*.

Por otro lado, el estudio de RMSD realizado en el DNA y la proteína por separado, refleja un comportamiento equivalente al del complejo, obteniendo un promedio de 0.37 y 0.18 nm para la proteína y el DNA respectivamente, con una desviación estándar de 0.39 en el caso del primero y de 0.109 en el caso de la CARgbox (Fig. 15). Los valores de RMSD obtenidos al final de la simulación son de 0.38 nm para la proteína y de 0.077 en el caso del DNA, lo que sugiere estructuras muy similares a la semilla, sobre todo en el caso del DNA cuyo cálculo de RMSD es menor a 1Å. Estos resultados sugieren que la zona de interacción entre el factor de transcripción y su elemento cis es poco variable, ya que el ácido nucleico no experimenta grandes cambios durante la dinámica y los cambios (aunque mínimos, se acentúan en comparación con el DNA) presentes en la proteína puedan deberse principalmente al movimiento experimentado en la zona del C-terminal de la proteína donde se encuentra depletada la estructura del dominio MEF2A.

Gracias al análisis realizado anteriormente, podemos inferir que los datos aportados por la metodología elaborada para medir la ΔG_{bind} que está basada en la estructura arquetípica reportada en la determinación por resonancia magnética de Huang et al 2000, será efectiva porque se basa en una arquitectura representativa del

complejo, sobre todo en el caso del DNA cuya variación es mínima. El modelo arquetípico es representativo de de la estructura que posee la CArGbox y MEF2A durante la formación del complejo.

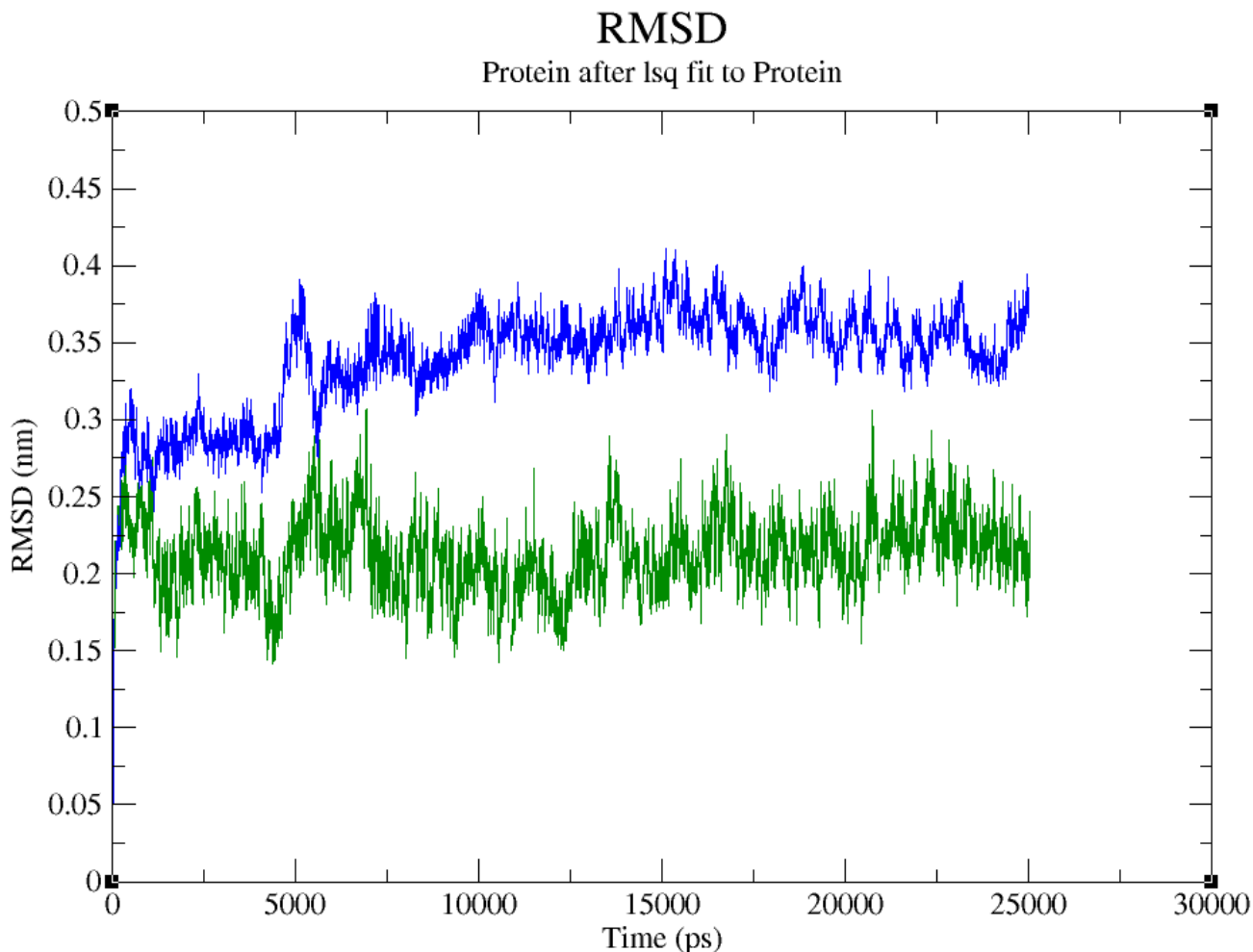


Fig.15. Valores de RMSD de MEF2A y su CArGbox registrados a lo largo de una dinámica de 25 ns. Los valores de RMSD fueron calculados cada 5 ns, se realizaron dos registros: uno correspondiente a la proteína, en azul, basado en el esqueleto de aminoácidos y el otro correspondiente a la CArGbox en verde, tomando en cuenta el esqueleto de fosfatos. El valor promedio de RMSD de la proteína es de 0.3722 y el del ácido nucleico es de 0.185. Los valores de RMSD se obtuvieron con ayuda de la paquetería de GROMACS y las gráficas fueron realizadas en *xmgrace*.

Si bien se ha demostrado anteriormente la inexistencia relativa de los cambios en términos del acomodo espacial, tanto en el DNA como en MEF2A y por consiguiente en el complejo, es debido marcar diferencia entre cambios cuantitativos: en los conjuntos de coordenadas atómicas presentes al inicio y al final de la dinámica expresados en una distancia media cuadrática mínima; y los cambios cualitativos que le acontecen al complejo o a uno de sus integrantes durante la misma. Es de esencial importancia explorar la estructura secundaria de la proteína a pesar de haber inmutabilidad relativa a lo largo del tiempo con respecto a sus posiciones generales.

2.9.5. Estudio sobre la evolución de la estructura secundaria de MEF2A a lo largo de la dinámica.

Uno de los análisis más significativos que podemos realizar acerca de la proteína, es sobre el cambio de estructura secundaria que experimentó debido al movimiento de los átomos en el polipéptido y su interacción entre las dos cadenas que conforman a MEF2A así como con el oligonucleótido al cual se encuentra asociado e

incluso con el solvente, ya que al suceder una redistribución de los átomos en regiones específicas, podría verse alterada la estabilidad del complejo formado con la CArGbox o del mismo dímero. Dado que los estudios realizados con anterioridad se basan en imágenes estáticas, será de gran utilidad observar a lo largo de un lapso de tiempo el comportamiento de las α hélices y las láminas β ya que tenemos la posibilidad de describir conformaciones que hasta el momento no se han observado y aún más importante, las consecuencias que tienen éstos cambios sobre la interacción entre MEF2A y su CArGbox.

Las determinaciones de estructura secundaria fueron tomadas cada 2.5 ns para la cadena A y B, del dímero. Como se ha expuesto en la sección de resultados correspondiente a este análisis, se localizaron cambios puntuales en la estructura secundaria, éstos pueden ser considerados como mínimos dentro del entorno de cambios posibles a lo largo de 25 ns, por lo que el hecho anterior nos da una idea de la estabilidad de la estructura cuaternaria de MEF2A, misma que se ve reflejada durante la formación del complejo. Ahondando sobre esta última afirmación, surge una cuestión interesante: corroborar si la estabilidad que posee el FT es debida a su unión con el oligonucleótido o si en su ausencia aumentaría su plasticidad. Debido a la acotación de este trabajo al respecto del complejo por ahora es una pregunta que queda para el análisis posterior.

Si bien se ha juzgado como mínimos los cambios que existen en la estructura, algunos de ellos podrían estar relacionados con el reconocimiento y la unión del dímero a su CArGbox, por encontrarse en zonas que durante este trabajo fueron reconocidas como fundamentales para la unión del complejo. Tomemos en cuenta las regiones accesibles al solvente de la zona de unión entre ambas moléculas, ya que será este parámetro de accesibilidad el que nos hará referencia de los sitios que también son accesibles para la interacción con la CArGbox (para este estudio sólo se analizó a la proteína libre, por lo que las regiones reportadas como accesibles al solvente, corresponden en algunos casos a contactos con el DNA). Tanto Gly1 como Arg2 mantienen valores constantes de 9 en ambas cadenas -el valor más alto de accesibilidad al solvente- a lo largo de toda la dinámica, lo que se corresponde con las interacciones de puentes de hidrógeno entre el factor de transcripción y el oligonucleótido. ya que ambos aminoácidos son fundamentales para la estabilidad del complejo.

El α hélice I es un foco de interacciones en el complejo, ya que es el sitio (junto con algunos de los aminoácidos del extremo amino terminal) donde se realiza la unión entre el polipéptido y el oligonucleótido (conservado en toda la familia MADS). Este hecho bien conocido en la literatura (Huang, et al. 2000, Santelli & Richmond 2000, Wu, Et al. 2010, He, Et al. 2011) se observa en el caso de los aminoácidos más expuestos al solvente (debido a las vueltas dentro del α hélice), dígase por ejemplo de la Lys22, Lys29 y Lys30.

Uno de los resultados más significativos de éste estudio, fue el registro de un cambio de estructura secundaria en el α hélice I en los residuos de Lys22, Arg23 y Lys24, ya que se relaciona directamente con la interacción de puentes de hidrógeno existente entre MEF2A y su CArGbox. Las observaciones de éste tipo se dataron a los 7.5, 10, 15, 17.5 y 22 ns de dinámica molecular, por lo tanto es un cambio intermitente de estructura, ya que nunca se sostuvo a lo largo de la dinámica; sin embargo, este cambio se presentó en tres aminoácidos que forman puentes de H con el oligonucleótido: Lys22 de la cadena A con G15 y A14 de la

cadena B del DNA (Cfr. La Fig. 9 de la sección 2.8.3) como se observa en la Fig. 16.

Nótese que a los cambios de estructura les corresponde una reducción en el número de puentes de hidrógeno (ya sea con G 15 o A 14) como se observa claramente a los 7.5, 10, 15, 17.5 y 22.5 ns; lo que puede deberse a la pérdida en la estructura del α hélice, es decir, cuando el residuo Lys22 posee esta conformación, es capaz de formar dos puentes de hidrógeno con diferentes nucleótidos y al perder esta estructura también pierde dicha capacidad. La pérdida de estructura puede ser causada por la intimidad en la interacción con los nucleótidos ya que al mantenerse aferrados a este residuo, le provocan una pérdida de su contexto conformacional al exigir una cercanía mayor entre el nucleótido y la Lys, en comparación a la sostenida con los residuos adyacentes con los que conforman la hélice. La relevancia de estas conjeturaciones sólo podrá advertirse en un estudio con mayores puntos de sondeo a lo largo de la dinámica, sin embargo es importante resaltar resultados que pueden ser la base del comportamiento de éste evento para que con la profundización de su estudio pueda concretarse ésta visión inicial al incluir los factores de los que se carecen en el presente estudio.

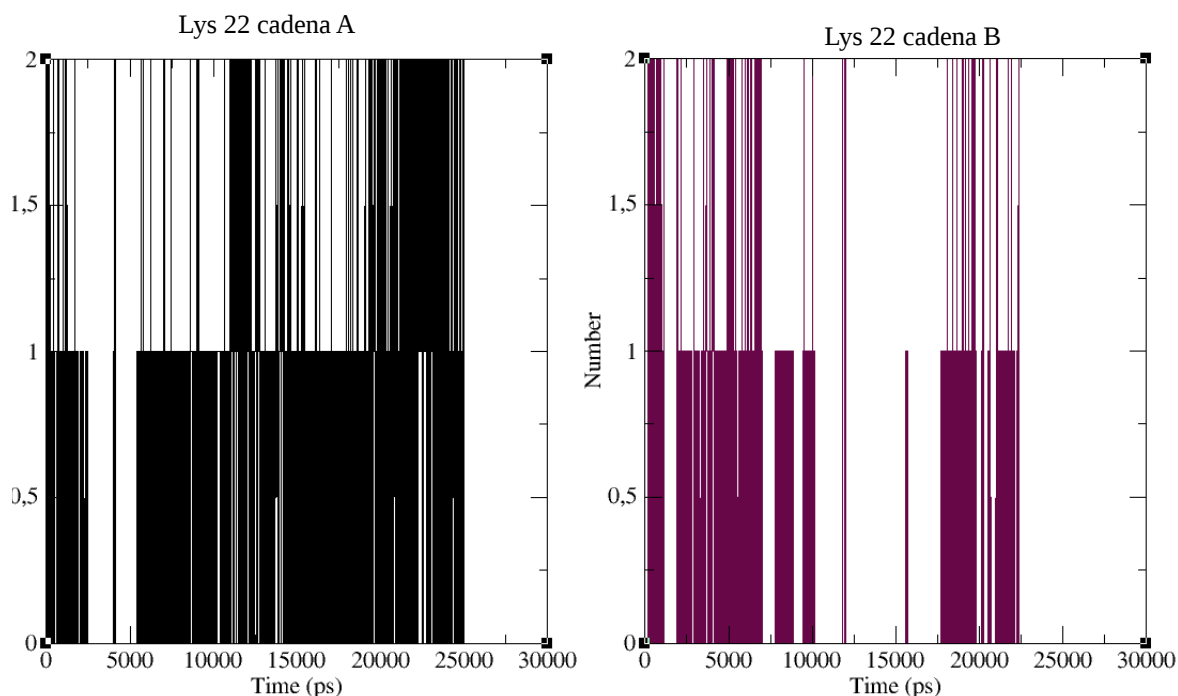


Fig.16. Puentes de hidrógeno formados entre los residuos de Lys 22 de las cadenas A y B de de MEF2A y ciertos nucleótidos presentes en las cadenas I y II de la CArGbox a lo largo de 25 ns. Los registros de puentes de hidrógeno individuales observados en la sección 2.8.1 fueron conjugados, por lo que en la gráfica que corresponde a la Lys 22 de la cadena A se graficaron en color negro las interacciones sostenidas con los nucleótidos A14 y G15 de la cadena II, y para el residuo de Lys 22 de la cadena B se registraron en color uva las interacciones sostenidas con T 13, A 14 y G 15 de la cadena I de la CArGbox. En la gráfica que corresponde a los puentes de hidrógeno formados por la Lys 22 de la cadena A se observan interacciones más duraderas que en su contraparte de la cadena B. El eje y corresponde al número de interacciones y el eje x al tiempo de duración de la dinámica.

La cadena B del dímero MEF2A posee tres asociaciones por puente de Hidrógeno con T 13, A 14 y G 15 de la cadena A de la CArGbox, y como observamos en las representaciones de estructura secundaria, el α -hélice de la cadena en cuestión nunca pierde su estructura, lo que iría a favor de la hipótesis que considera que al permanecer la hélice sin cambios, la posibilidad de interacción se acentúa y al verse interrumpida las interacciones son más específicas e individualizadas.

La Arg23 es otro de los aminoácidos que se ve inmiscuido en el cambio estructural ocurrido en el hélice α , estando en la cadena A interactúa con T 13 y A 14 de la cadena II del oligonucleótido su contraparte de la cadena B forma puentes de hidrógeno con los mismos nucleótidos pero de la cadena I. Al igual que en el caso de Lys23 las interacciones están sostenidas por mayor tiempo a lo largo de la dinámica formando un menor número de interacciones durante el mismo lapso de tiempo. Nótese en la Fig. 17 la ocurrencia de tres puentes de hidrógeno en el residuo Arg23 de la cadena B a lo largo de la dinámica al contrario de lo ocurrido en la cadena A en donde prevalecen la formación de dos puentes de hidrógeno.

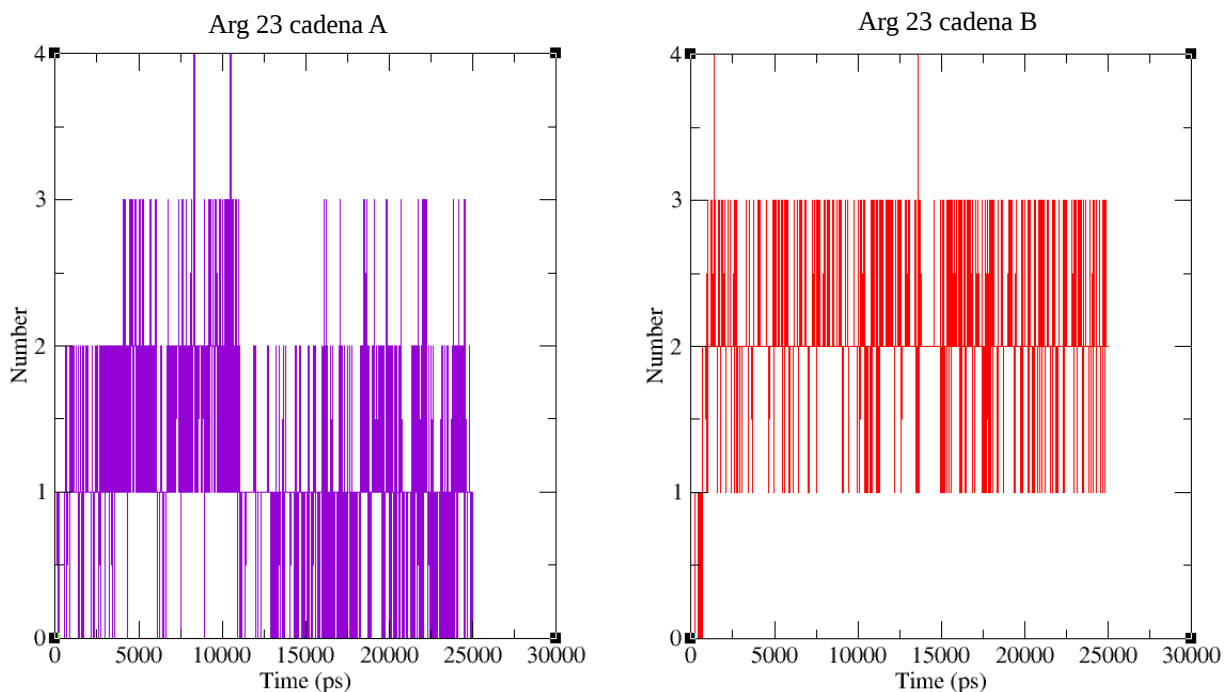


Fig.17. Puentes de hidrógeno formados entre los residuos de Arg 23 de las cadenas A y B de MEF2A y ciertos nucleótidos presentes en las cadenas I y II de la GarGbox a lo largo de 25 ns. Los registros de puentes de hidrógeno individuales observados en la sección 6.6... fueron conjugados, por lo que en la gráfica que corresponde a la Lys 22 de la cadena A en color violeta se graficaron las interacciones sostenidas con los T 13 y A 14 de la cadena II, y para el residuo de Lys 22 de la cadena B se registraron en color rojo las interacciones sostenidas con T 13, y A 14 de la cadena I de la CArGbox. Se observa la diferencia en el número de puentes de hidrógenos establecidos por Arg 23 en la cadena A y B, registrándose un menor número de éstos a lo largo de la dinámica en la cadena A que en la B, pese a que en ambas existe interacción de puentes de hidrógeno con MEF2A a lo largo de los 25 ns. El eje y corresponde al número de interacciones y el eje x al tiempo de duración de la dinámica.

El residuo Lys24 de la cadena A experimentó un cambio estructural a los 17.5 ns, único tiempo en el que fue registrado y en el que se observa una clara relación con la formación de un puente de hidrógeno formado con el residuo de A 13 presente en la cadena II de la CArGbox, por lo que se nota claramente la influencia que éste cambio tuvo para la formación de la interacción a partir de su aparición.

Al comparar el porcentaje total en los tiempos de duración de los puentes de hidrógeno a lo largo de la dinámica de los residuos de Lys22 y Arg23 de la cadena A con los mismos residuos en la cadena B, observamos que las interacciones formadas con los aminoácidos de la cadena A, ocupan tiempos más largos que en el otro péptido lo cual sugiere que la pérdida momentánea de la estructura en la α -hélice probablemente se deba a la

inestabilidad en los puentes de hidrógeno intercadena que participan en sostener dicha estructura, por lo que la disolución de éstos estaría ligada con la formación de nuevos nexos con los nucleótidos presentes en el DNA (que al parecer logran interrumpir las aportaciones energéticas necesarias para sostener la estructura de la doble hélice) derivando en una aportación energética mayor para la sustentación del complejo. Cabe resaltar que durante éste proceso se observa un equilibrio entre la aparición y pérdida en la estructura de la hélice, ya que como se ha mencionado anteriormente el evento sucede de forma intermitente, ya que probablemente un cambio definitivo afectaría la estabilidad estructural de la proteína afectándole así en su funcionamiento.

Finalmente los cambios observados en las láminas β I y II en ambas cadenas de ME2A así como los registrados en el hélice α II, fueron más sutiles y corresponden probablemente en mayor grado a la interacción entre ambas cadenas que están encargadas de sostener la estructura cuaternaria del dímero.

2.9.6. Distribución de los puentes de hidrógeno en las interacciones aminoácido-ácido nucleico.

En numerosos reportes se ha asentado la conformación canónica que posee la CARGbox a la que se une el factor de transcripción MEF2A (Kradolfer et al., 2013). Las estructuras arquetípicas de los elementos-cis, provienen de una construcción estadística basada en las recurrencias por sitio de los nucleótidos observados en dichas posiciones como resultado de múltiples trabajos experimentales, a partir de éstas observaciones se conjuntan los datos obtenidos por sitio y se toman en cuenta la recurrencia de los nucleótidos por sitio, al conjuntar suficientes datos, podrá determinarse la probabilidad de que cierto nucleótido ocupe un lugar determinado, de este modo tendremos una idea de la secuencia de bases que posee el sitio de unión de un factor de transcripción.

Sin duda, debajo de la construcción estadística de estas secuencias, subyacen los principios químicos, físicos y biológicos que dan lugar en última consecuencia, a la selección particular de la frecuencia de cada una de las bases que conforman el elemento-cis y que se traduce en la especificidad que posee un TF. Al comprender estos principios, seremos capaces no solamente de reconocer el alfabeto de nuestra secuencia blanco, sino también de reconocer la naturaleza de la interacción entre TF y sus sitios de unión a DNA.

Una de las fuerzas principales que permiten la formación de los complejos TF-elemento-cis son los puentes de hidrógeno (como en el caso de la estructura secundaria de las proteínas y de la doble hélice de DNA). Se ha puesto énfasis en su estudio obteniendo así un mapa de interacciones que descende de un modelo dinámico, en éste se aprecia su disposición a lo largo de la CARGbox, en la que se distingue una distribución no aleatoria de las interacciones. Si se pone atención en el número de interacciones que posee cada uno de los pares de base, se notará que en los primeros cuatro sitios que componen la CARGbox, uno de los nucleótidos del par contiene mayor número de interacciones que su base complementaria, si partimos en dirección 3' \rightarrow 5' (en ambas hebras) observaremos que los residuos de G15, A14, T13 y A12 poseen un número de interacciones mayor que el de su base complementaria; en contraste a los nucleótidos que componen la parte central de la CARGbox (residuos 10 y 11) que poseen un número equivalente de puentes de hidrógeno en cada uno de los nucleótidos del par.

Para explicar la distribución de los puentes de hidrógeno, debemos remitirnos al acomodo espacial de los

átomos dentro del complejo, el ácido nucleico y la proteína poseen dos zonas de interacción: el surco mayor y el surco menor del DNA. En el primero se concentran las interacciones que se establecen con los aminoácidos que están dentro del α -hélice I presentes en ambas cadenas de MEF2A, posicionadas en las bases G15, A14, T13 y A12, por lo que estos sitios de contacto sólo tienen lugar en el surco mayor, al contrario de los aminoácidos cercanos al N-terminal que están en contacto con el surco menor.

Es importante tomar en cuenta las distinciones particulares que muestran los pares de bases según la dirección en la que se encuentren, ya que gracias a estas diferencias MEF2A puede reconocer el par A-T del par T-A lo mismo que con uno de G-C y C-G como se observa en la Fig. 18.

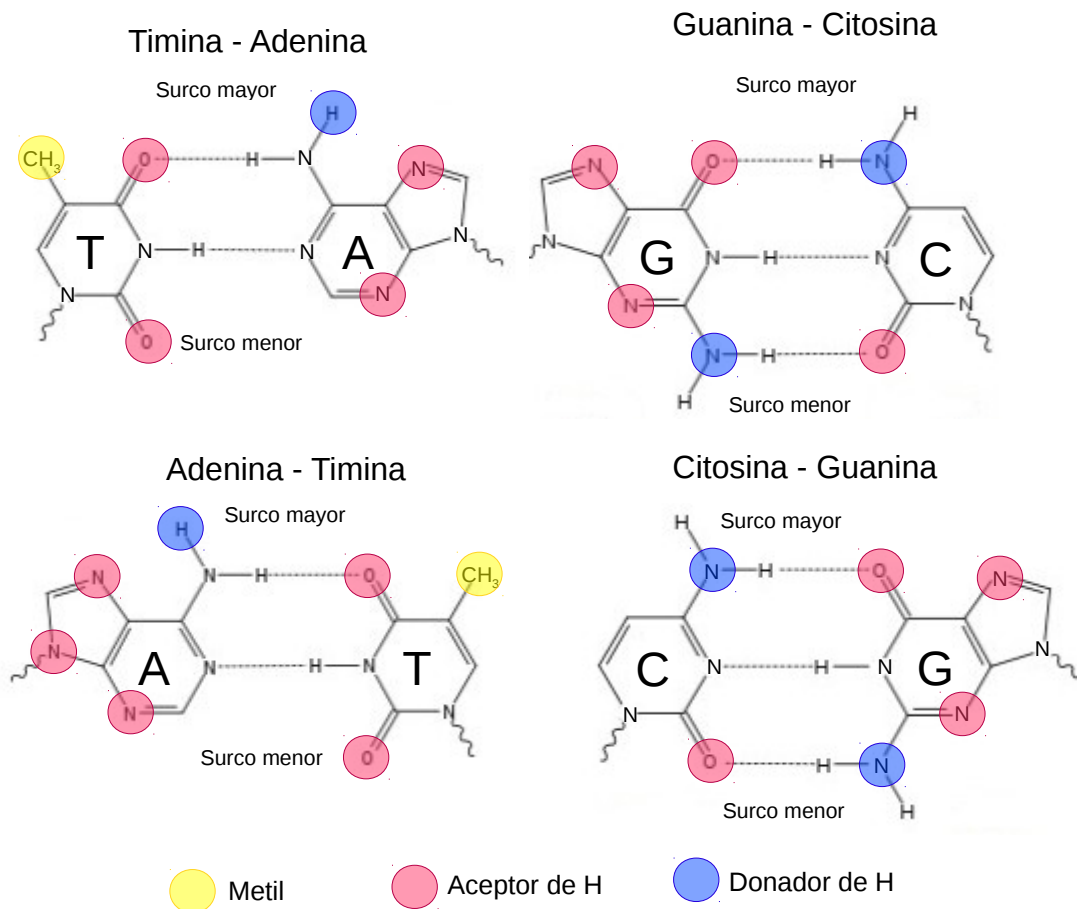


Fig. 18. Diferenciación estructural del par A-T del T-A y del par G-C del C-G. La exposición de los grupos metilo (en amarillo), los aceptores de H (en rojo) y los donadores de H (en azul) poseen un acomodo bien diferenciado en el surco mayor según el acomodo que posea el par de bases. Nótese que el par T-A expone a su lado izquierdo en el surco mayor un grupo metilo mientras que el par A-T, lo expone a su derecha, del mismo modo se nota la diferencia en la disposición que existe del grupo amino en el caso de los pares G-C y C-G. Del lado izquierdo de la figura se resalta con un esquema de colores a los grupos metilo, los donadores de H y los receptores de H.

Las diferencias presentes en el surco mayor respecto al acomodo en los aceptores, donadores de H y grupos funcionales actúan como guías para el correcto ensamblaje de los factores de transcripción sobre el DNA (Huret JL, 2006). Debido a que MEF2A se une a la parte interior del surco mayor y menor reconociendo grupos aceptores, donadores y metilos hidrofóbicos, posee cuatro patrones de reconocimiento posibles en el surco mayor y dos en el surco menor, debido a ello observamos un incremento en el número de puentes de Hidrógeno

que establecen los nucleótidos G15, A14, T13 y A12 dirección 3' → 5' (en ambas hebras) en comparación con sus bases complementarias de cada par.

Como habíamos establecido con anterioridad ésta disparidad en el número de interacciones presentes en los nucleótidos de cada par, desaparece en la región media de la CArGbox (nucleótidos 10 y 11 dirección 5' → 3'), este caso se explica atendiendo nuevamente a su acomodo espacial, ya que los grupos aceptores que quedan disponible se encuentran en el surco menor (un átomo de N y otro de O en el caso de T-A y A-T) y en éste caso no existen diferencias en los patrones de reconocimiento como los expuestos en el surco mayor, por lo tanto, ambos nucleótidos comparten con mayor equidad las interacciones de puente de hidrógenos con los aminoácidos del amino terminal.

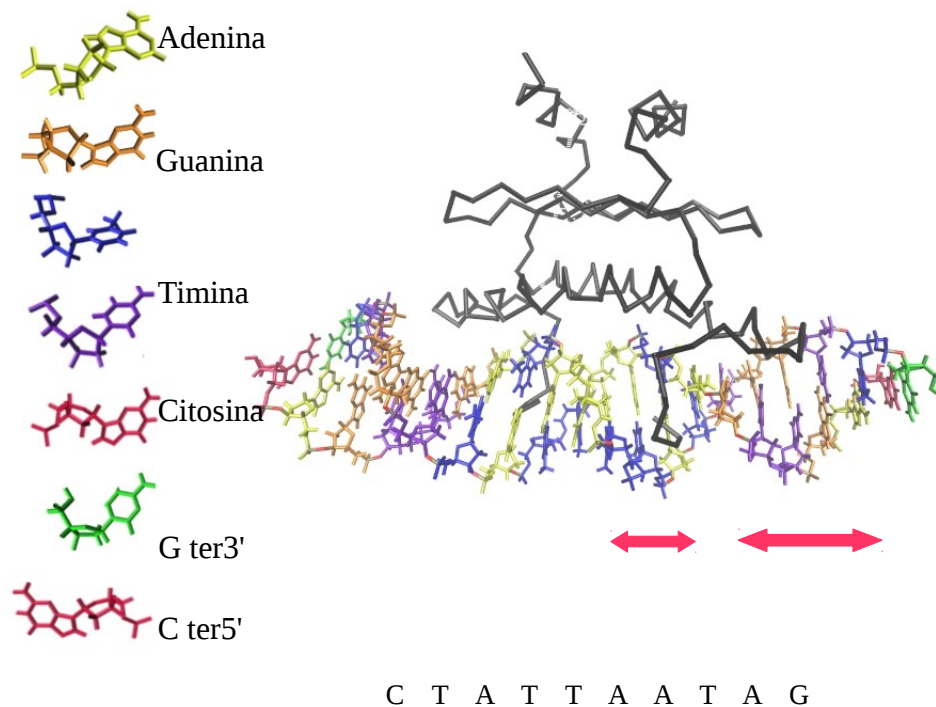


Fig.19. Acomodo de MEF2A sobre los surcos del DNA. Imagen obtenida del archivo .pdb con No de id:1C7U, del lado izquierdo de la figura se observa el color que le corresponde a cada una de las bases representadas en la CArGbox incluyendo los extremos 5' y 3' que corresponden a una G y una C respectivamente, la proteína se halla representada en color negro y se marca con dos flecha de color rojo las zonas que corresponden al surco mayor (flecha grande) y al surco menor (flecha pequeña); en la parte inferior de la figura se marcan con la nomenclatura de las bases, las regiones aproximadas en las que se ubican. En ésta imagen se aprecia claramente la inclusión del extremos amino terminal en el surco menor y la interacción del α -hélice con el surco mayor. La representación se obtuvo con el visualizador de proteínas VMD.

Al relacionar las frecuencias observadas en cada una de las posiciones de la CArGbox, notaremos que la zona más variable se encuentra en el centro, y la zona invariable (en la que es menos probable la ocurrencia de otras bases que no sean C, T, A dirección 5' → 3), debido a que los nucleótidos colocados al centro no poseen un patrón diferenciable de reconocimiento en cuanto a sus átomos donadores, de tal manera poseen menor presión selectiva, que permite una variación mayor de nucleótidos, ya que el patrón de reconocimiento A-T y T-A en el surco menor es idéntico para ambos casos, a diferencia de las encontradas en el surco mayor, en consecuencia estos últimos poseen una mayor presión de selección, y están limitados al cambio para conservar el patrón de

donadores y aceptores con el fin de que el factor de transcripción MEF2A los seleccione.

2.9.7. Interacciones entre la CArGbox con residuos de Lys y Arg.

Es importante señalar que de los 19 aminoácidos identificados en esta investigación, responsables de la unión a la CArGbox, 14 corresponden a residuos de Lys y Arg, motivo de que la selección natural actúa en el sentido de MEF2A para conservar dichos residuos; caso similar a lo que ocurre en las histonas, cuyo índice de enriquecimiento Lys/Arg llega a conformar hasta el 22% de su estructura (Badeux et al. 2013), otorgando sitios de unión de reconocimiento específico en los surcos mayores, ya que los residuos básicos de Lys y Arg se unen a la hendidura mayor del DNA (Acosta-Viana et al 1996), a pH intracelular estos residuos que poseen carga positiva, facilitan la interacción con los grupos fosfatos del DNA, de esta manera las cargas no sólo facilitan la unión, sino que promueven el reconocimiento entre la proteína y el elemento-cis. La distribución de los residuos en el surco mayor y menor, es fundamental sobre todo en el caso de la Arginina (ésta no encaja en el surco mayor de los pares G-C).



CAPÍTULO III

Identificación de nuevas CArGbox a las que potencialmente se une el factor de transcripción MEF2A en el genoma de *Homo sapiens*

3.1.. Planteamiento del problema del capítulo III.

La expresión genética está modulada en tiempo y espacio con la finalidad de asegurar la función de diversos módulos metabólicos presentes en los organismos, no puede entenderse sin un conocimiento detallado de características que hacen posible la unión al DNA con los TF's para la producción o represión de productos genómicos, que llevan a cabo una tarea determinada en un tiempo establecido. Este proceso ha sido blanco de un fino modelaje evolutivo que hace frente al reto de reconocer un substrato (elemento-cis), dentro de un contexto invadido de blancos posibles en la molécula de DNA, sin embargo, se sabe que los factores de transcripción son capaces de discernir entre un gran número de blancos posibles. Esta capacidad fundamental para las proteínas de unión a DNA, es el principal tema a tratar de esta tesis. Con anterioridad se han descrito algunos de los principales elementos que subyacen en el reconocimiento, la unión y la permanencia del complejo formado entre el factor de transcripción y su secuencia blanco. A partir de estas premisas, se realizaron predicciones que contribuyeron a la labor de reconocer las propiedades necesarias con la que cuenta una secuencia de DNA para postularse como una CArGbox, ligando del factor de transcripción MEF2A. Esto se logró a partir del cálculo de la Energía Libre de Unión (ΔG_{bind}), con la que se determinaron las secuencias que tienen cualidades fisicoquímicas, adecuadas para ser afines a la unión de MEF2A.

3.2. Objetivo del capítulo III

Determinar nuevas CArGbox que se unan al factor de transcripción MEF2A en el genoma de *Homo sapiens*, por medio de la determinación de la ΔG_{bind} .

3.3. Hipótesis del capítulo III

Si determinamos la ΔG_{bind} derivada de la unión de MEF2A con diferentes secuencias de DNA, como modelo de cuantificación de afinidad, se podrán identificar oligonucleótidos que poseen una interacción equivalente a la estudiada en el arquetipo de CArGbox, reportada en el archivo *.pdb* con No de id:1C7U.

3.4. Sistema experimental del capítulo III

- Estructura cuaternaria de MEF2A contenida en el archivo *.pdb* con No de id:1C7U
- Modelos tridimensionales de CArGbox.

3.5. Recursos informáticos del capítulo III.

Software

- *Ipython v.2.6*
- *MultiCArG.py*
- *APBS*
- *X3DNA*
- *MolProvity*

Servidores

- WebLogo (<http://weblogo.berkeley.edu/>)
- RSAT (<http://rsat.ulb.ac.be/>)
- MEME (<http://meme.nbcr.net>)

3.6. INTRODUCCIÓN AL CAPITULO III.

3.6.1 Calculo de la ΔG_{bind} .

La paquetería Adaptive Poisson-Boltzmann Solver (*APBS*) (Nathan A. Baker, 2010) creada para realizar calculos electrostáticos en sistemas macromoleculares, requiere como entrada un archivo en formato *pqr*, que complementa la información de coordenadas atómicas presentes en el archivo PDB con el pK_a , asigna cargas expresadas en [e] y el radio atómico (Å). Los cálculos se realizan tomando en cuenta las especificaciones del campo de fuerzas AMBER. *MultiCARG*, ejecuta una determinación de energía libre de solvatación para la *CARGbox*, y para el complejo en cada ciclo. *APBS* requiere los archivos en formato *.pqr*: el que corresponde al complejo, el de las moléculas que lo integran por separado. El archivo *.pqr* correspondiente al dímero MEF2A, se encuentra dentro de las librerías de *MultiCARG*, con la finalidad de minimizar el tiempo de cómputo y evitar el cálculo en cada ciclo del programa, ya que lo único variable en el complejo es el ácido nucleico.

La cuantificación de la energía libre de unión ($\Delta G_{\text{binding}}$) se resolvió mediante el esquema MM-PBSA aceptado ampliamente en la literatura (Tzoupis *et al.*, 2013; Millán-Pacheco *et al.*, 2009; Srinivasan *et al.*, 1998) por medio de la paquetería *APBS v.10*, que agrega componentes de mecánica molecular ajenos al tratamiento Poisson-Boltzmann (PB), acorde a la siguiente ecuación:

$$\Delta G = E_{MM} + PB + \gamma(ASA) - TS_{\text{soluto}} \quad (1)$$

La energía libre (ΔG) puede descomponerse en: E_{MM} , representa la energía interna de la molécula. La ecuación PB describe las interacciones electrostáticas entre moléculas dentro de una solución para modelar el efecto de solvatación $\gamma(ASA)$, que integra las contribuciones no polares en un modelo de accesibilidad al solvente multiplicado por el coeficiente de tensión γ 0.005 kcal/Å²; y finalmente TS_{soluto} , que representa la entropía de la molécula según los factores vibracionales de la misma.

Los elementos de la ecuación representan en su mayoría interacciones Coulombicas. Las contribuciones energéticas de mecánica molecular están descritas por:

$$E_{MM} = E_{\text{interna}} + E_{\text{electrostatica}} + E_{\text{vdW}} \quad (2)$$

En donde E_{interna} es la suma de $E_{\text{enlaces}} + E_{\text{angulos}} + E_{\text{tors}}$ que en conjunto describen la energía contenida en las uniones entre átomos, por otro lado, los terminos $E_{\text{electrostatica}}$ y E_{vdW} se refieren a las fuerzas de no enlace. Así se resume como E_{MM} la energía contenida dentro del soluto.

La ecuación Poisson-Boltzman es el clásico tratamiento para modelar la interacción en solución, y está descrita de la siguiente manera:

$$\nabla^2 [\epsilon(r) \Phi(r)] - \epsilon(r) \kappa(r)^2 \sinh[\Phi(r)] + 4\pi \rho^f(r) / kT = 0 \quad (3)$$

Donde $\phi(r)$ es el potencial electrostático (adimensional), (kT/q) representa la temperatura absoluta y la carga del protón en unidades de la constante de Boltzmann (k_B), ϵ es la constante dieléctrica y ρ^f es la densidad de carga (en unidades de carga por protón) del soluto. El término κ descende de la ecuación $\kappa^2 = 1/\lambda^2 = 8\rho q^2 I / \epsilon kT$, donde λ es el radio de Debye. Las variables ϕ , ϵ , κ y ρ son todas funciones del vector r .

Cuando $\phi(r)$ es pequeño en relación con kT , el término \sinh puede ser linearizado (lo cual ahorra tiempo de cómputo) mediante el establecimiento de $\sinh(\phi) = \phi$ produciendo la PB lineal.

Es importante señalar que en éste modelo se asume una constante dieléctrica invariable para todo el solvente, valoración relevante para cuando se realiza un cálculo de interacción sin fuerzas iónicas incluidas, ya que a causa de esto el agua es mucho más polarizable por un campo eléctrico como el que procede de la mayoría de las proteínas y ácidos nucleicos, por lo menos dos ϵ 's son requeridas para capturar la física fundamental de moléculas polares en solución acuosa, por lo que esta formula encaja para el cálculo de interacción entre dos solutos.

El efecto de tener dos constantes dieléctricas esta expresado en la ecuación PB a través de la derivada de $\epsilon(r)$ en el primer término, ésta región corresponde a la superficie molecular, donde se halla una discontinuidad dieléctrica entre el menor valor de $[\epsilon]$ del soluto y el alto $[\epsilon]$ del solvente.

La ecuación P-B puede estimarse como un modelo dieléctrico simple, donde la carga de polaridad opuesta al $\phi(r)$ es inducida en la superficie molecular. Un ejemplo simple sería el caso de una carga positiva en el centro de una esfera de baja ϵ , donde se induce una distribución de cargas negativas en la superficie de la esfera. El potencial positivo del solvente causado por la fijación de la carga positiva, puede ser detectado por la carga superficial negativa, así como el valor más alto de ϵ del solvente y la carga mayor de la superficie (la cual produce una reacción de campo) en la base de los modelos de solvatación continua.

Con estos datos puede lograrse una idea muy clara de la distribución de cargas y el potencial electrostático (observable en isosuperficies). Se puede entender la inducción de carga superficial como un alineamiento de dipolos de agua en la superficie molecular, producida por un exceso de carga local. Sin embargo, describir la carga de la superficie en los sistemas moleculares más complejos, toma en cuenta la polarización de todo el solvente.

La molécula de soluto se concibe como una cavidad de baja ϵ , inmersa en un medio con diferentes valores de ϵ (representados por el paso de una solución acuosa). Dado que las soluciones analíticas de la ecuación PB, sólo están disponibles para elementos geométricos simples, se requirió una simplificación de la forma del soluto y su distribución de carga, esto se logró al describir la molécula como un continuo en la fase de interacción con el solvente.

El término $\gamma(\text{ASA})$ incluido en la ecuación de la energía libre (G), utilizado en la especificación de las contribuciones no polares de la solvatación (características geométricas que permitieron ingresar o no al solvente

en ciertas regiones de la superficie molecular) estuvieron dadas por el área de accesibilidad al solvente (ASA), definida por la superficie del soluto, el radio de las moléculas de agua y un coeficiente de tensión (γ) que delimita un área de tensión superficial en la interfase de la disolución.

El programa APBS fue el elegido para computar el cálculo de la energía libre de unión con el método MM-PBSA que puede resumirse conceptualmente de la siguiente manera:

$$\Delta G_{bind} = G_{complejo} - [G_{proteina} + G_{ADN}] \quad (4)$$

$$\Delta G_{bind} = \Delta H - T \Delta S \quad (5)$$

Donde ΔH se compone de cambios de entalpía en la fase gaseosa de la formación de los complejos (ΔE_{MM}) y la contribución de la energía libre de solvatación (Δg_{sol}), donde $T\Delta S$ se refiere a la contribución entrópica de la unión. APBS obtiene un aproximado de la ecuación 5 utilizando la fórmula 1 y desplegando E_{MM} como se indica en la fórmula 2.

Es preciso recordar que cada una de las ΔG incluye la energía libre de solvatación la cual es calculada para cada una de las moléculas por separado (complejo, proteína y DNA), en dos bloques que comprenden a las contribuciones energéticas de la solvatación polar y apolar. La primera es calculada con respecto a la siguiente ecuación:

$$\Delta_{solvP} G = \Delta_{qsolv} G - \Delta_{qvacio} G$$

Donde $\Delta_{qsolv}G$ es igual a la carga de la proteína en solución y $\Delta_{qvacio}G$ indica la carga la proteína en el vacío. Es muy importante mencionar que la carga electrostática de estas energías libres, incluyen los término de interacciones electrostáticas intramoleculares (E_{MM}) en el cómputo realizado por APBS; de este modo $\Delta_{solvP}G$ es un diferencial que da como resultado el costo energético de llevar una molécula del vacío a un estado solvatado.

La energía de solvatación no polar esta determinada por:

$$\Delta_{solvNP} G = (\Delta_{Dss} G - \Delta_{Dsv} G) + \Delta_{ASA} G$$

Donde en $\Delta_{Dss}G$ se introdujeron las fuerzas de dispersión entre el soluto y el solvente y con $\Delta_{Dsv}G$ se especificaron las fuerzas de dispersión entre el soluto y el vacío. $\Delta_{ASA}G$ representa la presión que ejerce el solvente sobre la superficie del soluto. Como se observa en esta parte, se introduce el término γASA de la ecuación de energía libre G .

Una vez que se obtuvieron la energías libres por separado, se procede a incluirlas en la ecuación ΔG_{bind} .

3.7. METODOLOGIA DEL CAPITULO III.

El cálculo de la ΔG_{bind} requiere de las estructuras tridimensionales de la proteína y el oligonucleótido, por lo que son necesarias las coordenada moleculares de otras cadenas de DNA distintas al arquetipo contenido en el archivo *.pdb* 1C7U. Esta necesidad obligó al desarrollo de una metodología secuencial que parte de la obtención de la construcción de estructuras tridimensionales del complejo y finaliza con la determinación de la ΔG_{bind} .

3.7.1. El programa *MultiCARG*.

Escrito en *ipython* v.2.6, se compone de varios módulos divididos según la tarea específica que les asignó, estos son:

- Módulo I : Creación del universo 4^{10} .
- Módulo II: Anexión de las bases complementarias a la secuencia lineal de nucleótidos.
- Módulo III: Construcción del archivo “*bp_step2.par*”.
- Módulo IV: Construcción del archivo *.pdb* (contiene la estructura tridimensional de la nueva CARGbox).
- Módulo V: Formación del complejo MEF2A-CARGbox.
- Módulo VI: Calculo de la ΔG_{bind} .

3.7.1.a Creación del Universo 4^{10} .

El primer módulo tiene por objetivo construir un archivo en formato *.txt*, contiene todas las secuencias de nucleótidos posibles, con una longitud de 10 pares de bases, construidas al permutar los cuatro nucleósidos que componen al DNA, considerados como elementos de un alfabeto con el cual se construyeron palabras de 10 caracteres (el tamaño de la CARGbox a la que se une MEF2A). Por lo tanto se resolvió la combinatoria del conjunto $\text{DNA} = \{A, T, G, C\}$ en el que cada uno de sus elementos puede repetirse hasta 10 veces o incluso no aparecer en la secuencia, el número de permutaciones resultantes de esta combinatoria es de 1,048,526, cuyo cálculo está denotado por el número 4^{10} , resuelto con el programa *UniversoDeOligos.py* presentado como “Código I” en el Apéndice D. Cabe mencionar que este código no se incluye en el programa *MultiCARG.py*, ya que éste forma un bucle al final de su ejecución y vuelve a su primera línea de comando, redirigiendo a un archivo que contiene el universo de secuencias 4^{10} , por lo que su construcción es necesaria en una ocasión.

Con base a las secuencias que se obtuvieron y se leen en dirección 5'→3', se conformó el archivo “*Universooligos.txt*” del que se tomaron secuencias utilizadas como entrada del programa *MultiCARG.py*.

3.7.1.b. Anexión de las bases complementarias a la secuencia lineal de nucleótidos.

El programa *MultiCARG.py* comienza con la selección de la primera línea del archivo *Universooligos.txt* (Si se desea la lectura de una secuencia en específico, asegúrese de que ésta sea la primera o única línea del

archivo) para continuar con el siguiente oligonucleótido al final del cálculo de ΔG_{bind} y así sucesivamente.

De regreso al paso inicial para elegir una secuencia, el programa añade la base complementaria de cada nucleótido, como preparación para la elaboración del fichero “*bp_step2.par*” (archivo de entrada del programa *X3DNA*), por medio de una serie de condicionales a través de las cuales se identificó cada nucleótido en su posición específica, para reescribirse como un un par de bases, y así conservar el sentido biológico de cada unidad, como se observa en la configuración del oligonucleótido ubicado en la línea 922211 del archivo “*Universoligos.txt*”: El primer nucleótido del par está ubicado en dirección 5'→3':

(T-A, 'G-C', 'A-T', 'C-G', 'A-T', 'G-C', 'C-G', 'G-C', 'A-T', 'G-C)

3.7.1.c. Construcción del archivo “*bp_step2.par*”.

Una vez que la secuencia de nucleótidos adquirió el formato requerido para ser incorporado al archivo “*bp_step2.par*” (Fig. 20), se realizó la sustitución de la nueva secuencia de nucleótidos por la secuencia original hallada en dicho fichero, ésta última corresponde a los veinte nucleótidos que conforman la CARGbox obtenida después de una minimización con el algoritmo *steepest-descent* de 9318 ps, que registró una fuerza máxima de 90.570236 kJ mol⁻¹ nm⁻¹ dentro del sistema. (Cfr. Dinámica molecular Capítulo II). Este archivo se obtuvo con el programa *analyze* incluido en la *suite* de *X3DNA* y contiene parámetros estructurales tanto de translación como de rotación en bases simples y pares de bases. Estos valores sólo corresponden a los átomos que forman las bases nitrogenadas presentes en el DNA excluyendo sus átomos de Hidrógeno.

```
| 20 # base-pairs
0 # ***local base-pair & step parameters***
#      Shear  Stretch  Stagger  Buckle  Prop-Tw  Opening  Shift  Slide  Rise  Tilt  Roll  Twist
C-G    0.18   -0.19   -0.18    0.84    2.57    -2.34    0.00    0.00    0.00    0.00    0.00
T-A   -0.01   -0.14    0.30    4.25   -7.87   -4.00   -0.01   -0.58    4.00   -2.07   -5.09   27.90
C-G    0.14   -0.22    0.25    0.68   -6.07   -2.31    0.14   -0.17    4.05   -1.04    4.44   32.69
G-C   -0.15   -0.22    0.51    7.55  -13.03   -2.74   -0.13    0.63    3.71    0.46   -3.71   32.02
G-C   -0.14   -0.21    0.26    5.61  -11.06   -3.07    0.28    0.99    3.87    1.29   -0.36   41.19
C-G    0.15   -0.25    0.45   -0.23  -12.45   -3.06    0.35    0.25    4.02   -2.48   10.32   36.10
T-A   -0.01   -0.12    0.04    3.58    0.10   -3.04   -0.35   -0.66    3.60    4.84  -17.75   29.33
A-T    0.02   -0.13    0.24   -0.22  -16.88   -3.08   -0.66    0.84    3.98   -1.07   10.20   41.05
T-A   -0.01   -0.12    0.38    1.22  -14.25   -3.53   -0.05    0.64    3.67   -0.81  -12.77   38.35
T-A   -0.02   -0.12    0.43   -4.52  -20.35   -4.02   -0.64    1.60    3.91    1.57   -1.01   39.95
A-T    0.02   -0.12    0.43    4.08  -20.56   -4.08   -0.00    1.92    3.22   -0.05  -12.25   42.48
A-T    0.01   -0.12    0.37   -0.99  -14.54   -3.44    0.66    1.59    3.88   -1.27   -1.18   39.97
T-A   -0.02   -0.13    0.24    0.36  -16.68   -3.10    0.06    0.64    3.67    0.88  -12.61   38.33
A-T    0.01   -0.12    0.04   -3.56    0.09   -3.03    0.66    0.84    3.99    1.11   10.32   41.05
G-C   -0.15   -0.25    0.45    0.26  -12.40   -3.04    0.35   -0.66    3.61   -4.79  -17.67   29.31
C-G    0.14   -0.21    0.26   -5.59  -11.10   -3.08   -0.36    0.25    4.02    2.42   10.25   36.13
C-G    0.15   -0.22    0.52   -7.60  -13.06   -2.76   -0.27    0.99    3.87   -1.31   -0.39   41.20
G-C   -0.14   -0.22    0.25   -0.68   -6.07   -2.30    0.14    0.63    3.70   -0.42   -3.67   32.02
A-T    0.01   -0.14    0.30   -4.27   -7.87   -3.99   -0.14   -0.17    4.05    1.06    4.44   32.70
G-C   -0.18   -0.19   -0.18   -0.85    2.58   -2.34    0.01   -0.58    4.00    2.03   -5.10   27.89
```

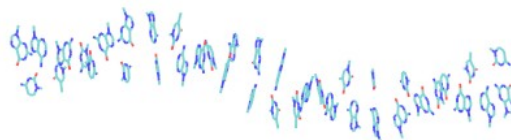
Fig.20. Archivo *bp_setep.par* de la CARGbox CTATTAATAG. El archivo se construyó en base a la CARGbox reportada en el archivo *.pdb* con No de id:1C7U después de una minimización de 9318 ps, en la que se incluye también el contexto de DNA en el que está inmerso el sitio de unión a MEF2A. La CARGbox esta resaltada dentro de un marco rojo. Los parámetros estructurales que se incluyen corresponden a los valores de translación y rotación de las bases nitrogenadas, se excluyen los grupos fosfato, y los átomos de Hidrógeno. Los valores fueron obtenidos con el programa *analyze* incluido en la paquetería de *X3DNA*.

Una vez obtenido el archivo “*bp_sep2.par*” inicial se sustituyeron los diez nucleótidos que conforman a la CARGbox por la secuencia tomada al inicio de la ejecución del programa *MultiCARG.py*, de este modo se

respetar la posición en el modelo tridimensional de cada uno de los pares de bases y las coordenadas del modelo de CArGbox resuelto por resonancia magnética nuclear, minimizado posteriormente para disponer a los átomos en posiciones de mínima energía. Cabe acentuar que los ácidos nucleicos que no pertenecen a la CArGbox, es decir su contexto genómico, no es manipulado en ningún proceso del programa.

3.7.1.d. Construcción del archivo .pdb que contiene la estructura tridimensional de la nueva CArGbox.

Este módulo inicia con el programa *rebuild* incluido en la paquetería de *X3DNA* que construye el archivo “*CArGrpda.pdb*” con las nuevas coordenadas atómicas de la CArGbox mutada, construida a partir del archivo “*bp_step2.par*”. La obtención del archivo *pdb* es fundamental para la manipulación y el reconocimiento de la molécula en diversas paqueterías, como APBS. La molécula obtenida hasta el momento se muestra en la Fig. 21.



AAAAAGCAT

Fig.21. CArGbox contenida en el archivo “*CArGrpda.pdb*”. Visualización de la molécula de DNA construida con *X3DNA*, ésta cuenta con los parámetros espaciales de la CArGbox resuelta por RMN y procesada con una minimización energética. La secuencia de la hebra líder de la CArGbox se especifica en la parte inferior de la figura debajo de los nucleótidos correspondientes en la molécula. Representación realizada con el método *line* en *VMD*.

Con el objetivo de obtener un modelo más realista, fue necesario incorporar moléculas de desoxirribosa, fosfatos y todos los átomos de hidrógeno. Por lo tanto del archivo “*CArGrpda*” se tomó la porción que pertenece a la CArGbox mutante y se insertó en el “fantasma” del archivo original “*CArGoriginal.pdb*”, realizado al asignarle a una variable “*b*” al contenido de éste documento, con el fin de asegurar su integridad en cada ciclo de ejecución de *MultiCArG.py*. Las sustituciones se realizaron con las particularidades propias de las 8 transversiones y 4 transiciones posibles durante un suceso de mutación, que ocurrieron tanto en la hebra líder como en la hebra complementaria. A continuación se especifican 6 de los 12 casos que resolvió *MultiCArG.py*, que ejemplifican la forma de ejecución del programa con el que hasta ahora el oligonucleótido contuvo azúcares y un esqueleto de fosfatos.

3.7.1.d.a. Mutación de A por T (Hebra líder).

- i) Seleccionar del archivo *pdb* que contiene a la CArG_box mutante los átomos correspondientes a la Timina (de C1' a C6). Dado que cada nucleótido está numerado del 1 al 40, solo es necesario reconocer el número que corresponde a la base mutada para seleccionar todos sus átomos.
- ii) Sustituir la Adenina en el complejo original MEF2A-CArG_box a partir del átomo C1' hasta H2 por los átomos de la Timina. Esta sustitución permitirá conservar los fosfatos y la desoxirribosa.

- iii) Sustituir la etiqueta de identificación de la Adenina (DA) por DT en los átomos correspondientes al azúcar y los fosfatos (de P a C2') y en los átomos correspondientes a la Timina sustituir T por DT.
- iv) Sustituir la cadena de A a C a partir del átomo C1' a C4.
- v) Cambiar la numeración por la del residuo al que se sustituyó.

3.7.1.d.b. Mutación de T por A (Hebra complementaria)

- i) Seleccionar del archivo pdb que contiene a la CArG_box mutante los átomos correspondientes a la Adenina (de C1' a C4).
- ii) Sustituir la Timina en el complejo original MEF2A-CArG_box a partir del átomo C1' hasta H6 por los átomos de Adenina.
- iii) Sustituir la etiqueta de identificación de la Timina (DT) por DA en los átomos correspondientes al azúcar y los fosfatos (de P a C2') y en los átomos correspondientes a la Adenina sustituir A por DA.
- iv) Sustituir la cadena de B a D a partir del átomo C1' a C4.
- v) Cambiar la numeración por la del residuo al que se sustituyó.

3.7.1.d.c. Mutación de G por A (Hebra líder)

- i) Seleccionar del archivo pdb que contiene a la CArG_box mutante los átomos correspondientes a la Guanina (de C1' a C4)
- ii) Sustituir la Adenina en el complejo MEF2A-CArG_box a partir del átomo C1' hasta H2 por los átomos de Guanina.
- iii) Sustituir la etiqueta de identificación de la Adenina (DA) por DG en los átomos correspondientes al azúcar y los fosfatos (de P a C2') y en los átomos correspondientes a la Guanina sustituir G por DG.
- iv) Sustituir la cadena de A a C a partir del átomo C1' a C4.
- v) Cambiar la numeración por la del residuo al que se sustituyó.

3.7.1.d.d. Mutación de C por T (Hebra complementaria).

- i) Seleccionar del archivo pdb que contiene a la CArG_box mutante los átomos correspondientes a la Citosina (de C1' a C6).
- ii) Sustituir la Timina en el complejo MEF2A-CArG_box a partir del átomo C1' hasta H6 por los átomos de Citosina.
- iii) Sustituir la etiqueta de identificación de la Timina (DT) por DC en los átomos correspondientes al azúcar y los fosfatos (de P a C2') y en los átomos correspondientes a la Citosina sustituir C por DC.
- vi) Sustituir la cadena de B a D a partir del átomo C1' a C4.
- vii) Cambiar la numeración por la del residuo al que se sustituyó.

3.7.1.d.e. Mutación de una C por A (Hebra líder)

- i) Seleccionar del archivo pdb que contiene a la CArG_box mutante los átomos correspondientes a la Citosina (de C1' a C6).
- ii) Sustituir la Adenina en el complejo MEF2A-CArG_box a partir del átomo C1' hasta H2 por los átomos de Citosina.
- iii) Sustituir la etiqueta de identificación de la Adenina (DA) por DC en los átomos correspondientes a el azúcar y los fosfatos (de P a C2') y en los átomos correspondientes a la Citosina sustituir C por DC.
- iv) Sustituir la cadena de A por C a partir del átomo C1' a C6.
- v) Cambiar la numeración por la del el residuo al que se sustituyó.

3.7.1.d.f. Mutación de una G por T (Hebra complementaria).

- i) Seleccionar del archivo pdb que contiene a la CArG_box mutante los átomos correspondientes a la Guanina (de C1' a C4).
- ii) Sustituir la Timina en el complejo MEF2A-CArG_box a partir del átomo C1' hasta H6 por los átomos de Guanina.
- iii) Sustituir la etiqueta de identificación de la Timina (DT) por DG en los átomos correspondientes a el azúcar y los fosfatos (de P a C2') y en los átomos correspondientes a la Guanina sustituir G por DG.
- iv) Sustituir la cadena de B por D a partir del átomo C1' a C4.
- v) Cambiar la numeración por la del el residuo al que se sustituyó.

3.7.1.e. Adición de los átomos de H.

Las interacciones entre factores de transcripción y sus secuencias blanco dependen de la formación de puentes de hidrógeno, por lo que fue importante añadirlos a los ácidos nucleicos, que se construyeron *de novo*.

Debido a que cada nucleótido de la CArGbox posee las mismas coordenadas atómicas, propias de su estructura (cuando se le coloca en determinada posición, sin importar el contexto en el que éste inmerso), fue posible construir una base de datos de coordenadas atómicas con los hidrógenos ausentes para cada nucleótido en las 10 posiciones. Se construyeron secuencias poli-Adenina, poli-Timina, poli-Citosina y poli-Guanina, en las que se sustituyeron todas las posiciones de la CArGbox por una secuencia repetitiva compuesta de cada uno de los nucleótidos presentes en el DNA. Una vez que se obtuvo el archivo “.pdb” de cada decámero, se convirtieron en archivo de entrada de la paquetería *MolProbit*, en las que se adicionó los átomos de hidrógeno faltantes de las cuatro secuencias repetitivas, tomando en cuenta principios fisicoquímicos, que aseguraron la colocación de cada hidrógeno en posiciones atómicas de mínima energía.

Las coordenadas de hidrógenos se resguardan en ficheros de la librería ATOM_H con las siguientes consideraciones: la correspondencia con una base determinada, la posición que ésta ocupa dentro de la CArGbox y si pertenecía a la hebra líder o a su complementaria; de modo que la librería posee cuatro ficheros por cada posición de la CArGbox que corresponden a los cuatro nucleótidos con probabilidad de ocupar esa posición, de

igual modo se cuenta con cuatro ficheros más por cada posición de la CArGbox que corresponden a las coordenadas atómicas de las bases de la hebra complementaria, como resultado se obtuvieron 80 ficheros disponibles para completar los átomos ausentes en los *.pdb*, contruidos en los módulos anteriores.

3.7.1.f. Formación del complejo MEF2A-CArGbox.

Una vez disponible la molécula de DNA con átomos de hidrógeno y con la topología propia de nuestro arquetipo, el programa *MultiCARG.py* realizó la conjunción del factor de transcripción con su putativa secuencia blanco a través de la suma de los archivos “*MEF2A.pdb*” y “*CarGmutnuevo.pdb*”, para constituir el fichero “*complex.pdb*” que posee la estructura del complejo. La complementariedad de ambas moléculas se logró por medio de la reproducción de las coordenadas atómicas originales (reportadas por la determinación de resonancia magnética) fundamentales a lo largo del proceso de reescritura que dio lugar a la secuencia de una nueva CArGbox, con este principio se aseguró la formación del complejo con los parámetros aceptados hasta el momento.

3.7.1.g. Computo de la energía libre de unión.

El cálculo de ΔG_{bind} APBS sigue tres pasos básicos para su determinación:

- 1) Cálculo de la energía electrostática del complejo (incluyen interacciones intramoleculares)
- 2) Cálculo de la energía electrostática de la proteína (incluyen interacciones intramoleculares)
- 3) Cálculo de la energía electrostática del DNA (incluyen interacciones intramoleculares)

Los parámetros de entrada que se le otorgan a APBS para la realización de los cálculos están presentes en el documento *Eucalc.in*, al cual llama *MultiCARG.py*, que a continuación se describe :

```
read
mol pqr CARG.pqr      #Nombre del archivo .pqr correspondiente al oligonucleótido
mol pqr MEF2A.pqr    #Nombre del archivo .pqr correspondiente a la proteína
mol pqr complex.pqr  #Nombre del archivo .pqr correspondiente al complejo
end
```

APBS localiza los documentos de entrada que corresponden al factor de transcripción a la CArGbox en turno y al complejo formado en el módulo anterior.

Los siguientes elementos del *script* “*Eucalc.in*” se enuncian a detalle en la sección que corresponde a los parámetros de la CArGbox con la finalidad de detallar bien el método utilizado en este trabajo y hacerlo reproducible, o también para usarlo como base en la resolución de un problema particular de la misma naturaleza. La explicación de cada uno de los parámetros en la sección del ácido nucleico, puede extrapolarse a los utilizados en el caso de la proteína MEF2A y el complejo.

```

elec name CARG          #Primer calculo energético sobre el oligonucleótido
  mg-auto              #Indica que la PBE se realizará de forma automática
  dime 161 129 193    #Especifica el número de puntos que tendrá la malla
  cglen 95.2258 79.9054 127.8536 # Dimensiones de la malla gruesa
  cgcent mol 3        #Especifica el centro de la malla gruesa
  fglen 76.0152 67.0032 95.2080 #Tamaño de la malla fina.
  fgcent mol 3        # Centro de la malla fina.
  Mol 1                #Especifica la molécula para la cual se resolverá la PBE
  lpbe                 #Indica que la PBE será linearizada
  bcfl sdh             # Especifica las condiciones de límite dentro del medio en el
                       # que esta envuelta la proteína.

  Pdie 2.0             # Constante dieléctrica del soluto
  sdie 78.54           # Constante dieléctrica del solvente
  chgm spl0            # Indica el método por el cual las cargas puntuales de la
                       # moléculas serán mapeadas dentro de la rejilla.

  srfm mol             # especifica el modelo utilizado para la construcción del
                       # dieléctrico y el coeficiente de accesibilidad iónica

  srad 1.4             # Radio del solvente
  swin 0.3             # Tasa de cambio molecular.
  Sdens 10.0           # Accesibilidad del solvente
  temp 300             # Temperatura
  gamma 0.105         # tensión superficial.
  calcenergy total
end

```

```

elec name MEF2A        #Proteína
  mg-auto
  dime 161 129 193
  cglen 95.2258 79.9054 127.8536
  cgcent mol 3
  fglen 76.0152 67.0032 95.2080
  fgcent mol 3
  mol 2
  lpbe
  bcfl sdh

```

pdie 2.0
sdie 78.54
chgm spl0
srfm mol
srad 1.4
swin 0.3
sdens 10.0
temp 300
gamma 0.105
calcenergy total
end

elec name complex # Complejo

mg-auto
dime 161 129 193
cglen 95.2258 79.9054 127.8536
cgcent mol 3
fglen 76.0152 67.0032 95.2080
fgcent mol 3
mol 3
lpbe
bcfl sdh
pdie 2.0
sdie 78.54
chgm spl0
srfm mol
srad 1.4
swin 0.3
sdens 10.0
temp 300
gamma 0.105
calcenergy total
end

quit

Finalizado el cálculo de Energía libre de unión por parte del programa *APBS*, se registró en el documento “*history.txt*”, del cual *MultiCARG* extrae la línea en la que se encuentra reportada la Energía Libre de Unión y la reescribe en el documento “*cantidad.txt*” junto con la secuencia de nucleótidos con la que se formó el complejo.

De este modo, finalizó la ejecución completa de todos los módulos del programa, para que de forma automática *MultiCARG.py* comenzara nuevamente por el primer módulo utilizando de semilla la secuencia de nucleótidos en turno, por lo que no fue necesaria la intervención del usuario para la determinación de la energía libre de unión.

3.7.2. Búsqueda de nuevas CARGbox.

Con ayuda del módulo “random” de Python, se extrajeron 339672 decámeros de DNA al azar del archivo “*Univsooligos.txt*” con el que se creó el archivo *CARGenrg.txt*, a los que se les calculó su ΔG_{bind} , según el criterio de APBS.

Una vez determinado este valor, se extrajeron 7 subpoblaciones del archivo “*CARGenrg.txt*” que se dividieron considerando valores de ΔG_{bind} y que dan como resultado los siguientes conjuntos:

- CARGrango = $\{x \mid x > 3.8e+02 \text{ kJ/mol}, x < 4.5e+02 \text{ kJ/mol}\}$.
- $-3.9e+02 \text{ kJ/mol}$ = $\{x \mid x > 3.8e+02 \text{ kJ/mol}, x < 4.0e+02 \text{ kJ/mol}\}$.
- $-4.0e+02 \text{ kJ/mol}$ = $\{x \mid x > 3.9e+02 \text{ kJ/mol}, x < 4.1e+02 \text{ kJ/mol}\}$.
- $-4.1e+02 \text{ kJ/mol}$ = $\{x \mid x > 4.0e+02 \text{ kJ/mol}, x < 4.2e+02 \text{ kJ/mol}\}$.
- $-4.2e+02 \text{ kJ/mol}$ = $\{x \mid x > 4.1e+02 \text{ kJ/mol}, x < 4.3e+02 \text{ kJ/mol}\}$.
- $-4.3e+02 \text{ kJ/mol}$ = $\{x \mid x > 4.2e+02 \text{ kJ/mol}, x < 4.4e+02 \text{ kJ/mol}\}$.
- $-4.4e+02 \text{ kJ/mol}$ = $\{x \mid x > 4.3e+02 \text{ kJ/mol}, x < 4.5e+02 \text{ kJ/mol}\}$.

Esto con el propósito de facilitar el análisis de la ΔG_{bind} como indicador de secuencias CARGbox, ya que el único resultado de ésta naturaleza en el que se basa la designación de un decámero como CARGbox putativa, es la de la secuencia arquetípica 5' CTATTAATAG 3'. El establecimiento de estas poblaciones pretende establecer un rango al rededor del valor propio de la secuencia antes mencionada, también facilitar el análisis por grupos para entender la naturaleza que subyace en el cambio de los diferentes niveles energéticos, según la composición de las secuencias.

Posteriormente se realizaron diagramas de logo, que provenían de secuencias contenidas en cada población, con auxilio del programa *Weblogo* (Crooks *et al.*, 2004) bajo el método *frequency*, que contabiliza el número de ocurrencias de cada nucleótido en una posición dada para dividir estas repeticiones entre el total de secuencias ingresadas.

3.7.3. Arreglo de datos.

Debido a que el arreglo de datos influye directamente en la forma de visualizar las gráficas y sobre todo en la aparición de patrones, es importante puntualizar cómo se realizó este acomodo, ya que se encuentra directamente relacionado con la forma en la que se construyó el archivo “*Universooligos.txt*”, (que posee el resultado de la combinatoria 4^{10}). La primera línea que posee este documento, corresponde a una secuencia de poli-Adenina; la segunda entrada posee A de los sitios 1 al 9 y C en la última posición, la cual cambia a G y T para las secuencias que ocupan las líneas 3 y 4 respectivamente. Una vez que todos los nucleótidos han ocupado la última posición, varió el sitio 9 en el mismo orden de sucesión y este método se repitió hasta llegar al primer lugar del décamero. La Fig. 22 muestra las sucesiones antes descritas partiendo de la primer línea de las secuencias localizada en el archivo “*Universooligos.txt*”.

	A	C	G	T
A	('A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A')	('A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'C')	('A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'G')	('A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'T')
C	('A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'T', 'X')	('A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'T', 'X', 'X')	('A', 'A', 'A', 'A', 'A', 'A', 'T', 'X', 'X', 'X')	('A', 'A', 'A', 'A', 'T', 'X', 'X', 'X', 'X')
G	('A', 'A', 'A', 'A', 'T', 'X', 'X', 'X', 'X', 'X')	('A', 'A', 'A', 'T', 'X', 'X', 'X', 'X', 'X', 'X')	('A', 'A', 'T', 'X', 'X', 'X', 'X', 'X', 'X', 'X')	('A', 'T', 'X', 'X', 'X', 'X', 'X', 'X', 'X', 'X')
T	('T', 'X', 'X', 'X', 'X', 'X', 'X', 'X', 'X', 'X')			

Fig. 22. Modo en el que se lleva a cabo la sucesión de bases en la combinatoria 4^{10} . En las primeras cuatro secuencias de la figura solo varía la posición 10 siguiendo el orden $A \rightarrow C \rightarrow G \rightarrow T$, en las líneas subsecuentes se ha introducido el elemento “X” con el objetivo de representar la variación constante de éste sitio, también a partir de la línea 4 el último nucleótido en aparecer será la T, ya que ésta posee el último turno de la sucesión. El gradiente de colores que aparece sobre las secuencias indica que conforme el azul se intensifica el contenido de Adenina disminuye en las secuencias; éste gradiente también puede apreciarse en la parte superior e izquierda de la figura, en la que se especifica como el contenido de cada una de las bases se hace más alto mientras el gradiente tiende a tomar un color más oscuro.

3.7.4. Búsqueda de genes utilizando motivos descubiertos.

Para explorar el genoma de *Homo sapiens* en búsqueda de nuevos genes regulados por el factor de transcripción MEF2A, se utilizaron como motor de búsqueda CArGbox putativas, designadas como tales según su ΔG_{bind} y designadas como la población CArGrango. Se usó el servidor (<http://rsat.ulb.ac.be/rsat/>) *regulatory sequence analysis tools* (RSATools) adaptando el protocolo expuesto por el mismo servidor (Turatsinze *et al.*, 2008) para cubrir las necesidades de éste análisis.

Se comenzó por elegir el genoma de *Homo sapiens* como blanco de búsqueda, sin especificar algún gen en particular para que todas sus secuencias promotoras fuesen blanco de investigación, consideradas desde el sitio -1 del inicio de la transcripción hasta el -5500. Según las recomendaciones para organismos eucariontes se previno el solapamiento de genes vecinos, se admitieron posiciones imprecisas y se activó “*Mask repeats*”

para enmascarar los elementos repetitivos que pudiesen obstaculizar la localización de motivos. Una vez que este primer conjunto de instrucciones fue admitido por el servidor, se alimentó al programa “*matrix-scan*” (por turnos) con cada una de las matrices que se elaboraron para cada grupo de secuencias definido en la sección 3.7.5 de metodologías del presente capítulo.

La construcción del *background model* que se refiere al modelo que refleja la composición de bases presente en el entorno genómico, es decir la configuración que no corresponde al motivo; se eligió un modelo Markoviano de segundo orden, reportado como altamente robusto para la designación de secuencias como elementos-cis a comparación del modelo de Bernoulli (Turatsinze *et al.*, 2008), utilizado en el protocolo original.

Para la construcción de el *background model* también se tomaron en cuenta las precalibraciones conocidas para el caso del contexto genómico de *Homo sapiens*, las cuales reportan frecuencias de 0.251, 0.241, 0.247 y 0.259, para las bases A, C, G y T respectivamente.

Para la exploración de nuevos elementos-cis, se decidió tomar en cuenta la cadena líder y su complementaria solicitando el puntaje de peso ($W_s > 1$) y el P-value ($1e-4$) como indicadores de la probabilidad de una secuencia para ser designada como CArGbox. En ambos casos el valor se considera como altamente confiable (Turatsinze *et al.*, 2008).

3.7.5. Construcción de matrices

El servidor RSATools basa la localización de elementos-cis en matrices de puntuación por posición específica (PSSMs), por lo que se construyeron seis matrices en formato TRANSFAC en base a cada uno de los conjuntos definidos en la sección 3.7.2 de métodos. Las matrices fueron fabricadas al determinar la frecuencia de aparición de cada nucleótido por posición específica. Se logró contabilizando “a mano” el contenido total de A, C, G ó T en cada sitio de las secuencias que forman cada subgrupo energético (que va de -3.9 kJ/mol a - 4.4 kJ/mol). Se realizaron las PSSMs bajo el formato TRANSFAC por ser de fácil lectura y construcción, además de ser aceptado por la IUPAC. Las matrices resultantes se exponen en la sección de resultados del presente capítulo.

3.8. RESULTADOS DEL CAPÍTULO III

3.8.1. El programa MultiCArG.

Su ejecución permite de forma automática la construcción de mutantes en base a una secuencia de DNA conservando los parámetros estructurales de un modelo determinado por RMN, cristalografía o *docking*. Es capaz de elaborar complejos y con base en éstos determinar la energía libre de unión, por lo que se considera una herramienta complementaria en el estudio de la interacción de proteínas con unión a DNA y de sus secuencias blanco (*Cfrt.* apéndice D). Las secciones más importantes del programa se resaltan a continuación:

3.8.2. Archivo “*Universooligos.txt*”

Este archivo posee 1,048,526 secuencias que corresponden a todos los oligonucleótidos probables a los

cuales puede unirse el factor de transcripción MEF2A que conforman una base de datos que puede ser semilla del programa *MultiCarG.py*, y que asegura su completa automatización para evitar la intervención del usuario con el programa para ahorrar tiempo de cómputo. A continuación se muestra una sección de las secuencias obtenidas localizadas en las primeras 15 líneas del fichero y en las que se observa claramente la sustitución de bases en las últimas posiciones de la CArGbox:

```
(A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'A')
(A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'C')
(A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'G')
(A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'T')
(A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'C, 'A')
(A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'C, 'C')
(A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'C, 'G')
(A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'C, 'T')
(A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'G, 'A')
(A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'G, 'C')
(A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'G, 'G')
(A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'G, 'T')
(A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'T, 'A')
(A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'T, 'C')
(A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'A, 'T, 'G')
```

3.8.3. Librerías de átomos de hidrógenos.

Se cuenta con una librería de 80 archivos que corresponden a las coordenadas de átomos de hidrógeno pertenecientes a los nucleótidos introducidos en las secuencias de DNA por eventos simulados de transversión y transición a los que se sometió el arquetipo de CArGbox. Ésta librería permite la adición de átomos de hidrógeno de cualquier nucleótido y su base complementaria, con la premisa de adicionar los átomos con distancias de enlaces y ángulos en las posiciones canónicas para cada base. (Chen *et al.*, 2010).

3.8.4. Obtención de estructuras tridimensionales de DNA en formato *.pdb*.

La obtención de estructuras tridimensionales de las secuencias CArGbox construidas *de novo*, constituyó un reto informático, ya que hasta el momento no se cuenta con ninguna paquetería disponible para simular procesos de mutación en una molécula de DNA, conservando una estructura molecular definida, como la proveniente de resonancia Magnética Nuclear o Cristalografía. El programa *MultiCarG.py* manipula funciones de la paquetería *X3DNA* y lo complementa con líneas de código para crear mutantes de la CArGbox. El programa es apto para simular procesos de mutación en una estructura de DNA y conserva sus parámetros de coordenadas originales en un formato altamente manipulable como el *.pdb*.

3.8.5. Complejos MEF2A-CArGbox.

Fue posible obtener nuevos complejos MEF2A-CArGbox (no determinados por RMN o cristalografía) reproduciendo las características estructurales canónicas reportadas en el archivo *.pdb* con No. de Id: 1C7U, que pueden ser sometidas a procesos experimentales de dinámica molecular, determinación de energía libre, entre otros. En la Fig. 23 se observa el complejo formado con la CArGbox mutante ATGGTACATT como ejemplo de estructuras que pueden obtenerse al manipular el programa *MultiCarG.py*:



Fig. 23. Complejo MEF2A-CARgbox(ATGGTACATT). Utilizando el programa *MultiCARg.py* se simularon 6 procesos de transversión en los sitios 1, 4, 7, 8, 9 y 10 de la CARgbox y una transición en el sitio 3, pese a éstos cambios los parámetros estructurales de la CARgbox igualan a los descritos en las determinaciones de RMN y Cristalografía, por lo que ésta mutante se acopla de forma idónea al factor de transcripción MEF2A, lo que permite el estudio de ambas moléculas como un complejo. El dímero de MEF2A está representado en color gris, los nucleótidos se distinguen por colores, la adenina de color rojo, la guanina de color anaranjado, la timina de color azul y la citosina de color violeta. En la parte inferior de la figura se resalta la secuencia de la caja CARg especificando su secuencia. La visualización se obtuvo con el programa *VMD*.

3.8.6. Determinación de la energía de unión.

El programa *MultiCARg.py* tuvo por objeto primordial determinar la energía libre de unión generada entre MEF2A y una serie de veinteámeros de DNA que contienen una secuencia de diez nucleótidos en su parte central, a la que con base a la ΔG_{bind} , se pretendió caracterizar como CARgbox.

3.8.7. Determinación de la energía libre de unión generada en diversos complejos MEF2A unidos a diversas secuencias de DNA.

La secuencia 5' CTATTAATAG 3' se acepta ampliamente por su comprobación experimental como sitio de unión del factor de transcripción MEF2A (McKinsey *et al.*, 2000); su Δg_{bind} es un parámetro que se consideró para designar a otras secuencias como CARgbox; por lo tanto la ΔG_{bind} obtenida del complejo arquetípico (tomado del *.pdb* con No de id: 1C7U) al que se le aplicó una minimización que reportó energías inferiores a 100 $\text{kJ mol}^{-1} \text{ nm}^{-1}$, fue de: $-4.241505721755e+02 \text{ kJ/mol}$.

En la exploración de nuevas secuencias CARgbox probables, se calculó el valor de ΔG_{bind} para 339672 secuencias (Cfr. en el archivo "*CarGenerg.txt*"). Estos valores pueden apreciarse en la Fig. 24).

ΔG_{bind} CORRESPONDIENTE A LA POBLACIÓN TOTAL DE 339672 SECUENCIAS

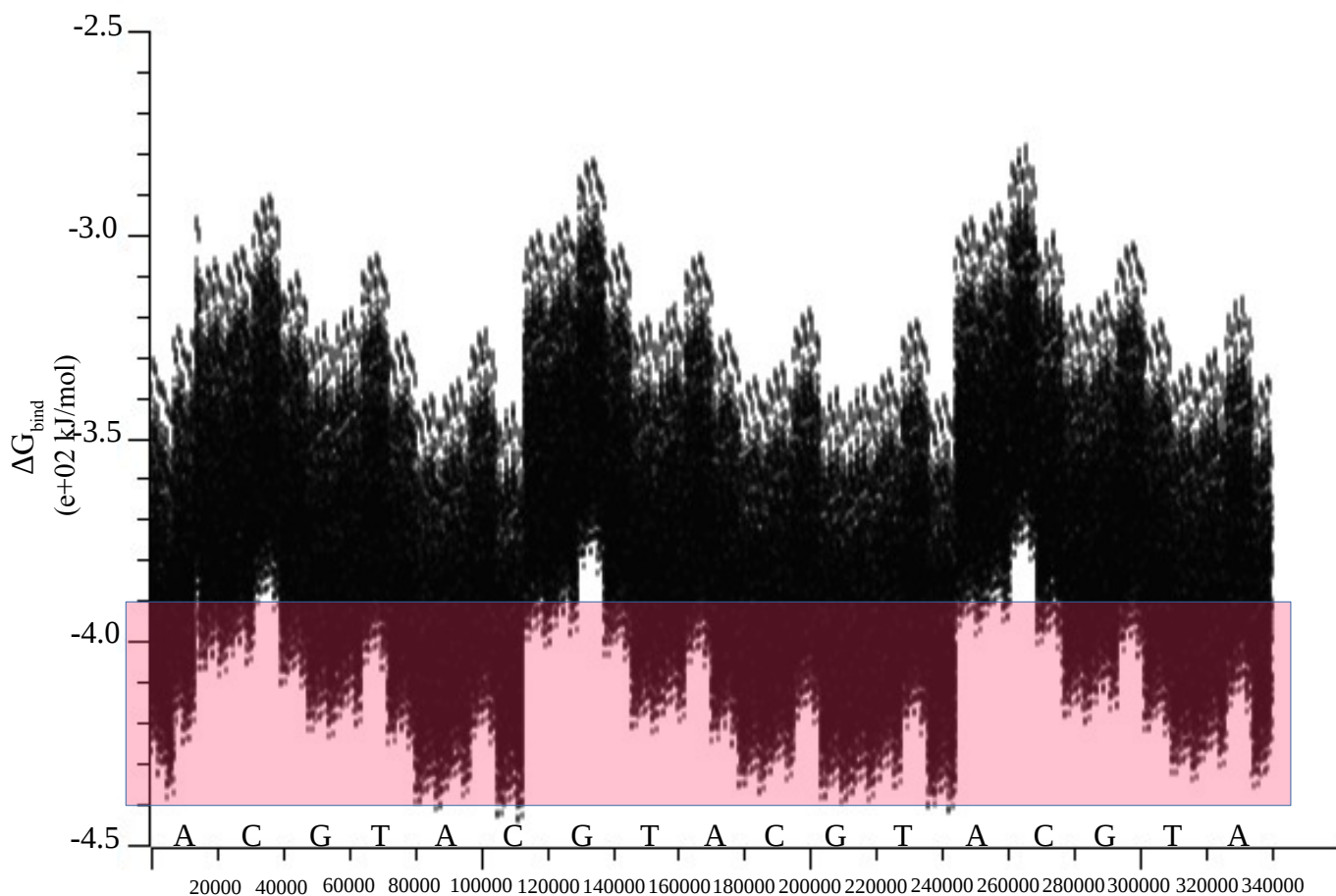


Fig. 24. Valores de ΔG_{bind} correspondientes a el total de secuencias estudiadas. Los valores de Valores de ΔG_{bind} presentados provienen de las 339672 secuencias tomadas al azar de la población de oligos resultante de la combinatoria 4^{10} (sección 3.7.2). El eje de las y corresponde a la ΔG_{bind} , al eje x pertenecen el número de secuencias y la progresión de nucleótidos que indica que base aparece en el primer sitio de los decámeros: como puede notarse cada 20000 secuencias cambia la primera base en el orden A, C, G y T, las secuencias están acomodadas según el arreglo descrito en la sección 3.7.3 de los métodos. El resto de las posiciones sustituye nucleótidos en el mismo orden, por lo que se observa un segmento de la combinatoria 4^{10} graficado según estas progresiones, nótese la aparición de un patrón en la distribución de los datos estudiados. La sección resaltada en rojo corresponde a la población de secuencias a analizar y cuyo conjunto está definido por: $\{x | x > 3.8e+02 \text{ kJ/mol}, x < 4.5e+02 \text{ kJ/mol}\}$. El gráfico fue obtenido con el programa *SciDAVis*.

Aunado al diagrama de dispersión (Fig. 25) se obtuvieron las frecuencias de aparición de los nucleótidos A, C, G y T en cada posición presente en el decámero que abarca la CArGbox, reportados como porcentaje de incidencia. Es importante acentuar que el histograma que se presenta a continuación, se construyó utilizando las 339672 secuencias a las que se les calculó el valor de ΔG_{bind} .

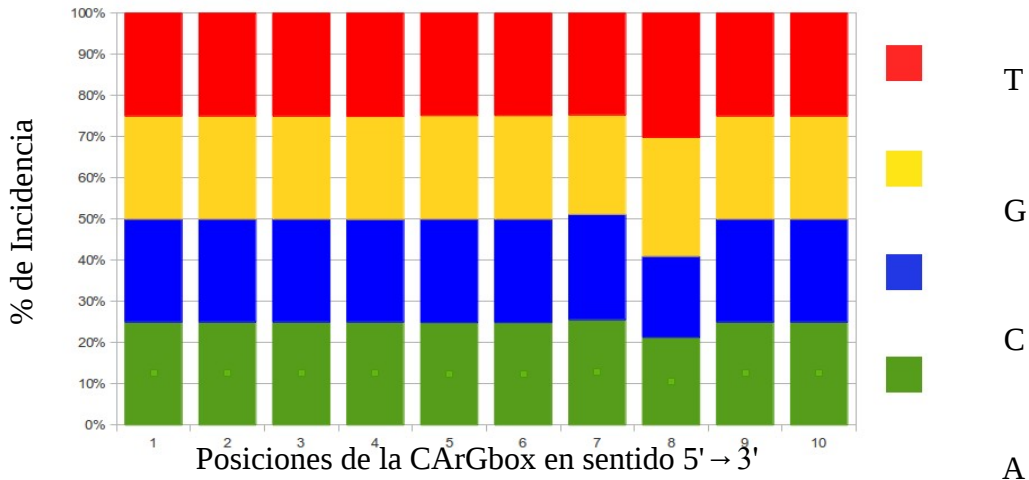


Fig.25. Incidencia de aparición de los nucleótidos presentes en las secuencias para las que se obtuvo el valor de ΔG_{bind} . Para cada posición de las secuencias se realizó un conteo de los nucleótidos presentes en la sitio x, posteriormente se obtuvo se porcentaje de incidencia con respecto al 100% de las secuencias (339672). La ubicación de los nucleótidos está reportada en sentido 5' → 3', a cada nucleótido le corresponde un color específico como se aprecia en el extremo derecho de la figura.

Como se observa claramente en la Fig.25, la composición general de la población de secuencias tomadas al azar es homogénea, ya que las cuatro bases de DNA están representadas en cada posición oscilando alrededor de un valor promedio de $24.2\% \pm 1.9\%$. La posición 8 es la menos homogénea, reporta los siguientes valores de incidencia: A = 21.29 %, C = 19.67, G = 27.10% y T = 30.16%.

3.8.8. El conjunto CARGrango.

Se extrajeron las secuencias con valores de ΔG_{bind} entre $-3.9e+02$ kJ/mol y $-4.4e+02$ kJ/mol. De esta población se obtuvo un diagrama de logo que muestra las frecuencias de cada nucleótido por posición específica.

Comparando con las frecuencias iniciales obtenidas de la población total, se observa claramente (Fig. 26) un incremento en la incidencia de A y C en la primer posición, un descenso de C en las posiciones 2 a 3 y 5 a 9; en los sitios 2 al 9 se observa un incremento de nucleótidos A y T, mientras que la posición 10 contiene una frecuencia similar a la que se observa en la población que cuenta con el total de las secuencias estudiadas.

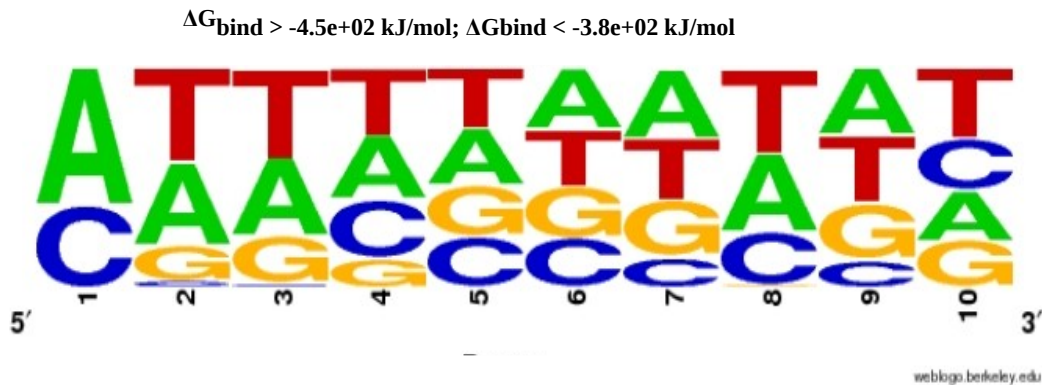


Fig. 26. Diagrama de logo del conjunto definido por: $\{x | x > -3.8e+02 \text{ kJ/mol}, x < -4.5e+02 \text{ kJ/mol}\}$. Se calcularon las frecuencia por sitio específico para cada base presente en el DNA, las posiciones se presentan en dirección 5' → 3'. Las bases con una frecuencia mayor a parecen en la cúspide del diagrama y las que poseen frecuencias menores aparecen en el fondo. Si la frecuencia de una base es desconsiderable o no presenta incidencias, no aparece reportado en el sitio. En general se observa un predominio de la A y T sobre la G Y C, siendo ésta última la menos representada en éste grupo de secuencias..

Existe una relación directamente proporcional entre los conjuntos de secuencias (definidos en la sección 3.7.2 de métodos) y sus valores de ΔG_{bind} . A continuación se enlista el número total de secuencias detectadas por cada grupo:

- CArGrango = $\{x \mid x > 3.8e+02 \text{ kJ/mol} , x < 4.5e+02 \text{ kJ/mol}\}$ (339672 secuencias).
- $-3.9e+2 \text{ kJ/mol} = \{x \mid x > 3.8e+02 \text{ kJ/mol} , x < 4.0e+02 \text{ kJ/mol}\}$, (43894 secuencias),
- $-4.0e+2 \text{ kJ/mol} = \{x \mid x > 3.9e+02 \text{ kJ/mol} , x < 4.1e+02 \text{ kJ/mol}\}$, (36510 secuencias).
- $-4.1e+2 \text{ kJ/mol} = \{x \mid x > 4.0e+02 \text{ kJ/mol} , x < 4.2e+02 \text{ kJ/mol}\}$, (23040 secuencias).
- $-4.2e+2 \text{ kJ/mol} = \{x \mid x > 4.1e+02 \text{ kJ/mol} , x < 4.3e+02 \text{ kJ/mol}\}$, (10012 secuencias).
- $-4.3e+2 \text{ kJ/mol} = \{x \mid x > 4.2e+02 \text{ kJ/mol} , x < 4.4e+02 \text{ kJ/mol}\}$, (1916 secuencias).
- $-4.4e+2 \text{ kJ/mol} = \{x \mid x > 4.3e+02 \text{ kJ/mol} , x < 4.5e+02 \text{ kJ/mol}\}$, (44 secuencias).

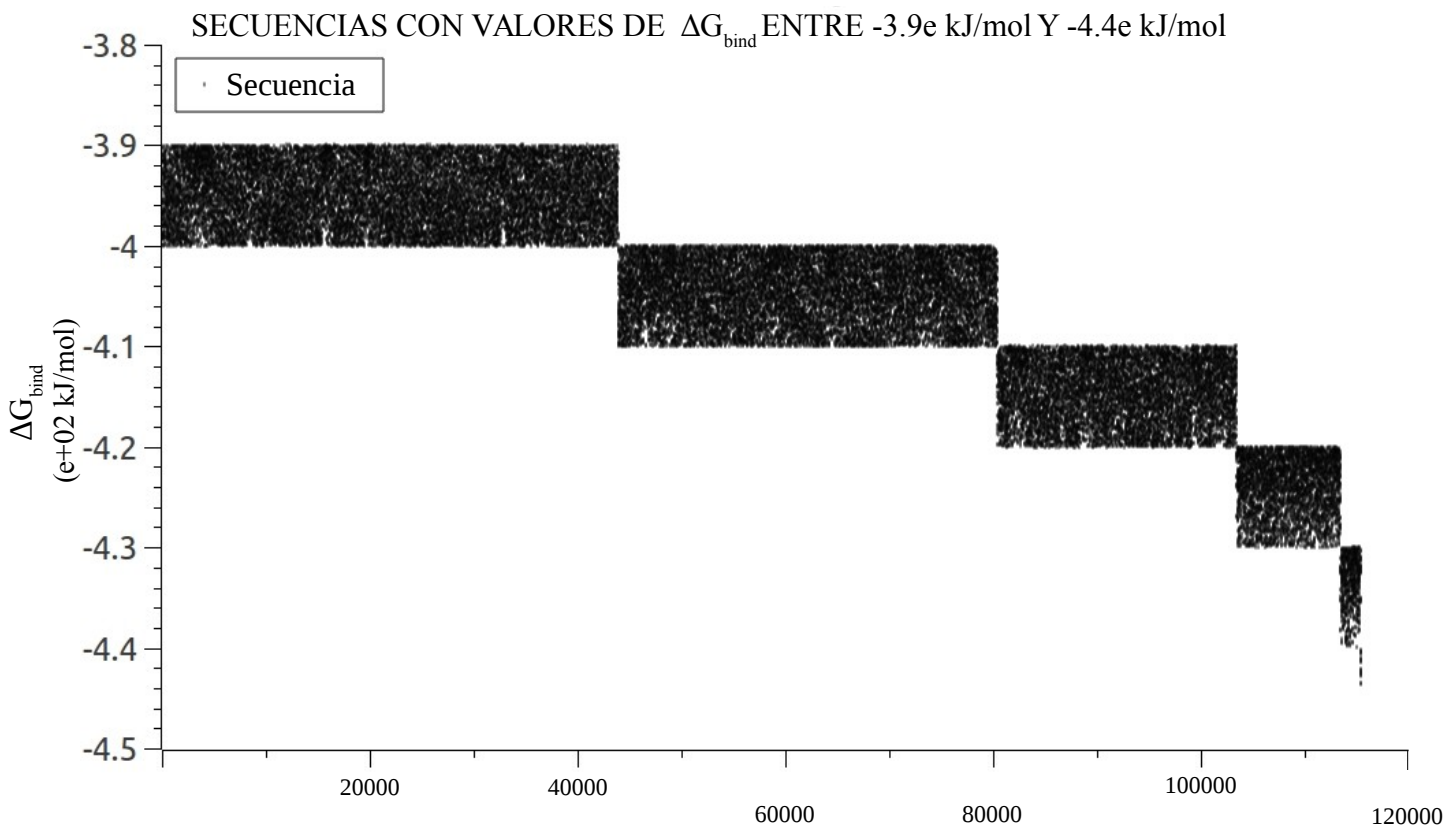


Fig. 27. Secuencias de DNA con valores de ΔG_{bind} entre $3.9e+02 \text{ kJ/mol}$ y $-4.4e+02 \text{ kJ/mol}$. Se muestran los valores de ΔG_{bind} del conjunto CarGenrgy compuesto por 115410 secuecnias representadas por un punto y colocadas en el eje de las x, el eje de las y corresponde a los valores de ΔG_{bind} . En éste gráfico se observa claramente la relación directamente proporcional entre el valor de ΔG_{bind} y el número de secuencias que posee cada escalón energético. Esta gráfica se construyó aislando por bloques cada uno de los conjuntos definidos en la sección XXX de métodos, por lo que no es posible apreciar el patrón de distribución observado en la gráfica dispuesta en la fig XXXX. Los datos fueron graficados en el programa *SciDAVis*.

Se muestra el extracto de datos correspondientes a la población CarGrango (Fig. 27), ordenados según la progresión de bases expuesta en la sección de métodos 3.7.2.

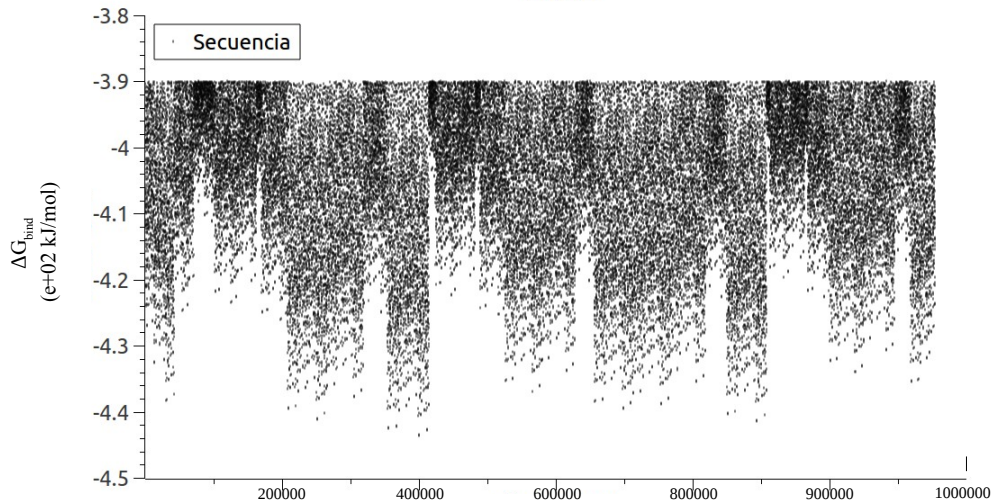


Fig.28. Secuencias de DNA con valores de ΔG_{bind} entre $-3.9\text{e}+02$ kJ/mol y $-4.4\text{e}+02$ kJ/mol presentadas según el acomodo de la sección 3.7.3 de métodos. El eje de las y corresponde a los valores de ΔG_{bind} y en el eje de las x se encuentran las secuencias de DNA las cuales están representadas por un punto. En la figura se logra intuir el mismo patrón cíclico observado en la Fig. 29 de la presente sección de resultados. El gráfico fue realizado con el programa *SciDAVis*.

En la Fig. 28 puede apreciarse un patrón generado por la presencia de altos contenidos de Timina, relacionado con bajos niveles energéticos.

Se obtuvieron seis subconjuntos de la población CArGrango, con la que se realizó una comparación entre las frecuencias iniciales y las que se obtuvieron de secuencias agrupadas por criterios energéticos.

El primer subconjunto está definido por: $\{x | x > -3.8\text{e}+02$ kJ/mol , $x < -4.0\text{e}+02$ kJ/mol $\}$. El diagrama de logo obtenido con base en esta población de secuencias (Fig. 29), muestra un incremento moderado en la incidencia de A en la posición 1, un dominio de T en las posiciones 2 y 3 atenuado por la presencia de G y A; los sitios 4 a 10 poseen frecuencias similares para cada uno de los nucleótidos, en cuyo caso no se observa predominancia por una base específica en este rango de energía para dichos sitios, sin embargo de manera general observamos un predominio de G y A, y una menor incidencia de C.

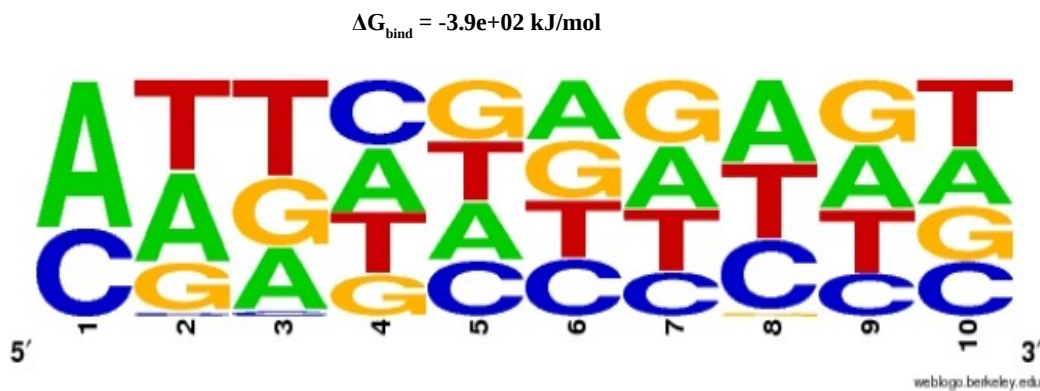


Fig. 29. Diagrama de logo de las secuencias que poseen un valor de ΔG_{bind} de $-3.9\text{e}+02$ kJ/mol. Se calcularon las frecuencia por sitio específico para cada base presente en el DNA, las posiciones se presentan en dirección $5' \rightarrow 3'$. Las bases con una frecuencia mayor aparecen en la cúspide del diagrama y las que poseen frecuencias menores aparecen en el fondo. Si la frecuencia de una base es despreciable o no presenta incidencias, no aparece en la posición. Con respecto a las frecuencias de la población que posee el rango completo de ΔG_{bind} reportadas, se observa un predominio de la A en la posición 1 y de la T en los sitios 2 y 3. El resto de los sitios permanece con frecuencias similares entre los nucleótidos.

El diagrama que se obtuvo para las secuencias con valor de $\Delta G_{\text{bind}} = -4.0\text{e}+02$ kJ/mol (Fig. 30) muestra un predominio de A y C en el sitio uno. De los sitios 2 al 9 se muestra un incremento en las frecuencias de T y A con respecto al paso energético $-3.9\text{e}+02$ kJ/mol y un decremento de la frecuencia de G. En la posición 10 no predomina alguna base particular, y en general la frecuencia de la C decae en los sitios 4 al 9.

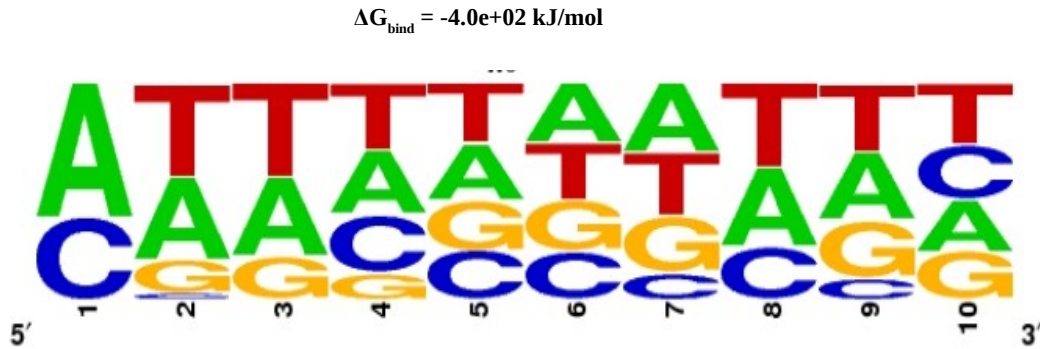


Fig. 30. Diagrama de logo de las secuencias que poseen un valor de ΔG_{bind} de $-4.0\text{e}+02$ kJ/mol. Se calcularon las frecuencia por sitio específico para cada base presente en el DNA, las posiciones se presentan en dirección $5' \rightarrow 3'$. Las bases con una frecuencia mayor aparecen en la cúspide del diagrama y las que poseen frecuencias menores aparecen en el fondo. En la figura se observa un incremento en la frecuencia de A y T en los sitios 2 al 5 y 8 al 9, el sitio 1 y 10 permanece sin cambios significativos con respecto al anterior paso energético de 3.9.

El diagrama de logo producido con base en el conjunto definido por $\{x | x < 4\text{e}+02 \text{ kJ/mol}, x > 4.2\text{e}+02 \text{ kJ/mol}\}$ (Fig. 31) presenta frecuencias similares al paso energético correspondiente a $\Delta G_{\text{bind}} = -4.0\text{e}+02$ kJ/mol; sin embargo se aprecia un decremento sutil de G y C en el sitio 2 al 10, así como un aumento de A en la posición 9. La frecuencia de bases es muy similar entre si en el sitio 10.

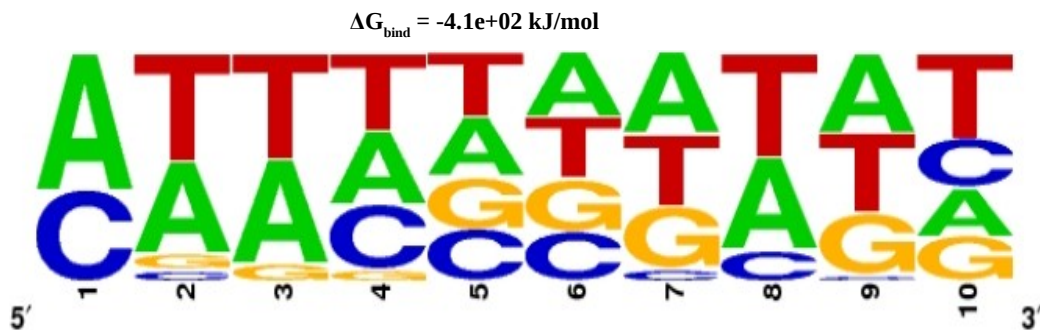


Fig. 31. Diagrama de logo de las secuencias que poseen un valor de ΔG_{bind} de $-4.1\text{e}+02$ kJ/mol. Se calcularon las frecuencia por sitio específico para cada base presente en el DNA, las posiciones se presentan en dirección $5' \rightarrow 3'$. Las bases con una frecuencia mayor aparecen en la cúspide del diagrama y las que poseen frecuencias menores aparecen en el fondo. Se observan frecuencias más altas de T y A de los sitios 2 al 9, y una frecuencia medior de C y G (que en algunos casos es mínima).

En el diagrama del logo de las secuencias con $\Delta G_{\text{bind}} = -4.2\text{e}+02$ kJ/mol (Fig. 32) se observa al sitio 1 sin cambios importantes con respecto a los pasos energéticos anteriores, existe un aumento en la frecuencia de A en el sitio 3 y 7, un aumento en la frecuencia de T en los sitios 2, 4, 8, 9 y 10, esta última posición comienza con una tendencia a favor de ser ocupada por T. Debido a los aumentos de A y T la frecuencia de C y G se ve minimizada.

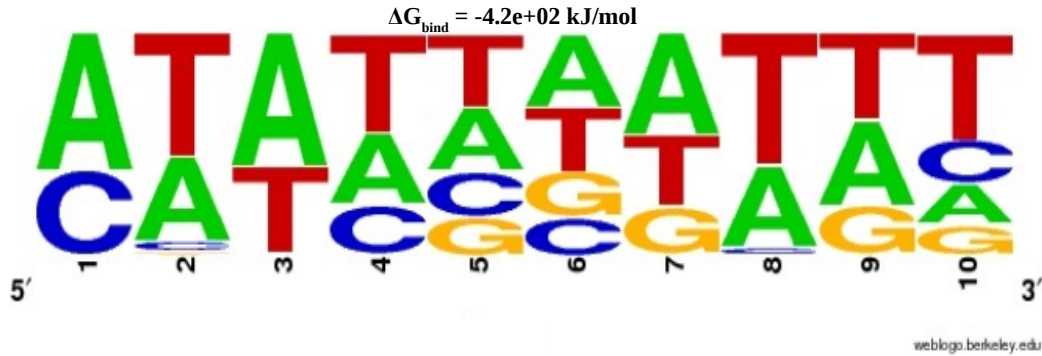


Fig. 32. Diagrama de logo de las secuencias que poseen un valor de ΔG_{bind} de $-4.2e+02 \text{ kJ/mol}$. Se calcularon las frecuencia por sitio específico para cada base presente en el DNA, las posiciones se presentan en dirección $5' \rightarrow 3'$. Las bases con una frecuencia mayor aparecen en la cúspide del diagrama y las que poseen frecuencias menores aparecen en el fondo. Las frecuencias de A y T dominan a lo largo de todos los sitios haciendo decaer el contenido de C y G. En el sitio 10 comienza a observarse una tendencia en el predominio de la T.

El gráfico de logo elaborado con las secuencias con un valor de $\Delta G_{\text{bind}} = -4.3e+02 \text{ kJ/mol}$ (Fig. 33), muestra un aumento generalizado en el contenido de T y A, lo que implica una composición más homogénea entre este grupo de decámeros. Los sitios 2,4, 8 y 10 están enriquecidos en T, mientras que los sitios 3, 7 y 9 lo están en A; la posición 1 conserva una alta frecuencia de C y A, mientras que la 5 y 6 poseen un contenido equitativas para los 4 nucleótidos.

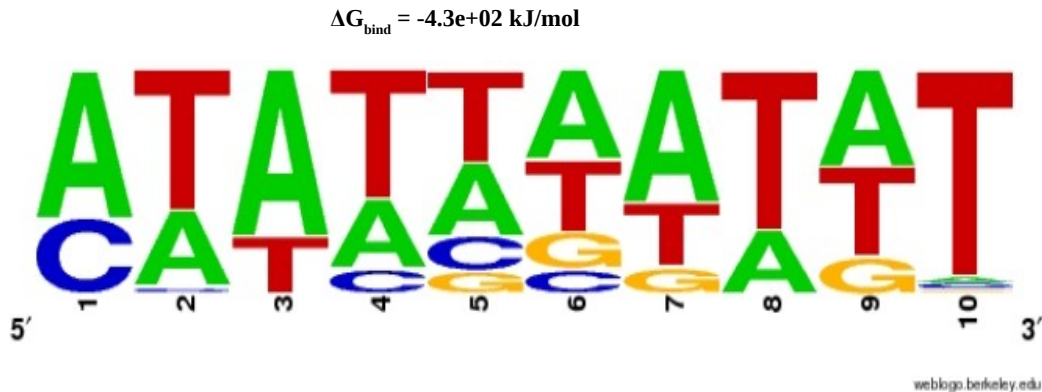


Fig. 33. Diagrama de logo de las secuencias que poseen un valor de ΔG_{bind} de $-4.3e+02 \text{ kJ/mol}$. Se calcularon las frecuencia por sitio específico para cada base presente en el DNA, las posiciones se presentan en dirección $5' \rightarrow 3'$. Las bases con una frecuencia mayor aparecen en la cúspide del diagrama y las que poseen frecuencias menores aparecen en el fondo. El contenido de T y A aumenta en este grupo de secuencias sobre todo en el caso del primer nucleótido en el sitio 2, 8 y 10.

El conjunto de secuencias definidas por valores de $\Delta G_{\text{bind}} = \{x \mid x > 4.5e+02 \text{ kJ/mol}, x < 4.3e+02 \text{ kJ/mol}\}$ componen la última subpoblación, este grupo de decámeros se encuentra altamente enriquecido en nucleótidos de A en las posiciones 1, 3 y 7, se observa una mayor frecuencia de aparición de T en las posiciones 2, 4, 5, 8 y 10 como se observa en la figura 34. Muchas de las secuencias alojadas en este conjunto poseen C en la primera posición y un contenido medianamente equilibrado de A/T en las posiciones 6 y 9. El contenido de G y C desaparece totalmente de los sitios 2 al 8 y en el sitio 10, sin embargo una minoría de secuencias contienen G en la posición 9.

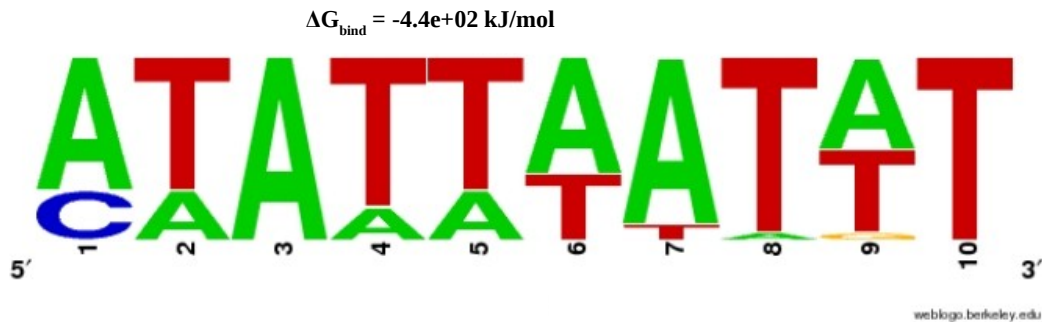


Fig. 34. Diagrama de logo de las secuencias que poseen un valor de ΔG_{bind} de $-4.4\text{e}+02 \text{ kJ/mol}$. Se calcularon las frecuencia por sitio específico para cada base presente en el DNA, las posiciones se presentan en dirección 5' \rightarrow 3'. Las bases con una frecuencia mayor aparecen en la cúspide del diagrama y las que poseen frecuencias menores aparecen en el fondo. Observamos una recurrencia significativa de A y T en todas las posiciones, dado este aumento la C y la T ya no se encuentran representadas en las secuencias de éste grupo, excepto por la C en la posición 1 y un reminecente de G en el sitio 9.

Con respecto al primer cálculo de las frecuencias del conjunto definido por todos los valores de ΔG_{bind} calculados (lo que se ha venido nombrando como la población CArGenerg), se observa una marcada tendencia en el aumento del contenido de T y A en todas las posiciones y un equilibrio en las frecuencias de A y C en el sitio 1, conforme los valores de ΔG_{bind} decaen de $-3.9\text{e}+02 \text{ kJ/mol}$ a $-4.4\text{e}+02 \text{ kJ/mol}$. Cabe resaltar el diagrama de logo correspondiente a las secuencias con un valor de $\Delta G_{\text{bind}} = -4.2\text{e}+02 \text{ kJ/mol}$, que muestra la distribución de frecuencias más parecida al dominio canónico 5' (CTA(T/A)₄TAG) 3' de la CArGbox de MEF2A aceptado hasta ahora ampliamente (ejemplo: el valor de $\Delta G_{\text{bind}} = -4.241505721755\text{e}+02 \text{ kJ/mol}$ de la secuencia que se tomó como arquetipo en esta investigación 5' CTATTAATAG 3').

3.8.9. Matrices de puntuación por posición específica (PSSMs).

La construcción de matrices es un paso fundamental para la descripción estadística de los factores de transcripción, éstas otorgan una descripción detallada de la estructura del elemento-cis al que se une el FT, y son herramientas fundamentales para la búsqueda de sitios de regulación dentro del genoma y la identificación de genes regulados por el factor de transcripción estudiado. Atendiendo a esta necesidad, se obtuvieron seis PSSMs producidas con base en la determinación de la ΔG_{bind} .

Las matrices que se muestran en la figura 35, están escritas en formato TRANSFAC y derivan de los grupos descritos en la sección 3.7.2 de métodos, pueden copiarse sin cambios para ser utilizadas.

```

; Matrices derivada de las secuencias con DGbind=-3.9kJ/mol
;
; MATRIX 1/1 : -3.9kJ_mol
;
AC      -3.9
XX
DT      Quetzalcoatl Escalante-Covarrubias
XX
P0      A  C  G  T
1  13629 8318  0  0
2   8318  613 4538 8478
3   5856  719 6418 8954
4   6035 6173 3966 5773
5   5423 5303 5648 5573
6   5574 5436 5483 5454
7   5877 4194 6016 5860
8   7733 6514  632 7068
9   5920 4126 6001 5900
10  5210 5150 5185 6402
XX
//

```

```

; Matrices derivada de las secuencias con DGbind=-4.0kJ/mol
;
; MATRIX 1/1 : -4.0kJ_mol
;
AC      -4.0
XX
DT      Quetzalcoatl Escalante-Covarrubias
XX
P0      A  C  G  T
1  11319 6936  0  0
2   6936  651 2821 7847
3   6697  35 3611 7912
4   5419 4881 2357 5598
5   4762 4191 4236 5066
6   5045 4048 4333 4829
7   5842 2377 4769 5267
8   6645 4474  23 7113
9   5733 1883 4886 5754
10  4423 4626 4096 5110
XX
//

```

```

; Matrices derivada de las secuencias con DGbind=-4.1kJ/mol
;
; MATRIX 1/1 : -4.1kJ_mol
;
AC      -4.1
XX
DT      Quetzalcoat Escalante-Covarrubias
XX
P0      A  C  G  T
1  6915 4605  0  0
2  4605 569  858 5488
3  5222  0  922 5376
4  3673 3264 630 3953
5  2995 2622 2641 3262
6  3296 2484 2670 3070
7  4114 596 3171 3639
8  4643 1618  0 5259
9  4050 287 3135 4048
10 2431 2490 2257 4342
XX
//

```

```

; Matrices derivada de las secuencias con DGbind=-4.2kJ/mol
;
; MATRIX 1/1 : -4.2kJ_mol
;
AC      -4.2
XX
DT      Quetzalcoat Escalante-Covarrubias
XX
P0      A  C  G  T
1  3102 1904  0  0
2  1904 243  70 2789
3  3009  0  52 1945
4  1627 1096 36 2247
5  1440 1015 877 1674
6  1673 850 1025 1458
7  2321 31 1063 1591
8  1851 186  0 2969
9  1936 3 1108 1959
10 912 984 689 2421
XX
//

```

```

; Matrices derivada de las secuencias con DGbind=-4.3kJ/mol
;
; MATRIX 1/1 : -4.3kJ_mol
;
AC      -4.3
XX
DT      Quetzalcoat Escalante-Covarrubias
XX
P0      A  C  G  T
1  633 325  0  0
2  325 31  0  602
3  714  0  0  244
4  296 110  0  552
5  313 151 99 395
6  383 105 158 312
7  572  0 115 271
8  274  0  0  684
9  418  0 158 382
10 34 32 14 878
XX
//

```

```

; Matrices derivada de las secuencias con DGbind=-4.4kJ/mol
;
; MATRIX 1/1 : -4.4kJ_mol
;
AC      -4.4
XX
DT      Quetzalcoat Escalante-Covarrubias
XX
P0      A  C  G  T
1  16 6 0 0
2  6 0 0 16
3  22 0 0 0
4  4 0 0 18
5  6 0 0 16
6  14 0 0 8
7  20 0 0 2
8  1 0 0 21
9  11 0 1 10
10 0 0 0 22
XX
//

```

Fig. 35. PSSMs construidas con secuencias de las que se obtuvo su ΔG_{bind} . Se agruparon en seis grupos cuyas secuencias poseen los valores de: $-3.9e+02$ kJ/mol, $-4.e+02$ kJ/mol, $-4.1e+02$ kJ/mol, $-4.2e+02$ kJ/mol, $-4.3e+02$ kJ/mol y $-4.4e+02$ kJ/mol. Para cada grupo se calculó el contenido de cada base en el total de las secuencias, éste corresponde al número que se indica bajo cada columna correspondiente a un nucleótido específico; recuérdese que todas las secuencias poseen una longitud de 10 pb, sus posiciones se indican en el extremo izquierdo de cada formato debajo de los nucleótidos y están descritas en dirección $5' \rightarrow 3'$.

3.8.10. Nuevos elementos-cis descubiertos en el genoma de *Homo sapiens*.

Los resultados de las búsquedas de elementos-cis en el servidor RSATools, basada en las frecuencias de las matrices construidas utilizando las secuencias con un valor de ΔG_{bind} entre $-3.9e+02$ kJ/mol y $-4.4e+02$ kJ/mol, se reportan en la Tabla 4. Cabe mencionar que los datos obtenidos en base a esta exploración exceden a los expuestos en el presente reporte debido a que RSATools no elimina sinónimos, por lo que fue necesario excluirlos con el fin de contemplar el alcance real de esta búsqueda.

Por medio de la metodología propuesta en la presente investigación se lograron identificar más de 500 genes cuya expresión probablemente esté siendo regulada por MEF2A. Utilizando la PSSMs construida con las secuencias que dieron un valor de $\Delta G_{bind} = -4.1e+02$ kJ/mol se detectaron 20 genes, convirtiéndose así en la matriz que detectó un menor número de genes, en comparación con las PSSMs construidas con las secuencias

que dieron un valor de $\Delta G_{\text{bind}} = -4.0\text{e}+02$ kJ/mol, $-4.3\text{e}+02$ kJ/mol y $-4.4\text{e}+02$ kJ/mol con las que se rastrearon al rededor de 200 genes. La PSMM construida con las secuencias que dieron un valor de $\Delta G_{\text{bind}} = -4.2\text{e}+02$ kJ/mol arrojó un total de 47 genes y la PSMM construida con las secuencias que dieron un valor de $\Delta G_{\text{bind}} = -3.9\text{e}+02$ kJ/mol arrojó un total de 90 genes.

Tabla. 4. Número de genes putativos rastreados por el método de derivación de la ΔG_{bind} . La primera columna de la tabla corresponde al grupo energético del cual se obtuvo la matriz de puntuación por posición específica que se utilizó como motor para la búsqueda de elementos-*cis* putativos del factor de transcripción MEF2A en el servidor RSAT, por cada grupo se especifica el número total de genes encontrados así como el promedio de los valores de W_s y P-value con su desviación estándar.

Valor de ΔG_{bind}	# Total de genes identificados por RSATools	$\bar{x} W_s$	\bar{x} P-value
-3.9e+02 kJ/mol	90	2.62 \pm 0.01	6.4e-05 \pm 2.36
-4e+02 kJ/mol	205	3.17 \pm 0.01	4.6e-05 \pm 2.46
-4.1e+02 kJ/mol	20	4.18 \pm 0.005	5.7e-05 \pm 2.81
-4.2e+02 kJ/mol	47	6.10 \pm 0.007	5.5e-05 \pm 2.61
-4.3e+02 kJ/mol	204	8.20 \pm 0.06	5.2e-05 \pm 2.66
-4.4e+02 kJ/mol	228	10.9 \pm 0	5.4e-05 \pm 2.62

Conforme descienden los niveles energéticos, se observa un aumento del promedio en los valores de W_s , proporcionado por RSAT que refleja un aumento en la probabilidad, para que las secuencias reportadas ocurran conforme al modelo propuesto por cada uno de los PSSMs y no al *background model*.

Con respecto a los valores de P-value, parecen descender a la par de ΔG_{bind} , aunque de forma poco significativa, ya que todos ellos corresponden a magnitudes en el orden de 1×10^{-5} .

En las tablas 5 a la 10 se reportan los genes encontrados con ayuda del servidor RSAT utilizando las frecuencias de las matrices construidas con las secuencias de valores de ΔG_{bind} entre $-3.9\text{e}+02$ kJ/mol y $-4.4\text{e}+02$ kJ/mol. Aunada a esta búsqueda se realizó una profunda investigación de cada gen asociado al elemento-*cis* localizado, esto con la finalidad de conocer el tejido específico en el que se expresa cada uno, y con ello obtener una perspectiva acerca de su rol biológico en *Homo sapiens*.

ENTRADA	NÚMERO DE ACCESO	DESCRIPCIÓN	TEJIDO	CIS-E	W_s	REFERENCIA
CALR3_HUMAN	9606.ENSPO0000269881	Proteína Calreticulina-3 funciona como chaperona independiente de lectina durante la espermatog	Testiculos	1	2.64	http://www.uniprot.org/uniprot/Q96L12
NP_115583.1	Q9H6X5-CS044_HUMAN	Proteína sin clasificar, interactúa con la proteína CBY2 durante el desarrollo testicular.	Testiculos	1	2.64	http://www.uniprot.org/uniprot/Q9H6X5
Q96HF8_HUMAN	P12036 - NFH_HUMAN	Neurofilamento polipeptídico, juega un importante papel en la maduración de los axones	Nervioso	1	2.64	http://www.uniprot.org/uniprot/P12036
NP_835223.1	Q5TH74 - STPG1_HUMAN	Metil-guanina inductor de la apoptosis en espermatozoides	Testiculos	2	2.63	Http://www.uniprot.org/uniprot/Q5TH74
CCD21_HUMAN	Q6P2H3 - CEP85_HUMAN	Proteína de centrosomas	Ubicuo	2	2.63	http://www.uniprot.org/uniprot/Q6P2H3#Q6P2H3-2
Q5Q9A2_HUMAN	Q5T013 - HYL_HUMAN	Hidroxipiruvato deshidrogenasa putativa	Ubicuo	2	2.63	http://www.uniprot.org/uniprot/Q5T013
Q56B4_HUMAN	Q8NH76 - Q56B4_HUMAN	Receptor olfativo	Nervioso	2	2.63	http://www.uniprot.org/uniprot/Q8NH76
FUT4_HUMAN	P22083 - FUT4_HUMAN	Alpha-(1,3)-fucosyltransferase 4	Immune	2	2.63	http://www.uniprot.org/uniprot/P22083
DDX6_HUMAN	P26196 - DDX6_HUMAN	Probable ATP-dependiente RNA helicase DDX6, involucrada en apoptosis	Apoptosis	2	2.63	http://www.uniprot.org/uniprot/P26196

Tabla. 5. Secuencias CARg putativas obtenidas por medio de las matrices construidas en base a las secuencias con un valor de ΔG_{bind} de $-3.9\text{e}+02$ kJ/mol. La primera columna de la tabla corresponde a la entrada con la que se localizó el gen, la segunda columna corresponde al número de acceso al cual se recurrió para explorarlo, la tercer columna corresponde a una descripción breve del gen, la cuarta columna corresponde al tejido en el que se expresa este gen, la siguiente columna (CIS-E) corresponde al número de veces que una CARgbox aparece en la sección regulatoria del gen, la sexta columna reporta el W_s calculado en el servidor RSAT y finalmente se incluye la referencia para que pueda consultarse la liga de donde se recabó la información presentada del gen. Esta lista cuenta con 90 genes identificados.

Q96HX2_HUMAN	Q9HCH0 - NCK5L_HUMAN	Nck-associated protein 5-like	Ubicuo	1	2.63	http://www.uniprot.org/uniprot/Q9HCH0
HXC8_HUMAN	P31273 - HXC8_HUMAN	Homeobox protein Hox-C8	Homeobox	1	2.63	http://www.uniprot.org/uniprot/P31273
ELK3_HUMAN	P41970 - ELK3_HUMAN	Regulador negativo que forma complejo ternario con SRF	Homeobox	1	2.63	http://www.uniprot.org/uniprot/P41970
Q5T911_HUMAN	Q5T911 - Q5T911_HUMAN	Factor de transcripción, cofactor de la ARN polimerasa II	FT	2	2.63	http://www.uniprot.org/uniprot/Q5T911
Q6PU16_HUMAN	Q6PJ77 - ZC3HE_HUMAN	Reconocimiento de secuencias poli-A en neuronas	Nervioso	2	2.63	http://www.uniprot.org/uniprot/Q6PJ77
ENST00000382039	A6NNK5 - A6NNK5_HUMAN	Tumor suppressor p53-binding protein 1	FT	2	2.63	http://www.uniprot.org/uniprot/A6NNK5
Q7Z2F6_HUMAN	Q7Z2F6 - ZN720_HUMAN	Proteína de unión a ADN	FT	1	2.63	http://www.uniprot.org/uniprot/Q7Z2F6
ATBF1_HUMAN	Q15911 - ZFH33_HUMAN	Zinc finger homeobox protein 3	Homeobox	2	2.63	http://www.uniprot.org/uniprot/Q15911
LOX12_HUMAN	P18054 - LOX12_HUMAN	Promueve la diferenciación de fibroblastos	Muscular	3	2.63	http://www.uniprot.org/uniprot/P18054
ANFY1_HUMAN	Q9P2R3 - ANFY1_HUMAN	Proteína controladora del tráfico intracelular	FT	1	2.63	http://www.uniprot.org/uniprot/Q9P2R3
MYH10_HUMAN	P35580 - MYH10_HUMAN	Miosina que estabiliza el colágeno tipo 1	Muscular	3	2.63	http://www.uniprot.org/uniprot/P35580
K1H8_HUMAN	O76015 - KRT38_HUMAN	Keratina del tipo cuticular	Muscular	1	2.63	http://www.uniprot.org/uniprot/O76015
NP_438170.1	Q9BXL6 - CAR14_HUMAN	Protege a las células contra la apoptosis	Apoptosis	2	2.63	http://www.uniprot.org/uniprot/Q9BXL6
PPIP2_HUMAN	Q9H939 - PPIP2_HUMAN	Promueve la formación de actina	Muscular	1	2.63	http://www.uniprot.org/uniprot/Q9H939
UBIQ_HUMAN	P0CCG47 - UBB_HUMAN	Promueve la endocitosis	Trafico	2	2.63	http://www.uniprot.org/uniprot/P0CCG47
TSH3_HUMAN	Q63HK5 - TSH3_HUMAN	Regula la formación de histonas	FT	1	2.63	http://www.uniprot.org/uniprot/Q63HK5
1433B_HUMAN	P31946 - 1433B_HUMAN	Factor de transcripción que ayuda a diferenciar osteocitos y juega un rol en la apoptosis	Oseo	8	2.63	http://www.uniprot.org/uniprot/P31946
SEM4C_HUMAN	Q9C0C4 - SEM4C_HUMAN	Receptor de membrana requerido para un desarrollo normal del cerebro	Nervioso	1	2.63	http://www.uniprot.org/uniprot/Q9C0C4
MARH7_HUMAN	Q9H992 - MARH7_HUMAN	Ubiquitina que participa en la formación de proteínas de desecho para promover endocitosis	Trafico	1	2.63	http://www.uniprot.org/uniprot/Q9H992
PDE11_HUMAN	Q9HCR9 - PDE11_HUMAN	regula la concentración de cAMP y cGMP, su ausencia se liga con la disfunción eréctil	Testículos	6	2.63	http://www.uniprot.org/uniprot/Q9HCR9
NP_057399.1	Q9UBP9 - GULP1_HUMAN	Proteína GULP1, participa en la fagocitosis durante la apoptosis	Trafico	2	2.63	http://www.uniprot.org/uniprot/Q9UBP9
Q8NHU2-6	Q8NHU2 - CFA61_HUMAN	Proteína 61 asociada a la formación del flagelo espermático	Testículos	3	2.63	http://www.uniprot.org/uniprot/Q8NHU2#Q8NHU2-6
NP_787110.2	P41235 - HNF4A_HUMAN	Factor 4-alpha del hepatocito	Hepatico	2	2.63	http://www.uniprot.org/uniprot/P41235
ITIH4_HUMAN	Q14624 - ITIH4_HUMAN	Inter-alpha-trypsin inhibitor heavy chain H4, envuelto en respuesta inflamatoria	Hepatico	1	2.63	http://www.uniprot.org/uniprot/Q14624
NPY2R_HUMAN	P49146 - NPY2R_HUMAN	Neuropeptide Y receptor type 2	Nervioso	2	2.63	http://www.uniprot.org/uniprot/P49146
GFRAL_HUMAN	Q6UXV0 - GFRAL_HUMAN	Receptor tipo ALFA GDNF implicado en la apoptosis neuronal.	Nervioso	1	2.63	http://www.uniprot.org/uniprot/Q6UXV0
HSF2_HUMAN	Q03933 - HSF2_HUMAN	Proteína activadora de factores del cheque térmico	FT	2	2.63	http://www.uniprot.org/uniprot/Q03933
NCOA2_HUMAN	NCOA2_HUMAN	Receptor nuclear responsable de la diferenciación de adipositos	Adiposo	1	2.63	http://www.uniprot.org/uniprot/NCOA2
FANCC_HUMAN	Q00597 - FANCC_HUMAN	Proteína FANCC envuelta en el mantenimiento y estabilidad del cromosoma	Cc	3	2.63	http://www.uniprot.org/uniprot/Q00597
Q5T535_HUMAN	Q7Z591 - AKNA_HUMAN	Factor de transcripción esencial en la diferenciación de linfocitos	Inmune	1	2.63	http://www.uniprot.org/uniprot/Q7Z591
GAGB1_HUMAN	O75459 - PAGE1_HUMAN	Antígeno P de la familia 1	Inmune	1	2.63	http://www.uniprot.org/uniprot/O75459
IL20_HUMAN	Q9NYY1 - IL20_HUMAN	Interleucina	Epidermis	3	2.61	www.uniprot.org/uniprot/Q9NYY1
P15529-11	P15529 - MCP_HUMAN	Cofactor de proteína de membrana CD46 relacionado con el reconocimiento durante la unión de gal	Testículos	1	2.61	http://www.uniprot.org/uniprot/P15529#P15529-11
Q6ZU20_HUMAN	Q2KJY2 - KIF26B_HUMAN	Proteína cinasa KIF26B	Riñón	2	2.61	http://www.uniprot.org/uniprot/Q6ZU20
Q5T098_HUMAN	Q5T098 - Q5T098_HUMAN	Utrofina, que diferencia el musculo	Muscular	2	2.61	http://www.uniprot.org/uniprot/Q5T098
NFL_HUMAN	P07196 - NFL_HUMAN	proteína constructora de neurofilamentos	Nervioso	2	2.61	http://www.uniprot.org/uniprot/P07196
Q7Z3N4_HUMAN	Q7Z3N4	Proteína hipotética DKFZp686C24272 ubicada en endosomas	Nervioso	1	2.61	http://www.uniprot.org/uniprot/Q7Z3N4
OR4F5_HUMAN	ENSG00000186092 - OR4F5	Receptor olfativo transmembranal	Nervioso	1	2.61	http://www.uniprot.org/uniprot/OR4F5
UB2J2_HUMAN	ENST00000349431 - Q8N2K1	Enzima Ubiquitina E2, J2	Ubicuo	1	2.61	http://www.uniprot.org/uniprot/UB2J2
NM_001303681	ENSM000780001955120-RM2	Sin clasificar	Sin Clasificar	1	2.61	http://www.uniprot.org/uniprot/NM_001303681
NADK_HUMAN	ENST00000341426	Cinasa localizada en fibroblastos	Conectivo	2	2.61	http://www.uniprot.org/uniprot/NADK
Q6PJ22_HUMAN	ENSG00000189301	- Q6PJ22Proteina sin identificar	Ubicuo	1	2.61	http://www.uniprot.org/uniprot/Q6PJ22
Q9UQV8_HUMAN	Q9UQV8 - Q9UQV8_HUMAN	Proteína RET involucrada en el desarrollo del tubo neural	Nervioso	1	2.61	http://www.uniprot.org/uniprot/Q9UQV8
RET_HUMAN	P07949 - RET_HUMAN	Proto-oncogen RET involucrado en la migración de células neuronales	Nervioso	1	2.61	http://www.uniprot.org/uniprot/P07949
Q6WN34-2	Q6WN34 - CRDL2_HUMAN	proteína CHRDL2 controla la formación del cartilago	Muscular	2	2.61	http://www.uniprot.org/uniprot/Q6WN34
VP33A_HUMAN	Q96AX1 - VP33A_HUMAN	Proteína de tráfico vesicular formadora de vacuolas VPS33A	Trafico	2	2.61	http://www.uniprot.org/uniprot/Q96AX1
DTD2_HUMAN	Q96FN9 - DTD2_HUMAN	Tirosina marcadora de vesículas	trafico	1	2.61	http://www.uniprot.org/uniprot/Q96FN9
DPF3_HUMAN	Q92784 - DPF3_HUMAN	factor mitótico para la formación del tubo neural	Nervioso	3	2.61	http://www.uniprot.org/uniprot/Q92784
ODO2_HUMAN	P36957 - ODO2_HUMAN	Proteína DLST mitocondrial	Adiposo	1	2.61	http://www.uniprot.org/uniprot/P36957
COR07_HUMAN	P57737 - COR07_HUMAN	Regulador de la actina-F, proteína Cronin-7	Trafico	2	2.61	http://www.uniprot.org/uniprot/P57737
Q13798_HUMAN	Q13798 - Q13798_HUMAN	Proteína de tráfico transmembranal	Trafico	1	2.61	http://www.uniprot.org/uniprot/Q13798
NP_689559.1	Q8N370 - LAT4_HUMAN	Proteína transportadora de aminoácidos neutros	trafico	3	2.61	http://www.uniprot.org/uniprot/NP_689559.1
Q8N4C8-3	Q8N4C8 - MINK1_HUMAN	Cinasa MINK1 envuelta en la transducción de señales en la neurona	Nervioso	2	2.61	http://www.uniprot.org/uniprot/Q8N4C8#Q8N4C8-3
CALR3_HUMAN	Q96L12 - CALR3_HUMAN	Proteína calreticulina-3 actúa como chaperona independiente de lectina durante la espermatogene	Testiculo	1	2.61	http://www.uniprot.org/uniprot/Q96L12
THBG_HUMAN	P05543 - THBG_HUMAN	proteína Serpina7 transportadora de la hormona tiroxina en sistema nervioso central	Nervioso	3	2.63	http://www.uniprot.org/uniprot/P05543
GAGB1_HUMAN	O75459 - PAGE1_HUMAN	Antígeno PAGE1	Inmune	1	2.63	http://www.uniprot.org/uniprot/O75459
Q5T535_HUMAN	Q7Z591 - AKNA_HUMAN	Factor de transcripción AKNA que activa la transcripción del receptor CD40 de los linfocitos	inmune	1	2.63	http://www.uniprot.org/uniprot/Q5T535
FANCC_HUMAN	Q00597 - FANCC_HUMAN	proteína FANCC, mantiene la estabilidad de los cromosomas	Cc	3	2.63	www.uniprot.org/uniprot/Q00597
NCOA2_HUMAN	Q15596 - NCOA2_HUMAN	Proteína NCOA2 que regula el metabolismo de glucosa en el tejido pardo	Adiposo	1	2.63	http://www.uniprot.org/uniprot/Q15596
Q8N157-3	Q8N157 - AH11_HUMAN	Juberina AH11 requerida para la diferenciación del tejido del sistema nervioso central	Nervioso	1	2.63	http://www.uniprot.org/uniprot/Q8N157#Q8N157-3
CO2_HUMAN	P06681 - CO2_HUMAN	Componente C2 que activa al factor C1 cuando surge la respuesta inmunitaria	Inmune	1	2.63	http://www.uniprot.org/uniprot/P06681
Q8N6L6_HUMAN	Q8N6L6 - Q8N6L6_HUMAN	Componente C2b que actúa al factor C1 cuando surge la respuesta inmunitaria	Inmune	1	2.63	http://www.uniprot.org/uniprot/Q8N6L6
NPY2R_HUMAN	P49146 - NPY2R_HUMAN	Neuropeptido receptor Y tipo 2 NPY2R	Nervio	1	2.63	http://www.uniprot.org/uniprot/P49146
Q14624-2	Q14624 - ITIH4_HUMAN	Participa en procesos inflamatorios y regenerativos del hígado	Hepatico	1	2.63	http://www.uniprot.org/uniprot/Q14624#Q14624-2
P41235-3	P41235 - HNF4A_HUMAN	Factor 4-alfa del hepatocito, controla la expresión genética	Hepatico	1	2.63	http://www.uniprot.org/uniprot/P41235#P41235-3
Q8NHU2	Q8NHU2 - CFA61_HUMAN	Proteína CFAP61 asociada al desarrollo del sistema nervioso	Nervioso	1	2.63	http://www.uniprot.org/uniprot/Q8NHU2#Q8NHU2-6
SEM4C_HUMAN	Q9C0C4 - SEM4C_HUMAN	Semaforina 4c requerida para el desarrollo normal del cerebro	Nervioso	1	2.63	http://www.uniprot.org/uniprot/Q9C0C4
NP_597812.1	Q9BY07 - S4A5_HUMAN	Cotransportador electrogenico de iones sodio y bicarbonato, mantiene el pH en hepatocitos	Hepatico	1	2.63	http://www.uniprot.org/uniprot/NP_597812.1
TSH3_HUMAN	Q63HK5 - TSH3_HUMAN	TF TSH3, regula el desarrollo de las neuronas envueltas en el control respiratorio	Nervioso	1	2.63	http://www.uniprot.org/uniprot/TSH3
Q92908	Q92908 - GATA6_HUMAN	Factor de transcripción GATA6 participa en la morfogenesis de hueso	Oseo	1	2.61	http://www.uniprot.org/uniprot/Q92908
PRAF1_HUMAN	Q8UI14 - PRAF1_HUMAN	Proteína receptora RABAC1 regula la formación de vesículas en el aparato de Golgi	Trafico	1	2.61	http://www.uniprot.org/uniprot/PRAF1
VISL1_HUMAN	P62760 - VISL1_HUMAN	Proteína VSNL1 proteína inhibidora de la fosforilación de rodopsina, controla el fluido cerebroespinal	Nervioso	8	2.61	http://www.uniprot.org/uniprot/P62760
NP_786951.1	Q7RTU1 - TCF23_HUMAN	Factor de transcripción 23 inhibidor de la miogenesis	Musculo	2	2.61	http://www.uniprot.org/uniprot/NP_786951.1
WNT7A_HUMAN	O00755 - WNT7A_HUMAN	Proteína WNT7A ayuda a diferenciar los conductos de Müller	Testículos	1	2.61	http://www.uniprot.org/uniprot/WNT7A
STAG1_HUMAN	Q8VWM7 - STAG1_HUMAN	Proteína STAG1 ayuda a cohesionar las cromátidas hermanas durante la replicación del ADN	Cc	2	2.61	http://www.uniprot.org/uniprot/Q8VWM7
PGFRA_HUMAN	P16234 - PGFRA_HUMAN	Factor de crecimiento PDGFRA se expresa durante el desarrollo embrionario y ayuda a diferenciar	Oseo	6	2.61	http://www.uniprot.org/uniprot/P16234
OSMR_HUMAN	Q99650 - OSMR_HUMAN	Subunidad beta del receptor OSMR, participa en la diferenciación de musculo	Musculo	6	2.61	http://www.uniprot.org/uniprot/Q99650
HCN1_HUMAN	O60741 - HCN1_HUMAN	Mediador de la respuesta al estímulo de la sensibilidad hacia lo agrio	Nervio	1	2.61	http://www.uniprot.org/uniprot/O60741
NP_115485.1	Q5XKR4 - OTP_HUMAN	Proteína homeobox OTP, participa en la diferenciación de las células del hipotálamo	Nervioso	3	2.61	http://www.uniprot.org/uniprot/Q5XKR4
IL4_HUMAN	P05112 - IL4_HUMAN	Interleucina 4	Inmune	3	2.61	http://www.uniprot.org/uniprot/P05112
DDX46_HUMAN	Q7L014 - DDX46_HUMAN	Helicasa DDX46 ayuda a la formación de los cuerpos de Cajal durante la replicación	Cc	3	2.61	http://www.uniprot.org/uniprot/Q7L014
NP_075252.2	P22455 - FGFR4_HUMAN	Factor de crecimiento del fibroblasto FGFR4	Digestivo	1	2.61	http://www.uniprot.org/uniprot/P22455

Tabla. 6. Secuencias CARg putativas obtenidas por medio de las matrices construidas en base a las secuencias con un valor de ΔG_{bind} de $-4.0e+02$ kJ/mol. La primer columna de la tabla corresponde a la entrada con la que se localizó el gen, la segunda columna corresponde al número de acceso al cual se recurrió para explorarlo, la tercer columna corresponde a una descripción breve del gen, la cuarta columna corresponde al tejido en el que se expresa este gen, la siguiente columna (CIS-E) corresponde al número de veces que una CARg aparece en la sección regulatoria del gen, la sexta columna reporta el W_s calculado en el servidor RSAT y finalmente se incluye la referencia para que pueda consultarse la liga de donde se recabó la información presentada del gen. Esta lista cuenta con 205 genes identificados.

ENTRADA	NÚMERO DE ACCESO	DESCRIPCIÓN	TEJIDO	CIS-E	Ws	REFERENCIA
VAMP3_HUMAN	Q15836 - VAMP3_HUMAN	Proteína VAMP3 envuelta en el transporte vesicular	Trafico	1	3.19	http://www.uniprot.org/uniprot/Q15836
Q15381_HUMAN	Q15381 - Q15381_HUMAN	Proteína de reconocimiento al cromosoma Y	Testículo	110	3.19	http://www.uniprot.org/uniprot/Q15381
VCY2_HUMAN	O14599 - VCY2_HUMAN	Proteína BP2 asociada al espermatozoides	Testículo	9	3.19	http://www.uniprot.org/uniprot/O14599
Q658T2_HUMAN	ENSP00000384573	Proteína DAZ1 reguladora de la espermatozoides	Testículo	54	3.19	http://pax-db.org/#protein/Q658T2
Q86SG3-2	Q86SG3 - DAZ4_HUMAN	Proteína DAZ4 reguladora de la espermatozoides	Testículo	12	3.19	http://www.uniprot.org/uniprot/Q86SG3#Q86SG3-2
PARK7_HUMAN	Q99497 - PARK7_HUMAN	Proteína DJ-1 regula el estrés oxidativo en las neuronas dopaminérgicas	Nervioso	10	3.19	http://www.uniprot.org/uniprot/Q99497
SPSB1_HUMAN	Q96BD6 - SPSB1_HUMAN	Proteína SOCS 1, media la ubiquitinación y la posterior degradación del proteosoma	Trafico	1	3.19	http://www.uniprot.org/uniprot/Q96BD6
NP_115691.1	Q9BSK2 - S2533_HUMAN	Proteína acareadora de soluto familia 25 miembro 33, participa en la fagocitosis de células dendríticas	Immune	4	3.19	http://www.uniprot.org/uniprot/Q9BSK2
Q13117-2	Q13117 - DAZ2_HUMAN	Proteína DAZ2 reguladora de la espermatozoides	Testículo	4	3.19	http://www.uniprot.org/uniprot/Q13117#Q13117-2
NP_065097.2	Q9NR90 - DAZ3_HUMAN	Proteína DAZ3 reguladora de la espermatozoides	Testículo	4	3.19	http://www.uniprot.org/uniprot/Q9NR90
NP_689798.1	Q15415 - RBY1F_HUMAN	Proteína de unión a RNA del cromosoma Y, familia 1 miembro F1J, participa en espermatozoides	Testículo	40	3.19	http://www.uniprot.org/uniprot/Q15415
Q6NSB5_HUMAN	P0DUJ3 - RBY1A_HUMAN	Proteína de unión a RNA del cromosoma Y, familia 1 miembro A1, participa en espermatozoides	Testículo	4	3.19	http://www.uniprot.org/uniprot/P0DUJ3
NP_001006118.2	A6N6Q0 - RBY1E_HUMAN	Proteína de unión a RNA del cromosoma Y, familia 1 miembro E, participa en espermatozoides	Testículo	12	3.19	http://www.uniprot.org/uniprot/A6N6Q0
NP_001006121.1	A6NDE4 - RBY1B_HUMAN	Proteína de unión a RNA del cromosoma Y, familia 1 miembro B, participa en espermatozoides	Testículo	12	3.19	http://www.uniprot.org/uniprot/A6NDE4
Q13385_HUMAN	Q13385 - Q13385_HUMAN	Proteína de unión a RNA que participa en la espermatozoides	Testículo	11	3.19	http://www.uniprot.org/uniprot/Q13385
IF1AY_HUMAN	O14602 - IF1AY_HUMAN	Iniciador de la traducción	Testículo	6	3.19	http://www.uniprot.org/uniprot/O14602
Q9BY66-2	Q9BY66 - KDM5D_HUMAN	Dimetilasa específica de lisina 5D, metila durante la espermatozoides	Testículo	9	3.19	http://www.uniprot.org/uniprot/Q9BY66#Q9BY66-2
CY15A_HUMAN	Q9BZA5 - TXNG2_HUMAN	Putativa gama_taxina, regula el tráfico vesicular	Trafico	1	3.19	http://www.uniprot.org/uniprot/Q9BZA5
CD24_HUMAN	P25063 - CD24_HUMAN	Transductor de señales CD24 que modula la respuesta de los linfocitos B	Immune	9	3.19	http://www.uniprot.org/uniprot/P25063
HSFY1_HUMAN	Q9JL16 - HSFY1_HUMAN	Factor de transcripción que regula la diferenciación de las células de Sertoli	Testículo	140	3.19	http://www.uniprot.org/uniprot/Q9JL16
CDY2_HUMAN	Q9Y6F7 - CDY2_HUMAN	Histona acetil transferasa específica de cromosoma 2, se activa durante la espermatozoides	Testículos	27	3.19	http://www.uniprot.org/uniprot/Q9Y6F7
Q8N5B6_HUMAN	Q8NFZ3 - NLGN1_HUMAN	Neuroiglina 4, ayuda a las interacciones neurona_neurona	Nervioso	3	3.19	http://www.uniprot.org/uniprot/Q8NFZ3
ENST00000328089	Q96BD6 - SPSB1_HUMAN	SOCS box 1, responsable de la ubiquitinación durante la diferenciación del tejido nervioso	Nervioso	1	3.19	http://www.uniprot.org/uniprot/Q96BD6
NLGN1_HUMAN	Q8NFZ3 - NLGN1_HUMAN	Putativa Neuroiglina	Nervioso	34	3.19	http://www.uniprot.org/uniprot/Q8NFZ3
ENST00000382868	A6NML8 - A6NML8_HUMAN	Neuroiglina 4Y	Nerviosos	6	3.19	http://www.uniprot.org/uniprot/A6NML8
Q8NFZ3-2	Q8NFZ3 - NLGN1_HUMAN	Neuroiglina 4Y Isoforma 2	Nerviosos	4	3.19	http://www.uniprot.org/uniprot/Q8NFZ3#Q8NFZ3-2
VCY1_HUMAN	O14598 - VCY1_HUMAN	Proteína VCY1, ayuda a mediar el proceso de espermatozoides	Testículos	10	3.19	http://www.uniprot.org/uniprot/O14598
TYB4Y_HUMAN	O14604 - TYB4Y_HUMAN	Timosina beta 4, ayuda a estructurar el citoesqueleto de las neuronas	Nervio	9	3.19	http://www.uniprot.org/uniprot/O14604
ENST00000382893	A6NCS7 - A6NCS7_HUMAN	Histona dimetilasa UTY, trabaja en el cromosoma Y durante la masculinización	Testículo	2	3.19	http://www.uniprot.org/uniprot/A6NCS7
UTY_HUMAN	O14607 - UTY_HUMAN	Histona dimetilasa, cataliza trimetilaciones en el cromosoma Y de la histona 3	Testículo	6	3.19	http://www.uniprot.org/uniprot/O14607
DDX3Y_HUMAN	O15523 - DDX3Y_HUMAN	Helicasa DDX3Y, modula la espermatozoides	Testículo	54	3.19	http://www.uniprot.org/uniprot/O15523
Q13381_HUMAN	Q13381 - Q13381_HUMAN	Factor de transcripción	Testículo	15	3.19	http://www.uniprot.org/uniprot/Q13381
Q01534-2	Q01534 - TSPY1_HUMAN	Proteína codificadora en testículos	Testículo	9	3.19	http://www.uniprot.org/uniprot/Q01534#Q01534-2
ENST00000383020	A6NDE4 - RBY1B_HUMAN	Factor de transcripción RBMY1B, regulador de la espermatozoides	Testículos	18	3.19	http://www.uniprot.org/uniprot/A6NDE4
PRKY_HUMAN	Q43930 - PRKY_HUMAN	Proteína Cíncas PRKY, inmiscuida en la transsexualización en hembras	Testículos	6	3.19	http://www.uniprot.org/uniprot/Q43930
TBL1Y_HUMAN	Q9BQ87 - TBL1Y_HUMAN	Proteína TBL1Y, regula la degradación del proteosoma del complejo 19S	Cc	20	3.19	http://www.uniprot.org/uniprot/Q9BQ87
AMELY_HUMAN	Q99218 - AMELY_HUMAN	Amelogenina isoforma Y, juega un papel importante en la mineralización de los dientes	Oseo	18	3.19	http://www.uniprot.org/uniprot/Q99218
ENST00000383042	A6NGT6 - A6NGT6_HUMAN	Proteína codificadora en testículos que promueva la proliferación de espermatozoides	Testículo	4	3.19	http://www.uniprot.org/uniprot/A6NGT6
Q9BZA8-3	Q9BZA8 - PC1Y_HUMAN	Proteína codificadora en testículos que promueva la proliferación de espermatozoides	Testículo	15	3.19	http://www.uniprot.org/uniprot/Q9BZA8#Q9BZA8-3
TF2LY_HUMAN	Q8JUE0 - TF2LY_HUMAN	Proteína Homeobox TGIF2L1, reguladora de la transcripción en tejido testicular	Testículo	9	3.19	http://www.uniprot.org/uniprot/Q8JUE0
SRY_HUMAN	Q05066 - SRY_HUMAN	Regulador transcripcional esencial en la diferenciación del sexo en machos	Testículos	28	3.19	http://www.uniprot.org/uniprot/Q05066
IL9R_HUMAN	Q01113 - IL9R_HUMAN	Receptor 9 de interleucina, participa en la formación de vesículas	Trafico	2	3.19	http://www.uniprot.org/uniprot/Q01113
P51809-3	P51809 - VAMP7_HUMAN	Proteína asociada vesicular de membrana 7	Cc	3	3.19	http://www.uniprot.org/uniprot/P51809#P51809-3
SPY3_HUMAN	O43610 - SPY3_HUMAN	Factor de crecimiento de fibroblastos	Piel	9	3.19	http://www.uniprot.org/uniprot/O43610
Q9NVH6-2	Q9NVH6 - TMLH_HUMAN	Trimetil Lisina Dioxigenasa mitocondrial, asociada al desarrollo del autismo	Nervio	3	3.19	http://www.uniprot.org/uniprot/Q9NVH6#Q9NVH6-2
CLIC2_HUMAN	O15247 - CLIC2_HUMAN	Canal de cloro con actividad dependiente del pH, su ausencia esta relacionada con el retraso mental	Nervio	1	3.19	http://www.uniprot.org/uniprot/O15247
MTCPA_HUMAN	P56277 - CMC4_HUMAN	Proteína 4 de unión a motivos Cx9C, participa en la proliferación de linfocitos T	Cc	15	3.19	http://www.uniprot.org/uniprot/P56277
Q9BWH2_HUMAN	Q9BWH2 - FUND2_HUMAN	Protooncogen FUND2	Oseo	4	3.19	http://www.uniprot.org/uniprot/Q9BWH2
NP_063916.1	P00451 - F8B_HUMAN	Factor de coagulación 4	Sanguíneo	4	3.19	http://www.uniprot.org/uniprot/P00451
ENST00000389529	A6NGZ8 - SMIM9_HUMAN	Proteína integral de membrana 9	Ovario	3	3.19	http://www.uniprot.org/uniprot/A6NGZ8
NP_542179.1	Q8MWW8 - GAB3_HUMAN	Proteína de diferenciación de macrófagos	Immune	1	3.19	http://www.uniprot.org/uniprot/Q8MWW8
CTG1B_HUMAN	P78358 - CTG1B_HUMAN	Antígeno testicular CTAG1A	Testículo	3	3.19	http://www.uniprot.org/uniprot/P78358
Q9NY13_HUMAN	Q8MWW8 - GAB3_HUMAN	Proteína GRB-2 asociada a la diferenciación de macrófagos	Immune	3	3.19	http://www.uniprot.org/uniprot/Q9NY13
SC6A8_HUMAN	P48029 - SC6A8_HUMAN	Proteína reguladora de creatinina en cerebro	Cerebro	1	3.19	http://www.uniprot.org/uniprot/P48029
Q9JUL41-2	Q9JUL41 - PNMMA3_HUMAN	Antígeno paraneoplásico MNA3	Immune	1	3.19	http://www.uniprot.org/uniprot/Q9JUL41#Q9JUL41-2
NP_443158.1	Q96PV4 - PNMMA5_HUMAN	Antígeno paraneoplásico PNMMA5	Immune	1	3.19	http://www.uniprot.org/uniprot/Q96PV4
GBRT_HUMAN	Q9UN88 - GBRT_HUMAN	Sub-unidad teta del receptor gama-acido aminobutírico, inhibe la señal de GABA en neurotransmisores	Nervioso	5	3.19	http://www.uniprot.org/uniprot/Q9UN88
GBRA3_HUMAN	P34903 - GBRA3_HUMAN	Sub-unidad alfa3 del receptor gama-acido aminobutírico, inhibe la señal de GABA en neurotransmisores	Nervioso	6	3.19	http://www.uniprot.org/uniprot/P34903
MAGAA_HUMAN	P43363 - MAGAA_HUMAN	Su expresión normal se da en espermatozoides y espermatozoides	Testículo	9	3.19	http://www.uniprot.org/uniprot/P43363
Q6PCD2_HUMAN	P78334 - GBRE_HUMAN	Sub-unidad epsilon del receptor gama-acido aminobutírico, inhibe la señal de GABA en neurotransmisores	Nervioso	1	3.19	http://www.uniprot.org/uniprot/P78334
MAGAA4_HUMAN	P43358 - MAGAA4_HUMAN	Antígeno asociado a melanoma, solo aparece en cancer perro Su expresión normal se da en espermatozoides	Testículo	14	3.19	http://www.uniprot.org/uniprot/P43358
CNGA2_HUMAN	Q16280 - CNGA2_HUMAN	Transductor de las señales olfatorias a neuronas	Nerviosos	2	3.19	http://www.uniprot.org/uniprot/Q16280
FATE1_HUMAN	Q969F0 - FATE1_HUMAN	Proteína transmembranal expresada en testículos adultos y en formación fetal	Testículo	2	3.19	http://www.uniprot.org/uniprot/Q969F0
ENST00000370361	Q3ZAQ7 - VMA21_HUMAN	Proteína vacuolar	Trafico	1	3.19	http://www.uniprot.org/uniprot/Q3ZAQ7
MTR1L_HUMAN	Q13585 - MTR1L_HUMAN	Receptor de melatonina	Nervioso	2	3.19	http://www.uniprot.org/uniprot/Q13585
Q8TAW2_HUMAN	Q8TC22 - C99L2_HUMAN	Antígeno CD99 expresado en linfocitos y en la respuesta inmune	Immune	1	3.19	http://www.uniprot.org/uniprot/Q8TC22
MTMR1_HUMAN	Q13613 - MTMR1_HUMAN	Proteína miotubular	Muscular	2	3.19	http://www.uniprot.org/uniprot/Q13613
MTM1_HUMAN	Q13496 - MTM1_HUMAN	Miotubularina requerida en la miogénesis del músculo esquelético	Muscular	1	3.19	http://www.uniprot.org/uniprot/Q13496
CG1_HUMAN	Q13495 - MAM1D1_HUMAN	Proteína MAM1D1 requerida en la formación de la uretra	Testículo	15	3.19	http://www.uniprot.org/uniprot/Q13495

MAGA9_HUMAN	P43362 - MAGA9_HUMAN	Antígeno 9 asociado a melanoma y aparece normalmente en testículos	Testículo	18	3.19	http://www.uniprot.org/uniprot/P43362
MAGAB_HUMAN	P43364 - MAGAB_HUMAN	Antígeno 11 asociado a melanoma y aparece normalmente en testículos	Testículo	4	3.19	http://www.uniprot.org/uniprot/P43364
Q7Z2F7_HUMAN	P51816 - AFF2_HUMAN	Proteína 4F4/FMR2 miembro 2, en su ausencia se expresan problemas del lenguaje y la lectura	Nervioso	9	3.19	http://www.uniprot.org/uniprot/P51816
FMR1N_HUMAN	Q8N0W7 - FMR1N_HUMAN	Proteína que en su ausencia se presentan problemas de retardo mental	Nervioso	4	3.19	http://www.uniprot.org/uniprot/Q8N0W7
CX001_HUMAN	Q96002 - TM257_HUMAN	Proteína transmembranal 257, se expresa en la zona del hipocampo	Nervioso	4	3.19	http://www.uniprot.org/uniprot/Q96002
SLIK2_HUMAN	Q9H156 - SLIK2_HUMAN	Proteína supresora del crecimiento neuronal	Nervioso	12	3.19	http://www.uniprot.org/uniprot/Q9H156
NP_001009614.1	Q5VSR9 - SPXN1_HUMAN	Proteína espermática asociada con los núcleos en el cromosoma X	Testículo	1	3.19	http://www.uniprot.org/uniprot/Q5VSR9
UE2NL_HUMAN	Q5JXB2 - UE2NL_HUMAN	Enzima ubiquitina conjugada E2 expresada en epidídimo	Testículo	30	3.19	http://www.uniprot.org/uniprot/Q5JXB2
SLIK4_HUMAN	Q81W52 - SLIK4_HUMAN	Proteína 4 NTRK supresora del crecimiento neuronal	Nervioso	6	3.19	http://www.uniprot.org/uniprot/Q81W52
MAGC2_HUMAN	Q8UBF1 - MAGC2_HUMAN	Antígeno C2 asociado a melanoma y desarrollo de testículos	Testículo	2	3.19	http://www.uniprot.org/uniprot/Q8UBF1
NP_803251.1	Q8TD91 - MAGC3_HUMAN	Antígeno C3 asociado a melanoma y desarrollo de testículos	Testículo	3	3.19	http://www.uniprot.org/uniprot/Q8TD91
SPNXD_HUMAN	Q9BXN6 - SPNXD_HUMAN	Proteína espermática del cromosoma D	Testículo	3	3.19	http://www.uniprot.org/uniprot/Q9BXN6
LDOC1_HUMAN	Q95751 - LDOC1_HUMAN	Proteína LDOC1, asociada a desarrollo de cáncer cerebral	Nervioso	9	3.19	http://www.uniprot.org/uniprot/Q95751
NP_004056.2	P51861 - CDR1_HUMAN	Antígeno relacionado a la degeneración del cerebelo	Nervioso	27	3.19	http://www.uniprot.org/uniprot/P51861
Q8NB49-2	Q8NB49 - AT11C_HUMAN	Proteína transportadora de fosfolípidos implicada en la formación de vesículas	Trafico	3	3.19	http://www.uniprot.org/uniprot/Q8NB49-2
P10911-4	P10911 - MCF2_HUMAN	Protooncogen DBL isoforma 1 en cerebelo	Nervioso	5	3.19	http://www.uniprot.org/uniprot/P10911-4
FA9_HUMAN	P00740 - FA9_HUMAN	Factor de coagulación IX	Sanguíneo	2	3.19	http://www.uniprot.org/uniprot/P00740
FGF13_HUMAN	Q92913 - FGF13_HUMAN	Factor de crecimiento del fibroblasto 13	Nervioso	3	3.19	http://www.uniprot.org/uniprot/Q92913
OR4F5_HUMAN	Q8NH21 - OR4F5_HUMAN	Receptor olfatorio 4F5	Nervioso	15	3.19	http://www.uniprot.org/uniprot/Q8NH21
UBP26_HUMAN	Q9BXU7 - UBP26_HUMAN	Hidrolasa 26, se expresa en espermatogonias	Testículos	13	3.19	http://www.uniprot.org/uniprot/Q9BXU7
Q96MM7-2	Q96MM7 - H6S72_HUMAN	Geparin-sulfotransferasa	Sanguíneo	12	3.19	http://www.uniprot.org/uniprot/Q96MM7-2
Q8IUR4_HUMAN	Q9NUK0 - MBNL3_HUMAN	Proteína de unión a genes musculares Muscletin-like protein 3	Muscular	4	3.19	http://www.uniprot.org/uniprot/Q9NUK0
RAP2C_HUMAN	Q9Y3L5 - RAP2C_HUMAN	Ras-related protein Rap-2c lleva su función principal en músculo esquelético	Ubicuo	12	3.19	http://www.uniprot.org/uniprot/Q9Y3L5
ENST00000370879	X6R7S7 - X6R7S7_HUMAN	FERM domain-containing protein 7	Nervios	1	3.19	http://www.uniprot.org/uniprot/X6R7S7
NP_001035917.1	Q9P289 - STK26_HUMAN	Modulador de la apoptosis Serine/threonine-protein kinase 26	Cc	3	3.19	http://www.uniprot.org/uniprot/Q9P289
O13H1_HUMAN	Q8NG32 - O13H1_HUMAN	Receptor olfativo OR13H1	Nervioso	1	3.19	http://www.uniprot.org/uniprot/Q8NG32
NP_991402.1	Q8N6C5 - GSF1_HUMAN	Immunoglobulin superfamily member 1, se expresa principalmente en testículos	Testículo	9	3.19	http://www.uniprot.org/uniprot/Q8N6C5
Q6ZRI8-2	Q6ZRI8 - RHG36_HUMAN	Rho GTPase-activating protein 36	Ubicuo	2	3.19	http://www.uniprot.org/uniprot/Q6ZRI8-2
COVA1_HUMAN	Q16206 - ENOX2_HUMAN	Ecto-NOX disulfide-thiol exchanger 2, formador de vesículas	Trafico	5	3.19	http://www.uniprot.org/uniprot/Q16206
FA45B_HUMAN	Q8NSW5 - FA45B_HUMAN	Protein FAM45B, se expresa en pituitaria	Nervioso	4	3.19	http://www.uniprot.org/uniprot/Q8NSW5
Q95258-3	Q95258 - UCP5_HUMAN	Brain mitochondrial carrier protein 1, expresión exclusiva en células cerebrales	Nervioso	3	3.19	http://www.uniprot.org/uniprot/Q95258-3
Q9UJ2_HUMAN	Q9UJ2 - Q9UJ2_HUMAN	Zinc finger protein 280C	Ubicuo	3	3.19	http://www.uniprot.org/uniprot/Q9UJ2
Q95831-2	Q95831 - AIFM1_HUMAN	Apoptosis-inducing factor 1, mitochondria	Muscular	3	3.19	http://www.uniprot.org/uniprot/Q95831-2
RB33A_HUMAN	Q14088 - RB33A_HUMAN	Ras-related protein Rab-33A, expresado en linfocitos	Immune	3	3.19	http://www.uniprot.org/uniprot/Q14088
Q9BVJ6-2	Q9BVJ6 - UT14A_HUMAN	U3 small nucleolar RNA-associated protein 14 homolog A	Ubicuo	2	3.19	http://www.uniprot.org/uniprot/Q9BVJ6-2
ZDH9C_HUMAN	Q9Y397 - ZDH9C_HUMAN	Palmitoyltransferase ZDH9C9	Ubicuo	8	3.19	http://www.uniprot.org/uniprot/Q9Y397
ENST00000356892	O75995 - SASH3_HUMAN	SAM and SH3 domain-containing protein 3, expresado en linfocitos	Immune	2	3.19	http://www.uniprot.org/uniprot/O75995
Q01968-2	Q01968 - OCRL_HUMAN	Inositol polyphosphate 5-phosphatase OCRL-1	Ubicuo	10	3.19	http://www.uniprot.org/uniprot/Q01968-2
P28370-2	P28370 - SMC1A_HUMAN	Probable global transcription activator SNF2L1, potencia el crecimiento de las neuronas	Nervioso	15	3.19	http://www.uniprot.org/uniprot/P28370-2
TEN1_HUMAN	Q9UKZ4 - TEN1_HUMAN	Teurina-1, Promueve las conexiones del sistema nervioso	Nervioso	5	3.19	http://www.uniprot.org/uniprot/Q9UKZ4
O60880-5	O60880 - SH2A_HUMAN	SH2 domain-containing protein 1A, receptora de neurotrofinas	Nervioso	3	3.19	http://www.uniprot.org/uniprot/O60880-5
STAG2_HUMAN	Q8N3U4 - STAG2_HUMAN	Cohesin subunit SA-2, se requiere para la formación de las cromátidas hermanas	Cc	18	3.19	http://www.uniprot.org/uniprot/Q8N3U4
ENST00000245838	Q8N1Z7 - THOC2_HUMAN	THO complex subunit 2, necesario en el splicing.	Ubicuo	9	3.19	http://www.uniprot.org/uniprot/Q8N1Z7
GRIA3_HUMAN	P42263 - GRIA3_HUMAN	Glutamate receptor 3, juega un rol importante en la excitación de las neuronas	Nervioso	16	3.19	http://www.uniprot.org/uniprot/P42263
CT47_HUMAN	Q5JQC4 - CT47A_HUMAN	Antígeno 47A de cancer testicular	immune	33	3.19	http://www.uniprot.org/uniprot/Q5JQC4
NP_054779.1	Q9JULC - MCTS1_HUMAN	Malignant T-cell-amplified sequence 1	Immune	2	3.19	http://www.uniprot.org/uniprot/Q9JULC
Q13620-2	Q13620 - CUL4B_HUMAN	Cullin-4B	Nervioso	5	3.19	http://www.uniprot.org/uniprot/Q13620-2
P13473-2	P13473 - LAMP2_HUMAN	Lysosome-associated membrane glycoprotein 2	Trafico	4	3.19	http://www.uniprot.org/uniprot/P13473-2
Q9UN42-2	Q9UN42 - AT1B4_HUMAN	Protein ATP1B4, coregulador transcripcional que interviene en el desarrollo del musculo	Muscular	1	3.19	http://www.uniprot.org/uniprot/Q9UN42-2
PEPP1_HUMAN	Q8NHV9 - RHXF1_HUMAN	Rhox homeobox family member 1	Testículo	5	3.19	http://www.uniprot.org/uniprot/Q8NHV9
NKAP_HUMAN	Q8NF57 - NKAP_HUMAN	NF-kappa-B-activating protein	Immune	2	3.19	http://www.uniprot.org/uniprot/Q8NF57
NKRF_HUMAN	O15226 - NKRF_HUMAN	NF-kappa-B-repressing factor	Immune	5	3.16	http://www.uniprot.org/uniprot/O15226
3UBE2A_HUMAN	P49459 - UBE2A_HUMAN	Ubiquitin-conjugating enzyme E2 A	Nervioso	3	3.16	http://www.uniprot.org/uniprot/P49459
PGRCl_HUMAN	O00264 - PGRCl_HUMAN	Membrane-associated progesterone receptor component 1	Higado	1	3.16	http://www.uniprot.org/uniprot/O00264
Q6ZV70_HUMAN	Q6ZV70 - Q6ZV70_HUMAN	NTERLEUKIN-13 RECEPTOR ALPHA-1 CHAIN (IL-13R-ALPHA-1) (IL-13RA-1)	Immune	3	3.16	http://www.uniprot.org/uniprot/Q6ZV70
NP_653259.3	Q5JSL3 - DOC11_HUMAN	Dedicator of cytokinesis protein 11	higado	1	3.16	http://www.uniprot.org/uniprot/Q5JSL3
Q5JSH3-2	Q5JSH3 - WDR44_HUMAN	WD repeat-containing protein 44	Trafico	3	3.16	http://www.uniprot.org/uniprot/Q5JSH3-2
KLH13_HUMAN	Q9P2N7 - KLH13_HUMAN	Kelch-like protein 13	Cc	48	3.16	http://www.uniprot.org/uniprot/Q9P2N7
NP_001017978.1	Q5H943 - KKL1C1_HUMAN	Kita-kyushu lung cancer antigen 1	Immune	3	3.16	http://www.uniprot.org/uniprot/Q5H943
S6A14_HUMAN	Q9UN76 - S6A14_HUMAN	Sodium- and chloride-dependent neutral and basic amino acid transporter B(0+)	Ubicuo	2	3.16	http://www.uniprot.org/uniprot/Q9UN76
AGTR2_HUMAN	P50052 - AGTR2_HUMAN	Type-2 angiotensin II receptor	Nervioso	4	3.16	http://www.uniprot.org/uniprot/P50052
PLST_HUMAN	P13797 - PLST_HUMAN	Plastin-3	Oseo	15	3.16	http://www.uniprot.org/uniprot/P13797
NP_057467.1	Q9P127 - LUZP4_HUMAN	Leucine zipper protein 4	Testículo	2	3.16	http://www.uniprot.org/uniprot/Q9P127
CX055_HUMAN	Q8N7X1 - RMXL3_HUMAN	RNA-binding motif protein, X-linked-like-3	Ubicuo	7	3.16	http://www.uniprot.org/uniprot/Q8N7X1
I13R2_HUMAN	Q14627 - I13R2_HUMAN	Interleukin-13 receptor subunit alpha-2	Immune	6	3.16	http://www.uniprot.org/uniprot/Q14627
5HT2C_HUMAN	P28335 - 5HT2C_HUMAN	5-hydroxytryptamine receptor 2C	Nervioso	15	3.16	http://www.uniprot.org/uniprot/P28335
ENST00000371958	A6NP16 - A6NP16_HUMAN	Angiotensin	Vascular	4	3.16	http://www.uniprot.org/uniprot/A6NP16
LHPL1_HUMAN	Q86W10 - LHPL1_HUMAN	Lipoma HMGIC fusion partner-like 1 protein	Ubicuo	6	3.16	http://www.uniprot.org/uniprot/Q86W10
ZCH16_HUMAN	Q6ZR62 - ZCH16_HUMAN	Zinc finger CCHC domain-containing protein 16	Ubicuo	5	3.16	http://www.uniprot.org/uniprot/Q6ZR62
TRPC5_HUMAN	Q9JL62 - TRPC5_HUMAN	Short transient receptor potential channel 5	Nervioso	1	3.16	http://www.uniprot.org/uniprot/Q9JL62
G7281_HUMAN	Q9NP73 - ALG13_HUMAN	Putative bifunctional UDP-N-acetylglucosamine transferase and deubiquitinase ALG13	nervioso	1	3.16	http://www.uniprot.org/uniprot/Q9NP73
DCX_HUMAN	O43602 - DCX_HUMAN	Neuronal migration protein doublecortin	Nervioso	2	3.16	http://www.uniprot.org/uniprot/O43602
CAN6_HUMAN	Q9Y6Q1 - CAN6_HUMAN	Calpain-6	Placenta	2	3.16	http://www.uniprot.org/uniprot/Q9Y6Q1
PAK3_HUMAN	O75914 - PAK3_HUMAN	Serine/threonine-protein kinase PAK 3	Nervioso	18	3.16	http://www.uniprot.org/uniprot/O75914
CRDL1_HUMAN	Q9BU40 - CRDL1_HUMAN	Chordin-like protein 1	Nervioso	3	3.16	http://www.uniprot.org/uniprot/Q9BU40
RGAG1_HUMAN	Q8NET4 - RGAG1_HUMAN	Retrotransposon gag domain-containing protein 1	Nervioso	4	3.16	http://www.uniprot.org/uniprot/Q8NET4
Q6ZNC2_HUMAN	Q6ZNC2 - Q6ZNC2_HUMAN	Receptor del sabor	Nervioso	2	3.16	http://www.uniprot.org/uniprot/Q6ZNC2
Q9Y4X0-3	Q9Y4X0 - AMMR1_HUMAN	AMME syndrome candidate gene 1 protein	Nervioso	7	3.16	http://www.uniprot.org/uniprot/Q9Y4X0-3
AMER1_HUMAN	Q5JTC6 - AMER1_HUMAN	APC membrane recruitment protein 1	ubicuo	7	3.16	http://www.uniprot.org/uniprot/Q5JTC6

Q9P119_HUMAN	Q9P119 - Q9P119_HUMAN	Proteína que ayuda a la diferenciación de los hepatocitos en el feto	Higado	7	3.16	http://www.uniprot.org/uniprot/Q9P119
O60488-2	O60488 - ACSL4_HUMAN	Long-chain-fatty-acid-CoA ligase 4	Nervioso	24	3.16	http://www.uniprot.org/uniprot/O60488#O60488-2
Q9NPJ8-2	Q9NPJ8 - NXT2_HUMAN	NTF2-related export protein 2	trafico	2	3.16	http://www.uniprot.org/uniprot/Q9NPJ8#Q9NPJ8-2
GUC2F_HUMAN	P51841 - GUC2F_HUMAN	Retinal guanylyl cyclase 2	Nervioso	5	3.16	http://www.uniprot.org/uniprot/P51841
NP_003695.1	O14654 - IRS4_HUMAN	Insulin receptor substrate 4	Muscular	30	3.16	http://www.uniprot.org/uniprot/O14654
Q14031-2	Q14031 - CO4A6_HUMAN	Collagen alpha-6(IV) chain	higado	3	3.16	http://www.uniprot.org/uniprot/Q14031#Q14031-2
ENST00000372232	Q8WYNO - ATG4A_HUMAN	Cysteine protease ATG4A	Muscular	6	3.16	http://www.uniprot.org/uniprot/Q8WYNO
TX13B_HUMAN	Q9BXU2 - TX13B_HUMAN	Testis-expressed sequence 13B protein	Testiculo	1	3.16	http://www.uniprot.org/uniprot/Q9BXU2
Q9UUV3-2	Q9UUV3 - TRIM1_HUMAN	Probable E3 ubiquitin-protein ligase MID2	Testiculo	2	3.16	http://www.uniprot.org/uniprot/Q9UUV3#Q9UUV3-2
T22D3_HUMAN	Q99576 - T22D3_HUMAN	TSC22 domain family protein 3	Immune	4	3.16	http://www.uniprot.org/uniprot/Q99576
FRPD3_HUMAN	Q5JV73 - FRPD3_HUMAN	FERM and PDZ domain-containing protein 3	Nervioso	1	3.16	http://www.uniprot.org/uniprot/Q5JV73
MYCL2_HUMAN	P12525 - MYCP1_HUMAN	Putative myc-like protein MYCLP1	Testiculo	1	3.16	http://www.uniprot.org/uniprot/P12525
RBM41_HUMAN	Q96I25 - RBM41_HUMAN	RNA-binding protein 41	Placenta	1	3.16	http://www.uniprot.org/uniprot/Q96I25
NP_942682.1	Q0IIM8 - TBC8B_HUMAN	TBC1 domain family member 8B	Nervioso	4	3.16	http://www.uniprot.org/uniprot/Q0IIM8
RN128_HUMAN	Q8TEB7 - RN128_HUMAN	E3 ubiquitin-protein ligase RNF128	Immune	3	3.16	http://www.uniprot.org/uniprot/Q8TEB7
CX057_HUMAN	Q6NSI4 - CX057_HUMAN	Uncharacterized protein CXorf57	Nervioso	2	3.16	http://www.uniprot.org/uniprot/Q6NSI4
MUML1_HUMAN	Q6H9M0 - MUML1_HUMAN	PWWP domain-containing protein MUM1L1	iESTICULO	31	3.16	http://www.uniprot.org/uniprot/Q6H9M0
THBG_HUMAN	P05543 - THBG_HUMAN	Thyroxine-binding globulin	higado	8	3.16	http://www.uniprot.org/uniprot/P05543Cc
NRK_HUMAN	Q7Z2Y5 - NRK_HUMAN	Nik-related protein kinase	Placenta	11	3.16	http://www.uniprot.org/uniprot/Q7Z2Y5
TX13A_HUMAN	Q9BXU3 - TX13A_HUMAN	Testis-expressed sequence 13A protein	Testiculo	3	3.16	http://www.uniprot.org/uniprot/Q9BXU3
IRPL2_HUMAN	Q9NP60 - IRPL2_HUMAN	X-linked interleukin-1 receptor accessory protein-like 2	Nervioso	6	3.16	http://www.uniprot.org/uniprot/Q9NP60
ESX1L_HUMAN	Q8N693 - ESX1_HUMAN	Homeobox protein ESX1	testiculo	1	3.16	http://www.uniprot.org/uniprot/Q8N693
CX039_HUMAN	Q6PEV8 - F199X_HUMAN	Protein FAM199X	Placenta	16	3.16	http://www.uniprot.org/uniprot/Q6PEV8
MCAR6_HUMAN	Q6H9E4 - S2553_HUMAN	Solute carrier family 25 member 53	Nervioso	7	3.16	http://www.uniprot.org/uniprot/Q6H9E4
H2BFM_HUMAN	P0C1H6 - H2BFM_HUMAN	Histone H2B type F-M	Cc	2	3.16	http://www.uniprot.org/uniprot/P0C1H6
RAB9B_HUMAN	Q9NP90 - RAB9B_HUMAN	Ras-related protein Rab-9B	Trafico	1	3.16	http://www.uniprot.org/uniprot/Q9NP90
P60201-2	P60201 - MYPR_HUMAN	Myelin proteolipid protein	Nervioso	1	3.16	http://www.uniprot.org/uniprot/P60201#P60201-2
TMM31_HUMAN	Q5JXX7 - TMM31_HUMAN	Transmembrane protein 31	Testiculo	2	3.16	http://www.uniprot.org/uniprot/Q5JXX7
MO4L2_HUMAN	Q15014 - MO4L2_HUMAN	Mortality factor 4-like protein 2	Cc	20	3.16	http://www.uniprot.org/uniprot/Q15014
Q15170-2	Q15170 - TCAL1_HUMAN	Transcription elongation factor A protein-like 1	Ubcuo	18	3.16	http://www.uniprot.org/uniprot/Q15170#Q15170-2
TCAL3_HUMAN	Q969E4 - TCAL3_HUMAN	Transcription elongation factor A protein-like 3	Ubcuo	6	3.16	http://www.uniprot.org/uniprot/Q969E4
TCAL4_HUMAN	Q96E15 - TCAL4_HUMAN	Transcription elongation factor A protein-like 4	ubcuo	5	3.16	http://www.uniprot.org/uniprot/Q96E15
RB40A_HUMAN	Q8WXH6 - RB40A_HUMAN	Ras-related protein Rab-40A	Trafico	24	3.16	http://www.uniprot.org/uniprot/Q8WXH6
NP_996800.1	Q00994 - BEX3_HUMAN	Protein BEX3	testiculo	30	3.16	http://www.uniprot.org/uniprot/Q00994
WPB5_HUMAN	Q9UHQ7 - WPB5_HUMAN	WW domain-binding protein 5	muscular	27	3.16	http://www.uniprot.org/uniprot/Q9UHQ7
TCAL5_HUMAN	Q5H9L2 - TCAL5_HUMAN	Transcription elongation factor A protein-like 5	Nervioso	1	3.16	http://www.uniprot.org/uniprot/Q5H9L2
Q5H9L1_HUMAN	Q5H9L1 - Q5H9L1_HUMAN	Transcription elongation factor A protein-like 8	Nervioso	2	3.16	http://www.uniprot.org/uniprot/Q5H9L1
BEX4_HUMAN	Q9NWD9 - BEX4_HUMAN	Protein BEX4	Nervioso	3	3.16	http://www.uniprot.org/uniprot/Q9NWD9
BEX1_HUMAN	Q9HBH7 - BEX1_HUMAN	Protein BEX1	Nervioso	2	3.16	http://www.uniprot.org/uniprot/Q9HBH7
RB40L_HUMAN	P0C0E4 - RB40L_HUMAN	Ras-related protein Rab-40A-like	Nervioso	21	3.16	http://www.uniprot.org/uniprot/P0C0E4
NP_085142.1	Q6P177 - BHLH9_HUMAN	Protein BHLH9	Nervioso	42	3.16	http://www.uniprot.org/uniprot/Q6P177
GASP1_HUMAN	Q5JY77 - GASP1_HUMAN	G-protein coupled receptor-associated sorting protein 1	Nervioso	8	3.16	http://www.uniprot.org/uniprot/Q5JY77
ARMX5_HUMAN	Q6P1M9 - ARMX5_HUMAN	Armadillo repeat-containing X-linked protein 5	Ubcuo	18	3.16	http://www.uniprot.org/uniprot/Q6P1M9
Q5H9J9_HUMAN	Q5H9J9 - T11X2_HUMAN	T-complex protein 11 X-linked protein 2	Ubcuo	6	3.16	http://www.uniprot.org/uniprot/Q5H9J9
NXF2_HUMAN	Q9CZY0 - NXF2_HUMAN	Nuclear RNA export factor 2	testiculo	76	3.16	http://www.uniprot.org/uniprot/Q9CZY0
BEX5_HUMAN	Q5H9J7 - BEX5_HUMAN	Protein BEX5	Nervioso	5	3.16	http://www.uniprot.org/uniprot/Q5H9J7
Q9H1B4-3	Q9H1B4 - NXF5_HUMAN	Nuclear RNA export factor 5	Testiculo	3	3.16	http://www.uniprot.org/uniprot/Q9H1B4#Q9H1B4-3
ARMX2_HUMAN	Q7L311 - ARMX2_HUMAN	Armadillo repeat-containing X-linked protein 2	Ubcuo	14	3.16	http://www.uniprot.org/uniprot/Q7L311
ARMX3_HUMAN	Q9UH62 - ARMX3_HUMAN	Armadillo repeat-containing X-linked protein 3	Ubcuo	1	3.16	http://www.uniprot.org/uniprot/Q9UH62
ARMX6_HUMAN	Q7L4S7 - ARMX6_HUMAN	Protein ARMX6	Ubcuo	1	3.16	http://www.uniprot.org/uniprot/Q7L4S7
ARMX1_HUMAN	Q9P291 - ARMX1_HUMAN	Armadillo repeat-containing X-linked protein 1	Ubcuo	7	3.16	http://www.uniprot.org/uniprot/Q9P291
Q5H9R4-2	Q5H9R4 - ARMX4_HUMAN	Armadillo repeat-containing X-linked protein 4	Ubcuo	7	3.16	http://www.uniprot.org/uniprot/Q5H9R4#Q5H9R4-2
RL36A_HUMAN	P83881 - RL36A_HUMAN	60S ribosomal protein L36a	testiculo	1	3.16	http://www.uniprot.org/uniprot/P83881
Q3MS92_HUMAN	Q3MS92 - Q3MS92_HUMAN	Dominant-negative kinase-deficient Brutons tyrosine kinase isoform 6	Immune	1	3.16	http://www.uniprot.org/uniprot/Q3MS92
Q9H5R0_HUMAN	Q5H9L4 - TAF7_HUMAN	Transcription initiation factor TFIIID subunit 7-like	Testiculo	9	3.16	http://www.uniprot.org/uniprot/Q9H5R0
DRP2_HUMAN	Q13474 - DRP2_HUMAN	Dystrophin-related protein 2	Nervioso	2	3.16	http://www.uniprot.org/uniprot/Q13474
NP_079193.2	Q96GJ1 - TRM2_HUMAN	tRNA (uracil(54)-C(5))-methyltransferase homolog	Immune	42	3.16	http://www.uniprot.org/uniprot/Q96GJ1

Tabla. 7. Secuencias CARg putativas obtenidas por medio de las matrices construidas en base a las secuencias con un valor de ΔG_{bind} de $-4.1e+02$ kJ/mol. La primer columna de la tabla corresponde a la entrada con la que se localizó el gen, la segunda columna corresponde al número de acceso al cual se recurrió para explorarlo, la tercer columna corresponde a una descripción breve del gen, la cuarta columna corresponde al tejido en el que se expresa este gen, la siguiente columna (CIS-E) corresponde al número de veces que una CARgbox aparece en la sección regulatoria del gen, la sexta columna reporta el W_s calculado en el servidor RSAT y finalmente se incluye la referencia para que pueda consultarse la liga de donde se recabó la información presentada del gen. Esta lista cuenta con 20 genes identificados.

ENTRADA	NÚMERO DE ACCESO	DESCRIPCIÓN	TEJIDO	CIS-E	W_s	REFERENCIA #20
PO2F1_HUMAN	P14859 - PO2F1_HUMAN	POU domain, class 2, transcription factor 1	Immune	76	4.19	http://www.uniprot.org/uniprot/P14859
Q14902_HUMAN	Q14902 - Q14902_HUMAN	Myelin protein zero	Nervioso	3	4.19	http://www.uniprot.org/uniprot/Q14902
JAM1_HUMAN	Q9Y624 - JAM1_HUMAN	Junctional adhesion molecule A	Immune	16	4.19	http://www.uniprot.org/uniprot/Q9Y624
O10T2_HUMAN	Q8NGX3 - O10T2_HUMAN	Olfactory receptor 10T2	Nervioso	12	4.19	http://www.uniprot.org/uniprot/Q8NGX3
Q5V797_HUMAN	Q5V797 - SYDE2_HUMAN	Rho GTPase-activating protein SYDE2	Immune	6	4.19	http://www.uniprot.org/uniprot/Q5V797
P22694-2	P22694 - KAPCB_HUMAN	cAMP-dependent protein kinase catalytic subunit beta isoforma 4	Nervioso	11	4.19	http://www.uniprot.org/uniprot/P22694#P22694-2
TEKT2_HUMAN	Q9UIF3 - TEKT2_HUMAN	Tektin-2	Testiculo	7	4.19	http://www.uniprot.org/uniprot/Q9UIF3
TINAL_HUMAN	Q9GZM7 - TINAL_HUMAN	Tubulointerstitial nephritis antigen-like	Immune	9	4.19	http://www.uniprot.org/uniprot/Q9GZM7
FCERA_HUMAN	P12319 - FCERA_HUMAN	High affinity immunoglobulin epsilon receptor subunit alpha	Immune	57	4.18	http://www.uniprot.org/uniprot/P12319

O10K1_HUMAN	Q8NGX5 - O10K1_HUMAN	Olfactory receptor 10K1	Nervioso	9	4.18	http://www.uniprot.org/uniprot/Q8NGX5
ARHGB_HUMAN	O15085 - ARHGB_HUMAN	Rho guanine nucleotide exchange factor 11	Nervioso	8	4.18	http://www.uniprot.org/uniprot/O15085
Q59FE7_HUMAN	Q59FE7 - Q59FE7_HUMAN	Ecotropic viral integration site 5 variant	Nervioso	4	4.18	http://www.uniprot.org/uniprot/Q59FE7
Q5TB30-4	Q5TB30 - DEP1A_HUMAN	DEP domain-containing protein 1A	Testiculo	17	4.18	http://www.uniprot.org/uniprot/Q5TB30#Q5TB30-4
Q8NC51-2	Q8NC51 - PAIRB_HUMAN	Plasminogen activator inhibitor 1 RNA-binding protein	Musculo	11	4.18	http://www.uniprot.org/uniprot/Q8NC51#Q8NC51-2
P48357-3	P48357 - LEPR_HUMAN	Leptin receptor	Immune	8	4.18	http://www.uniprot.org/uniprot/P48357#P48357-3
Q8NBB3_HUMAN	Q86E52 - OMA1_HUMAN	Metalloendopeptidase OMA1, mitochondrial	Musculo	11	4.18	http://www.uniprot.org/uniprot/Q86E52
ELAV4_HUMAN	P26378 - ELAV4_HUMAN	ELAV-like protein 4	Nervioso	6	4.18	http://www.uniprot.org/uniprot/P26378
PYRG1_HUMAN	P17812 - PYRG1_HUMAN	CTP synthase 1	Immune	4	4.18	http://www.uniprot.org/uniprot/P17812
Q86UV5-5	Q86UV5 - UBP48_HUMAN	Ubiquitin carboxyl-terminal hydrolase 48	Nervioso	10	4.18	http://www.uniprot.org/uniprot/Q86UV5#Q86UV5-5
NP_115691.1	Q9BSK2 - S2533_HUMAN	Solute carrier family 25 member 33	Nervioso	5	4.18	http://www.uniprot.org/uniprot/Q9BSK2

Tabla. 8. Secuencias CARG putativas obtenidas por medio de las matrices construidas en base a las secuencias con un valor de ΔG_{bind} de $-4.2e+02$ kJ/mol. La primer columna de la tabla corresponde a la entrada con la que se localizó el gen, la segunda columna corresponde al número de acceso al cual se recurrió para explorarlo, la tercer columna corresponde a una descripción breve del gen, la cuarta columna corresponde al tejido en el que se expresa este gen, la siguiente columna (CIS-E) corresponde al número de veces que una CARGbox aparece en la sección regulatoria del gen, la sexta columna reporta el W_s calculado en el servidor RSAT y finalmente se incluye la referencia para que pueda consultarse la liga de donde se recabó la información presentada del gen. Esta lista cuenta con 47 genes identificados.

ENTRADA	NÚMERO DE ACCESO	DESCRIPCIÓN	TEJIDO	CIS-E	Ws	REFERENCIA #47
CD3G_HUMAN	P09693 - CD3G_HUMAN	T-cell surface glycoprotein CD3 gamma chain	Immune	2	6.11	http://www.uniprot.org/uniprot/P09693
EVA1_HUMAN	Q60487 - MPZL2_HUMAN	Myelin protein zero-like protein 2	Nervioso	2	6.11	http://www.uniprot.org/uniprot/Q60487
U2966_HUMAN	Q6UWV2 - MPZL3_HUMAN	Myelin protein zero-like protein 3	Nervioso	7	6.11	http://www.uniprot.org/uniprot/Q6UWV2
Q86YT9-2	Q86YT9 - JAML1_HUMAN	Junctional adhesion molecule-like	Immune	4	6.11	http://www.uniprot.org/uniprot/Q86YT9#Q86YT9-2
ATNG_HUMAN	P54710 - ATNG_HUMAN	Sodium/potassium-transporting ATPase subunit gamma	Nervioso	2	6.11	http://www.uniprot.org/uniprot/P54710
DSCL1_HUMAN	Q8TD84 - DSCL1_HUMAN	Down syndrome cell adhesion molecule-like protein 1	Nervioso	9	6.11	http://www.uniprot.org/uniprot/Q8TD84
NP_055771.4	Q9UPV0 - CE164_HUMAN	Centrosomal protein of 164 kDa	Ubicuo	1	6.11	http://www.uniprot.org/uniprot/Q9UPV0
Q9P0D2_HUMAN	Q9P0D2 - Q9P0D2_HUMAN	HSPC104, endopeptidase activity	Placenta	2	6.11	http://www.uniprot.org/uniprot/Q9P0D2
BACE1_HUMAN	P56817 - BACE1_HUMAN	Beta-secretase 1	Nervioso	2	6.11	http://www.uniprot.org/uniprot/P56817
RN214_HUMAN	P56817 - BACE2_HUMAN	Beta-secretase 2	Nervioso	1	6.11	http://www.uniprot.org/uniprot/P56817
Q3C1X1_HUMAN	Q16549 - PCSK7_HUMAN	Proprotein convertase subtilisin/kexin type 7	Ubicuo	1	6.11	http://www.uniprot.org/uniprot/Q16549
Q8NBJ9-2	Q8NBJ9 - SIDT2_HUMAN	SID1 transmembrane family member 2	Pancreas	7	6.11	http://www.uniprot.org/uniprot/Q8NBJ9#Q8NBJ9-2
PA1B2_HUMAN	P68402 - PA1B2_HUMAN	Platelet-activating factor acetylhydrolase IB subunit beta	Ubicuo	3	6.11	http://www.uniprot.org/uniprot/P68402
Q9Y2K2-2	Q9Y2K2 - SIK3_HUMAN	Serine/threonine-protein kinase SIK3	Oseo	6	6.11	http://www.uniprot.org/uniprot/Q9Y2K2#Q9Y2K2-2
APOA5_HUMAN	Q6Q788 - APOA5_HUMAN	Apolipoprotein A-V	Higado	3	6.11	http://www.uniprot.org/uniprot/Q6Q788
Q8N2F4_HUMAN	Q8BY67 - CADM1_HUMAN	Cell adhesion molecule 1	Immune	4	6.11	http://www.uniprot.org/uniprot/Q8BY67
ORN_HUMAN	Q9Y3B8 - ORN_HUMAN	Oligonucleotidase, mitochondrial	Musculo	3	6.11	http://www.uniprot.org/uniprot/Q9Y3B8
Q05516-2	Q05516 - ZBT16_HUMAN	Zinc finger and BTB domain-containing protein 16	Immune	1	6.11	http://www.uniprot.org/uniprot/Q05516#Q05516-2
5HT3R_HUMAN	P46098 - 5HT3A_HUMAN	5-hydroxytryptamine receptor 3A	Nervioso	9	6.11	http://www.uniprot.org/uniprot/P46098
NP_006019.1	Q95264 - 5HT3B_HUMAN	5-hydroxytryptamine receptor 3B	Nervioso	5	6.11	http://www.uniprot.org/uniprot/Q95264
ZW10_HUMAN	Q43264 - ZW10_HUMAN	Centromere/kinetochore protein zw10 homolog	Trafico	7	6.11	http://www.uniprot.org/uniprot/Q43264
P14416-2	P14416 - DRD2_HUMAN	D(2) dopamine receptor	Nervioso	4	6.11	http://www.uniprot.org/uniprot/P14416#P14416-2
ANKK1_HUMAN	Q8NFD2 - ANKK1_HUMAN	Ankyrin repeat and protein kinase domain-containing protein 1	Muscular	2	6.11	http://www.uniprot.org/uniprot/Q8NFD2
Q9H892-2	Q9H892 - TTC12_HUMAN	Tetrapeptide repeat protein 12	Immune	2	6.11	http://www.uniprot.org/uniprot/Q9H892#Q9H892-2
PTPS_HUMAN	Q03393 - PTPS_HUMAN	6-pyruvoyl tetrahydrobiopterin synthase	Nervioso	1	6.11	http://www.uniprot.org/uniprot/Q03393
Q8IUS0_HUMAN	Q8BYV7 - BCDO2_HUMAN	Beta,beta-carotene 9',10'-oxygenase	Ubicuo	6	6.11	http://www.uniprot.org/uniprot/Q8BYV7
TEX12_HUMAN	Q9BXU0 - TEX12_HUMAN	Testis-expressed sequence 12 protein	Testiculo	9	6.11	http://www.uniprot.org/uniprot/Q9BXU0
IL18_HUMAN	Q14116 - IL18_HUMAN	Interleukin-18	Immune	3	6.11	http://www.uniprot.org/uniprot/Q14116
ODP2_HUMAN	P10515 - ODP2_HUMAN	Dihydropyridyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex.	Ubicuo	5	6.11	http://www.uniprot.org/uniprot/P10515
NP_219493.1	Q155Q3 - DIXC1_HUMAN	Positive effector of the Wnt signaling pathway; activates WNT3A signaling via DVL2	musculo	6	6.11	http://www.uniprot.org/uniprot/Q155Q3
2AAB_HUMAN	P30154 - 2AAB_HUMAN	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A beta isoform	testiculo	2	6.11	http://www.uniprot.org/uniprot/P30154
SN1L2_HUMAN	Q9H0K1 - SIK2_HUMAN	Serine/threonine-protein kinase SIK2	Higado	4	6.11	http://www.uniprot.org/uniprot/Q9H0K1
OBF1_HUMAN	Q16633 - OBF1_HUMAN	POU domain class 2-associated factor 1	Immune	5	6.11	http://www.uniprot.org/uniprot/Q16633
NP_997312.1	Q6ZS62 - COLC1_HUMAN	Colorectal cancer-associated protein 1	Immune	5	6.11	http://www.uniprot.org/uniprot/Q6ZS62
Q9P2F6-2	Q9P2F6 - RHG20_HUMAN	Rho GTPase-activating protein 20	Nervioso	6	6.11	http://www.uniprot.org/uniprot/Q9P2F6#Q9P2F6-2
ADX_HUMAN	P10109 - ADX_HUMAN	Adrenodoxin, mitochondrial	Testiculo	1	6.11	http://www.uniprot.org/uniprot/P10109
Q9C0D7_HUMAN	Q9C0D7 - ZC12C_HUMAN	Probable ribonuclease ZC3H12C	Immune	8	6.11	http://www.uniprot.org/uniprot/Q9C0D7
DDX10_HUMAN	Q13206 - DDX10_HUMAN	Probable ATP-dependent RNA helicase DDX10	testiculo	1	6.11	http://www.uniprot.org/uniprot/Q13206
SLC2B_HUMAN	Q8NEV8 - EXPH5_HUMAN	Exophilin-5	Piel	6	6.09	http://www.uniprot.org/uniprot/Q8NEV8
Q7Z4H8-3	Q7Z4H8 - KDEL2_HUMAN	KDEL motif-containing protein 2	Testiculo	6	6.09	http://www.uniprot.org/uniprot/Q7Z4H8#Q7Z4H8-3
ATM_HUMAN	Q13315 - ATM_HUMAN	Serine-protein kinase ATM	Immune	24	6.09	http://www.uniprot.org/uniprot/Q13315
Q86FG8_HUMAN	P24752 - THIL_HUMAN	Acetyl-CoA acetyltransferase, mitochondrial	Nervioso	2	6.09	http://www.uniprot.org/uniprot/P24752
CUL5_HUMAN	Q93034 - CUL5_HUMAN	Cullin-5	Trafico	6	6.09	http://www.uniprot.org/uniprot/Q93034
RB39A_HUMAN	Q14964 - RB39A_HUMAN	Ras-related protein Rab-39A	trafico	4	6.09	http://www.uniprot.org/uniprot/Q14964
Q14963_HUMAN	Q8IXU6 - S35F2_HUMAN	Solute carrier family 35 member F2	Trafico	1	6.09	http://www.uniprot.org/uniprot/Q8IXU6
SARCO_HUMAN	O00631 - SARCO_HUMAN	Sarcolipin	Muscular	5	6.09	http://www.uniprot.org/uniprot/O00631
ELMD1_HUMAN	Q8N336 - ELMD1_HUMAN	ELMO domain-containing protein 1	Nervioso	4	6.09	http://www.uniprot.org/uniprot/Q8N336#P1

Tabla. 9. Secuencias CaRG putativas obtenidas por medio de las matrices construidas en base a las secuencias con un valor de ΔG_{bind} de $-4.3e+02$ kJ/mol. La primer columna de la tabla corresponde a la entrada con la que se localizó el gen, la segunda columna corresponde al número de acceso al cual se recurrió para explorarlo, la tercer columna corresponde a una descripción breve del gen, la cuarta columna corresponde al tejido en el que se expresa este gen, la siguiente columna (CIS-E) corresponde al número de veces que una CaRGbox aparece en la sección regulatoria del gen, la sexta columna reporta el W_s calculado en el servidor RSAT y finalmente se incluye la referencia para que pueda consultarse la liga de donde se recabó la información presentada del gen. Esta lista cuenta con 204 genes identificados.

ENTRADA	NÚMERO DE ACCESO	DESCRIPCIÓN	TEJIDO	CIS-E	Ws	REFERENCIA #204
Q15381_HUMAN	Q15381 - Q15381_HUMAN	Y-chromosome RNA recognition motif protein	Testículo	104	8.35	www.uniprot.org/uniprot/Q15381
CDY1_HUMAN	Q9Y6F8 - CDY1_HUMAN	Testis-specific chromodomain protein Y 1	Testículo	30	8.35	http://www.uniprot.org/uniprot/Q9Y6F8
Q96RM8_HUMAN	Q96RM8 - Q96RM8_HUMAN	Golgi autoantigen golgin subfamily a2-like	Testículo	3	8.35	http://www.uniprot.org/uniprot/Q96RM8
VCY2_HUMAN	O14599 - VCY2_HUMAN	Testis-specific basic protein Y 2	Testículo	18	8.35	http://www.uniprot.org/uniprot/O14599
Q658T2_HUMAN	Q658T2 - Q658T2_HUMAN	Putative uncharacterized protein DKFZp666C074	Testículo	60	8.35	http://www.uniprot.org/uniprot/Q658T2
Q96P42_HUMAN	Q96P42 - Q96P42_HUMAN	Deleted in azoospermia 1	Testículo	180	8.35	http://www.uniprot.org/uniprot/Q96P42
Q86SG3-2	Q86SG3 - DAZ4_HUMAN	Deleted in azoospermia protein 4	Testículo	9	8.35	http://www.uniprot.org/uniprot/Q86SG3#Q86SG3-2
ENST00000358944	F2ZZU1 - F2ZZU1_HUMAN	RNA-binding motif protein, Y chromosome, family 1 member A1	Testículo	4	8.35	http://www.uniprot.org/uniprot/F2ZZU1
HSFY1_HUMAN	Q9H606 - PRORY_HUMAN	Proline-rich protein, Y-linked	Intestino	248	8.35	http://www.uniprot.org/uniprot/Q9H606
Q96L6-3	Q96L6 - HSFY1_HUMAN	Heat shock transcription factor, Y-linked	Testículo	144	8.35	http://www.uniprot.org/uniprot/Q96L6#Q96L6-3
CDY2_HUMAN	Q9Y6F7 - CDY2_HUMAN	Testis-specific chromodomain protein Y 2	Testículo	60	8.35	http://www.uniprot.org/uniprot/Q9Y6F7
NLGN4_HUMAN	Q8NFZ3 - NLGN4_HUMAN	Neuroigin-4, Y-linked	Nervioso	39	8.35	http://www.uniprot.org/uniprot/Q8NFZ3
ENST00000382868	A6NMU8 - A6NMU8_HUMAN	Neuroigin 4, Y-linked, isoform CRA_c	Nervioso	7	8.35	http://www.uniprot.org/uniprot/A6NMU8
TYB4Y_HUMAN	O14604 - TYB4Y_HUMAN	Thymosin beta-4, Y-chromosomal	Ubicuo	6	8.35	http://www.uniprot.org/uniprot/O14604
ENST00000382893	A6NCS7 - A6NCS7_HUMAN	Histone demethylase UTY	Testículo	4	8.35	http://www.uniprot.org/uniprot/A6NCS7
UTY_HUMAN	O14607 - UTY_HUMAN	Histone demethylase UTY2	Testículo	12	8.35	http://www.uniprot.org/uniprot/O14607
DDX3Y_HUMAN	O15523 - DDX3Y_HUMAN	ATP-dependent RNA helicase DDX3Y	Testículo	30	8.35	http://www.uniprot.org/uniprot/O15523
Q8TEH8_HUMAN	Q8TEH8 - Q8TEH8_HUMAN	FLJ00219 protein	Immune	3	8.35	http://www.uniprot.org/uniprot/Q8TEH8
Q13381_HUMAN	Q13381 - Q13381_HUMAN	RNA binding motif	Ubicuo	30	8.35	www.uniprot.org/uniprot/Q13381
Q01534-2	Q01534 - TSPY1_HUMAN	Proteína codificadora en testículos	Testículo	12	8.35	http://www.uniprot.org/uniprot/Q01534#Q01534-2
A6NGL4	A6NGL4 - A6NGL4_HUMAN	Testis-specific Y-encoded protein 8	Testículo	4	8.35	http://www.uniprot.org/uniprot/A6NGL4
ENST00000383020	A6NDE4 - RBY1B_HUMAN	Proteína de unión a RNA del cromosoma Y, familia 1 miembro B, participa en espermatogenesis	Testículo	7	8.35	http://www.uniprot.org/uniprot/A6NDE4
PRKY_HUMAN	O43930 - PRKY_HUMAN	Proteína Cinesa PRKY, inmiscuida en la transexualización en hembras	Testículos	21	8.22	http://www.uniprot.org/uniprot/O43930
AMELY_HUMAN	Q99218 - AMELY_HUMAN	Amelogenina isoforma Y, juega un papel importante en la mineralización de los dientes	Oseo	18	8.22	http://www.uniprot.org/uniprot/Q99218
Q13381_HUMAN	Q13381 - Q13381_HUMAN	Factor de transcripción que controla la espermatogenesis	Testículo	30	8.22	http://www.uniprot.org/uniprot/Q13381
ENST00000383042	A6NGT6 - A6NGT6_HUMAN	Proteína codificadora en testículos que promueve la proliferación de esperma	Testículo	7	8.22	http://www.uniprot.org/uniprot/A6NGT6
Q9BZA8-2	Q9BZA8 - PC11Y_HUMAN	Protocaderina 11, dependiente de calcio se le asocia con enfermedades del aparato reproductor masculino	Testículo	27	8.22	http://www.uniprot.org/uniprot/Q9BZA8#Q9BZA8-2
TFZLY_HUMAN	Q8IUE0 - TFZLY_HUMAN	Proteína Homeobox TGIF2L1, reguladora de la transcripción en tejido testicular	Testículo	15	8.22	http://www.uniprot.org/uniprot/Q8IUE0
ZFY_HUMAN	P08048 - ZFY_HUMAN	Zinc finger Y-chromosomal protein	Testículo	6	8.22	http://www.uniprot.org/uniprot/P08048
SRY_HUMAN	Q05066 - SRY_HUMAN	Regulador transcripcional esencial en la diferenciación del sexo en machos	Testículos	27	8.22	http://www.uniprot.org/uniprot/Q05066
IL9R_HUMAN	Q01113 - IL9R_HUMAN	Receptor 9 de interleucina, participa en la formación de vesículas	Trafico	1	8.22	http://www.uniprot.org/uniprot/Q01113
P51809-2	P51809 - VAMP7_HUMAN	Proteína asociada vesicular de membrana 7	Cc	24	8.22	http://www.uniprot.org/uniprot/P51809#P51809-2
SPY3_HUMAN	O43610 - SPY3_HUMAN	Factor de crecimiento de fibroblastos	Piel	27	8.22	http://www.uniprot.org/uniprot/O43610
Q9NVH6-2	Q9NVH6 - TMLH_HUMAN	Trimetil Lisina Dioxigenasa mitocondrial, asociada al desarrollo del autismo	Nerviso	7	8.22	http://www.uniprot.org/uniprot/Q9NVH6#Q9NVH6-2
CLIC2_HUMAN	O15247 - CLIC2_HUMAN	Canal de cloro con actividad dependiente del pH, su ausencia esta relacionada con el retraso mental	Nerviso	1	8.22	http://www.uniprot.org/uniprot/O15247
RB39B_HUMAN	Q96DA2 - RB39B_HUMAN	Ras-related protein Rab-39B	Nervioso	6	8.22	http://www.uniprot.org/uniprot/Q96DA2
PFD3_HUMAN	P61758 - PFD3_HUMAN	Prefoldin subunit 3	Ubicuo	2	8.22	http://www.uniprot.org/uniprot/P61758
MTCPA_HUMAN	P56277 - CMC4_HUMAN	Cx9C motif-containing protein 4, coadyuvante en la proliferación celular	Immune	21	8.22	http://www.uniprot.org/uniprot/P56277
Q9BWH2_HUMAN	Q9BWH2 - FUNDC2_HUMAN	Protooncogen FUNDC2	Oseo	3	8.22	http://www.uniprot.org/uniprot/Q9BWH2
FA8_HUMAN	P00451 - FA8_HUMAN	Factor de coagulación 4	Sanguíneo	3	8.22	http://www.uniprot.org/uniprot/P00451
ENST00000369529	A6NGZ8 - SMIM9_HUMAN	Small integral membrane protein 9	Ubicuo	4	8.22	http://www.uniprot.org/uniprot/A6NGZ8
DKC1_HUMAN	O60832 - DKC1_HUMAN	HACA ribonucleoprotein complex subunit 4, requerida para la biogénesis del ribosoma	CC	1	8.22	www.uniprot.org/uniprot/O60832
CTG1B_HUMAN	P78358 - CTG1B_HUMAN	Cancer/testis antigen 1	Testículo	6	8.22	http://www.uniprot.org/uniprot/P78358
Q9NY13_HUMAN	Q8WWW8 - GAB3_HUMAN	Proteína GRB-2 asociada a la diferenciación de macrofagos	Immune	6	8.22	http://www.uniprot.org/uniprot/Q8WWW8
Q9BU27_HUMAN	Q9BU27 - Q9BU27_HUMAN	AM3A protein proliferadora de peroxisomas	Trafico	5	8.22	http://www.ncbi.nlm.nih.gov/gene/60343
FAM3A_HUMAN	P98173 - FAM3A_HUMAN	Protein FAM3A, se activa en procesos de defensa contra hongos	Immune	30	8.22	http://www.uniprot.org/uniprot/P98173
P3_HUMAN	P09131 - P3_HUMAN	P3 protein, transportador de sodio	Nervioso	270	8.22	http://mp3juices.to/ind?q=goldfrapp&page=2
MECP2_HUMAN	P51608 - MECP2_HUMAN	Methyl-CpG-binding protein 2, Coadyuvante en el proceso de metilación del ADN	CC	4	8.22	http://www.uniprot.org/uniprot/P51608
DUS9_HUMAN	Q99956 - DUS9_HUMAN	Dual specificity protein phosphatase 9	Ubicuo	2	8.22	http://www.uniprot.org/uniprot/Q99956
Q8N4A4_HUMAN	Q8N4A4 - Q8N4A4_HUMAN	Similar to dual specificity phosphatase 9	Ubicuo	1	8.22	http://www.uniprot.org/uniprot/Q8N4A4
MAGA1_HUMAN	P43355 - MAGA1_HUMAN	Melanoma-associated antigen 1	Immune	3	8.22	http://www.uniprot.org/uniprot/P43355
Q8N1R8_HUMAN	O15231 - ZNF185_HUMAN	Zinc finger protein 185	Placenta	2	8.22	http://www.uniprot.org/uniprot/O15231
CSAG2_HUMAN	Q9Y5P2 - CSAG2_HUMAN	Chondrosarcoma-associated gene 2/3 protein	Riñón	15	8.22	http://www.uniprot.org/uniprot/Q9Y5P2
GBRA3_HUMAN	P34903 - GBRA3_HUMAN	Sub-unidad alfa3 del receptor gama-acido aminobutirico, inhibe la señal de GABA en neurotransmisores	Nervioso	6	8.22	http://www.uniprot.org/uniprot/P34903
MAGAA_HUMAN	P43363 - MAGAA_HUMAN	Su expresión normal se da en spermatogonias y espermatozoides	Testículo	9	8.22	http://www.uniprot.org/uniprot/P43363
CNGA2_HUMAN	Q16280 - CNGA2_HUMAN	Transductor de las señales olfatorias a neuronas	Nervisos	4	8.22	http://www.uniprot.org/uniprot/Q16280
MTR1L_HUMAN	Q13585 - MTR1L_HUMAN	Receptor de melatonina	Nervioso	1	8.22	http://www.uniprot.org/uniprot/Q13585
CG1_HUMAN	Q13495 - MAMLD1_HUMAN	Proteína MAMLD1 requerida en la formación de la uretra	Testículo	11	8.22	http://www.uniprot.org/uniprot/Q13495
MAGA9_HUMAN	P43362 - MAGA9_HUMAN	Antígeno 9 asociado a melanoma y aparece normalmente en testículos	Testículo	15	8.22	http://www.uniprot.org/uniprot/P43362
MAGAB_HUMAN	P43364 - MAGAB_HUMAN	Antígeno 11 asociado a melanoma y aparece normalmente en testículos	Testículo	2	8.22	http://www.uniprot.org/uniprot/P43364
T185A_HUMAN	Q8NFB2 - T185A_HUMAN	Transmembrane protein 185A	Nervioso	2	8.22	http://www.uniprot.org/uniprot/Q8NFB2
O60597_HUMAN	O60597 - O60597_HUMAN	Iduronate 2-sulfatase	Trafico	4	8.22	http://www.uniprot.org/uniprot/O60597
Q7ZF2F_HUMAN	P51816 - AFF2_HUMAN	Proteína AF4/FMR2 miembro 2, en su ausencia se expresan problemas del lenguaje y la lectura	Nervioso	7	8.22	http://www.uniprot.org/uniprot/P51816
FMR1N_HUMAN	Q8NDW7 - FMR1N_HUMAN	Proteína que en su ausencia se presentan problemas de retraso mental	Nervioso	3	8.22	http://www.uniprot.org/uniprot/Q8NDW7
Q8NAQ8_HUMAN	Q8NAQ8 - YP023_HUMAN	Putative uncharacterized protein FLJ34945	Desconocido	4	8.22	http://www.uniprot.org/uniprot/Q8NAQ8
CX001_HUMAN	Q96002 - TM257_HUMAN	Proteína transmembranal 257, se expresa en la zona del hipocampo	Nervioso	11	8.22	http://www.uniprot.org/uniprot/Q96002
SLIK2_HUMAN	Q8H156 - SLIK2_HUMAN	Proteína supresora del crecimiento neural	Nervioso	24	8.22	http://www.uniprot.org/uniprot/Q8H156
UE2NL_HUMAN	Q5JXB2 - UE2NL_HUMAN	Enzima ubiquitina conjugada E2 expresada en epidídimo	Testículo	45	8.22	http://www.uniprot.org/uniprot/Q5JXB2

SLIK4_HUMAN	Q8IW52 - SLIK4_HUMAN	Proteína 4 NTRK supresora del crecimiento neuronal	Nervioso	9	8.22	http://www.uniprot.org/uniprot/Q8IW52
MAGC1_HUMAN	Q60732 - MAGC1_HUMAN	Melanoma-associated antigen C1	Testículos	1	8.22	http://www.uniprot.org/uniprot/Q60732
MAGC3_HUMAN	Q8TD91 - MAGC3_HUMAN	Melanoma-associated antigen C3	Testículos	10	8.22	http://www.uniprot.org/uniprot/Q8TD91
SPNXD_HUMAN	Q9BXN6 - SPNXD_HUMAN	Proteína espermática del cromosoma D	Testículo	5	8.22	http://www.uniprot.org/uniprot/Q9BXN6
LDOC1_HUMAN	Q95751 - LDOC1_HUMAN	Proteína LDOC1, asociada a desarrollo de cáncer cerebral	Nervioso	6	8.22	http://www.uniprot.org/uniprot/Q95751
SPNXB_HUMAN	Q9NS25 - SPNXB_HUMAN	Sperm protein associated with the nucleus on the X chromosome B1F	Testículo	1	8.22	http://www.uniprot.org/uniprot/Q9NS25
NP_004066.2	P51861 - CDR1_HUMAN	Cerebellar degeneration-related antigen 1	Nervioso	39	8.22	http://www.uniprot.org/uniprot/P51861
NP_001013421.1	Q5JRM2 - CX066_HUMAN	Uncharacterized protein CXorf66	Desconocido	12	8.22	http://www.uniprot.org/uniprot/Q5JRM2
Q8NB49-2	Q8NB49 - AT11C_HUMAN	Proteína transportadora de fosfolípidos implicada en la formación de vesículas	Trafico	3	8.22	http://www.uniprot.org/uniprot/Q8NB49#Q8NB49-2
ENST00000370557	A0A067XG54_HUMAN	Phospholipid-transporting ATPase	Trafico	2	8.22	http://www.uniprot.org/uniprot/A0A067XG54
P10911-4	P10911 - MCF2_HUMAN	Protooncogen DBL isoforma 1 en cerebelo	Nervioso	5	8.22	http://www.uniprot.org/uniprot/P10911#P10911-4
Q92913-2	Q92913 - FGF13_HUMAN	Fibroblast growth factor 13	Nervioso	5	8.22	http://www.uniprot.org/uniprot/Q92913#Q92913-2
Q969H1_HUMAN	Q969H1 - Q969H1_HUMAN	Proteína desconocida	Desconocido	4	8.22	http://www.uniprot.org/uniprot/Q969H1
ARHG6_HUMAN	Q15052 - ARHG6_HUMAN	Rho guanine nucleotide exchange factor 6	Ubicuo	2	8.22	http://www.uniprot.org/uniprot/Q15052#Q15052-2
Q3L8U2_HUMAN	Q3L8U2 - Q3L8U2_HUMAN	CD154 protein, participa en la respuesta inmune	Inmune	6	8.22	http://www.uniprot.org/uniprot/Q3L8U2
CD40L_HUMAN	P29965 - CD40L_HUMAN	Involved in immunoglobulin class switching	Inmune	6	8.22	http://www.uniprot.org/uniprot/P29965
BRS3_HUMAN	P32247 - BRS3_HUMAN	Bombesin receptor subtype-3, Role in sperm cell division, maturation, or function	Testículo	2	8.22	http://www.uniprot.org/uniprot/P32247
GP112_HUMAN	Q81ZF6 - GP112_HUMAN	Probable G-protein coupled receptor 112	Hgado	7	8.22	http://www.uniprot.org/uniprot/Q81ZF6
Q5JX16_HUMAN	Q81WC1 - MA7D3_HUMAN	MAP7 domain-containing protein 3	Cc	2	8.22	http://www.uniprot.org/uniprot/Q81WC1
Q13642-3	Q13642 - FHL1_HUMAN	Four and a half LIM domains protein 1. May have an involvement in muscle development or hypertrophy	Muscular	4	8.22	http://www.uniprot.org/uniprot/Q13642#Q13642-3
SL9A6_HUMAN	Q92581 - SL9A6_HUMAN	Electroneutral exchange of protons for Na+ and K+ across the early and recycling endosome membranes	Nervioso	5	8.22	http://www.uniprot.org/uniprot/Q92581
Q5JSJ4-4	Q5JSJ4 - DX26B_HUMAN	Protein DDX26B	Desconocido	12	8.22	http://www.uniprot.org/uniprot/Q5JSJ4#Q5JSJ4-4
Q3ZM62_HUMAN	Q3ZM62 - Q3ZM62_HUMAN	Putative uncharacterized protein	Desconocido	2	8.22	http://www.uniprot.org/uniprot/Q3ZM62
Q96HG1_HUMAN	Q96HG1 - SIM10_HUMAN	Small integral membrane protein 10	Ubicuo	8	8.22	http://www.uniprot.org/uniprot/Q96HG1
Q9LUG1-3	Q9LUG1 - MSPD1_HUMAN	Motile sperm domain-containing protein 1	Testículo	5	8.22	http://www.uniprot.org/uniprot/Q9LUG1#Q9LUG1-3
F222C_HUMAN	Q6P4D5 (F222C_HUMAN)	Función desconocida	Desconocido	9	8.22	http://www.uniprot.org/uniprot/Q6P4D5
Q6P187_HUMAN	Q6P187 (Q6P187_HUMAN)	Función desconocida	Desconocido	9	8.22	http://www.uniprot.org/uniprot/Q6P187
PLAC1_HUMAN	Q9HBJ0 (PLAC1_HUMAN)	May play a role in placental development.	Plaventa	4	8.22	http://www.uniprot.org/uniprot/Q9HBJ0
HPRT_HUMAN	P00492 (HPRT_HUMAN)	Hypoxanthine-guanine phosphoribosyltransferase	Ubicuo	4	8.22	http://www.uniprot.org/uniprot/P00492
Q5JRC6_HUMAN	Q5JRC6 (Q5JRC6_HUMAN)	PHD finger protein 6	TF	5	8.22	http://www.uniprot.org/uniprot/Q5JRC6
Q81WS0-2	Q81WS0 (PHF6_HUMAN)	Transcriptional regulator that associates with ribosomal RNA promoters and suppresses ribosomal RNA.	TF	5	8.22	http://www.uniprot.org/uniprot/Q81WS0#Q81WS0-2
GPC3_HUMAN	P51654 (GPC3_HUMAN)	Glycan-3. Cell surface proteoglycan that bears heparan sulfate.	Ubicuo	7	8.22	http://www.uniprot.org/uniprot/P51654
UBP26_HUMAN	Q9BXU7 (UBP26_HUMAN)	Ubiquitin carboxyl-terminal hydrolase 26	Testículo	18	8.22	http://www.uniprot.org/uniprot/Q9BXU7
Q96MM7-3	Q96MM7 (H6S72_HUMAN)	Heparan-sulfate 6-O-sulfotransferase 2	Trafico	1	8.22	http://www.uniprot.org/uniprot/Q96MM7#Q96MM7-3
Q81UR4_HUMAN	Q9NLU0 (MBNL3_HUMAN)	Muscleblind-like protein 3	Músculo	4	8.22	http://www.uniprot.org/uniprot/Q9NLU0
RAP2C_HUMAN	Q9Y3L5 (RAP2C_HUMAN)	Ras-related protein Rap-2c	Ubicuo	36	8.22	http://www.uniprot.org/uniprot/Q9Y3L5
FRMD7_HUMAN	Q6ZUT3 (FRMD7_HUMAN)	FERM domain-containing protein 7. Plays a role in neurite development	Nervioso	3	8.22	http://www.uniprot.org/uniprot/Q6ZUT3
Q8NBY1_HUMAN	Q8NBY1 (Q8NBY1_HUMAN)	Serine/threonine-protein kinase 26	Ubicuo	5	8.22	http://www.uniprot.org/uniprot/Q8NBY1
MST4_HUMAN	Q9P289 (STK26_HUMAN)	Serine/threonine-protein kinase 26	Cc	5	8.22	http://www.uniprot.org/uniprot/Q9P289
NP_991402.1	Q8N6C5 (IGSF1_HUMAN)	Immunoglobulin superfamily member 1	Inmune	20	8.22	http://www.uniprot.org/uniprot/Q8N6C5
Q6ZRI8-2	Q6ZRI8 (RHG36_HUMAN)	Rho GTPase-activating protein 36	Ubicuo	2	8.22	http://www.uniprot.org/uniprot/Q6ZRI8#Q6ZRI8-2
COVA1_HUMAN	Q16206 (ENOX2_HUMAN)	Ecto-NOX disulfide-thiol exchanger 2	Inmune	13	8.22	http://www.uniprot.org/uniprot/Q16206
FA45B_HUMAN	Q6NSW5 (FA45B_HUMAN)	Protein FAM45B	Ubicuo	6	8.22	http://www.uniprot.org/uniprot/Q6NSW5
Q5JY83_HUMAN	Q5JY83 (Q5JY83_HUMAN)	RNA-binding motif protein, X-linked 2	Ubicuo	4	8.22	http://www.uniprot.org/uniprot/Q5JY83
RBMX2_HUMAN	Q9Y388 (RBMX2_HUMAN)	RNA-binding motif protein, X-linked 2	Ubicuo	4	8.22	http://www.uniprot.org/uniprot/Q9Y388
GP119_HUMAN	Q8TDV5 (GP119_HUMAN)	Glucose-dependent insulinotropic receptor	Pancreas	1	8.22	http://www.uniprot.org/uniprot/Q8TDV5
Q95258-3	Q95258 (UCP5_HUMAN)	Brain mitochondrial carrier protein 1	Nervioso	3	8.22	http://www.uniprot.org/uniprot/Q95258#Q95258-3
Q9LUJ2_HUMAN	Q9LUJ2 (Q9LUJ2_HUMAN)	Zinc finger protein 280C	Ubicuo	4	8.15	http://www.uniprot.org/uniprot/Q9LUJ2
Q95831-2	Q95831 (AIFM1_HUMAN)	Apoptosis-inducing factor 1, mitochondrial	Musculo	4	8.15	http://www.uniprot.org/uniprot/Q95831#Q95831-2
RB33A_HUMAN	Q14088 (RB33A_HUMAN)	Ras-related protein Rab-33A	Inmune	4	8.15	http://www.uniprot.org/uniprot/Q14088
Q5H9F3_HUMAN	Q5H9F3 (BCORL_HUMAN)	BCL-6 corepressor-like protein 1	Testículo	2	8.15	http://www.uniprot.org/uniprot/Q5H9F3
Q9BVJ6-2	Q9BVJ6 (UT14A_HUMAN)	U3 small nucleolar RNA-associated protein 14 homolog A	Ubicuo	1	8.15	http://www.uniprot.org/uniprot/Q9BVJ6#Q9BVJ6-2
ZDH9_HUMAN	Q9Y397 (ZDH9_HUMAN)	Palmitoyltransferase ZDHHC9	Ubicuo	9	8.15	http://www.uniprot.org/uniprot/Q9Y397
SLY_HUMAN	O75995 (SASH3_HUMAN)	SAM and SH3 domain-containing protein 3	Inmune	3	8.15	http://www.uniprot.org/uniprot/O75995
Q01968-2	Q01968 (OCLR_HUMAN)	Inositol polyphosphate 5-phosphatase OCLR-1	Ubicuo	10	8.15	http://www.uniprot.org/uniprot/Q01968#Q01968-2
P28370-2	P28370 (SMCA1_HUMAN)	Probable global transcription activator SNF2L1	Nervioso	9	8.15	http://www.uniprot.org/uniprot/P28370#P28370-2
NP_001013650.1	Q5VW00 (DC122_HUMAN)	DDB1- and CUL4-associated factor 12-like protein 2	Ubicuo	35	8.15	http://www.uniprot.org/uniprot/Q5VW00
TEN1_HUMAN	Q9UKZ4 (TEN1_HUMAN)	Teneurin-1	Nervioso	7	8.15	http://www.uniprot.org/uniprot/Q9UKZ4
Q60880-5	Q60880 (SH21A_HUMAN)	SH2 domain-containing protein 1A	Nervioso	5	8.15	http://www.uniprot.org/uniprot/Q60880#Q60880-5
STAG2_HUMAN	Q8N3U4 (STAG2_HUMAN)	Cohesin subunit SA-2	Ubicuo	23	8.15	http://www.uniprot.org/uniprot/Q8N3U4
BIRC4_HUMAN	P98170 (XIAP_HUMAN)	E3 ubiquitin-protein ligase XIAP	Ubicuo	1	8.15	http://www.uniprot.org/uniprot/P98170
GRIA3_HUMAN	P42263 (GRIA3_HUMAN)	Glutamate receptor 3	Nervioso	10	8.15	http://www.uniprot.org/uniprot/P42263
CT47_HUMAN	Q5JQC4 (CT47A_HUMAN)	Cancer/testis antigen 47A	Testículo	74	8.15	http://www.uniprot.org/uniprot/Q5JQC4
Q8NH5_HUMAN	Q8NH5 (Q8NH5_HUMAN)	Putative uncharacterized protein	Ubicuo	8	8.15	http://www.uniprot.org/uniprot/Q8NH5
Q13620-2	Q13620 (CUL4B_HUMAN)	Cullin-4B	Ubicuo	12	8.15	http://www.uniprot.org/uniprot/Q13620#Q13620-2
Q5JRV8-2	Q5JRV8 (T255A_HUMAN)	Transmembrane protein 255A	Ubicuo	1	8.15	http://www.uniprot.org/uniprot/Q5JRV8#Q5JRV8-2
KAISO_HUMAN	Q86T24 (KAISO_HUMAN)	Transcriptional regulator Kaiso	Vascular	2	8.15	http://www.uniprot.org/uniprot/Q86T24
Q8N9T2-2	Q8N9T2 (CX042_HUMAN)	Putative uncharacterized protein CXorf42	Ubicuo	2	8.15	http://www.uniprot.org/uniprot/Q8N9T2#Q8N9T2-2
PEPP1_HUMAN	Q8NHV9 (RHXF1_HUMAN)	Rhox homeobox family member 1	Testículo	6	8.15	http://www.uniprot.org/uniprot/Q8NHV9
NKAP_HUMAN	Q8N5F7 (NKAP_HUMAN)	NF-kappa-B-activating protein	Inmune	1	8.15	http://www.uniprot.org/uniprot/Q8N5F7

Q9BZ17-2	Q9BZ17 (REN3B_HUMAN)	Regulator of nonsense transcripts 3B	Ubicuo	2	8.15	http://www.uniprot.org/uniprot/Q9BZ17#Q9BZ17-2
ANKRD58	A6NJG2 (SWAHD_HUMAN)	Ankyrin repeat domain-containing protein SOWAHD	Ubicuo	1	8.15	http://www.uniprot.org/uniprot/A6NJG2
NKRF_HUMAN	O15226 (NKRF_HUMAN)	NF-kappa-B-repressing factor	Immune	2	8.15	http://www.uniprot.org/uniprot/O15226
Q8N8L5_HUMAN	Q8WUT9 (S2543_HUMAN)	Solute carrier family 25 member 43	Trafico	5	8.15	http://www.uniprot.org/uniprot/Q8WUT9
PGRC1_HUMAN	O00264 (PGRC1_HUMAN)	Membrane-associated progesterone receptor component 1	Higado	1	8.15	http://www.uniprot.org/uniprot/O00264
K1210_HUMAN	Q9ULL0 (K1210_HUMAN)	Uncharacterized protein KIAA1210	Ubicuo	3	8.15	http://www.uniprot.org/uniprot/Q9ULL0
I13R1_HUMAN	P78552 (I13R1_HUMAN)	Interleukin-13 receptor subunit alpha-1	Immune	3	8.15	http://www.uniprot.org/uniprot/P78552
Q8N397_HUMAN	Q8N397 (Q8N397_HUMAN)	Putative uncharacterized protein DKFZp761M142	Ubicuo	6	8.15	http://www.uniprot.org/uniprot/Q8N397
Q5JSH3-2	Q5JSH3 (WDR44_HUMAN)	WD repeat-containing protein 44	Trafivo	11	8.15	http://www.uniprot.org/uniprot/Q5JSH3#Q5JSH3-2
KLH13_HUMAN	Q9P2N7 (KLH13_HUMAN)	Kelch-like protein 13	Ubicuo	52	8.15	http://www.uniprot.org/uniprot/Q9P2N7
S6A14_HUMAN	Q9UN76 (S6A14_HUMAN)	Sodium- and chloride-dependent neutral and basic amino acid transporter B(0+)	Pulmón	1	8.15	http://www.uniprot.org/uniprot/Q9UN76
AGTR2_HUMAN	P50052 (AGTR2_HUMAN)	Type-2 angiotensin II receptor	Ovario	9	8.15	http://www.uniprot.org/uniprot/P50052
6YX002_HUMAN	Q9BTK2 (YX002_HUMAN)	Putative uncharacterized protein LOC642776	Ubicuo	6	8.15	http://www.uniprot.org/uniprot/Q9BTK2
PLST_HUMAN	P13797 (PLST_HUMAN)	Plastin-3	Músculo	19	8.15	http://www.uniprot.org/uniprot/P13797
NP_057467.1	Q9P127 (LUZP4_HUMAN)	Leucine zipper protein 4	Testiculo	3	8.15	http://www.uniprot.org/uniprot/Q9P127
CX055_HUMAN	Q8N7X1 (RMXL3_HUMAN)	RNA-binding motif protein, X-linked-like-3	Ubicuo	10	8.15	http://www.uniprot.org/uniprot/Q8N7X1
LRC2_HUMAN	Q5VUJ6 (LRC2_HUMAN)	Leucine-rich repeat and calponin homology domain-containing protein 2	Ubicuo	8	8.15	http://www.uniprot.org/uniprot/Q5VUJ6
I13R2_HUMAN	Q14627 (I13R2_HUMAN)	Interleukin-13 receptor subunit alpha-2	Immune	7	8.15	http://www.uniprot.org/uniprot/Q14627
5HT2C_HUMAN	P28335 (5HT2C_HUMAN)	5-hydroxytryptamine receptor 2C	Nervioso	6	8.15	http://www.uniprot.org/uniprot/P28335
AMOT_HUMAN	Q4VCS5 (AMOT_HUMAN)	Angiomotin. Plays a central role in tight junction maintenance via the complex formed with ARHGAP17.	Músculo	5	8.15	http://www.uniprot.org/uniprot/Q4VCS5
LHPL1_HUMAN	Q86W10 (LHPL1_HUMAN)	Lipoma HMGIC fusion partner-like 1 protein	Ubicuo	15	8.15	http://www.uniprot.org/uniprot/Q86W10
ZCH16_HUMAN	Q6ZR62 (ZCH16_HUMAN)	Zinc finger CCHC domain-containing protein 16	Ubicuo	6	8.15	http://www.uniprot.org/uniprot/Q6ZR62
TRPC5_HUMAN	Q9UL62 (TRPC5_HUMAN)	Short transient receptor potential channel 5	Nervioso	1	8.15	http://www.uniprot.org/uniprot/Q9UL62
CX045_HUMAN	Q9NP73 (ALG13_HUMAN)	Putative bifunctional UDP-N-acetylglucosamine transferase and deubiquitinase ALG13	Ubicuo	3	8.15	http://www.uniprot.org/uniprot/Q9NP73
DCX_HUMAN	Q43602 (DCX_HUMAN)	Neuronal migration protein doublecortin	Nervioso	6	8.15	http://www.uniprot.org/uniprot/Q43602
PAK3_HUMAN	O75914 (PAK3_HUMAN)	Senenithreonine-protein kinase PAK 3	Nervioso	9	8.15	http://www.uniprot.org/uniprot/O75914
CRDL1_HUMAN	Q9BU40 (CRDL1_HUMAN)	Chordin-like protein 1	Nervioso	22	8.15	http://www.uniprot.org/uniprot/Q9BU40
RGAG1_HUMAN	Q8NET4 (RGAG1_HUMAN)	Retrotroposon gag domain-containing protein 1	Nervioso	8	8.15	http://www.uniprot.org/uniprot/Q8NET4
Q9Y4X0-3	Q9Y4X0 (AMMR1_HUMAN)	AMME syndrome candidate gene 1 protein	Ubicuo	5	8.15	http://www.uniprot.org/uniprot/Q9Y4X0#Q9Y4X0-3
AMER1_HUMAN	Q5JTC6 (AMER1_HUMAN)	APC membrane recruitment protein 1	Nervioso	5	8.15	http://www.uniprot.org/uniprot/Q5JTC6
TM164_HUMAN	Q5JUC3 (TM164_HUMAN)	Transmembrane protein 164	Ubicuo	3	8.15	http://www.uniprot.org/uniprot/Q5JUC3
O60488-2	Q60488 (ACSL4_HUMAN)	Long-chain-fatty-acid-CoA ligase 4	Ubicuo	18	8.15	http://www.uniprot.org/uniprot/Q60488#Q60488-2
Q5JYV7_HUMAN	Q9NPJ8 (NXT2_HUMAN)	NTF2-related export protein 2	Ubicuo	4	8.15	http://www.uniprot.org/uniprot/Q9NPJ8
GUC2F_HUMAN	P51841 (GUC2F_HUMAN)	Retinal guanylyl cyclase 2	Nervioso	2	8.15	http://www.uniprot.org/uniprot/P51841
Q14031-2	Q14031 (C0446_HUMAN)	Collagen alpha-6(IV) chain	Ubicuo	4	8.15	http://www.uniprot.org/uniprot/Q14031#Q14031-2
TX13B_HUMAN	Q9BXU2 (TX13B_HUMAN)	Testis-expressed sequence 13B protein	Testiculo	2	8.15	http://www.uniprot.org/uniprot/Q9BXU2
Q9UJV3-2	Q9UJV3 (TRIM1_HUMAN)	Probable E3 ubiquitin-protein ligase MID2	Ovario	6	8.15	http://www.uniprot.org/uniprot/Q9UJV3#Q9UJV3-2
T22D3_HUMAN	Q99576 (T22D3_HUMAN)	TSC22 domain family protein 3	Immune	5	8.15	http://www.uniprot.org/uniprot/Q99576
FRPD3_HUMAN	Q5JV73 (FRPD3_HUMAN)	FERM and PDZ domain-containing protein 3	Ubicuo	1	8.15	http://www.uniprot.org/uniprot/Q5JV73
MYCL2_HUMAN	P12525 (MYCP1_HUMAN)	Putative myc-like protein MYCLP1	Testiculo	2	8.15	http://www.uniprot.org/uniprot/P12525
RBM41_HUMAN	Q96JZ5 (RBM41_HUMAN)	RNA-binding protein 41	Ubicuo	4	8.15	http://www.uniprot.org/uniprot/Q96JZ5
A1YR24_HUMAN	Q8TE76 (MORC4_HUMAN)	MORC family CW-type zinc finger protein 4	Placenta	4	8.15	http://www.uniprot.org/uniprot/Q8TE76
NP_942582.1	Q0IIM8 (TBC8B_HUMAN)	TBC1 domain family member 8B	Ubicuo	7	8.15	http://www.uniprot.org/uniprot/Q0IIM8
RN128_HUMAN	Q8TEB7 (RN128_HUMAN)	E3 ubiquitin-protein ligase RNF128	Nervioso	4	8.15	http://www.uniprot.org/uniprot/Q8TEB7
MUML1_HUMAN	Q5H9M0 (MUML1_HUMAN)	PWWP domain-containing protein MUM1L1	Ubicuo	44	8.15	http://www.uniprot.org/uniprot/Q5H9M0
NRK_HUMAN	Q7Z2Y5 (NRK_HUMAN)	Nik-related protein kinase	Embrion	8	8.15	http://www.uniprot.org/uniprot/Q7Z2Y5
TX13A_HUMAN	Q9BXU3 (TX13A_HUMAN)	Testis-expressed sequence 13A protein	Testiculo	2	8.15	http://www.uniprot.org/uniprot/Q9BXU3
IRPL2_HUMAN	Q9NP60 (IRPL2_HUMAN)	X-linked interleukin-1 receptor accessory protein-like 2	Nervioso	14	8.15	http://www.uniprot.org/uniprot/Q9NP60
CX039_HUMAN	Q6PEV8 (F199X_HUMAN)	Protein FAM199X	Ubicuo	15	8.15	http://www.uniprot.org/uniprot/Q6PEV8
MCAR6_HUMAN	Q5H9E4 (S2553_HUMAN)	Solute carrier family 25 member 53	Ubicuo	8	8.15	http://www.uniprot.org/uniprot/Q5H9E4
H2BFM_HUMAN	P0C1H6 (H2BFM_HUMAN)	Histone H2B type F-M	Ubicuo	2	8.15	http://www.uniprot.org/uniprot/P0C1H6
RAB9B_HUMAN	Q9NP90 (RAB9B_HUMAN)	Ras-related protein Rab-9B	Ubicuo	9	8.15	http://www.uniprot.org/uniprot/Q9NP90
TMM31_HUMAN	Q5JXX7 (TMM31_HUMAN)	Transmembrane protein 31	Ubicuo	4	8.11	http://www.uniprot.org/uniprot/Q5JXX7
MO4L2_HUMAN	Q15014 (MO4L2_HUMAN)	Mortality factor 4-like protein 2	Ubicuo	12	8.11	http://www.uniprot.org/uniprot/Q15014
Q15170-2	Q15170 (TCAL1_HUMAN)	Transcription elongation factor A protein-like 1	Ubicuo	50	8.11	http://www.uniprot.org/uniprot/Q15170#Q15170-2
TCAL3_HUMAN	Q969E4 (TCAL3_HUMAN)	Transcription elongation factor A protein-like 3	Ubicuo	28	8.11	http://www.uniprot.org/uniprot/Q969E4
TCAL4_HUMAN	Q96E15 (TCAL4_HUMAN)	Transcription elongation factor A protein-like 4	Ubicuo	8	8.11	http://www.uniprot.org/uniprot/Q96E15
RB40A_HUMAN	Q8WVH6 (RB40A_HUMAN)	Ras-related protein Rab-40A	Trafivo	24	8.11	http://www.uniprot.org/uniprot/Q8WVH6
BEX3_HUMAN	Q00994 (BEX3_HUMAN)	Protein BEX3	Testiculos	7	8.11	http://www.uniprot.org/uniprot/Q00994
WPB5_HUMAN	Q9UHQ7 (WPB5_HUMAN)	WW domain-binding protein 5	Ubicuo	21	8.11	http://www.uniprot.org/uniprot/Q9UHQ7
TCAL7_HUMAN	Q9BRU2 (TCAL7_HUMAN)	Transcription elongation factor A protein-like 7	Nervioso	24	8.11	http://www.uniprot.org/uniprot/Q9BRU2
BEX2_HUMAN	Q9BXY8 (BEX2_HUMAN)	Protein BEX2	Nervioso	9	8.11	http://www.uniprot.org/uniprot/Q9BXY8
Q5H9L1_HUMAN	Q5H9L1 (Q5H9L1_HUMAN)	Transcription elongation factor A protein-like 8	Ubicuo	7	8.11	http://www.uniprot.org/uniprot/Q5H9L1
RB40L_HUMAN	P0C0E4 (RB40L_HUMAN)	Ras-related protein Rab-40A-like	Nervioso	15	8.11	http://www.uniprot.org/uniprot/P0C0E4
NP_085142.1	Q6PI77 (BHLH9_HUMAN)	Protein BHLH9. May play a role in the control of cellular aging and survival.	Nervioso	42	8.11	http://www.uniprot.org/uniprot/Q6PI77
GASP2_HUMAN	Q96D09 (GASP2_HUMAN)	G-protein coupled receptor-associated sorting protein 2	Nervioso	1	8.11	http://www.uniprot.org/uniprot/Q96D09

Tabla. 10. Secuencias CArG putativas obtenidas por medio de las matrices construidas en base a las secuencias con un valor de ΔG_{bind} de $-4.4e+02$ kJ/mol. La primer columna de la tabla corresponde a la entrada con la que se localizó el gen, la segunda columna corresponde al número de acceso al cual se recurrió para explorarlo, la tercer columna corresponde a una descripción breve del gen, la cuarta columna corresponde al tejido en el que se expresa este gen, la siguiente columna (CIS-E) corresponde al número de veces que una CArGbox aparece en la sección regulatoria del gen, la sexta columna reporta el W_s calculado en el servidor RSAT y finalmente se incluye la referencia para que pueda consultarse la liga de donde se recabó la información presentada del gen. Esta lista cuenta con 228 genes identificados.

ENTRADA	NÚMERO DE ACCESO	DESCRIPCIÓN	TEJIDO	CIS-e	W_s	REFERENCIA
Q15381_HUMAN	Q15381 - Q15381_HUMAN	Unión y procesamiento de RNA del gen YRRM, cromosoma Y	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q15381
Q8NSB5_HUMAN	PODJ03 - RBY1A_HUMAN	Proteína de unión a mRNA, involucrada en splicing, de cromosoma Y, espermatogonia	Testículos	3	10.9	http://www.uniprot.org/uniprot/PODJ03
NP_001006118.2	A6NEQ0 - RBY1E_HUMAN	Proteína de unión a mRNA involucrada en splicing, de cromosoma Y, espermatogonia	Testículos	3	10.9	http://www.uniprot.org/uniprot/A6NEQ0
NP_001006120.2	POC7P1 - RBY1D_HUMAN	Proteína de unión a mRNA involucrada en splicing, de cromosoma Y, espermatogonia	Testículos	2	10.9	http://www.uniprot.org/uniprot/POC7P1
ENST00000358944	F22ZU1 - F22ZU1_HUMAN	Proteína de unión a RNA, cromosoma Y	No especificado	2	10.9	http://www.uniprot.org/uniprot/F22ZU1
NP_001006121.1	A6NDE4 - RBY1B_HUMAN	Proteína de unión a mRNA, involucrada en splicing, de cromosoma Y, espermatogonia	Testículos	2	10.9	http://www.uniprot.org/uniprot/A6NDE4
NP_001013421.1	Q5JRM2 - CX066_HUMAN	Proteína sin caracterizar, proteína de membrana	Trafico	1	10.9	http://www.uniprot.org/uniprot/Q5JRM2
Q5JRC6_HUMAN	Q5JRC6 - Q5JRC6_HUMAN	Proteína de unión a Zinc, ubicada en nucleo.	Tf	1	10.9	http://www.uniprot.org/uniprot/Q5JRC6
Q8IWS0-2	Q8IWS0 - PHF6_HUMAN	Isoforma de regulador transcripcional, involucrado en Síndrome B-F-L	Tf	1	10.9	http://www.uniprot.org/uniprot/Q8IWS0#Q8IWS0-2
PHF6_HUMAN	Q8IWS0 - PHF6_HUMAN	Regulador transcripcional, involucrado en Síndrome de Boerjeson-Forsman-Lehmann	Tf	2	10.9	http://www.uniprot.org/uniprot/Q8IWS0
UCP5_HUMAN	Q95258 - UCP5_HUMAN	Proteína mitocondrial neuronal que participa en movimiento de protones	Nervioso	2	10.9	http://www.uniprot.org/uniprot/Q95258
T22D3_HUMAN	Q95976 - T22D3_HUMAN	Inhibidor transcripcional que protege diversas células inmunitarias de apoptosis	Tf	3	10.9	http://www.uniprot.org/uniprot/Q95976
MUML1_HUMAN	Q5H9M0 - MUML1_HUMAN	Exosoma vesicular extracelular, unido a membrana y antigeno mutado de melanoma	Trafico	1	10.9	http://www.uniprot.org/uniprot/Q5H9M0
NP_079193.2	Q96GJ1 - TRM2_HUMAN	Metiltransferasa de tRNA, involucrado en estabilización/maduración de tRNA	No especificado	5	10.9	http://www.uniprot.org/uniprot/Q96GJ1
FA46D_HUMAN	Q8NEK8 - FA46D_HUMAN	Proteína FAM46D, familia de antígenos cancerosos/testiculares	Testículos	1	10.9	http://www.uniprot.org/uniprot/Q8NEK8
ENST00000276105	Q9UMY4 - SNX12_HUMAN	Proteína posiblemente involucrada varios pasos de trafico intracelular, unión a PI3P	Trafico	2	10.9	http://www.uniprot.org/uniprot/Q9UMY4
Q9UMY4-2	Q9UMY4 - SNX12_HUMAN	Proteína posiblemente involucrada varios pasos de trafico intracelular, unión a PI3P	Trafico	1	10.9	http://www.uniprot.org/uniprot/Q9UMY4#Q9UMY4-2
NP_001033066.1	P22557 - HEMO_HUMAN	Enzima mitocondrial ALAS2, de ruta de síntesis de grupo HEMO	Circulatorio	2	10.9	http://www.uniprot.org/uniprot/P22557
ENST00000376358	A6NM71 - A6NM71_HUMAN	Proteína sin caracterizar	No especificado	1	10.9	http://www.uniprot.org/uniprot/A6NM71
WIPI4_HUMAN	Q9Y484 - WIPI4_HUMAN	Proteína importante en el proceso de autofagia, unión a PI3P	Cc	2	10.9	http://www.uniprot.org/uniprot/Q9Y484
Q9Y484-2	Q9Y484 - WIPI4_HUMAN	Isoforma 2. Proteína importante en el proceso de autofagia, unión a PI3P	Cc	1	10.9	http://www.uniprot.org/uniprot/Q9Y484#Q9Y484-2
DMD_HUMAN	P11532 - DMD_HUMAN	Distrofina presente en complejos sinápticos, transmisión, estabiliza sarcolema	Muscular	1	10.9	http://www.uniprot.org/uniprot/P11532
ENST00000399173	Q6ZNI6 - M3K15_HUMAN	Cinasa ampliamente distribuida involucrada en señalización activada en estrés celular	Apoptosis	1	10.9	http://www.uniprot.org/uniprot/Q6ZNI6
SCML2_HUMAN	Q9UQR0 - SCML2_HUMAN	Proteína PcG, que forma complejos proteicos para reprimir transcripción en desarrollo	Tf	1	10.9	http://www.uniprot.org/uniprot/Q9UQR0
Q5JQ76_HUMAN	Q02040 - AK17A_HUMAN	Factor regulador de splicing de pre-mRNA de manera PKA-dependiente	Tf	2	10.9	http://www.uniprot.org/uniprot/Q02040
Q02040-2	Q02040 - AK17A_HUMAN	Factor regulador de splicing de pre-mRNA de manera PKA-dependiente	Tf	1	10.9	http://www.uniprot.org/uniprot/Q02040#Q02040-2
NP_066652.1	Q15485 - FCN2_HUMAN	Ficolina 2, participa en la inmunidad innata activando vía de lectinas	Immune	1	10.9	http://www.uniprot.org/uniprot/Q15485
NP_443133.2	Q8KCM7 - SCMC2_HUMAN	Acameador mitocondrial de metabolitos, nucleótidos, cofactores, dependiente de Ca	Trafico	1	10.9	http://www.uniprot.org/uniprot/Q8KCM7
ZBT34_HUMAN	Q8N0N2 - ZBT34_HUMAN	Dedo de Zinc, posiblemente sea un represor transcripcional	Tf	1	10.9	http://www.uniprot.org/uniprot/Q8N0N2
Q5VZW9_HUMAN	Q5VZW9 - Q5VZW9_HUMAN	Receptor tirosin-cinasa muscular y esquelético, unión a ATP	Musculo	1	10.9	http://www.uniprot.org/uniprot/Q5VZW9
Q5T0B5_HUMAN	Q5T0B5 - Q5T0B5_HUMAN	Receptor tirosin-cinasa, del gen MUSK	Musculo	1	10.9	http://www.uniprot.org/uniprot/Q5T0B5
MUSK_HUMAN	Q15146 - MUSK_HUMAN	Receptor tirosin-cinasa, participa en formación y mantenimiento sinapsis neuromuscular	Musculo	3	10.9	http://www.uniprot.org/uniprot/Q15146
Q5T7X2_HUMAN	Q09161 - NCBP1_HUMAN	Subunidad del complejo CBC, de unión al extremo 5'-cap de pre-mRNA	Tf	1	10.9	http://www.uniprot.org/uniprot/Q09161
FA22G_HUMAN	Q5VZR2 - NTM2G_HUMAN	Proteína de la familia NUT miembro 2G, del cromosoma 9	No especificado	2	10.9	http://www.uniprot.org/uniprot/Q5VZR2
Q96N04_HUMAN	Q8I241 - RASEF_HUMAN	Proteína de unión predominante a GDP, también GTP	Cc	2	10.9	http://www.uniprot.org/uniprot/Q8I241
NP_001001670.1	Q6ZQ02 - S31D1_HUMAN	Proteína de membrana que puede estar asociada a espermatogénesis	Testículos	1	10.9	http://www.uniprot.org/uniprot/Q6ZQ02
NP_986831.1	Q9HCF6 - TRPM3_HUMAN	Canal de Calcio, presente primordialmente en riñón, y zonas específicas de cerebro	Renal	2	10.9	http://www.uniprot.org/uniprot/Q9HCF6
FRDA_HUMAN	Q16595 - FRDA_HUMAN	Frataxina mitocondrial, promueve formación de grupo HEMO, función oxidoreductasa	Corazón	1	10.9	http://www.uniprot.org/uniprot/Q16595
OR2S1_HUMAN	Q9NQN1 - OR2S1_HUMAN	Receptor olfatorio 2S2 acoplado a proteína G, percepción de aroma	Nervioso	1	10.9	http://www.uniprot.org/uniprot/Q9NQN1
Q6ZNR2_HUMAN	Q6ZNR2 - DUTL_HUMAN	Proteína no identificada, posiblemente dUTPasa	Cc	1	10.9	http://www.uniprot.org/uniprot/Q6ZNR2
DMRTA_HUMAN	Q5VZB9 - DMRTA_HUMAN	Factor de transcripción, involucrado en diferenciación sexual; diversos organos	Tf	1	10.9	http://www.uniprot.org/uniprot/Q5VZB9
Q5T284_HUMAN	Q5T284 - Q5T284_HUMAN	Proteína de subunidad 3 del complejo de activación de snRNA	Tf	1	10.9	http://www.uniprot.org/uniprot/Q5T284
Q5H8C1-4	Q5H8C1 - FREM1_HUMAN	Proteína de MEC, función en diferenciación/adhesión epidermal en desarrollo embrionario	Epidemis	1	10.9	http://www.uniprot.org/uniprot/Q5H8C1#Q5H8C1-4
Q5VW31_HUMAN	Q5VW31 - Q5VW31_HUMAN	Factor transcripcional de unión específica a secuencias DNA, para replicación	Tf	1	10.9	http://www.uniprot.org/uniprot/Q5VW31
Q4F7X0_HUMAN	Q4F7X0 - Q4F7X0_HUMAN	Proteína integral, adenilato ciclasa, señalización IC de transcripción	Cc	1	10.9	http://www.uniprot.org/uniprot/Q4F7X0
ADCY8_HUMAN	P40145 - ADCY8_HUMAN	Proteína integral, adenilato ciclasa de unión a membrana, Ca dependiente	Nervioso	1	10.9	http://www.uniprot.org/uniprot/P40145
CH032_HUMAN	Q96HA8 - NTAQ1_HUMAN	N-terminal glutamina aminohidrolasa, involucrada en degradación de proteínas	Cc	1	10.9	http://www.uniprot.org/uniprot/Q96HA8
TR11B_HUMAN	Q00300 - TR11B_HUMAN	Receptor de TNF, actúa sobre osteoclastos en homeostasis ósea, ampliamente distribuido	Oseo	1	10.9	http://www.uniprot.org/uniprot/Q00300
OXR1_HUMAN	Q8N573 - OXR1_HUMAN	Proteína posiblemente involucrada en protección contra estrés oxidativo	Trafico	1	10.9	http://www.uniprot.org/uniprot/Q8N573
WD21C_HUMAN	Q8NA75 - DC4L2_HUMAN	Enzima y rutas no especificadas	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q8NA75
Q8NM7_HUMAN	Q49A92 - CH034_HUMAN	Proteína no caracterizada	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q49A92
CP7A1_HUMAN	P22680 - CP7A1_HUMAN	Enzima involucrada en metabolismo de lípidos y homeostasis del colesterol	Higado	1	10.9	http://www.uniprot.org/uniprot/P22680
CH041_HUMAN	Q8NVR4 - TTI2_HUMAN	Regulador de respuesta de resistencia a daño de DNA	Trafico	1	10.9	http://www.uniprot.org/uniprot/Q8NVR4
ADA1A_HUMAN	P35348 - ADA1A_HUMAN	Receptor alfa adrenérgico, señal nuclear en miocitos cardiacos	Corazón	1	10.9	http://www.uniprot.org/uniprot/P35348
RPO3D_HUMAN	P05423 - RPO3D_HUMAN	Subunidad de RNA polimerasa III, participando en procesos diversos procesos	Cc	1	10.9	http://www.uniprot.org/uniprot/P05423
ARY2_HUMAN	P11245 - ARY2_HUMAN	Arylamino N-acetiltransferasa, participa en detoxificación y activación de carcinógenos	Trafico	1	10.9	http://www.uniprot.org/uniprot/P11245
Q86TP7_HUMAN	A4D0S4 - LAMB4_HUMAN	Subunidad de Laminina, adhesión, migración y organización celular de distintos tejidos	Trafico	1	10.9	http://www.uniprot.org/uniprot/A4D0S4
PR57_HUMAN	P35998 - PR57_HUMAN	Subunidad de proteasa dependiente de ATP, interacción en infecciones víricas	Cc	1	10.9	http://www.uniprot.org/uniprot/P35998
Q6P5S9_HUMAN	Q6P5S9 - Q6P5S9_HUMAN	Proteína chaperona	Cc	1	10.9	http://www.uniprot.org/uniprot/Q6P5S9
Q96543-2	Q96543 - DNJC2_HUMAN	Componente del complejo RAC, actúa tanto como chaperona como regulador de transcripción	Tf	1	10.9	http://www.uniprot.org/uniprot/Q96543#Q96543-2
Q9UD82_HUMAN	P54289 - CA2D1_HUMAN	Subunidad alfa-2/delta de canal de calcio dependiente de voltaje, regula activación	Muscular	1	10.9	http://www.uniprot.org/uniprot/P54289
Q92556-2	Q92556 - ELMO1_HUMAN	Proteína involucrada en remodelación de membrana en fagocitosis y movilidad	Immune	1	10.9	http://www.uniprot.org/uniprot/Q92556#Q92556-2
JAZF1_HUMAN	Q86VZ6 - JAZF1_HUMAN	Potencial factor de transcripción, proteína yuxtuesta a dedo de zinc	Tf	1	10.9	http://www.uniprot.org/uniprot/Q86VZ6
SNX10_HUMAN	Q9Y5X0 - SNX10_HUMAN	Proteína de unión a PI, involucrada en separación de proteínas y trafico de endosomas	Oseo	1	10.9	http://www.uniprot.org/uniprot/Q9Y5X0
Q96SQ3_HUMAN	Q8N2E2 - VVDE_HUMAN	Proteína secretada, factor D von Willebrand y dominio EGF	Sanguíneo	1	10.9	http://www.uniprot.org/uniprot/Q8N2E2
ARI1B_HUMAN	Q8NFD5 - ARI1B_HUMAN	Involucrado en transcripción y remodelación DNA, complejo nBAF en células nerviosas	Corazón	1	10.9	http://www.uniprot.org/uniprot/Q8NFD5
Q8NFD5-3	Q8NFD5 - ARI1B_HUMAN	Involucrado en transcripción y remodelación DNA, complejo nBAF en células nerviosas	Corazón	1	10.9	http://www.uniprot.org/uniprot/Q8NFD5#Q8NFD5-3
NOX3_HUMAN	Q9HBY0 - NOX3_HUMAN	NAPDH oxidasa, biogénesis de estructuras de oído interno, productor de superóxido	Apoptosis	1	10.9	http://www.uniprot.org/uniprot/Q9HBY0

HDC_HUMAN	Q8UB19 - HDC_HUMAN	Puede tener papel en cánceres humanos, regulador de desarrollo de redes epiteliales	MULTIPLES	1	10.9	http://www.uniprot.org/uniprot/Q8UB19
OR2A4_HUMAN	Q95047 - OR2A4_HUMAN	Receptor olfatorio	Nervioso	2	10.9	http://www.uniprot.org/uniprot/Q95047
CF058_HUMAN	Q6P5S2 - LEG1H_HUMAN	Proteína secretasa que puede estar involucrada en el desarrollo temprano de hígado	Trafico	1	10.9	http://www.uniprot.org/uniprot/Q6P5S2
GOPC_HUMAN	Q9HD26 - GOPC_HUMAN	Proteína asociada a tráfico proteico y degradación, presente en citoplasma y Golgi	Tf	1	10.9	http://www.uniprot.org/uniprot/Q9HD26
Q9HD26-2	Q9HD26 - GOPC_HUMAN	Proteína asociada a tráfico proteico y degradación, presente en citoplasma y Golgi	Tf	1	10.9	http://www.uniprot.org/uniprot/Q9HD26#Q9HD26-2
O510W8_HUMAN	Q5TCS8 - KAD9_HUMAN	Involucrada interconversión de nucleosido fosfatasa, actividad cinasa	Cc	1	10.9	http://www.uniprot.org/uniprot/Q5TCS8
NP_071374.1	Q8N3C0 - ASOC3_HUMAN	Subunidad de DNA Helicasa ATP dependiente involucrada en reparación de acetilación	Cc	2	10.9	http://www.uniprot.org/uniprot/Q8N3C0
Q8WVX4_HUMAN	Q8WVX4 - Q8WVX4_HUMAN	Expresada en desarrollo de hipotálamo fetal, mandaselina	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q8WVX4
ENS00000369293	X6R7A2 - X6R7A2_HUMAN	Glucoproteína mannosidasa, mandaselina	No especificado	1	10.9	http://www.uniprot.org/uniprot/X6R7A2
MANEA_HUMAN	Q5SR19 - MANEA_HUMAN	Glucoproteína mannosidasa, en aparato de Golgi	Hígado	1	10.9	http://www.uniprot.org/uniprot/Q5SR19
Q5W1L7_HUMAN	Q5W1L7 - Q5W1L7_HUMAN	Transportador de nucleosidos presente en membrana de Golgi	Trafico	1	10.9	http://www.uniprot.org/uniprot/Q5W1L7
S35A1_HUMAN	P78382 - S35A1_HUMAN	Transportador de CMP-ácido siálico desde citosol a vesículas del Golgi	Trafico	1	10.9	http://www.uniprot.org/uniprot/P78382
Q8N675_HUMAN	Q8N675 - Q8N675_HUMAN	Proteína CLIC5	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q8N675
Q5VW79_HUMAN	Q5VW79 - Q5VW79_HUMAN	Promotor de transcripción por RNA polimerasa II	Tf	1	10.9	http://www.uniprot.org/uniprot/Q5VW79
Q81WV8-3	Q81WV8 - UBR2_HUMAN	Promueve la ubiquitinación N-terminal, inactivación y transcripción nivel histona	Musculo	1	10.9	http://www.uniprot.org/uniprot/Q81WV8#Q81WV8-3
Q9P146_HUMAN	Q9P146 - Q9P146_HUMAN	No especificado	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q9P146
Q96QC4_HUMAN	Q96QC4 - Q96QC4_HUMAN	Polipéptido relacionado a MHC clase I	Immune	1	10.9	http://www.uniprot.org/uniprot/Q96QC4
Q6UXQ7_HUMAN	Q6UXQ7 - Q6UXQ7_HUMAN	No especificado	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q6UXQ7
MAS1L_HUMAN	P35410 - MAS1L_HUMAN	Receptor acoplado a proteína G	No especificado	3	10.9	http://www.uniprot.org/uniprot/P35410
OR2J1_HUMAN	Q9GZK6 - OR2J1_HUMAN	Receptor olfatorio, acoplado a proteína G	Nervioso	2	10.9	http://www.uniprot.org/uniprot/Q9GZK6
OR2B3_HUMAN	Q76000 - OR2B3_HUMAN	Receptor olfatorio putativo	Nervioso	2	10.9	http://www.uniprot.org/uniprot/Q76000
ENS00000378054	E2QR9F9 - E2QR9F9_HUMAN	Corepresor de transcripción, regulador negativo de ciclo celular, organogenesis	Tf	1	10.9	http://www.uniprot.org/uniprot/E2QR9F9
ENS00000378059	H7C608 - H7C608_HUMAN	Corepresor de transcripción, regulador negativo de ciclo celular, organogenesis	Tf	1	10.9	http://www.uniprot.org/uniprot/H7C608
ENS00000320066	Q75496 - GEM1_HUMAN	Inhibe replicación DNA, regulador negativo de ciclo celular, organogenesis	Tf	2	10.9	http://www.uniprot.org/uniprot/Q75496
FAB8A1_HUMAN	Q8UBU6 - FAB8A1_HUMAN	Proteína integral altamente conservada	Testiculos	1	10.9	http://www.uniprot.org/uniprot/Q8UBU6
Q96FF1_HUMAN	Q96FF1 - Q96FF1_HUMAN	Proteína ATXN1	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q96FF1
Q96C81_HUMAN	Q96C81 - Q96C81_HUMAN	Proteína ATXN1	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q96C81
ATX1_HUMAN	P54253 - ATX1_HUMAN	Factor de unión a cromatina, promotor de transcripción, corepresor de CBF1	Tf	1	10.9	http://www.uniprot.org/uniprot/P54253
Z354A_HUMAN	Q60765 - Z354A_HUMAN	Dedo de zinc, factor de transcripción promotor	Riñon	1	10.9	http://www.uniprot.org/uniprot/Q60765
GLRA1_HUMAN	P23415 - GLRA1_HUMAN	Receptor de glicina, canal de cloro	MULTIPLES, Musculo	1	10.9	http://www.uniprot.org/uniprot/P23415
SG3A2_HUMAN	Q96PL1 - SG3A2_HUMAN	Proteína extracelular mediadora de receptor para endocitosis	Pulmon	1	10.9	http://www.uniprot.org/uniprot/Q96PL1
JKIP2_HUMAN	Q96A48 - JKIP2_HUMAN	Cinasa que interactúa con microtubulos en Golgi	Cerebro	2	10.9	http://www.uniprot.org/uniprot/Q96A48
KCD16_HUMAN	Q68DU8 - KCD16_HUMAN	Subunidad auxiliar de receptores GABA-B, regula señalización de proteína G	Nervioso	1	10.9	http://www.uniprot.org/uniprot/Q68DU8
NFIP1_HUMAN	Q9BT67 - NFIP1_HUMAN	Activa ligasas, mediadora de ubiquitinación, control de crecimiento celular	Immune	1	10.9	http://www.uniprot.org/uniprot/Q9BT67
Q9BT67-2	Q9BT67 - NFIP1_HUMAN	Activa ligasas, mediadora de ubiquitinación, control de crecimiento celular	Immune	1	10.9	http://www.uniprot.org/uniprot/Q9BT67#Q9BT67-2
SNX2_HUMAN	Q60749 - SNX2_HUMAN	Involucrada en varias etapas de tráfico celular, unión a receptores, transporte	Trafico	1	10.9	http://www.uniprot.org/uniprot/Q60749
YTDC2_HUMAN	Q9H6S0 - YTDC2_HUMAN	Probable helicasa RNA dependiente de ATP	Immune	1	10.9	http://www.uniprot.org/uniprot/Q9H6S0
Q08499-9	Q08499 - PDE4D_HUMAN	Isoforma 9, Fosfodiesterasa, hidroliza cAMP	Musculo	1	10.9	http://www.uniprot.org/uniprot/Q08499#Q08499-9
CE029_HUMAN	Q8N292 - GAPT_HUMAN	Regula negativamente la proliferación de linfocitos B, mantenimiento de zona marginal	Immune, Bazo	2	10.9	http://www.uniprot.org/uniprot/Q8N292
FST_HUMAN	P19883 - FST_HUMAN	Antagonista de activina por unión, Inhibidor de síntesis y secreción de FSH	Trafico	1	10.9	http://www.uniprot.org/uniprot/P19883
NP_001005217.1	Q64ET8 - FRG2_HUMAN	Proteína FRG2	Musculo	1	10.9	http://www.uniprot.org/uniprot/Q64ET8
DCTD_HUMAN	P32321 - DCTD_HUMAN	Suple de sustrato nucleotídico a timidilato sintetasa, dCMP desaminasa	Cc	1	10.9	http://www.uniprot.org/uniprot/P32321
VEGFC_HUMAN	P49767 - VEGFC_HUMAN	VEGFC, factor de angiogenesis y sistema vascular linfático, actuando en embriogenesis	Linfatico	1	10.9	http://www.uniprot.org/uniprot/P49767
NPY5R_HUMAN	Q15761 - NPY5R_HUMAN	Receptor de neuropeptido Y, inhibe actividad adenilato ciclasa, desordenes alimenticios	Cerebro	1	10.9	http://www.uniprot.org/uniprot/Q15761
NP_115493.1	Q9BWT6 - MMD1_HUMAN	Requerido en pareamiento de cromosomas homólogos y recombinación en meiosis	Cc	1	10.9	http://www.uniprot.org/uniprot/Q9BWT6
SMAD1_HUMAN	Q15797 - SMAD1_HUMAN	Modulador transcripcional modulado por receptores tipo 1 BMP	Musculo	1	10.9	http://www.uniprot.org/uniprot/Q15797
NP_055945.2	Q6ZT07 - TBCD9_HUMAN	Puede actuar como GTPasa, familia RAB	Cc	1	10.9	http://www.uniprot.org/uniprot/Q6ZT07
UCP1_HUMAN	P25874 - UCP1_HUMAN	Transportador mitocondrial que genera un gradiente de protones	Adiposo Pardo	1	10.9	http://www.uniprot.org/uniprot/P25874
Q15723-1	Q15723 - ELF2_HUMAN	Isoforma 1, activa promotores para la transcripción de promotor BLK, RNA polimerasa III	Tf	1	10.9	http://www.uniprot.org/uniprot/Q15723
Q59FC9_HUMAN	Q59FC9 - Q59FC9_HUMAN	No especificado	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q59FC9
P37840-3	P37840 - SYUA_HUMAN	Puede que regule liberación y transporte de dopamina, lleva a baja activación caspasa3	Cerebro	1	10.9	http://www.uniprot.org/uniprot/P37840
Q06330-6	Q06330 - SUH_HUMAN	Regulador transcripcional en señalización Notch, regula destino celular	Tf	1	10.9	http://www.uniprot.org/uniprot/Q06330#Q06330-6
Q9P1K3_HUMAN	Q9P1K3 - Q9P1K3_HUMAN	No especificado	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q9P1K3
ADCL1_HUMAN	Q6PIU2 - NCEH1_HUMAN	Esterhidrolasa, metabolismo/degradación lipídica, relacionada a aterosclerosis	Immune	1	10.9	http://www.uniprot.org/uniprot/Q6PIU2
Q0D2K5_HUMAN	Q0D2K5 - EGFEM_HUMAN	No especificado, proteína de pseudogen, EGF-like	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q0D2K5
PP2CL_HUMAN	Q5SGD2 - PPM1L_HUMAN	Protein fosfatasa, actúa como supresor de vías de señalización de SAPK	Corazon	1	10.9	http://www.uniprot.org/uniprot/Q5SGD2
Q3KQ56_HUMAN	Q3KQ56 - Q3KQ56_HUMAN	Metaenoldehidrogenasa, cromosoma 3	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q3KQ56
NEP_HUMAN	P08473 - NEP_HUMAN	Peptidasa específica, que degrada opioides, FNA, angiotensina 1-9	Nervioso	1	10.9	http://www.uniprot.org/uniprot/P08473
TM108_HUMAN	Q6UXF1 - TM108_HUMAN	Proteína transmembranal 108.	Nervioso	1	10.9	http://www.uniprot.org/uniprot/Q6UXF1
PXR_HUMAN	Q75469 - NR1I2_HUMAN	Receptor nuclear, transcripción de genes de metabolismo y secreción de xenobioticos	Hígado	2	10.9	http://www.uniprot.org/uniprot/Q75469
LSAMP_HUMAN	Q13449 - LSAMP_HUMAN	Regula crecimiento neuronal específico y axonal, remodela circuitos de sistema límbico	Nervioso	1	10.9	http://www.uniprot.org/uniprot/Q13449
SYNPR_HUMAN	Q8TBG9 - SYNPR_HUMAN	Proteína de membrana de vesículas sinápticas.	Nervioso	1	10.9	http://www.uniprot.org/uniprot/Q8TBG9
NP_940865.1	Q6ZV76 - CO067_HUMAN	Proteína no caracterizada C3orf67, de cromosoma 3	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q6ZV76
Q7Z2D7_HUMAN	Q7Z2D7 - Q7Z2D7_HUMAN	Proteína relacionada a peroxisomas	Trafico	1	10.9	http://www.uniprot.org/uniprot/Q7Z2D7
PEX26_HUMAN	Q7Z412 - PEX26_HUMAN	Probablemente requerida en importación de peroxisomas, importación de catalasa	Trafico	1	10.9	http://www.uniprot.org/uniprot/Q7Z412
O11H1_HUMAN	Q8NG94 - O11H1_HUMAN	Receptor olfatorio	Nervioso	1	10.9	http://www.uniprot.org/uniprot/Q8NG94
Q3L168_HUMAN	Q3L168 - KR222_HUMAN	Proteína asociada a queratina, presente en matriz de corteza capilar.	Epitelial	1	10.9	http://www.uniprot.org/uniprot/Q3L168
TCQP_HUMAN	P50990 - TCQP_HUMAN	Chaperona, asiste plegado de proteínas mediante acción ATPasa	Trafico	1	10.9	http://www.uniprot.org/uniprot/P50990
Q96B65_HUMAN	Q96B65 - Q96B65_HUMAN	Ubiquitinil hidrolasa, de cromosoma 21	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q96B65
Q9UHP3-2	Q9UHP3 - UBP25_HUMAN	Desubiquitinasa, reciclado de ubiquitina o edición de poliubiquitina, isoforma USP25a	MULTIPLE	1	10.9	http://www.uniprot.org/uniprot/Q9UHP3#Q9UHP3-2
UBP25_HUMAN	Q9UHP3 - UBP25_HUMAN	Desubiquitinasa, reciclado de ubiquitina o edición de poliubiquitina	Musculo	1	10.9	http://www.uniprot.org/uniprot/Q9UHP3
Q9NSJ0_HUMAN	Q9NSJ0 - Q9NSJ0_HUMAN	Factor de crecimiento de fibroblastos (FGF)	Cc	1	10.9	http://www.uniprot.org/uniprot/Q9NSJ0
SNAI_HUMAN	Q95863 - SNAI1_HUMAN	Involucrado en transición epitelial-mesenchimal, formación de mesodermo, factor trans.	Riñon	1	10.9	http://www.uniprot.org/uniprot/Q95863
ARP5_HUMAN	Q9H9F9 - ARP5_HUMAN	Parte de complejo regulador transcripcional, replicación y posible reparo de DNA	Cc	1	10.9	http://www.uniprot.org/uniprot/Q9H9F9
CTBL1_HUMAN	Q8WYA6 - CTBL1_HUMAN	Parte de complejo requerido para splicing pre-mRNA, puede inducir apoptosis	Musculo	1	10.9	http://www.uniprot.org/uniprot/Q8WYA6
Q8WYA6-3	Q8WYA6 - CTBL1_HUMAN	Parte de complejo requerido para splicing pre-mRNA, puede inducir apoptosis, isoforma3	Musculo	1	10.9	http://www.uniprot.org/uniprot/Q8WYA6#Q8WYA6-3
Q95330_HUMAN	Q95330 - Q95330_HUMAN	No caracterizada, posible actividad en transporte proteico de vesículas en Golgi-RE	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q95330
ERG3_HUMAN	Q9Y282 - ERG3_HUMAN	Posible rol en transporte compartimental entre Golgi-RE	Trafico	1	10.9	http://www.uniprot.org/uniprot/Q9Y282
Q9Y282-3	Q9Y282 - ERG3_HUMAN	Posible rol en transporte compartimental entre Golgi-RE, isoforma 3	Trafico	1	10.9	http://www.uniprot.org/uniprot/Q9Y282#Q9Y282-3
Q5JWS1_HUMAN	Q5JWS1 - Q5JWS1_HUMAN	Proteína presente en Golgi y RE	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q5JWS1

Q8Y282-2	Q8Y282 - ERG3_HUMAN	Posible rol en transporte compartimental entre Golgi-RE, isoforma 2	Tráfico	1	10.9	http://www.uniprot.org/uniprot/Q8Y282#Q8Y282-2
ENST00000379276	Q5JWB9 - Q5JWB9_HUMAN	Proteína transmembranal 230	Cc	1	10.9	http://www.uniprot.org/uniprot/Q5JWB9
SMOX_HUMAN	Q9NWM0 - SMOX_HUMAN	Flavoenzima, oxidación de espermina a espermidina, concentración IC de poliaminas	Tumoral	1	10.9	http://www.uniprot.org/uniprot/Q9NWM0
Q9NWM0-3	Q9NWM0 - SMOX_HUMAN	Flavoenzima, oxidación de espermina a espermidina, la mayor afinidad por 3 substratos	Tumoral	1	10.9	http://www.uniprot.org/uniprot/Q9NWM0#Q9NWM0-3
Q9NWM0-2	Q9NWM0 - SMOX_HUMAN	Flavoenzima, oxidación de espermina a espermidina, baja afinidad por poliaminas	Tumoral	1	10.9	http://www.uniprot.org/uniprot/Q9NWM0#Q9NWM0-2
Q9NWM0-4	Q9NWM0 - SMOX_HUMAN	Flavoenzima, oxidación de espermina a espermidina, mayor afinidad espermidina, bajos Km	Tumoral	1	10.9	http://www.uniprot.org/uniprot/Q9NWM0#Q9NWM0-4
Q9NWJ7_HUMAN	Q9NWJ7 - Q9NWJ7_HUMAN	Proteína cDNA FLJ20802.fis	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q9NWJ7
NUPL_HUMAN	P52594 - AGFG1_HUMAN	Requerida para fusión de vesículas, involucrada en tráfico de RNA y VIH replicación	Tráfico	1	10.9	http://www.uniprot.org/uniprot/P52594
P52594-2	P52594 - AGFG1_HUMAN	Requerida para fusión de vesículas, involucrada en tráfico de RNA y VIH replicación	Tráfico	1	10.9	http://www.uniprot.org/uniprot/P52594#P52594-2
Q2VIK4_HUMAN	Q2VIK4 - Q2VIK4_HUMAN	Endopeptidasa involucrada en catálisis proteica celular, parte del proteosoma	Tráfico	1	10.9	http://www.uniprot.org/uniprot/Q2VIK4
CB021_HUMAN	Q8N2C7 - UNC80_HUMAN	Componente homólogo de canal de sodio NALCN	Nervioso	1	10.9	http://www.uniprot.org/uniprot/Q8N2C7
NP_001074014.1	Q8WV17 - PPR1C_HUMAN	Subunidad de inhibidor de protein-fosfatasa 1, promueve progreso de G1/S	Musculo	1	10.9	http://www.uniprot.org/uniprot/Q8WV17
SSFA2_HUMAN	P28290 - SSFA2_HUMAN	Antígeno 2 específico de espermia, unión a actina	Testiculos	1	10.9	http://www.uniprot.org/uniprot/P28290
P28290-3	P28290 - SSFA2_HUMAN	Antígeno 2 específico de espermia, unión a actina, isoforma 3	Testiculos	1	10.9	http://www.uniprot.org/uniprot/P28290#P28290-3
SCN8A_HUMAN	Q15858 - SCN8A_HUMAN	Subunidad alfa de canal de Na dependiente de voltaje, rol en inflamación y dolor	Nervioso	1	10.9	http://www.uniprot.org/uniprot/Q15858
Q15858-2	Q15858 - SCN8A_HUMAN	Subunidad alfa de canal de Na dependiente de voltaje, rol en inflamación y dolor	Nervioso	1	10.9	http://www.uniprot.org/uniprot/Q15858#Q15858-2
Q86VD6_HUMAN	Q86VD6 - Q86VD6_HUMAN	No especificado, proteína LOC339742	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q86VD6
Q580R0-2	Q580R0 - CB027_HUMAN	Proteína no caracterizada C2orf27	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q580R0#Q580R0-2
SL9A2_HUMAN	Q9UBY0 - SL9A2_HUMAN	Regulación de pH por intercambio de Na/H, absorción de sodio	Musculo	1	10.9	http://www.uniprot.org/uniprot/Q9UBY0
Q5MD60_HUMAN	Q5MD60 - Q5MD60_HUMAN	Cinasa treonina/serina, regulador GTPasa	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q5MD60
MAK4_HUMAN	Q95819 - MAK4_HUMAN	HCK-S, serina/treonina cinasa, respuesta a estrés y TNF-alfa	MULTIPLES	1	10.9	http://www.uniprot.org/uniprot/Q95819
Q95819-4	Q95819 - MAK4_HUMAN	HCK-S, serina/treonina cinasa, respuesta a estrés y TNF-alfa, isoforma 4	Musculo	1	10.9	http://www.uniprot.org/uniprot/Q95819#Q95819-4
Q95819-5	Q95819 - MAK4_HUMAN	HCK-S, serina/treonina cinasa, respuesta a estrés y TNF-alfa, isoforma 5	Cerebro	1	10.9	http://www.uniprot.org/uniprot/Q95819#Q95819-5
Q95819-3	Q95819 - MAK4_HUMAN	HCK-S, serina/treonina cinasa, respuesta a estrés y TNF-alfa, isoforma 3	MULTIPLE	1	10.9	http://www.uniprot.org/uniprot/Q95819#Q95819-3
GCF_HUMAN	P16383 - GCF2_HUMAN	Factor represor de transcripción, unión a promotores de EPOF y beta-actina, splicing	TF	1	10.9	http://www.uniprot.org/uniprot/P16383
MSH6_HUMAN	P52701 - MSH6_HUMAN	Parte del sistema reparador post-replicativo, junto con MSH2, en G1 e inicio de S	TF	1	10.9	http://www.uniprot.org/uniprot/P52701
Q8LAI4_HUMAN	Q8LAI4 - Q8LAI4_HUMAN	Proteína espermática, involucrada en reparación de desajustes de DNA	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q8LAI4
DPY30_HUMAN	Q9C005 - DPY30_HUMAN	Parte del complejo MLL1/MLL, involucrada en metilación de histonas H3, epigenético	TF	2	10.9	http://www.uniprot.org/uniprot/Q9C005
Q9NZM3-2	Q9NZM3 - ITSN2_HUMAN	Proteína adaptadora, unión indirecta entre tráfico endocítico y producción de actina	TF	1	10.9	http://www.uniprot.org/uniprot/Q9NZM3#Q9NZM3-2
WDR35_HUMAN	Q9P2L0 - WDR35_HUMAN	Componente del complejo IFT-A, para degradación cililar y cilio genesis, apoptosis TNF	Apoptosis	1	10.9	http://www.uniprot.org/uniprot/Q9P2L0
Q9P2L0-2	Q9P2L0 - WDR35_HUMAN	Componente del complejo IFT-A, para degradación cililar y cilio genesis, apoptosis TNF	Apoptosis	2	10.9	http://www.uniprot.org/uniprot/Q9P2L0#Q9P2L0-2
ZNZ74_HUMAN	Q96GC6 - ZNZ74_HUMAN	Probable represor transcripcional, unión y acción específica en SETDB1 y TRIM28/KAP1	TF	1	10.9	http://www.uniprot.org/uniprot/Q96GC6
Q96GC6-2	Q96GC6 - ZNZ74_HUMAN	Probable represor transcripcional, unión y acción específica en SETDB1 y TRIM28/KAP1	TF	1	10.9	http://www.uniprot.org/uniprot/Q96GC6#Q96GC6-2
NP_689571.2	Q6UXV1 - IZUM2_HUMAN	Proteína integral espermática de unión a membrana de óvulo	Testiculos	1	10.9	http://www.uniprot.org/uniprot/Q6UXV1
CRX_HUMAN	Q43186 - CRX_HUMAN	Factor transcripcional sinérgico con NRL, RORB y RAX para células fotorreceptoras	Retina	1	10.9	http://www.uniprot.org/uniprot/Q43186
NP_001027546.1	Q9NYT6 - ZN26_HUMAN	Dedo de zinc, posible regulación transcripcional	TF	1	10.9	http://www.uniprot.org/uniprot/Q9NYT6
O75294_HUMAN	O75294 - O75294_HUMAN	Apertura de marco de lectura de cromosoma 19, de complejo CatSper	TF	2	10.9	http://www.uniprot.org/uniprot/O75294
CS015_HUMAN	Q6ZRH7 - CTSRG_HUMAN	Involucrado con CATSPER1 en motilidad de espermia y su hiperactivación	Testiculos	1	10.9	http://www.uniprot.org/uniprot/Q6ZRH7
ZN682_HUMAN	Q95780 - ZN682_HUMAN	Proteína dedo de zinc, probable regulador de transcripción	TF	1	10.9	http://www.uniprot.org/uniprot/Q95780
Q5JPJ7_HUMAN	Q9HCS2 - CP4FC_HUMAN	Citocromo P450 4F12, hidroxilación de leucotrieno B4, ácido araquidónico y ebastina	Intestino delgado	1	10.9	http://www.uniprot.org/uniprot/Q9HCS2
DUS3L_HUMAN	Q96G46 - DUS3L_HUMAN	Cataliza síntesis de dihidrouridina, encontrada en tRNA	Cc	1	10.9	http://www.uniprot.org/uniprot/Q96G46
Q95323_HUMAN	Q95323 - Q95323_HUMAN	Proteína no caracterizada, tRNA dihidrouridina sintasa	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q95323
Q96G46-3	Q96G46 - DUS3L_HUMAN	Cataliza síntesis de dihidrouridina, encontrada en tRNA, isoforma 3	Cc	1	10.9	http://www.uniprot.org/uniprot/Q96G46#Q96G46-3
Q8TDF5-2	Q8TDF5 - NETO1_HUMAN	Rol en desarrollo/mantenimiento de redes neuronales, memoria y aprendizaje visual	Retina	1	10.9	http://www.uniprot.org/uniprot/Q8TDF5#Q8TDF5-2
CAD19_HUMAN	Q9H159 - CAD19_HUMAN	Caheirina 19, adhesión celular	Cc	1	10.9	http://www.uniprot.org/uniprot/Q9H159
SPB10_HUMAN	P48695 - SPB10_HUMAN	Inhibidor de proteasas, podría regularlas durante hematopoyesis y apoptosis por TNF	Medula Osea	1	10.9	http://www.uniprot.org/uniprot/P48695
GALT1_HUMAN	Q10472 - GALT1_HUMAN	N-acetil-D-galactosamintransferasa, paso inicial de biosíntesis de oligosacáridos	Cc	1	10.9	http://www.uniprot.org/uniprot/Q10472
TTYH_HUMAN	P02766 - TTYH_HUMAN	Unión a hormonas tiroideas, posible transporte de tiroxina desde circulación a cerebro	Suero	1	10.9	http://www.uniprot.org/uniprot/P02766
Q12830-4	Q12830 - BPTF_HUMAN	Componente de NJRF, unión a histonas, desarrollo embrionario y cerebral	Testiculos	1	10.9	http://www.uniprot.org/uniprot/Q12830#Q12830-4
Q12830-2	Q12830 - BPTF_HUMAN	Componente de NJRF, unión a histonas, desarrollo embrionario y cerebral	Testiculos	1	10.9	http://www.uniprot.org/uniprot/Q12830#Q12830-2
EFCB3_HUMAN	Q8N7B9 - EFCB3_HUMAN	Proteína de dominio EF-hand, unión a Calcio	Muscular	1	10.9	http://www.uniprot.org/uniprot/Q8N7B9
Q8IVZ7_HUMAN	A2RUB1 - CQ104_HUMAN	Proteína sin caracterizar C17orf104	No especificado	1	10.9	http://www.uniprot.org/uniprot/A2RUB1
ENST00000318329	H7BXR9 - H7BXR9_HUMAN	Proteína 4-8 asociada a queratina	Epitelial	1	10.9	http://www.uniprot.org/uniprot/H7BXR9
ENST00000377723	H7BYC4 - H7BYC4_HUMAN	Proteína 9-2 asociada a queratina	Epitelial	1	10.9	http://www.uniprot.org/uniprot/H7BYC4
KR410_HUMAN	Q9BYQ7 - KRA41_HUMAN	Proteína 4-1 asociada a queratina, presente en matriz de corteza capilar	Folículo piloso	1	10.9	http://www.uniprot.org/uniprot/Q9BYQ7
ENST00000378634	H9KV77 - H9KV77_HUMAN	Metilación de histonas H4-R3	TF	1	10.9	http://www.uniprot.org/uniprot/H9KV77
Q9UHR3_HUMAN	Q9UHR3 - Q9UHR3_HUMAN	Proteína LZ16 de susceptibilidad a carcinoma nasofaríngeo	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q9UHR3
Q69GC3_HUMAN	Q69GC3 - Q69GC3_HUMAN	Componente de exosoma extracelular de vesícula	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q69GC3
NP_037407.4	Q6UB99 - ANR11_HUMAN	Desarrollo embrionario, oseo, morfogenesis facial, odontogenesis	TF	1	10.9	http://www.uniprot.org/uniprot/Q6UB99
Q6UX55_HUMAN	Q5XG92 - EST4A_HUMAN	Probable carboxilesterasa, secretada	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q5XG92
RHG17_HUMAN	Q68EM7 - RHG17_HUMAN	Proteína Rho GTPasa, unión entre células epiteliales y desarrollo de su polaridad	Corazón	1	10.9	http://www.uniprot.org/uniprot/Q68EM7
Q68EM7-3	Q68EM7 - RHG17_HUMAN	Proteína Rho GTPasa, unión entre células epiteliales y desarrollo de su polaridad	Corazón	1	10.9	http://www.uniprot.org/uniprot/Q68EM7#Q68EM7-3
Q68EM7-2	Q68EM7 - RHG17_HUMAN	Proteína Rho GTPasa, unión entre células epiteliales y desarrollo de su polaridad	Corazón	1	10.9	http://www.uniprot.org/uniprot/Q68EM7#Q68EM7-2
NP_665988.1	Q9NWB1 - RFOX1_HUMAN	Unión a RNA, regulando splicing, en especial durante entropoyesis	Musculo	1	10.9	http://www.uniprot.org/uniprot/Q9NWB1
ADCY9_HUMAN	Q60503 - ADCY9_HUMAN	Puede funcionar en situaciones de intercambio intracelular de Ca y cAMP, pulmones	Musculo	1	10.9	http://www.uniprot.org/uniprot/Q60503
Q38SD2-4	Q38SD2 - LRRK1_HUMAN	Serina/treonina cinasa	Cc	1	10.9	http://www.uniprot.org/uniprot/Q38SD2#Q38SD2-4
PDE8A_HUMAN	Q60658 - PDE8A_HUMAN	Hidrolasa de cAMP, posible regulación de cAMP en desarrollo de células germinativas	Testiculos/ovarios	1	10.9	http://www.uniprot.org/uniprot/Q60658
Q60658-4	Q60658 - PDE8A_HUMAN	Hidrolasa de cAMP, posible regulación de cAMP en desarrollo de células germinativas	Testiculos/ovarios	1	10.9	http://www.uniprot.org/uniprot/Q60658#Q60658-4
Q60658-2	Q60658 - PDE8A_HUMAN	Hidrolasa de cAMP, posible regulación de cAMP en desarrollo de células germinativas	Testiculos/ovarios	1	10.9	http://www.uniprot.org/uniprot/Q60658#Q60658-2
NP_001073904.1	Q8TF30 - WHAMM_HUMAN	Remodelación de membrana por polimerización de actina, morfología y posición de Golgi	Cerebro	1	10.9	http://www.uniprot.org/uniprot/Q8TF30
NP_001007123.1	A1L4K1 - FSD2_HUMAN	Dominio de fibronectina III y SPRY	No especificado	1	10.9	http://www.uniprot.org/uniprot/A1L4K1
NP_065722.1	Q5SYB0 - FRPD1_HUMAN	Estabiliza la unión a membrana de GPSM1, y promueve la interacción con GNAH1	Tráfico	1	10.9	http://www.uniprot.org/uniprot/Q5SYB0
NP_063960.1	Q9NQN1 - OR2S1_HUMAN	Receptor olfatorio	Nervioso	1	10.9	http://www.uniprot.org/uniprot/Q9NQN1
NP_001017969.1	Q6HYC2 - K2026_HUMAN	Proteína sin identificar KIAA2026	No especificado	1	10.9	http://www.uniprot.org/uniprot/Q6HYC2
NP_066968.2	Q9Y222 - DMTF1_HUMAN	Activador transcripcional en señalización Ras-Raf	Cc	1	10.9	http://www.uniprot.org/uniprot/Q9Y222
NP_059989.1	Q8NFD5 - AR1B_HUMAN	Involucrado en transcripción y remodelación DNA, complejo npBAF en células nerviosas	Corazón	1	10.9	http://www.uniprot.org/uniprot/Q8NFD5
NP_001010891.2	Q5HYI7 - MTX3_HUMAN	Podría funcionar en transporte de proteínas dentro de la mitocondria	Tráfico	1	10.9	http://www.uniprot.org/uniprot/Q5HYI7
NP_067008.2	Q6ZRH7 - CTSRG_HUMAN	Involucrado con CATSPER1 en motilidad de espermia y su hiperactivación	Testiculos	1	10.9	http://www.uniprot.org/uniprot/Q6ZRH7
NP_835229.2	Q8N9I5 - FADS6_HUMAN	Involucrada en metabolismo lipídico y biosíntesis, desaturasa	Cc	1	10.9	http://www.uniprot.org/uniprot/Q8N9I5
NP_060875.1	Q9NQ92 - COPRS_HUMAN	Proteína de unión a histonas H4, diferenciación en células musculares por PRMT5	Muscular	1	10.9	http://www.uniprot.org/uniprot/Q9NQ92

Basados en la literatura actual, fue posible construir una modesta lista de genes regulados por MEF2A, cuya detección se logró utilizando metodologías experimentales del tipo inmunoprecipitación de cromatina (ChIP). Todos los genes incluidos en dicha lista también fueron localizados por la metodología *in silico* desarrollada en la presente tesis. A continuación se mencionan las proteínas que resultan de la transcripción de dichos genes, las cuales fueron extraídas de la búsqueda basada en las matrices construidas utilizando las secuencias con un valor de ΔG_{bind} entre -3.9×10^2 kJ/mol y -4.4×10^2 kJ/mol, estas son : Nur77 -también nombrada como NR4A1- (Scheschonka, *et al.* 2007), MAML1 (Shen, *et al.* 2006), algunas proteínas del tipo HDAC's (Haberland *et al.* 2007), los factores de transcripción DLX5 y 6 (Gottlieb *et al.*, 2002), las proteínas COLQ y RUNX que participan en el desarrollo del sistema óseo (Arnold *et al.*, 2007), la proteína MMP10 (Chang *et al.*, 2006), la proteína TCRG1 inductora de la apoptosis (Woronicz *et al.*, 1995). También se encontraron por esta metodología las proteínas NKX, la cuál según el reporte de Karamboulas *et al.*, 2006, ve afectada su expresión al aumentar los niveles de los genes *mef2*- en el organismo. Estos resultados se muestran en la tabla 11.

Tabla 11. Proteínas cuya expresión por MEF2A se demostró mediante metodologías experimentales tipo ChIP y también fueron halladas determinando la ΔG_{bind} del complejo MEF2A-CARgbox. En la primer columna se reporta el valor de ΔG_{bind} que resultó de la unión entre el TF MEF2A y el elemento-cis cuya secuencia aparece en la segunda columna; también se reporta la ubicación corriente arriba del gen donde se localiza la putativa CARgbox, en la cuarta columna podemos ver el número de acceso con el que se puede ingresar al servidor de Uniprot en el que se recabo información de las proteínas que resultan de la transcripción de los genes en los que se localizaron las CARgbox reportadas, a esta información puede accederse por medio del *link* que aparece en la columna de referencia. En la quinta columna se describe el tejido en el cual se expresan las proteínas reportadas; en la sexta columna bajo el título de Cis-e se describen el número de cajas CARg putativas localizadas dentro del gen que codifica las proteínas y finalmente en la séptima columna se reporta el valor Ws que corresponde a la probabilidad que posee cada elementos-cis para ocurrir conforme al modelo propuesto por las PSSMs reportados en la sección 3.8.9 del capítulo III. Como puede observarse en la presente tabla, los resultados de la búsqueda basada en la determinación de la ΔG_{bind} calculada para cada complejo MEF2A-CARgbox son congruentes con los resultados obtenidos por metodologías experimentales como la precipitación de cromatina.

Valor de ΔG_{bind}	CARgbox	Ubicación	Número de acceso	Tejido	Cis-e	Ws	Referencia
-4.0e+02 kJ/mol	ATTTAAAAAT	-4723 a -4714	P22736 (NR4A1_HUMAN)	Nervioso	8	3.18	http://www.uniprot.org/uniprot/P22736
-4.3e+02 kJ/mol	ATATAAATTT	-5420 a -5411	P43354 (NR4A2_HUMAN)	Nervioso	10	8.02	http://www.uniprot.org/uniprot/P43354
-4.3e+02 kJ/mol	AAATTAATTT	-670 a -661	Q92570 (NR4A3_HUMAN)	Nervioso	4	7.64	http://www.uniprot.org/uniprot/Q92570
-3.9e+02 kJ/mol	ATTTTTATTT	-1396 a -1387	Q92585 (MAML1_HUMAN)	Óseo	1	2.61	http://www.uniprot.org/uniprot/Q92585
-4.3e+02 kJ/mol	ATATAATTTT	-3684 a -3675	Q13547 (HDAC1_HUMAN)	Músculo	2	7.24	http://www.uniprot.org/uniprot/Q13547
-4.3e+02 kJ/mol	AAATATATTT	-2756 a -2747	Q92769 (HDAC2_HUMAN)	Músculo	5	7.2	http://www.uniprot.org/uniprot/Q92769
-4.0e+02 kJ/mol	ATTTTAATTC	-3469 a -3460	P56524 (HDAC4_HUMAN)	Músculo	6	3.13	http://www.uniprot.org/uniprot/P56524
-4.3e+02 kJ/mol	ATATATATTT	-5459 a -5450	Q9UQL6 (HDAC5_HUMAN)	Músculo	3	7.78	http://www.uniprot.org/uniprot/Q9UQL6
-4.3e+02 kJ/mol	CTATAATAT	-831 a -822	Q9UBN7 (HDAC6_HUMAN)	Músculo	1	7.72	http://www.uniprot.org/uniprot/Q9UBN7
-4.3e+02 kJ/mol	CTAATAATTT	-2296 a -2287	Q9UKV0 (HDAC9_HUMAN)	Músculo	7	7	http://www.uniprot.org/uniprot/Q9UKV0
-4.0e+02 kJ/mol	ATTA AAAATC	-4962 a -4953	P56178 (DLX5_HUMAN)	Nervioso	7	3.07	http://www.uniprot.org/uniprot/P56178
-4.0e+02 kJ/mol	ATTACAATAT	-1206 a -1197	P56179 (DLX6_HUMAN)	Nervioso	1	3.07	http://www.uniprot.org/uniprot/P56179
-4.2e+02 kJ/mol	ATATAAATAT	-3506 a -3497	Q8TAU0 (NKX23_HUMAN)	Cardíaco	2	6	http://www.uniprot.org/uniprot/Q8TAU0
-4.0e+02 kJ/mol	ATTATAATAT	-4202 a -4193	Q9H2Z4 (NKX24_HUMAN)	Cardíaco	10	3.19	http://www.uniprot.org/uniprot/Q9H2Z4
-4.3e+02 kJ/mol	ATAAATATTT	-1773 a -1764	A6NCS4 (NKX26_HUMAN)	Cardíaco	2	7.19	http://www.uniprot.org/uniprot/A6NCS4
-4.0e+02 kJ/mol	ATTTAAAAAC	-3571 a -3562	O15522 (NKX28_HUMAN)	Cardíaco	5	3.1	http://www.uniprot.org/uniprot/O15522
-4.0e+02 kJ/mol	AATATAATTT	-4767 a -4758	Q99801 (NKX31_HUMAN)	Cardíaco	3	3.07	http://www.uniprot.org/uniprot/Q99801
-4.0e+02 kJ/mol	ATTTTAATTT	-1688 a -1679	P78367 (NKX32_HUMAN)	Cardíaco	16	3.16	http://www.uniprot.org/uniprot/P78367
-4.0e+02 kJ/mol	ATTTTAATTT	-5327 a -5318	P78426 (NKX61_HUMAN)	Cardíaco	4	3.16	http://www.uniprot.org/uniprot/P78426
-4.4e+02 kJ/mol	CTATAAATTT	-1742 a -1733	Q9Y215 (COLQ_HUMAN)	Óseo	1	8.9	http://www.uniprot.org/uniprot/Q9Y215
-4.0e+02 kJ/mol	ATTTTAATAT	-5094 a -5085	Q01196 (RUNX1_HUMAN)	Óseo	30	3.19	http://www.uniprot.org/uniprot/Q01196
-4.2e+02 kJ/mol	CTATAAATTT	-4955 a -4946	P09238 (MMP10_HUMAN)	Vascular	2	5.64	http://www.uniprot.org/uniprot/P09238
-4.0e+02 kJ/mol	ATTTTAATAT	-823 a -814	O14776 (TCRG1_HUMAN)	Inmune	3	3.19	http://www.uniprot.org/uniprot/O14776

Si bien son pocos los reportes experimentales que han logrado comprobar la participación de MEF2A como regulador de la expresión de ciertos genes dentro del genoma de *Homo sapiens*; existen investigaciones realizadas en otros organismos en los que se reportan proteínas cuya expresión esta regulada por MEF2A, como es el caso de las proteínas tipo SMAD's y MD1L1 homólogas a los genes *Medea* y la proteína *MAD* respectivamente, ambas pertenecientes a *Drosophila melanogaster* (Nguyen and Xu, 1998; Cripps *et al.*, 2004).

En ratón se ha demostrado que cuando se elimina el gen *mef2c* disminuyen los niveles de la proteína C57BL/6 (idcode: P04370 - MBP_MOUSE) que ayuda a la formación de músculo esquelético (Potthoff *et al.*, 2007) y en este trabajo se localizó a su homóloga humana la proteína MBP_HUMAN.

Se incluyen también las proteínas: WNT's de humano homólogas de Wingless, las proteínas humanas IHH y SHH homólogos de Hedgehog, la proteína humana FGFP homóloga de Fibroblast growth factor, TPM3

homóloga de TM1 (Lin, M. H. *et al.*, 1997) y EGF homólogo de Epidermal growth factor (Sandmann *et al.*, 2006; Taylor, 2000; Ruiz-Gomez *et al.*, 2002). Todas las proteínas mencionadas en este párrafo son homólogas con respecto a las reportadas en *Drosophila melanogaster* y se sabe por ensayos de ChIP que su expresión es regulada por MEF2A. En la Tabla 12 se muestran los resultados de este sondeo.

Tabla 12. Proteínas de *Homo sapiens* homólogas a secuencias de *Drosophila melanogaster* y *Mus musculus* cuya expresión por MEF2A se demostró mediante metodologías experimentales tipo ChIP y también fueron halladas determinando la ΔG_{bind} del complejo MEF2A-CARgbox. De los resultados expuestos en las tablas 5 a la 10, se extrajeron las proteínas de *Homo sapiens* que cuentan con homólogos de *Drosophila melanogaster* y *Mus musculus* cuya expresión por MEF2A se demostró mediante metodologías experimentales tipo ChIP. En la primera columna se reporta el valor de ΔG_{bind} que resultó de la unión entre el TF MEF2A y el elemento-cis cuya secuencia aparece en la segunda columna; también se reporta la ubicación corriente arriba de la sección regulatoria del gen donde se localiza la putativa CARgbox, en la cuarta columna podemos ver el número de acceso con el que se puede ingresar al servidor de Uniprot en el que se recaba información de las proteínas que resultan de la transcripción de los genes en los que se localizaron las CARgbox reportadas, a esta información puede accederse por medio del *link* que aparece en la columna de referencia. En la quinta columna se describe el tejido en el cual se expresan las proteínas reportadas; en la sexta columna bajo el título de Cis-e se describen el número de CARgbox putativas localizadas dentro del gen que codifica las proteínas y finalmente en la séptima columna se reporta el valor Ws que corresponde a la probabilidad que posee cada elemento-cis para ocurrir conforme al modelo propuesto por las PSSMs reportados en la sección 3.8.9 del capítulo III. Como puede observarse en la presente tabla, los resultados de la búsqueda basada en la determinación de la ΔG_{bind} calculada para cada complejo MEF2A-CARgbox en congruencia con los resultados encontrados en la literatura incluso para aquellas proteínas de *Homo sapiens* en las que aún no se demuestra por medio de metodologías tipo ChIP, su regulación bajo la participación de MEF2A, pero que sin embargo si se cuenta con reportes en la literatura de genes homólogos en los que se ha demostrado esta regulación con metodología experimental.

Valor de ΔG_{bind}	CARgbox	Ubicación	Número de acceso	Tejido	Cis-e	Ws	Referencia
-4.4e+02 kJ/mol	CTATTAATTTT	-4082 a -4073	Q9Y6D9 (MD1L1_HUMAN)	Músculo	1	7.6	http://www.uniprot.org/uniprot/Q9Y6D9
-4.3e+02 kJ/mol	CTATTAATAT	-1936 a -1927	Q15797 (SMAD1_HUMAN)	Músculo	6	8.35	http://www.uniprot.org/uniprot/Q15797
-4.3e+02 kJ/mol	ATATAAATAT	-4025 a -4016	Q15796 (SMAD2_HUMAN)	Músculo	10	8.15	http://www.uniprot.org/uniprot/Q15796
-4.3e+02 kJ/mol	ATAAATATTT	-3450 a -3441	O15105 (SMAD7_HUMAN)	Músculo	3	7.19	http://www.uniprot.org/uniprot/O15105
-4.3e+02 kJ/mol	ATATAATTTT	-1560 a -1551	O15198 (SMAD9_HUMAN)	Músculo	3	7.24	http://www.uniprot.org/uniprot/O15198
-4.0e+02 kJ/mol	ATTATTAATTT	-2291 a -2282	P56705 (WNT4_HUMAN)	Músculo	1	3.08	http://www.uniprot.org/uniprot/P56705
-4.0e+02 kJ/mol	ATTTTAAATT	-1570 a -1561	Q9H1J7 (WNT5B_HUMAN)	Músculo	4	3.13	http://www.uniprot.org/uniprot/Q9H1J7
-4.0e+02 kJ/mol	ATTAATAATTC	-2019 a -2010	Q9Y6F9 (WNT6_HUMAN)	Músculo	1	3.11	http://www.uniprot.org/uniprot/Q9Y6F9
-4.0e+02 kJ/mol	ATTTTAAATT	-1902 a -1893	Q93098 (WNT8B_HUMAN)	Músculo	3	3.13	http://www.uniprot.org/uniprot/Q93098
-4.0e+02 kJ/mol	ATTATAAAAT	-5373 a -5364	O14905 (WNT9B_HUMAN)	Músculo	1	3.16	http://www.uniprot.org/uniprot/O14905
-4.0e+02 kJ/mol	ATTATTATTC	-2483 a -2474	Q9UBV4 (WNT16_HUMAN)	Músculo	8	3.06	http://www.uniprot.org/uniprot/Q9UBV4
-4.0e+02 kJ/mol	ATTTTAAATAC	-3354 a -3345	Q14623 (IHH_HUMAN)	Músculo	2	3.16	http://www.uniprot.org/uniprot/Q14623
-4.0e+02 kJ/mol	ATTTAAAATC	-167 a -158	Q15465 (SHH_HUMAN)	Músculo	1	3.07	http://www.uniprot.org/uniprot/Q15465
-4.3e+02 kJ/mol	ATATTAATTT	-4226 a -4217	Q14512 (FGFP1_HUMAN)	Músculo	6	8.22	http://www.uniprot.org/uniprot/Q14512
-4.3e+02 kJ/mol	ATAATAATAT	-785 a -776	Q9BYJ0 (FGFP2_HUMAN)	Músculo	20	7.76	http://www.uniprot.org/uniprot/Q9BYJ0
-4.2e+02 kJ/mol	ATAAAAATAT	-1464 a -1455	Q8TAT2 (FGFP3_HUMAN)	Músculo	1	5.71	http://www.uniprot.org/uniprot/Q8TAT2
-4.3e+02 kJ/mol	ATATATATGT	-2319 a -2310	P22455 (FGFR4_HUMAN)	Músculo	13	7.95	http://www.uniprot.org/uniprot/P22455
-4.3e+02 kJ/mol	CTATATATAT	-328 a -319	P21781 (FGF7_HUMAN)	Músculo	12	7.28	http://www.uniprot.org/uniprot/P21781
-4.3e+02 kJ/mol	CTAATTATAT	-2588 a -2579	O15520 (FGF10_HUMAN)	Músculo	10	8.22	http://www.uniprot.org/uniprot/O15520
-4.3e+02 kJ/mol	CTATATATAT	-4016 a -4007	P61328 (FGF12_HUMAN)	Músculo	6	7.28	http://www.uniprot.org/uniprot/P61328
-4.3e+02 kJ/mol	ATATTAATTT	-2792 a -2783	Q92915 (FGF14_HUMAN)	Músculo	3	8.22	http://www.uniprot.org/uniprot/Q92915
-4.3e+02 kJ/mol	ATAATTATAT	-1783 a -1774	O76093 (FGF18_HUMAN)	Músculo	5	1.52	http://www.uniprot.org/uniprot/O76093
-4.3e+02 kJ/mol	ATATTCATTT	-1106 a -1097	Q9NP95 (FGF20_HUMAN)	Músculo	2	6.97	http://www.uniprot.org/uniprot/Q9NP95
-4.0e+02 kJ/mol	ATTTTAAAAT	-2159 a -2150	P01133 (EGF_HUMAN)	Músculo	3	3.16	http://www.uniprot.org/uniprot/P01133
-4.1e+02 kJ/mol	ATTTATATTT	-1028 a -1019	P06753 (TPM3_HUMAN)	Músculo	1	3.99	http://www.uniprot.org/uniprot/P06753
-4.4e+02 kJ/mol	CAATATATAT	-2880 a -2871	P02686 (MBP_HUMAN)	Vascular	1	7.6	http://www.uniprot.org/uniprot/P02686

3.9. DISCUSION DEL CAPITULO III.

3.9.1. Construcción de complejos MEF2A-CArGbox.

El tiempo consumido por 339672 simulaciones de *docking* para acoplar secuencias de DNA al dímero MEF2A, fue reemplazado por la construcción de complejos realizados por el programa *MultiCArG.py*, basados en la información biológica que contiene la determinación por RMN (archivo *.pdb* con No. de id:1C7U). Estas construcciones fueron basadas en un arquetipo ampliamente aceptado por la literatura y obtenido sin diferencias significativas por métodos cristalográficos (rmsd = 1.2 Å).

El contexto en el que se halla inmersa la CArGbox permaneció sin cambios, tanto a nivel estructural y a lo que el contenido de nucleótidos se refiere. En el caso de todos los complejos construidos, la secuencia de bases que flanquea el extremo 5' está compuesto por la secuencia CTCGG hasta topar con la CArGbox y por la secuencias CCGAG en el extremo 3'. Tres determinaciones de la estructura tridimensional del complejo (estudiadas a detalle en el capítulo I) poseen diferentes secuencias que flanquean a la CArGbox (Santinelli & Richmond, 2000; Wu *et al.*, 2010; He *et al.*, 2011), las cuales se reporta en la tabla 13.

Tabla 13. Regiones flanqueantes de la CArGbox pertenecientes a tres determinaciones estructurales independientes. Como puede observarse claramente, el contexto de la CArGbox varía tanto en número de nucleótidos como en composición de los mismos.

No de id	Extremo 5'	Extremo 3'
1egw	TAAG	CTT
3kov	AA	G
3p57	AAA	AA

No existen diferencias estructurales significativas en la disposición espacial del complejo al variar el contexto, por lo que se supuso que la forma en la que el programa *MultiCArG.py* construyó los complejos *de novo*, no afecta el cálculo de ΔG_{bind} .

Uno de los cambios más radicales a los que se sometió a la CArGbox durante la formación de los complejos, fue sin duda las mutaciones inducidas durante su construcción, éstas tuvieron lugar en todas sus posiciones. Durante este proceso se respetaron las coordenadas del esqueleto de fosfatos y azúcares, por otro lado los parámetros de rotación y translación de las bases reportadas en el arquetipo, también se mantuvieron sin cambios y se utilizaron como molde para la inserción de los nuevos nucleótidos.

La adición de átomos de H para cada base en todos los sitios de la CArGbox, fue realizado por el servidor *MolProbit* bajo la consigna de hacerlo con un mínimo energético, por lo que las librerías construidas con estas coordenadas, aseguran que la adición de los átomos no introduzcan impedimentos estéricos que provoquen aumentos de energía dentro del sistema. Además debe recordarse que el molde del cual se partió para la construcción de complejos poseía una minimización previa.

Tomando en cuenta los hechos antes expuestos, pueden calificarse de confiables los complejos formados,

que ofrecen una alternativa viable al *docking* y aportan nuevos modelos de estructuras tridimensionales que podrán ser utilizadas en estudios de caso, ya que incluso, sino se sigue el flujo del método propuesto, se tiene un procedimiento alternativo para la creación de complejos.

3.9.2. Determinación de la ΔG_{bind} .

Existe un consenso amplio en la literatura acerca del uso de APBS como una metodología confiable para el estudio de la interacción entre proteínas y DNA (Huang *et al.*, 2013; Rajagopalan *et al.*, 2013; Gao *et al.*, 2013). Debido a su rapidez de cálculo, facilidades en su compilación y la gran libertad que ofrece para modificar parámetros con los que se prepara el sistema para la determinación de la ΔG_{bind} , por lo que se integró esta paquetería al método para resolver la ecuación Poisson-Boltzman.

Durante el montaje del sistema se tuvieron en cuenta múltiples aspectos que permitieron su optimización, en este sentido se tuvo un especial cuidado en la potencia del campo eléctrico generado por el complejo para evitar polarizaciones en el medio dieléctrico (solvente) o incluso dentro del soluto que pudiesen malograr el sistema. Otra consideración que se tomó en cuenta dentro del soluto, consistió en evitar la suposición de valores de baja constante dieléctrica dentro del mismo, ya que en un principio crearon valores inadecuados en la interfaz del dieléctrico, por lo que se utilizaron valores dieléctricos propios del complejo, que fueron consistentes con la parametrización de los radios atómicos del soluto.

Como punto adicional se recomienda tener un extremo cuidado con la estructura del archivo que contiene las coordenadas atómicas en cuanto a su nomenclatura, ya que fueron constantes los siguientes errores: pérdida de cadenas laterales, protonaciones incorrectas en los residuos de K, desaparición de ácidos nucleicos y ruptura de los enlaces peptídicos. Una vez corregidas las ambigüedades de nomenclatura entre el campo de fuerzaz AMBER y el programa *PDb2PQR*, desaparecieron dichos errores.

3.9.3. Matrices de puntuación por posición específica obtenidas bajo los parámetros de éste estudio.

La utilidad que poseen las PSSMs en el estudio de los genes regulados por factores de transcripción y la comprensión de las redes metabólicas reguladas por estas proteínas, es fundamental, por lo que la elaboración de nuevas matrices contribuye al desarrollo de estos campos. Actualmente éstas son construidas con base en datos experimentales, las referentes a MEF2A se realizaron con un máximo de 104 secuencias, según búsquedas realizadas en los servidores TRANSFAC y *Oregano*.

Al establecer un motivo consenso con pocos datos, se corre el riesgo de no hacerlo representativo, lo que funcionaría como un motor de búsqueda poco robusto. Además debemos recordar que las búsquedas basadas en PSSMs están soportadas únicamente de forma estadística y que la única información biológica contenida intrínsecamente, es la proporcionada por la matriz. Las búsquedas basadas en PSSMs también asumen que la afinidad de los factores de transcripción por sus secuencias blanco siempre es la misma, lo que conlleva a errores estadísticos que proponen secuencias estadísticamente correctas, que en realidad carecen de propiedades fisicoquímicas adecuadas que promueven la unión con el TF.

Cabe mencionar que la obtención de matrices por el método creado en esta investigación, se basó en el reconocimiento de secuencias blanco que no requiere previo conocimiento del motivo, por lo que se justificó únicamente en las secuencias que se detectaron por medio del estudio de la ΔG_{bind} y permitió la construcción de matrices más robustas que poseen información biológica intrínseca, misma que refleja la fisicoquímica inmersa en la formación del complejo.

3.9.4. Identificación de regiones reguladas por MEF2A.

La determinación de la ΔG_{bind} que se realizó en las diferentes secuencias analizadas, dio un valor específico que refleja las características fisicoquímicas que subyacen en la unión de MEF2A con el DNA, ahora bien, aunque esta información por sí misma es valiosa, debemos concentrarnos en identificar cuáles de las secuencias estudiadas poseen las características necesarias que favorecen energéticamente la unión entre el TF y el DNA para ser propuestas como CArGbox.

El valor de $\Delta G_{\text{bind}} = -4.24\text{e}+02$ kJ/mol obtenido de la secuencia arquetipo reportada por Huang proporciona un parteaguas en la búsqueda de secuencias CArGbox, ya que podemos asumir de principio que todos los decámetros de DNA con un valor ΔG_{bind} de al menos $-4.2\text{e}+02$ kJ/mol son fuertes candidatos a ser nombrados elementos-cis de MEF2A, sin embargo, debemos ser consecuentes con la idea de realizar un estudio más exhaustivo con la finalidad de obtener un panorama general de la afinidad que la proteína tiene por otras secuencias ubicadas en diferentes niveles energéticos, por lo que vale la pena poner a prueba el estudio de caso al que se hace referencia, ya que es probable que el motivo consenso de CArGbox CTA(A/T)₄TAG, aceptado sin cambios desde su proposición (Andrés *et al.* 1995), solo sea representativo de una región energética particular y que por ello se hayan puesto de lado secuencias con diferentes valores de ΔG_{bind} .

De este modo el mapeo de diferentes regiones energéticas será muy útil en caso de carecer de la descripción de un motivo o cuando no se tenga identificada ninguna secuencia de DNA a la que se una el factor de transcripción estudiado.

Con este motivo se aislaron seis grupos de secuencias en poblaciones individuales según su valor de ΔG_{bind} localizadas en el rango de $-3.9\text{e}+02$ kJ/mol a $-4.4\text{e}+02$ kJ/mol. Con base en éste análisis se lograron identificar las siguientes relaciones que vale la pena discutir:

- Conforme disminuye el valor de ΔG_{bind} la frecuencia de cada base por posición específica se intensifica para una base en particular, lo que resulta en una pérdida de homogeneidad en el contenido de A, C, G ó T por sitio, observándose la tendencia del aumento en el contenido de A y T en todas las posiciones.
- El número de genes localizados aumenta considerablemente cuando se utilizan las PSSM's construida con un valor de $\Delta G_{\text{bind}} = -4.0\text{e}+02$ kJ/mol, $-4.3\text{e}+02$ kJ/mol y $-4.4\text{e}+02$ kJ/mol, en comparación con los hallados utilizando las matrices construidas en base a las secuencias con valores de $\Delta G_{\text{bind}} = -3.9\text{e}+02$ kJ/mol, $-4.1\text{e}+02$ kJ/mol y $-4.2\text{e}+02$ kJ/mol.

Las secuencias que poseen valores de $\Delta G_{\text{bind}} = -3.9\text{e}+02$ kJ/mol conforman un conjunto con un alto grado

de heterogeneidad, lo que se ve reflejado en los diagramas de logo obtenidos para las bases de dicha población, ya que las frecuencias de los sitios 2 al 10 son equiparables a las obtenidas en la población inicial de 339672 secuencias, es decir con una probabilidad de aparición de las bases A, T, C o G cercana a 0.25, sin reflejar un motivo estricto. Este hecho contrasta con la ganancia de homogeneidad que se observa conforme el valor de ΔG_{bind} disminuye, como se observa en el diagrama de logo realizado con las secuencias con valores de $-4.4\text{e}+02$ kJ/mol, en el que incluso puede percibirse claramente un motivo.

Con base en éstas observaciones, se plantea que las secuencias con valores bajos de ΔG_{bind} reflejan la presión selectiva que MEF2A ejerce sobre ellas, ya que como observamos en los diagramas de logo, existe una tendencia a la especificidad de las secuencias conforme vemos disminuir dicho valor, la cual es necesaria para llevar a cabo la fina regulación de la expresión genética, especificidad no reflejada en las secuencias con valores de $\Delta G_{\text{bind}} = -3.9\text{e}+02$ kJ/mol que no gozan de un motivo consenso (probablemente por ser muy comunes en el DNA humano) y que reflejan frecuencias similares a las de el modelo de *background* propuesto para *Homo sapiens* que nos indica una composición al azar. Si lo esperado es que un factor de transcripción se comporte de forma específica con respecto a la identificación de secuencias con composiciones canónicas, el grupo $-3.9\text{e}+02$ kJ/mol posee características energéticas inadecuadas para formar complejos con MEF2A por no ser diferenciables dentro del genoma, lo que las imposibilita como un aparato de regulación confiable.

Cómparese las 20 secuencias halladas en el rango de $-4.1\text{e}+02$ kJ/mol, o las 47 encontradas con un valor de $\Delta G_{\text{bind}} = -4.2\text{e}+02$ kJ/mol (que solo reflejan el 0.01295367% de secuencias del universo estudiado dentro de un contexto de 339672 decámeros de DNA) para notar la especificidad de MEF2A reflejada en este sondeo energético. La composición general de este último grupo no es homogénea en cuanto al contenido de bases por posición específica, lo que hace suponer a éstas secuencias como un buen modelo de CARGbox, ya que además poseen una composición muy similar a la canónica (sólo varía dramáticamente la posición 10 de G por T).

3.9.5. Variación del motivo consenso en posiciones 1 y 10 de la CARGbox.

Se observó una variación significativa del motivo consenso de CARGbox aceptado hasta el momento. Las bases C y G que ocupan las posiciones 1 y 10 respectivamente, se reportan con frecuencias de aparición del 100% en las matrices disponibles en los sitios TRANSFAC y *Oregano*, lo que se diferencia con las frecuencias obtenidas en cada nivel energético recabadas en la presente investigación (Tabla 14).

Tabla 14. Frecuencias de aparición en los sitios 1 y 10 por grupo energético, de los decámeros estudiados.

Grupo de secuencias	Frecuencia de aparición posición 1 (%)				Frecuencia de aparición posición 10 (%)			
	A	C	G	T	A	C	G	T
Literatura	0	100	0	0	0	0	100	0
$\Delta G_{\text{bind}} = -4\text{e}+02$ kJ/mol	62.004	37.99	0	0	24.22	23.34	22.43	27.99
$\Delta G_{\text{bind}} = -4.1\text{e}+02$ kJ/mol	60.02	39.97	0	0	21.10	21.61	19.59	37.69
$\Delta G_{\text{bind}} = -4.2\text{e}+02$ kJ/mol	61.96	38.034	0	0	18.21	19.65	13.76	48.36
$\Delta G_{\text{bind}} = -4.3\text{e}+02$ kJ/mol	66.07	33.92	0	0	3.54	3.34	1.46	91.64
$\Delta G_{\text{bind}} = -4.4\text{e}+02$ kJ/mol	72.72	27.27	0	0	0	0	0	100

En general, en la posición 1 (a diferencia de los reportes encontrados en la literatura) prevalece la A sobre la C; la G no ocupa un lugar significativo en la posición 10, sino que decae hasta que la T la sustituye en este sitio. En el caso de las de secuencias con valores de $\Delta G_{\text{bind}} = -4.2e+02$ kJ/mol, la G ocupa un 13.73% en la posición 10, caso reflejado en la CArGbox consenso.

Es importante atender la posibilidad de que el motivo propuesto actualmente en la literatura para la CArGbox de MEF2A surgiera de un grupo de secuencias con posiciones poco representativas, de ser el caso, dicha percepción limitaría el estudio de los genes regulados por este TF, ya que la probabilidad de encontrar nuevas CArGbox basados en los PSSMs con los que se cuenta hasta ahora, se encuentra mermada por basarse en frecuencias poco representativas.

Si los resultados de la presente investigación logran validarse por métodos no computacionales, habríamos de reconocer genes ocultos ante nuestra visión incompleta de la preferencia que MEF2A presenta por secuencias ricas en A y T en las posiciones 1 y 10..

3.9.6. Búsqueda de *elementoscis* con PSSMs contruidas por el método presentado en éste estudio.

Utilizando la metodología desarrollada en esta investigación pudieron ser localizados mas de 500 genes dentro del genoma de *Homo sapiens* que probablemente estén siendo regulados por MEF2A, si comparamos éste número con los 22 genes que hasta ahora se han detectado por medio de metodologías experimentales tipo ChIP, podríamos atribuirle una enorme capacidad de rastreo a la herramienta desarrollada en esta tesis, sin embargo aunque su autor se vea tentado a hacerlo, no deben tomarse estos resultados como definitivos, sino únicamente como una perspectiva que de ser tomada en cuenta, pueda orientar a la detección de genes regulados por MEF2A; sin que con ello pierda su validez y utilidad, pero si atribuyéndole un merecido grado de incertidumbre propio de cualquier metodología. Sin embargo este grado de incertidumbre puede vislumbrarse tomando en cuenta los reportes que hasta ahora podemos obtener de la literatura.

Si comparamos los resultados obtenidos entre el método *in silico* y los realizados en base a la inmunoprecipitación de cromatina, podemos observar su equivalencia, ya que determinando la ΔG_{bind} no solamente fue posible encontrar un gran número de CArGbox putativas, sino que también se redescubrieron genes que ya habían sido reportados en la literatura (Scheschonka, *et al.* 2007; Shen, *et al.* 2006; Haberland *et al.* 2007; Gottlieb *et al.*, 2002; Arnold *et al.*, 2007; Chang *et al.*, 2006; Woronicz *et al.*, 1995; Karamboulas *et al.*, 2006). Cabe resaltar el paralelismo entre los resultados obtenidos aún utilizando metodologías tan diferentes, ya que éste es uno de los puntos más fuertes a favor de la validación del método expuesto en la presente investigación.

Algunos de los resultados obtenidos por medio de la determinación de la ΔG_{bind} generada en el complejo MEF2A-CArGbox, pueden ser utilizados para ampliar el panorama que hasta ahora tenemos acerca del papel que desempeña el TF MEF2A durante la regulación en la expresión genética de *Homo sapiens*. Desde su descubrimiento se sabe que este factor de transcripción participa activamente en el desarrollo de los diferentes tipos musculares, incluso recientemente se hacen nuevos descubrimientos acerca de las severas miopatías que

surgen cuando MEF2A no desempeña su actividad correctamente, ya que se piensa que un gran número de genes se hallan regulados por este factor de transcripción (Bercin *et al.*, 2015), pero sin embargo se desconocen muchos de ellos. En éste trabajo se lograron rastrear más de 50 genes relacionados con la miogénesis de los diversos tipos musculares, que de ser tomados en cuenta podrían contribuir a una mayor comprensión del rol que juega MEF2A durante el desarrollo del músculo.

Si bien en la literatura se hace énfasis en el papel que juega MEF2A durante la miogénesis, actualmente se sabe que también participa como un regulador central durante el desarrollo de los huesos, el sistema nervioso, el sistema vascular y además contribuye a la diferenciación de los linfocitos T (Potthoff & Olson, 2007), estos descubrimientos recientes también se ven reflejados en los resultados obtenidos utilizando el método de determinación de la ΔG_{bind} , ya que se logró la detección de las proteínas: Amielogenina (AMELY_HUMAN) cuyo rol es crucial en la biomineralización de estructuras óseas, FMRD7 (FRMD7_HUMAN) que participa en el desarrollo neuronal, el regulador transcripcional Kaiso (KAISO_HUMAN) que contribuye a la regeneración de los endotelios del sistema vascular, y el transductor de señales CD24 (CD24_HUMAN) que promueve la proliferación de linfocitos B. Se hace mención de estos resultados con el afán de ejemplificar los obtenidos en la sección 3.8.10, de los cuales no se encuentran reportes en la literatura, pero que sin embargo son congruentes con los últimos descubrimientos realizados en MEF2A por medio de metodologías tipo ChIP.

En la presente investigación también se localizaron proteínas que presuntamente dependerían de MEF2A para expresarse y que contribuyen al desarrollo de tejidos y órganos que hasta ahora no se sabía eran regulados por MEF2A, tal es el caso de las más de 60 proteínas relacionadas con el desarrollo de los testículos. En una revisión realizada en la literatura únicamente se localizó un reporte que liga a MEF2A con el desarrollo de los testículos, realizado por Caroline Daems y publicado en el 2014. También se hallaron proteínas involucradas en el desarrollo de ovarios y placenta, así como reguladores del tráfico vesicular y el ciclo celular; con respecto a este último grupo de proteínas, se cuenta con algunos reportes que ligan a MEF2A con la regulación de la apoptosis en neuronas (McKinsey *et al.*; 2002; Xiaoming *et al.*, 2003).

Como hemos podido notar hasta ahora los resultados obtenidos en ésta investigación encajan con los reportes en la literatura y no sólo eso sino que también promueven la adquisición de un mayor conocimiento puesto que algunos de los genes encontrados representarían verdaderos descubrimientos si se logran confirmar los experimentos *in silico* con otro tipo de metodologías experimentales. Otro punto que vale la pena resaltar es el hecho de que los genes aquí expuestos poseen un inherente sentido biológico, es decir, los resultados de la búsqueda se centran solamente en grupos específicos de proteínas que participan en el desarrollo de sistemas y órganos concretos. La consecuencia de dicha observación nos lleva a pensar que las secuencias aquí detectadas no son producto de una búsqueda al azar, sino todo lo contrario, probablemente se logró recrear mediante métodos computacionales y por medio de la determinación de la ΔG_{bind} , un mecanismo que emula la selectividad que MEF2A posee por la CARGbox con un grado de eficiencia tal que permitió encontrar tanto genes localizados por técnicas de Inmunoprecipitación de cromatina, así como nuevos genes que no habían sido detectados; aunque dicho sea nuevamente, tales resultados se espera no sean tomados como definitivos y si como una

perspectiva valiosa para la investigación del rol que poseen los factores de transcripción dentro de las redes de expresión génica, y enunciado sea justamente esta última frase en plural, ya que el método aquí presentado puede ser utilizado no solo para MEF2A sino para cualquier factor de transcripción cuya estructura tridimensional sea conocida.

Por último y haciendo énfasis a la selectividad que tiene MEF2A por la CARGbox, debe recordarse que la determinación de la ΔG_{bind} , se realizó para 339672 secuencias de 10 pares de bases, con las cuales únicamente se localizaron al rededor de 500 genes, los cuales reflejan un número muy limitado de decámeros cuyo ΔG_{bind} resultó ser apto para proponerlos como blancos putativos de MEF2A. Esto nos sugiere que las características energéticas necesarias que debe poseer un oligonucleótido para ser CARGbox, son muy específicas y están limitadas a unas pocas secuencias; no sólo en el genoma de *Homo sapiens*, sino incluso en el total de secuencias probables a las que se puede unir el factor de transcripción (es decir, incluyendo secuencias hipotéticas). Sin duda es un hecho relevante que al parecer es el reflejo de la selección natural encargada de limitar el rango de receptividad de MEF2A para la formación de complejos con un patrón definido, refinando los medios por los cuales reconoce una secuencia para sólo ser receptivo de formar complejos con aquellas CARGbox que se adecuen a sus tan específicas necesidades energéticas; ya que como se acentúa en este estudio y como se percibe ampliamente en la literatura, los factores de transcripción son altamente específicos, pero ahora observamos la importancia que las cuestiones energéticas tienen en este acontecer.

3.9.7. ¿Por qué buscar nuevos métodos para el estudio de los factores de transcripción?

Sin duda los medios experimentales de biología molecular son fundamentales para el estudio del control de la expresión genética, desafortunadamente existen ocasiones en las que no se cuenta con condiciones experimentales o incluso de infraestructura para llevarlas a cabo, por lo que justificados en su rapidez y accesibilidad, se han desarrollado herramientas bioinformáticas enfocadas al reconocimiento de los sitios de unión de los factores de transcripción. Éstas últimas se basan en el reconocimiento de patrones por medio de PSSMs ¿Pero que se hace cuando no se tiene un motivo reconocido? En ese caso se recomienda el uso del presente método como un alternativa.

Sin embargo, la presente investigación no pretende sustituir a los métodos existentes, sino que busca ser utilizado para dar solución a problemas de base como la determinación de estructuras tridimensionales, la identificación de motivos, el reconocimiento de los factores de unión entre la proteína y el DNA, entre otros, pero sobre todo su incorporación a los métodos utilizados actualmente para generar complementariedad con las búsquedas basadas en PSSM (como se hizo en ésta tesis) y como un precedente para los estudios experimentales con los que se llevará a cabo la comprobación absoluta de lo recabado por la metodología bioinformática, ejemplo de ello es el sustento teórico que la presente investigación puede aportar al reporte de Matthew J. Potthoff and Eric N. Olson del 2007 en el que se plantea al factor de transcripción MEF2A como el regulador principal para la diferenciación de linfocitos, cuyos genes responsables de su diferenciación se han localizado con el método antes utilizado.

Por lo tanto se pone a disposición de los usuarios, la presente metodología para que sea utilizada como mejor convenga al trabajo de aquellos inmersos en el estudio de los factores de transcripción, ya que la regulación de la expresión genética es uno de los desafíos más grandes de la biología y la bioinformática, por lo que la identificación de los factores determinantes en la expresión de genes es preferiblemente, y en algunos casos necesariamente interdisciplinaria.

CONCLUSIONES

La obtención de las estructuras tridimensionales de las secuencias de CArGbox construidas *de novo*, constituyó un reto informático, ya que hasta el momento no se cuenta con ninguna paquetería disponible que simule procesos de mutación en una molécula de DNA conservando una estructura molecular determinada, las paqueterías disponibles hasta el momento de elaboración de esta tesis, únicamente están diseñadas para construir modelos tridimensionales igualando la configuración A, B o Z del DNA, sin la posibilidad de incluir consideraciones específicas al respecto de la estructura por parte de el usuario.

El programa *MultiCARG.py* se obtuvo como una salida alterna a la construcción de múltiples sistemas por medio del *docking* como método alternativo para la predicción de las coordenadas moleculares en la que habrían de estar MEF2A y su elemento-cis para formar un complejo estable, ya que para el caso de la determinación masiva del valor de ΔG_{bind} , calculado a través de las coordenadas tridimensionales, el *docking* se convierte en un problema de poder de computo que deriva en tiempos de espera antagónicos a la mayoría de investigaciones científicas; y tomando en cuenta que las estructuras moleculares determinadas por RNM y cristalografía poseen de forma intrínseca información biológica, se decidió hacer uso de éstas para conformar un método alternativo en el estudio de factores de transcripción y sus secuencias blanco; de éste modo :

- El protocolo desarrollado para la construcción *de novo* de la estructura tridimensional de los oligonucleótidos a los que se les hará interactuar con al proteína MEF2A en la determinación de la ΔG_{bind} basado en el modelo de MEF2A que va del aminoácido 1 al 73, es confiable por respetar la arquitectura del complejo MEF2A:CArGbox prevalente en diversas condiciones biológicas y experimentales.

En base a 25 ns de dinámica molecular se lograron recabar una serie de datos que permitieron establecer hechos fundamentales durante la interacción entre el factor de transcripción MEF2A y su CArGbox, tales como:

- Los puentes de hidrógeno encargados de mantener unidos a la proteína y al DNA son interacciones no estáticas con tiempos de permanencia definidos.
- Los puentes de hidrógenos formados entre la CArGbox y los residuos de Lys y Arg son primordiales durante el establecimiento y permanencia del complejo FT-CArGbox.
- Las bases 1, 2, 3 y 5 de ambas hebras de la CArGGbox, son las responsables de sostener un mayor número de interacciones con el factor de transcripción MEF2A.
- MEF2A esta sujeto a cambios de estructura secundaria durante su unión con la CArGbox.

Considerando lo anterior, fue posible construir secuencias CArGbox igualando las coordenadas atómicas del esqueleto de fosfatos y los azúcares presentados en el reporte de Huang *et al.*, 2000, cuyo mantenimiento fue esencial para igualar la forma que adquiere el DNA cuando se une a la proteína MEF2A, y así:

- Obtener un método alternativo para el estudio de los factores de transcripción y sus secuencias blanco, por medio del cual,
- Se obtuvieron PSSMs basadas en el reconocimiento de secuencias blanco por medio de la determinación de la ΔG_{bind} , que no requieren previo conocimiento del motivo al que se une el TF.
- Se establecieron 6 nuevos modelos alternativos de motivo CArGbox.
- Fue posible la determinación de nuevos genes regulados por el factor de transcripción MEF2A en el genoma de *Homo sapiens*.
- Los resultados de la búsqueda basada en la determinación de la ΔG_{bind} calculada para cada complejo MEF2A-CArGbox son equiparables con los resultados obtenidos por metodologías experimentales como la precipitación de cromatina.

PERSPECTIVAS

- Determinar el valor ΔG_{bind} para todas las secuencias del universo 4^{10} .

Contar con dichos valores permitirá la obtención de PSSMS mucho más robustas que permitirán identificar con mayor exactitud nuevos genes dentro del genoma de *Homo sapiens*. La realización de dicho estudio depende completamente de la obtención de más poder de cómputo, ya que la automatización de éste proceso se encuentra resuelta bajo el funcionamiento del programa MultiCARG.py, por lo que la expectativa de cumplir con esta perspectiva es muy alta.

- Realizar un servidor con el programa *MultiCARG.py*.

Contribuir con el estudio de los Factores de Transcripción se vislumbra como uno de los objetivos más altos de este trabajo, sin duda dejar accesible a todos los usuarios un código abierto del programa MultiCARG.py, promoverá al uso de esta metodología.

- Establecer un criterio específico para las ΔG_{bind} que deberán ser tomadas en cuenta en la búsqueda de nuevos elementos-cis.

Para el estudio presentado en este trabajo se contaba con un arquetipo de CARGbox obtenido de modo experimental al que se le midió su ΔG_{bind} , con ayuda de éste valor se construyó un rango a su alrededor con el fin de explorar otras energías libres en las que fuese posible la interacción de MEF2A con el DNA, pero cuándo no se tiene un motivo conocido al que pueda medírsele la ΔG_{bind} ¿Cómo decidiremos cuáles valores de energía libre tomaremos en cuenta? Una de las perspectivas más necesarias para esta investigación es el establecimiento de una metodología que nos ayude a resolver esta cuestión, sin embargo para este propósito se cuenta con valiosas herramientas de las cuáles echar mano. A continuación se describen algunas de éstas:

Todas las proteínas conocidas cuentan con homólogos en la naturaleza, por lo que sería muy razonable buscar aquellos que cuenten con un motivo conocido, para así obtener su ΔG_{bind} , bajo la suposición de que éste valor también se conservará en alto grado entre el TF del que no se cuentan con reportes de su motivo de unión y los parientes filogenéticos para los que sí se conoce éste motivo.

Uno de los hechos más contundentes para la resolución de este problema, surge de la observación acerca de la fuerza con la que se unen las secuencias CARG con MEF2A; en todos los casos los valores de ΔG_{bind} fueron los más altos encontrados durante el sondeo realizado en las secuencias hipotéticas tomadas del Universo 4^{10} , es decir dado este hecho podríamos hipotetizar que las secuencias de ADN que generen los valores más altos de ΔG_{bind} siempre serán las que corresponderán a los motivos reales a los cuales se une el factor de transcripción en la naturaleza. Dicho enunciado resulta confiable debido a que muy probablemente los factores de transcripción seleccionan aquellas secuencias con las que se genera una mejor unión para así asegurar la transcripción de los genes que controlan.

- Aplicar éste método en el caso de otros factores de transcripción y sus secuencias blanco.
- Construir el paisaje adaptativo de la CARGbox de MEF2A tomando como parámetro de adecuación la ΔG_{bind} .

BIBLIOGRAFÍA

- Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, Ribas DPL, Martinez-Castilla L, Yanofsky MF.** (2000) An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc Natl Acad Sci USA* 97:5328-5333.
- Andrés V., Cervera M., Mahdavi V.** (1995) Determination of the consensus binding site for MEF2 Expressed in muscle and brain tissue-specific sequence constraints. *Journal of Biological Chemistry*, Vol. 240, No. 40, issue 06: 23246-23249.
- Bachinski L., Sirito M., Böhme M., Baggerly K., Udd B. & Krahe R.** (2010) Altered MEF2 isoforms in myotonic dystrophy and other neuromuscular disorders. *Muscle & Nerve*, Vol. 42, issue 6: 856–863,
- Bayly, Cornell W., Koli.Man P.** (1993) A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: RESP Model. *J Phys Chem*, 97:10269-10280.
- Bates P., Kelley L., MacCallum R., Sternberg M.** (2002) Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins: Structure, Function, and Bioinformatics*, Vol.45, issue 5: 39–46
- Chipotc, Angyan J, Maigret B, Scheraga H** (1993) Modeling Amino-Acid Side-Chains. 2. Determination of Point Charges from Electrostatic Properties: Toward Transferable Point Charge Models. *J Phys Chem*, 97:9788 9796.
- Chipotc, Angyanj, Ferenczy G, Scheragah** (1993) Transferable Net Atomic Charges from a Distributed Multipole Analysis for the Description of Electrostatic Properties: A Case-Study of Saturated-Hydrocarbons. *J Phys Chem*, 97:6628--6636.
- Dodou E, Treisman R.** (1997) The *Saccharomyces cerevisiae* MADS-box transcription factor Rlm1 is a target for the Mpk1 mitogen-activated protein kinase pathway. *Mol. Cell. Biol.* 17:1848–59.
- Fickett J.** (1996) Coordinate positioning of MEF2 and myogenin binding sites . *The Journal of Biological Chemistry*, No. 271: 24927-24933.
- Gilson M.** (1993) Multiple-Site Titration and Molecular Modeling: 2 Rapid Methods for Computing Energies and Forces For Ionizable Groups in Proteins. *Proteins*, 15:266-282.
- Gong H., Xie J., Zhang N., Yao L.** (2011) MEF2A Binding to the Glut4 Promoter Occurs via AMPK[alpha]2-dependent Mechanism. *Medicine & Science in Sports & Exercise*, Vol.43, Issue 1.
- Gossett LA, Kelvin DJ, Sternberg EA, Olson EN.** (1989) A new myocyte-specific enhancer-binding factor that recognizes a conserved element associated with multiple muscle specific genes. *Mol. Cell. Biol.* 9:5022–33.
- Gramzow L., Ritz M. & Theiben G.** (2010) On the origin of MADS-domain transcription factors. *Trends in genetic*, Vol. 26, issue 4: 149-153
- Gribskov M, McLachlan M, Eisenberg D.** (1984) Profile Analysis: Detection of Distantly Related Proteins. *Proc Nat Acad Sci USA* , 84:4355-4358
- Haberer G., Wang Y., Mayer K.** (2011) The Non-coding Landscape of the Genome of *Arabidopsis thaliana*; *Plant Genetics and Genomics: Crops and Models*, Vol. 9: 67-121.
- Heffer A., Shultz J & Pick L.** (2010) Surprising flexibility in a conserved Hox transcription factor over 550 million years of evolution. *Proceedings of the National Academy of Science*, Vol. 108, issue 6: 2173-2179.
- Huang K., Louis J., Donaldson L., Lim F, Sharrocks A. & Glore G.** (2000) Solution structure of the MEF2A-DNA complex: structural basis for the modulation of DNA bending and specificity by MADS-box transcription factor. *The Journal of European Molecular Biology Organization*, Vol, 19, No 11: 2615-2628.
- HL Huang, FL Chang, SJ Ho, LS Shu** (2013) FRKAS: Knowledge Acquisition Using a Fuzzy Rule Base Approach to Insight of DNA-Binding Domains/Proteins *Protein Pept Lett.* Mar;20(3):299-308.
- de Folter S. & Angenent G** (2006) Trans meet cis in MADS science. *Elsevier*, Vol. 11, issue 5: 224-231
- De Bodt S., Raes J., Van de Peer Y., Theissen G.** (2003) And there where many : MADS goen genomic. *Trends plants, Sci* 8: 475-483.
- Deutsch.** (2010) Homeosis and beyond. What is the function of the Hox genes? *Advances in Experimental Medicine and*

Biology, Vol. 689: 155-165.

Pabo C. & Sauer R. (1992) Transcription factors: Structural families and principles of DNA recognition. Annual review of biochemistry, Vol. 61: 1053-1095.

Feng H., Cheng T., Steer J., Joyce D., Pavlos N., Leong C., Kular J., Liu J., Feng X., Zheng M., Xu J. (2009) MEF2 and mitf cooperate with nfatc1 to transactivate the v-atpase d2 promoter during RANKL-induced osteoclastogenesis .

Fiore F., Khudayberdiev S., Christensen M., Siegel G., Flavell S, Kim T., Greenberg M., Schrott G. (2009) MEF2-mediated transcription of the miR379–410 cluster regulates activity-dependent dendritogenesis by fine-tuning Pumilio2 protein levels. The EMBO Journal, Vol. 28: 697 – 710. Published online: 5 February 2009

Gao N, W Jiang, H Gao, Z Cheng (2013) Structural Basis of Human Transcription Factor Sry-related Box 17 Binding to DNA. Protein Pept Lett. 2013 Apr;20(4):481-8.

Jonas K. (1999) Protein secondary structure prediction based on position-specific scoring matrices. Journal of Molecular Biology, Vol. 292, issue 2: 195-202.

Kaufmann K., Melzer R., Theiben G. (2005) MIKC-type MADS-domain proteins: structural modulari1. M. Born, Z. Phys. 1, 45 (1920ty, protein interactions and network evolution in plants. Evolutionary Genomics, No 347: 183-198.

Kirkwood, J. (1939) Solvent Effects and Chemical Reactivity Chem. Phys. 7, 911.

Lilly B, Galewsky S, Firulli AB, Schulz RA, Ol-sonen. (1994) D-MEF2: a MADS box tran- scription factor expressed in differentiating mesoderm and muscle cell lineages during Drosophila embryogenesis. Proc. Natl. Acad. Sci. USA 91:5662–66

Liu, S.-H., Ma, J.-T., Yueh, A. Y., Lees-Miller, S. P., Anderson, C. W. Bayly C1eplakp, Cornellw, Koli.Manp: A Well-Behaved C. (1993) Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: RESP Model. J Phys Chem, 97:10269-10280.

Liu SH¹, Ma JT, Yueh AY, Lees-Miller SP, Anderson CW, Ng SY. (1993) The carboxyl-terminal transactivation domain of human serum response factor contains DNA-activated protein kinase phosphorylation sites, J. Bid. Chem. 268, 21 147-21 15

Lu X., Olson W. (2000) 3DNA: a versatile, integrated software system for the analysis, rebuild and visualization of three-dimensional nucleic-acis structures. Nature Protocols; Vol. 3, no. 7: 1213-1227.

Maerkl S., Quake S. (2011) A Systems Approach to Measuring the Binding Energy Landscapes of Transcription Factors , Science; Vol 315: 233-237

Marchler-Bauer & Bryant. (2004) CD-Search: protein domain annotations on the fly. Nucleids Acid Research; Vol.3suppl2: 327-331.

Martin, J. F., Miano, J. M., Hustad, C. M., Copeland, N. G., Jen- kins, N. A. & Olson, E. N. (1994) A Mef2 gene that generates a muscle-specific isoform via alternative mRNA splicing, Mol. Cell. Biol. 14, 1647-1656.

McDermott, J. C., Cardoso, M. C., Yu, Y.-T., Andres, V., Leifer, D., Krainc, D., Lipton, S. A. & Nadal-Ginard, B. (1993) MEF2C gene encodes skeletal muscle- and brain-specific tran- scription factors, Mol. Cell. Biol. 13, 2564-2577.

Morisaki T, Sermsuvitayawong K, Byun SH, Matsuda Y, Hidaki K, et al. 1997. Mouse MEF2B gene-unique member of MEF2 gene family. J. Biochem. 122:939–46.

McCammon, J. A.; Harvey, S. (1987) Dynamics of Proteins and Nucleic Acids; Cambridge University Press: Cambridge CrossRef.

Mueller, C. G. F. & Nordheim, A. (1991) A protein domain con-served between yeast MCMI and human SRF directs ternary complex formation, EMBO J. 10, 4219-4229

Norman, C., Runswick, M., Pollock, R. & Treisman, R. (1988) Isolation and properties of cDNA clones encoding SRF, a tran- scription factor that binds to the c-jos serum response element, Cell 55, 989-1003.

Olson E, Perry M, Schulz A. 1995. Regulation of muscle differentiation by the MEF2 family of MADS box transcription factors. Dev. Biol. 172:2–14.

Ornatsky O. & McDermott J. (1996) MEF2 Protein Expression, DNA Binding Specificity and Complex Composition, and Transcriptional Activity in Muscle and Non-muscle Cells. The Journal of Biological Chemistry, Vol. 271: 24927-24933.

David Kradolfer, Lars Hennig, Claudia Köhler (2002). Increased Maternal Genome Dosage Bypasses the Requirement

of the FIS Polycomb Repressive Complex 2 in Arabidopsis Seed Development Published: January 10, 2013 DOI: 10.1371/journal.pgen.1003163

Timothy A McKinsey, Chun Li Zhang, Eric N Olson (2002) MEF2: a calcium-dependent regulator of cell division, differentiation and death Volume 27, Issue 1, 1 January, Pages 40–47.

Huret JL . (2006) DNA: estructura molecular. Atlas Genet Cytogenet Oncol Haematol. September .

AI Badaeux, Y Shi (2013) Emerging roles for chromatin as a signal integration and storage platform Nature reviews Molecular cell biology.

Karla Y. Acosta-Viana, Jorge E. Zavala-Castro. (1996) Proteínas de unión a DNA. Rev Biomed; 7:163-172.163 Vol. 7/No. 3/Julio-Septiembre, 1996.

Prywes, R. & Zhu, H. (1992) In vitro squelching of activated transcription by serum response factor: Evidence for a common co-activator used by multiple transcriptional activators, Nucleic Acids Res. 20, 513-520.

Rajagopalan, SJ Teter, PH Zwart. (2013) Studies of Iscr reveal a unique mechanism for metal-dependent regulation of DNA binding specificity. Nat Struct Mol Biol. Jun;20(6):740-7. doi: 10.1038/nsmb.2568. Epub 2013 May 5.

Santucci A; Corradi B; Mancini M; Manetti F; Radi M; Shenone S & Botta M. (2009) C6-Unsubstituted Pyrazolo[3,4-d]pyrimidines Are Dual Src/Abl Inhibitors Effective against Imatinib Mesylate Resistant Chronic Myeloid Leukemia Cell Lines, Chemistry Enable Drug Discovery. olume 4, Issue 1, pages 118–126, January 12.

Sharrocks, A. D., Gille, H. & Shaw, P. E. (1993) Identification of amino acids essential for DNA binding and dimerization in ~ 67' Implications for a novel DNA-binding motif, Mol. Cell.

Shore, P. & Sharrocks, D. (1995) The MADS-box family of transcription factors, European Journal of Biochemistry, Vol. 229, issue 1: 1-13. Biol. 13, 123-132.

Steven P., Gregory T., Randolph E., Chantal C. and Bik-Kwoon T. (1988) Saccharomyces cerevisiae protein involved in plasmid maintenance is necessary for mating of MAT α cells, Vol. 204, issue 3: 593-606.

S1tkoff Sharp K, Honig B. (1993) Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. J Phys Chem, 98:1978--1988.

Strasser, A Wittmann, HJ (2013) How Valid is a Thermodynamic Discrimination - Modelling of GPCRs, 2013 – Springer DOI.

Taylor R & Kennard O. (1982) Crystallographic evidence for the existence of CH.cntdot.cntdot.cntdot.O, CH.cntdot.cntdot.cntdot.N and CH.cntdot.cntdot.cntdot.Cl hydrogen bonds. Journal of American. Chemistry. Society; Vol. 104, issue19: 5063–5070.

Turatsinze J., Thomas-Chollier M., Defrance M. & van Helden J. (2008) Using RSAT to scan genome sequences for transcription factor binding sites and cis-regulatory modules. Nature Publishing Group, doi:10.1038.

Van Dijk & Roeland van Ham (2010) Conserved and variable correlated mutations in the plant MADS protein network. Biomedical Central Genomics, Vol.11, suppl. 5: 607-6015.

Wang H., Caruso L., Downie B. and Perry S. (2004) The Embryo MADS Domain Protein AGAMOUS-Like 15 Directly Regulates Expression of a Gene Encoding an Enzyme Involved in Gibberellin Metabolism. Plant Cell, Vol. 16: 1206-1219.

Wu W., Huang X., Cheng., Li., Stefan de Folter, Huang Z., Jiang X., Hongxia Pang & Tao S. (2010) Conservation and evolution in and among SRF- and MEF2-type MADS domains and their binding sites. Molecular Biology and Evolution, Vol. 27 issue 8.

Zacharias M., Luty B., Davis M., McCammon A. (1992) Poisson-Boltzmann analysis of the X repressor-operator interaction . Journal of Biophysical; Vol. 63: 1280-1285.

Libros.

Taleisnik S. Receptores Celulares y La transducción de señales. 1ª ed. Córdoba. Encuentro Grupo Editor 2006.

Paqueterías.

APBS: Nathan A. Baker (nathanandrewbaker@gmail.com) Pacific Northwest National Laboratory. Additional contributing authors listed in the code documentation. Copyright (c) 2010-2011 Battelle Memorial Institute. Developed at the Pacific Northwest National Laboratory, operated by Battelle Memorial Institute, Pacific Northwest Division for the U.S. Department Energy. Portions Copyright (c) 2002-2010, Washington University in St. Louis. Portions Copyright (c) 2002-2010, Nathan A. Baker. Portions Copyright (c) 1999-2002, The Regents of the University of California. Portions Copyright (c) 1995, Michael Holst. All rights reserved.

Gromacs: Written by Emile Apol, Rossen Apostolov, Herman J.C. Berendsen, Aldert van Buuren, Pär Bjelkmar, Rudi van Drunen, Anton Feenstra, Gerrit Groenhof, Peter Kasson, Per Larsson, Pieter Meulenhoff, Teemu Murtola, Szilard Pall, Sander Pronk, Roland Schulz, Michael Shirts, Alfons Sijbers, Peter Tieleman, Berk Hess, David van der Spoel, and Erik Lindahl. Copyright (c) 1991-2000, University of Groningen, The Netherlands. Copyright (c) 2001-2010, The GROMACS development team at Uppsala University & The Royal Institute of Technology, Sweden. Check out <http://www.gromacs.org> for more information.

X3DNA: Andrew Colasanti, Xiang-Jun Lu & Wilma K. Olson Copyright (c) 2008-2013.

Xmgrace: Paul J Turner, Portland, OR Copyright (c) 1991-1995; Grace Development Team Copyright (c) 1996-2012.

Flujo de ejecución durante la preparación del sistema.

- I. Creación de ficheros *.gro* y *.top* a partir de una archivo *.pdb*. El archivo *.gro* contiene los átomos y sus coordenadas *x*, *y*, *z*, y el archivo *.top* posee la información de las masas atómicas, sus cargas y enlaces.

Archivo de entrada: *1C7Ugmx.pdb*

Archivos de salida: *1C7U.gro*, *1C7U.top*, *1C7U_DNA_chain_C.itp*, *posre_DNA_chain_C.itp*,
1C7U_DNA_chain_D.itp, *posre_DNA_chain_D.itp*, *1C7U_Protein_chain_A.itp*,
posre_Protein_chain_A.itp, *1C7U_Protein_chain_B.itp* y *posre_Protein_chain_B.itp*

Línea de código: *pdb2gmx -f 1C7Ugmx.pdb -o 1C7U -p 1C7U*

- II. Configuración de las dimensiones y el tipo de caja en el que se solvató la molécula.

Archivo de entrada: *1C7U.gro*

Archivo de salida: *1C7Ucaja*

Línea de código: *editconf -f 1C7U.gro -o 1C7Ucaja -bt dode -d 1.5*

- III. Solvatación.

Archivo de entrada: *1C7Ucaja.gro*

Archivo de salida: *1C7Usolv.gro*

Línea de código: *genbox -cp 1C7Ucaja.gro -cs spc216.gro -o 1C7Usolv*

Observaciones: se agregaron 27687 moléculas de agua.

- IV. Generación del primer archivo tpr.

Archivos de entrada: *1C7Usolv.gro*, *MINI.mdp (300 nsteps)* y *1C7U.top*

Salida: *1C7Ucompilado.tpr*, *mdout.mdp*

Línea de código: *grompp -f MINI.mdp -c 1C7Usolv.gro -o 1C7Ucompilado -p 1C7U.top*

Observaciones:NOTE 1 [file *1C7U.top*, line 50]:

System has non-zero total charge: -27.999995

Total charge should normally be an integer. See

http://www.gromacs.org/Documentation/Floating_Point_Arithmetic
for discussion on how close it should be to an integer.

- V. Obtención del archivo *.gro* al que se agregarán los iones.

Entrada: *1C7Ucompilado.tpr*

Salida: *1C7Upreion.gro*, *1C7Upreion.edr*, *1C7Upreion.trr*, *1C7Upreion.log*

Línea de código: *mdrun -s 1C7Ucompilado.tpr -deffnm 1C7Upreion -v*

- VI. Adición de iones al sistema para su neutralización con Na⁺

Entrada: *1C7Upreion.gro*

Salida: *1C7Uneutralizado.gro* y *genion.log*

Línea de código: *genion -s 1C7Ucompilado.tpr -o 1C7Uneutralizado -np 28 -nname NA -pq 1*

Observación: Se modificó el archivo 1C7U.top agregando los iones y restando su número en el soluto.

Flujo de ejecución de la minimización.

I. Compilación del archivo .tpr.

Archivos de entrada: *MINI.mdp (30000nsteps), 1C7U.top y 1C7Uneutralizado.gro*

Archivos de salida: *1C7Ucompilacion_para_minimizar.tpr y mdout.mdp*

Línea de código: *grompp -f MINI.mdp -c 1C7Uneutralizado.gro -p 1C7U.top -o 1C7Ucompilacion_para_minimizar.*

II. Corrimiento de la minimización.

Archivo de entrada: *1C7Ucompilacion_para_minimizar.tpr*

Archivo de salida: *MinimizaciónMOL.gro, MinimizaciónMOL.edr, MinimizaciónMOL.trr, MinimizaciónMOL.log*

Línea de código: *mdrun -s 1C7Ucompilacion_para_minimizar.tpr -deffnm MinimizaciónMOL -v*

OBSERVACIONES: Steepest Descents converged to Fmax < 100 in 9318 steps

Potential Energy = -1.4505045e+06

Maximum force = 9.0570236e+01 on atom 54593

Norm of force = 3.4548068e+00

Flujo de ejecución de la dinámica de posiciones restringidas.

I. Compilación del archivo .tpr.

Archivos de entrada: *PR.mdp(125000steps), MinimizaciónMOL.gro, MinimizaciónMOL.top*

Archivos de salida: *1C7Ucompilacion_para_PR.tpr y mdp.out*

Línea de código: *grompp -f PR.mdp -c MinimizaciónMOL.gro -p 1C7U.top -o 1C7Ucompilacion_para_PR.*

II. Desplazamiento de los átomos de solvente y iones con posiciones restringidas del soluto.

Archivo de entrada: *1C7Ucompilacion_para_PR.tpr*

Archivos de salida: *1C7U_PR.gro, 1C7U_PR.edr, 1C7U_PR.trr, 1C7U_PR.log*

Línea de código: *mdrun -s 1C7Ucompilacion_para_PR.tpr -deffnm 1C7U_PR*

2.7.8.a. Flujo de ejecución de la dinámica molecular de todos los átomos.

I. Compilación.

Archivos de entrada: *DM.mdp (12500000), 1C7U_PR.gro, 1C7U.top*

Archivos de salida: *1C7U_compilarDM_all.tpr y mdout.mdp*

Línea de código: *grompp -f DM.mdp -c 1C7U_PR.gro -p 1C7U.top -o 1C7U_compilarDM_tátomos.*

II Corrimiento dinámica de todos los átomos .

Archivos de entrada: *1C7U_compilarDM_todos los atomos.tpr y mdout.mdp*

Archivos de salida: *1C7U_DM.gro*, *1C7U_DM.edr*, *1C7U_DM.trr* y *1C7U_DM.log*

Línea de código: *mdrun -s 1C7U_compilarDM_all.tpr -deffnm 1C7U_DM*

El formato .pdb encontrado en las diferentes bibliotecas virtuales encargadas de resguardar dichos archivos, corresponde a una forma canónica de reportar las coordenadas de las moléculas allí reportadas, en cuanto a la estructura del propio archivo como a su gramática. Debido a ésta última cuestión debe tomarse en cuenta la nomenclatura propia del campo de fuerzas que se desee utilizar, ya que algunas veces es distinta a la expuesta en el archivo .pdb original, lo que conlleva a errores de lectura durante la conversión al formato nativo de GROMACS. Debido a estas inconsistencias gramaticales fue necesario realizar modificaciones de nomenclatura al archivo con No. de id: 1C7U para que pudiese ser reconocido por el campo de fuerzas AMBER99 protein, nucleic AMBER94 (Wang et al., J. Comp. Chem. 21, 1049-1074, 2000), los cambios realizados se muestran a continuación. Solo se incluye en ésta redacción las modificaciones realizadas, para mayor información acerca del estado original del archivo éste puede ser encontrado con el número de identificación anteriormente referido, se recomienda realizar ésta búsqueda a modo de ubicar los átomos que fueron modificados por medio de sus coordenadas. Las partes removidas del archivo se indican con (...)

Se especifica por cadena de aminoácidos el amino terminal con una N antes del nombre del residuo localizado en ésta posición. Los hidrógenos unidos al CA deben escribirse como HA1 y HA2 respectivamente.

ATOM	1 N	NGLY A	1	3.920	-15.519	-43.542	1.00	0.33	N
ATOM	2 CA	NGLY A	1	3.558	-16.910	-44.054	1.00	0.30	C
ATOM	3 C	NGLY A	1	4.301	-18.108	-43.568	1.00	0.31	C
ATOM	4 O	NGLY A	1	3.724	-19.145	-43.326	1.00	0.33	O
ATOM	5 H1	NGLY A	1	4.038	-15.548	-42.512	1.00	1.02	H
ATOM	6 H2	NGLY A	1	3.160	-14.852	-43.789	1.00	1.02	H
ATOM	7 H3	NGLY A	1	4.809	-15.211	-43.985	1.00	1.04	H
ATOM	8 HA1	NGLY A	1	2.398	-16.948	-44.157	1.00	0.32	H
ATOM	9 HA2	NGLY A	1	4.020	-17.096	-44.887	1.00	0.29	H
									(...)

Los hidrógenos de los residuos de Arginina, Lisina, Isoleucina, Glutamina, Treonina, Metionina, Ácido aspártico, Asparagina, Valina, Fenilalanina, Leucina, Alanina, Tirosina, Serina deben escribirse de la siguiente manera:

ATOM	23 1HB	ARG A	2	7.834	-17.771	-42.266	1.00	0.31	H
ATOM	24 2HB	ARG A	2	8.294	-19.449	-42.002	1.00	0.35	H
ATOM	25 1HG	ARG A	2	7.923	-18.303	-44.777	1.00	0.33	H
ATOM	26 2HG	ARG A	2	9.455	-18.203	-43.909	1.00	0.36	H
ATOM	27 1HD	ARG A	2	9.248	-20.754	-43.585	1.00	0.47	H
ATOM	28 2HD	ARG A	2	8.022	-20.680	-44.851	1.00	0.38	H
ATOM	29 HE	ARG A	2	10.115	-19.345	-45.994	1.00	1.02	H
ATOM	30 1HH1	ARG A	2	9.712	-22.326	-44.329	1.00	1.34	H
ATOM	31 2HH1	ARG A	2	10.945	-23.153	-45.221	1.00	1.73	H
ATOM	32 1HH2	ARG A	2	11.706	-20.416	-47.209	1.00	0.59	H
ATOM	33 2HH2	ARG A	2	12.073	-22.071	-46.858	1.00	0.80	H
									(...)
ATOM	45 1HB	LYS A	3	4.975	-19.430	-37.647	1.00	0.31	H
ATOM	46 2HB	LYS A	3	6.201	-19.942	-38.791	1.00	0.31	H
ATOM	47 1HG	LYS A	3	7.090	-17.884	-38.240	1.00	0.29	H
ATOM	48 2HG	LYS A	3	6.041	-17.365	-39.550	1.00	0.28	H
ATOM	49 1HD	LYS A	3	5.493	-15.835	-37.889	1.00	0.32	H
ATOM	50 2HD	LYS A	3	4.268	-17.062	-37.616	1.00	0.30	H
ATOM	51 1HE	LYS A	3	6.849	-16.749	-36.078	1.00	0.83	H

ATOM	52	2HE	LYS A	3	5.221	-16.381	-35.496	1.00	0.77	H
ATOM	53	1HZ	LYS A	3	5.781	-18.516	-34.695	1.00	1.95	H
ATOM	54	2HZ	LYS A	3	6.323	-19.032	-36.219	1.00	1.97	H
ATOM	55	3HZ	LYS A	3	4.669	-18.751	-35.955	1.00	1.98	H
					(...)					
ATOM	89	1HG1	ILE A	5	0.723	-13.657	-38.828	1.00	0.43	H
ATOM	90	2HG1	ILE A	5	0.364	-12.634	-37.439	1.00	0.39	H
ATOM	91	1HG2	ILE A	5	3.157	-13.734	-39.365	1.00	1.04	H
ATOM	92	2HG2	ILE A	5	3.363	-15.164	-38.391	1.00	1.01	H
ATOM	93	3HG2	ILE A	5	4.341	-13.745	-38.060	1.00	1.08	H
ATOM	94	1HD	ILE A	5	2.169	-12.020	-39.777	1.00	1.04	H
ATOM	95	2HD	ILE A	5	2.272	-11.119	-38.263	1.00	1.14	H
ATOM	96	3HD	ILE A	5	0.775	-11.107	-39.196	1.00	1.06	H
					(...)					
ATOM	108	1HB	GLN A	6	0.589	-14.846	-32.076	1.00	0.33	H
ATOM	109	2HB	GLN A	6	1.599	-13.437	-31.752	1.00	0.30	H
ATOM	110	1HG	GLN A	6	-0.216	-12.225	-30.799	1.00	0.90	H
ATOM	111	2HG	GLN A	6	-1.404	-13.371	-31.421	1.00	0.88	H
ATOM	112	1HE2	GLN A	6	-2.251	-14.385	-29.603	1.00	2.28	H
ATOM	113	2HE2	GLN A	6	-1.325	-15.088	-28.367	1.00	2.69	H
					(...)					
ATOM	144	1HG2	THR A	8	4.023	-10.555	-30.176	1.00	1.06	H
ATOM	145	2HG2	THR A	8	2.705	-11.117	-31.204	1.00	1.05	H
ATOM	146	3HG2	THR A	8	2.884	-11.757	-29.571	1.00	1.01	H
					(...)					
ATOM	200	1HB	MET A	11	10.509	-11.125	-27.144	1.00	0.93	H
ATOM	201	2HB	MET A	11	11.234	-9.925	-26.077	1.00	0.89	H
ATOM	202	1HG	MET A	11	8.903	-8.888	-25.892	1.00	1.89	H
ATOM	203	2HG	MET A	11	8.301	-10.354	-26.663	1.00	1.85	H
ATOM	204	1HE	MET A	11	8.941	-9.700	-22.347	1.00	2.00	H
ATOM	205	2HE	MET A	11	8.157	-8.790	-23.635	1.00	2.22	H
ATOM	206	3HE	MET A	11	9.912	-8.758	-23.486	1.00	2.32	H
					(...)					
ATOM	217	1HB	ASP A	12	12.571	-11.123	-31.871	1.00	0.31	H
ATOM	218	2HB	ASP A	12	14.329	-11.147	-31.826	1.00	0.33	H
					(...)					
ATOM	268	1HB	ASN A	15	13.614	-11.797	-34.685	1.00	0.29	H
ATOM	269	2HB	ASN A	15	12.255	-10.900	-34.031	1.00	0.27	H
ATOM	270	1HD2	ASN A	15	11.148	-13.080	-34.746	1.00	1.91	H
ATOM	271	2HD2	ASN A	15	10.532	-13.108	-36.327	1.00	1.15	H
					(...)					
ATOM	323	1HG1	VAL A	18	13.100	-8.987	-41.340	1.00	1.09	H
ATOM	324	2HG1	VAL A	18	11.877	-9.898	-40.452	1.00	1.02	H
ATOM	325	3HG1	VAL A	18	13.544	-10.470	-40.496	1.00	1.07	H
ATOM	326	1HG2	VAL A	18	15.008	-9.377	-38.098	1.00	1.11	H
ATOM	327	2HG2	VAL A	18	15.098	-7.752	-38.752	1.00	1.04	H
ATOM	328	3HG2	VAL A	18	15.290	-9.142	-39.814	1.00	1.04	H
					(...)					
ATOM	356	1HB	PHE A	20	8.544	-3.088	-37.616	1.00	0.45	H
ATOM	357	2HB	PHE A	20	10.023	-2.887	-38.558	1.00	0.40	H
					(...)					
ATOM	482	1HB	LEU A	27	2.427	-1.791	-44.639	1.00	0.31	H
ATOM	483	2HB	LEU A	27	2.935	-0.305	-45.434	1.00	0.28	H

ATOM	484	HG	LEU A 27	0.440	-0.091	-46.189	1.00	0.29	H
ATOM	485	1HD1	LEU A 27	0.281	-2.282	-44.120	1.00	1.05	H
ATOM	486	2HD1	LEU A 27	-0.558	-2.138	-45.662	1.00	1.13	H
ATOM	487	3HD1	LEU A 27	-1.048	-1.138	-44.298	1.00	0.99	H
ATOM	488	1HD2	LEU A 27	-0.043	1.251	-44.326	1.00	1.10	H
ATOM	489	2HD2	LEU A 27	1.718	1.201	-44.314	1.00	1.01	H
ATOM	490	3HD2	LEU A 27	0.806	0.184	-43.209	1.00	1.07	H
(...)									
ATOM	559	1HB	ALA A 31	0.271	3.048	-50.957	1.00	1.07	H
ATOM	560	2HB	ALA A 31	-0.461	2.758	-49.379	1.00	0.99	H
ATOM	561	3HB	ALA A 31	-1.166	3.950	-50.472	1.00	1.06	H
(...)									
ATOM	576	1HB	TYR A 32	1.574	1.903	-54.692	1.00	0.34	H
ATOM	577	2HB	TYR A 32	1.182	0.187	-54.799	1.00	0.32	H
ATOM	578	1HD	TYR A 32	1.393	3.365	-56.668	1.00	0.42	H
ATOM	579	2HD	TYR A 32	0.581	-0.844	-56.914	1.00	0.41	H
ATOM	580	1HE	TYR A 32	1.517	3.488	-59.140	1.00	0.54	H
ATOM	581	2HE	TYR A 32	0.709	-0.723	-59.387	1.00	0.52	H
ATOM	582	HH	TYR A 32	0.298	1.511	-61.142	1.00	1.08	H
(...)									
ATOM	625	1HB	SER A 35	-4.016	3.949	-54.354	1.00	0.59	H
ATOM	626	2HB	SER A 35	-3.224	3.039	-55.639	1.00	0.60	H

En el caso de la proteína MEF2A se halla un residuo de Prolina en el carboxilo terminal por lo que se deberá especificar como CPRO, además obsérvense los cambios en la nomenclatura de los hidrógenos.

ATOM	1239	N	CPRO A 74	-14.727	22.436	-35.949	1.00	1.29	N
ATOM	1240	CA	CPRO A 74	-16.072	22.202	-35.334	1.00	1.40	C
ATOM	1241	C	CPRO A 74	-17.208	22.612	-36.280	1.00	1.99	C
ATOM	1242	O	CPRO A 74	-18.307	22.103	-36.202	1.00	2.64	O
ATOM	1243	CB	CPRO A 74	-16.062	23.085	-34.070	1.00	2.33	C
ATOM	1244	CG	CPRO A 74	-15.051	24.150	-34.350	1.00	3.02	C
ATOM	1245	CD	CPRO A 74	-13.999	23.513	-35.254	1.00	2.29	C
ATOM	1246	HA	CPRO A 74	-16.180	21.166	-35.054	1.00	1.53	H
-ATOM	1247	1HB	CPRO A 74	-17.039	23.525	-33.901	1.00	2.74	H
ATOM	1248	2HB	CPRO A 74	-15.760	22.507	-33.209	1.00	2.64	H
ATOM	1249	1HG	CPRO A 74	-15.522	24.985	-34.856	1.00	3.68	H
ATOM	1250	2HG	CPRO A 74	-14.589	24.485	-33.433	1.00	3.64	H
ATOM	1251	1HD	CPRO A 74	-13.629	24.237	-35.954	1.00	2.61	H
ATOM	1252	2HD	CPRO A 74	-13.191	23.100	-34.671	1.00	2.64	H

Los cambios descritos anteriormente fueron realizados tanto para la cadena A como para la B.

En cuanto a la nomenclatura de los átomos presentes en los ácidos nucleicos, debe tomarse en cuenta que deben especificarse los extremos 5' y 3' colocando dicha indicación como se expone a continuación, notese también el cambio en la escritura de los hidrógenos en los nucleótidos de Citocina y Guanina, en el caso de ésta última póngase atención en la correcta escritura de los átomos de fosfatos.

ATOM	2507	O5'	DC5 C 201	27.596	-13.204	-28.337	1.00	0.35	O
ATOM	2508	C5'	DC5 C 201	27.806	-11.858	-27.903	1.00	0.36	C
ATOM	2509	C4'	DC5 C 201	27.571	-10.834	-29.022	1.00	0.36	C
ATOM	2510	O4'	DC5 C 201	28.507	-11.102	-30.095	1.00	0.35	O
ATOM	2511	C3'	DC5 C 201	26.195	-10.886	-29.671	1.00	0.34	C
ATOM	2512	O3'	DC5 C 201	25.698	-9.547	-29.765	1.00	0.36	O
ATOM	2513	C2'	DC5 C 201	26.285	-11.458	-31.071	1.00	0.33	C

ATOM	2514	C1'	DC5 C 201	27.799	-11.451	-31.313	1.00	0.34	C
ATOM	2515	N1	DC5 C 201	28.319	-12.764	-31.786	1.00	0.33	N
ATOM	2516	C2	DC5 C 201	29.028	-12.799	-32.981	1.00	0.34	C
ATOM	2517	O2	DC5 C 201	29.225	-11.761	-33.605	1.00	0.37	O
ATOM	2518	N3	DC5 C 201	29.492	-13.996	-33.423	1.00	0.35	N
ATOM	2519	C4	DC5 C 201	29.273	-15.120	-32.729	1.00	0.34	C
ATOM	2520	N4	DC5 C 201	29.740	-16.278	-33.192	1.00	0.36	N
ATOM	2521	C5	DC5 C 201	28.546	-15.095	-31.501	1.00	0.33	C
ATOM	2522	C6	DC5 C 201	28.090	-13.904	-31.068	1.00	0.33	C
ATOM	2523	H5'1	DC5 C 201	28.830	-11.760	-27.538	1.00	0.37	H
ATOM	2524	H5'2	DC5 C 201	27.119	-11.643	-27.085	1.00	0.37	H
ATOM	2525	H4'	DC5 C 201	27.749	-9.832	-28.638	1.00	0.37	H
ATOM	2526	H3'	DC5 C 201	25.512	-11.480	-29.059	1.00	0.35	H
ATOM	2527	H2'1	DC5 C 201	25.885	-12.473	-31.126	1.00	0.32	H
ATOM	2528	H2'2	DC5 C 201	25.776	-10.810	-31.785	1.00	0.33	H
ATOM	2529	H1'	DC5 C 201	28.024	-10.682	-32.060	1.00	0.35	H
ATOM	2530	H41	DC5 C 201	30.255	-16.306	-34.060	1.00	0.37	H
ATOM	2531	H42	DC5 C 201	29.578	-17.129	-32.673	1.00	0.37	H
ATOM	2532	H5	DC5 C 201	28.360	-16.010	-30.938	1.00	0.35	H
ATOM	2533	H6	DC5 C 201	27.530	-13.848	-30.135	1.00	0.33	H
ATOM	2534	H5T	DC5 C 201	27.865	-13.778	-27.612	1.00	0.68	H
(...)									
ATOM	3742	P	DG3 D 240	33.552	-16.229	-44.910	1.00	0.52	P
ATOM	3743	O1P	DG3 D 240	34.131	-15.615	-46.123	1.00	0.56	O
ATOM	3744	O2P	DG3 D 240	34.092	-17.504	-44.390	1.00	0.53	O
ATOM	3745	O5'	DG3 D 240	33.570	-15.168	-43.712	1.00	0.55	O
ATOM	3746	C5'	DG3 D 240	32.850	-13.941	-43.878	1.00	0.57	C
ATOM	3747	C4'	DG3 D 240	33.089	-12.994	-42.701	1.00	0.61	C
ATOM	3748	O4'	DG3 D 240	32.397	-13.480	-41.546	1.00	0.52	O
ATOM	3749	C3'	DG3 D 240	34.560	-12.899	-42.297	1.00	0.69	C
ATOM	3750	O3'	DG3 D 240	35.031	-11.559	-42.399	1.00	0.79	O
ATOM	3751	C2'	DG3 D 240	34.699	-13.362	-40.855	1.00	0.68	C
ATOM	3752	C1'	DG3 D 240	33.230	-13.270	-40.400	1.00	0.58	C
ATOM	3753	N9	DG3 D 240	32.828	-14.261	-39.387	1.00	0.52	N
ATOM	3754	C8	DG3 D 240	32.991	-15.616	-39.402	1.00	0.52	C
ATOM	3755	N7	DG3 D 240	32.477	-16.247	-38.391	1.00	0.48	N
ATOM	3756	C5	DG3 D 240	31.918	-15.225	-37.628	1.00	0.45	C
ATOM	3757	C6	DG3 D 240	31.212	-15.295	-36.400	1.00	0.41	C
ATOM	3758	O6	DG3 D 240	30.937	-16.294	-35.738	1.00	0.41	O
ATOM	3759	N1	DG3 D 240	30.816	-14.036	-35.966	1.00	0.40	N
ATOM	3760	C2	DG3 D 240	31.069	-12.853	-36.628	1.00	0.43	C
ATOM	3761	N2	DG3 D 240	30.599	-11.744	-36.053	1.00	0.43	N
ATOM	3762	N3	DG3 D 240	31.733	-12.783	-37.789	1.00	0.47	N
ATOM	3763	C4	DG3 D 240	32.127	-14.003	-38.228	1.00	0.48	C
ATOM	3764	H5'1	DG3 D 240	31.782	-14.175	-43.935	1.00	0.52	H
ATOM	3765	H5'2	DG3 D 240	33.160	-13.464	-44.810	1.00	0.61	H
ATOM	3766	H4'	DG3 D 240	32.720	-11.997	-42.949	1.00	0.65	H
ATOM	3767	H3'	DG3 D 240	35.131	-13.535	-42.986	1.00	0.72	H
ATOM	3768	H3T	DG3 D 240	35.982	-11.607	-42.541	1.00	1.17	H
ATOM	3769	H2'1	DG3 D 240	35.066	-14.386	-40.795	1.00	0.65	H
ATOM	3770	H2'2	DG3 D 240	35.333	-12.681	-40.286	1.00	0.77	H
ATOM	3771	H1'	DG3 D 240	33.013	-12.266	-40.026	1.00	0.65	H
ATOM	3772	H8	DG3 D 240	33.504	-16.134	-40.212	1.00	0.55	H

ATOM	3773	H1	DG3 D 240	30.290	-14.008	-35.104	1.00	0.37	H
ATOM	3774	H21	DG3 D 240	30.098	-11.803	-35.178	1.00	0.40	H
ATOM	3775	H22	DG3 D 240	30.743	-10.846	-36.493	1.00	0.46	H

(...)

Los átomos de fostato en la Citocina se escriben de la siguiente manera:

ATOM	3203	O1P	DC D 223	-26.612	-1.742	-56.452	1.00	0.36	O
ATOM	3204	O2P	DC D 223	-25.968	-3.750	-57.918	1.00	0.33	O

(...)

La forma correcta de escribir los átomos en los nucleótidos de Timina y Adenina es la siguiente (se resalta el cambio de nomenclatura en los fosfatos e hidrógenos):

ATOM	3520	P	DT D 233	3.579	-7.301	-47.567	1.00	0.27	P
ATOM	3521	OP1	DT D 233	2.950	-6.120	-46.928	1.00	0.28	O
ATOM	3522	OP2	DT D 233	4.758	-7.139	-48.449	1.00	0.29	O
ATOM	3523	O5'	DT D 233	3.926	-8.381	-46.414	1.00	0.27	O
ATOM	3524	C5'	DT D 233	2.842	-8.718	-45.541	1.00	0.27	C
ATOM	3525	C4'	DT D 233	3.081	-9.872	-44.553	1.00	0.28	C
ATOM	3526	O4'	DT D 233	3.361	-11.051	-45.338	1.00	0.28	O
ATOM	3527	C3'	DT D 233	4.261	-9.733	-43.597	1.00	0.29	C
ATOM	3528	O3'	DT D 233	3.824	-10.076	-42.272	1.00	0.30	O
ATOM	3529	C2'	DT D 233	5.357	-10.703	-43.967	1.00	0.30	C
ATOM	3530	C1'	DT D 233	4.598	-11.651	-44.901	1.00	0.29	C
ATOM	3531	N1	DT D 233	5.389	-12.051	-46.097	1.00	0.28	N
ATOM	3532	C2	DT D 233	5.585	-13.399	-46.303	1.00	0.28	C
ATOM	3533	O2	DT D 233	5.084	-14.248	-45.566	1.00	0.29	O
ATOM	3534	N3	DT D 233	6.369	-13.741	-47.383	1.00	0.28	N
ATOM	3535	C4	DT D 233	6.985	-12.867	-48.261	1.00	0.28	C
ATOM	3536	O4	DT D 233	7.659	-13.293	-49.199	1.00	0.28	O
ATOM	3537	C5	DT D 233	6.741	-11.473	-47.964	1.00	0.28	C
ATOM	3538	C7	DT D 233	7.458	-10.391	-48.772	1.00	0.30	C
ATOM	3539	C6	DT D 233	5.959	-11.113	-46.924	1.00	0.29	C
ATOM	3540	H5'	DT D 233	1.981	-8.966	-46.158	1.00	0.27	H
ATOM	3541	H5''	DT D 233	2.601	-7.827	-44.958	1.00	0.28	H
ATOM	3542	H4'	DT D 233	2.177	-10.034	-43.975	1.00	0.29	H
ATOM	3543	H3'	DT D 233	4.635	-8.709	-43.597	1.00	0.29	H
ATOM	3544	H2'	DT D 233	6.183	-10.211	-44.480	1.00	0.31	H
ATOM	3545	H2''	DT D 233	5.717	-11.235	-43.086	1.00	0.32	H
ATOM	3546	H1'	DT D 233	4.328	-12.546	-44.329	1.00	0.29	H
ATOM	3547	H3	DT D 233	6.486	-14.725	-47.562	1.00	0.28	H
ATOM	3548	H71	DT D 233	6.727	-9.733	-49.241	1.00	1.06	H
ATOM	3549	H72	DT D 233	8.077	-10.854	-49.541	1.00	1.04	H
ATOM	3550	H73	DT D 233	8.093	-9.808	-48.104	1.00	1.06	H
ATOM	3551	H6	DT D 233	5.766	-10.054	-46.746	1.00	0.29	H
ATOM	3552	P	DAD 234	4.805	-9.737	-41.043	1.00	0.31	P
ATOM	3553	OP1	DAD 234	3.977	-9.558	-39.827	1.00	0.33	O
ATOM	3554	OP2	DAD 234	5.692	-8.643	-41.478	1.00	0.34	O
ATOM	3555	O5'	DAD 234	5.706	-11.043	-40.854	1.00	0.29	O
ATOM	3556	C5'	DAD 234	5.077	-12.245	-40.456	1.00	0.28	C
ATOM	3557	C4'	DAD 234	6.078	-13.352	-40.211	1.00	0.26	C
ATOM	3558	O4'	DAD 234	6.669	-13.696	-41.463	1.00	0.26	O
ATOM	3559	C3'	DAD 234	7.232	-12.939	-39.313	1.00	0.27	C
ATOM	3560	O3'	DAD 234	7.320	-13.866	-38.254	1.00	0.26	O

ATOM	3561	C2'	DA D 234	8.536	-12.962	-40.101	1.00	0.26	C
ATOM	3562	C1'	DA D 234	8.093	-13.716	-41.367	1.00	0.26	C
ATOM	3563	N9	DA D 234	8.586	-13.110	-42.601	1.00	0.26	N
ATOM	3564	C8	DA D 234	8.484	-11.810	-43.020	1.00	0.27	C
ATOM	3565	N7	DA D 234	8.901	-11.584	-44.221	1.00	0.28	N
ATOM	3566	C5	DA D 234	9.326	-12.843	-44.634	1.00	0.27	C
ATOM	3567	C6	DA D 234	9.873	-13.296	-45.820	1.00	0.28	C
ATOM	3568	N6	DA D 234	10.145	-12.485	-46.836	1.00	0.29	N
ATOM	3569	N1	DA D 234	10.171	-14.596	-45.905	1.00	0.28	N
ATOM	3570	C2	DA D 234	9.943	-15.408	-44.881	1.00	0.28	C
ATOM	3571	N3	DA D 234	9.434	-15.084	-43.705	1.00	0.27	N
ATOM	3572	C4	DA D 234	9.140	-13.772	-43.656	1.00	0.27	C
ATOM	3573	H5'	DA D 234	4.381	-12.574	-41.236	1.00	0.28	H
ATOM	3574	H5''	DA D 234	4.534	-12.048	-39.543	1.00	0.28	H
ATOM	3575	H4'	DA D 234	5.592	-14.238	-39.796	1.00	0.26	H
ATOM	3576	H3'	DA D 234	7.051	-11.942	-38.896	1.00	0.28	H
ATOM	3577	H2'	DA D 234	8.872	-11.949	-40.343	1.00	0.28	H
ATOM	3578	H2''	DA D 234	9.314	-13.508	-39.579	1.00	0.26	H
ATOM	3579	H1'	DA D 234	8.379	-14.771	-41.306	1.00	0.26	H
ATOM	3580	H8	DA D 234	8.097	-11.024	-42.372	1.00	0.27	H
ATOM	3581	H61	DA D 234	10.540	-12.855	-47.687	1.00	0.29	H
ATOM	3582	H62	DA D 234	9.949	-11.500	-46.753	1.00	0.29	H
ATOM	3583	H2	DA D 234	10.231	-16.444	-45.013	1.00	0.28	H

(...)

Para un obtener mayores referencias acerca de la nomenclatura de los átomos revísense los documentos con extensión .rtp que se encuentran dentro de los archivos del campo de fuerza.

Archivo .mdp utilizado en el corrimiento de la minimización de energía.

Input Parameters:

```

integrator      =      steep
nsteps         =      30000
init_step      =      0
ns_type        =      Grid
nstlist        =      10
ndelta         =      2
nstcomm        =      10
comm_mode      =      Linear
nstlog         =      100
nstxout        =      100
nstvout        =      100
nstfout        =      0
nstcalcenergy =      -1
nstenergy      =      10
nstxtcout      =      0
init_t         =      0
delta_t        =      0.001
xtcprec        =      1000
nkx            =      90
nky            =      90
nkz            =      90
pme_order      =      4
ewald_rtol     =      1e-05
ewald_geometry =      0
epsilon_surface =      0
optimize_fft   =      FALSE
ePBC           =      xyz
bPeriodicMols =      FALSE
bContinuation =      FALSE
bShakeSOR      =      FALSE
etc            =      No
nsttcouple     =      -1
epc            =      No
epctype        =      Isotropic
nstpcouple     =      -1
tau_p          =      1

```

ref_p (3x3):

```

ref_p[ 0]={ 0.00000e+00, 0.00000e+00, 0.00000e+00}
ref_p[ 1]={ 0.00000e+00, 0.00000e+00, 0.00000e+00}
ref_p[ 2]={ 0.00000e+00, 0.00000e+00, 0.00000e+00}

```

compress (3x3):

```

compress[ 0]={ 0.00000e+00, 0.00000e+00, 0.00000e+00}
compress[ 1]={ 0.00000e+00, 0.00000e+00, 0.00000e+00}
compress[ 2]={ 0.00000e+00, 0.00000e+00, 0.00000e+00}

```

refcoord_scaling = No

posres_com (3):

```

posres_com[0]= 0.00000e+00
posres_com[1]= 0.00000e+00

```

```

posres_com[2]= 0.00000e+00
posres_comB (3):
posres_comB[0]= 0.00000e+00
posres_comB[1]= 0.00000e+00
posres_comB[2]= 0.00000e+00
andersen_seed    = 815131
rlist            = 1
rlistlong        = 1
rtpi             = 0.05
coulombtype      = PME
rcoulomb_switch  = 0
rcoulomb         = 1
vdwtype          = Cut-off
rvdw_switch      = 0
rvdw             = 1
epsilon_r        = 1
epsilon_rf       = 1
tabext           = 1
implicit_solvent = No
gb_algorithm     = Still
gb_epsilon_solvent = 80
nstgbradii      = 1
rgbradii        = 1
gb_saltconc     = 0
gb_obc_alpha    = 1
gb_obc_beta     = 0.8
gb_obc_gamma    = 4.85
gb_dielectric_offset = 0.009
sa_algorithm     = Ace-approximation
sa_surface_tension = 2.05016
DispCorr        = No
free_energy     = no
init_lambda     = 0
delta_lambda    = 0
n_foreign_lambda = 0
sc_alpha        = 0
sc_power        = 0
sc_sigma        = 0.3
sc_sigma_min    = 0.3
nstdhdl         = 10
separate_dhdl_file = yes
dhdl_derivatives = yes
dh_hist_size    = 0
dh_hist_spacing = 0.1
nwall           = 0
wall_type       = 9-3
wall_atomtype[0] = -1
wall_atomtype[1] = -1
wall_density[0]  = 0
wall_density[1]  = 0
wall_ewald_zfac  = 3
pull            = no
disre           = No

```

```

disre_weighting = Conservative
disre_mixed     = FALSE
dr_fc          = 1000
dr_tau         = 0
nstdisreout    = 100
orires_fc      = 0
orires_tau     = 0
nstorireout    = 100
dihre-fc       = 1000
em_stepsize    = 0.01
em_tol         = 100
niter          = 20
fc_stepsize    = 0
nstcgsteep     = 1000
nbfgscorr      = 10
ConstAlg       = Lincs
shake_tol      = 0.0001
lincs_order    = 4
lincs_warnangle = 30
lincs_iter     = 1
bd_fric        = 0
ld_seed        = 1993
cos_accel      = 0
deform (3x3):
  deform[ 0]={ 0.00000e+00, 0.00000e+00, 0.00000e+00}
  deform[ 1]={ 0.00000e+00, 0.00000e+00, 0.00000e+00}
  deform[ 2]={ 0.00000e+00, 0.00000e+00, 0.00000e+00}
userint1       = 0
userint2       = 0
userint3       = 0
userint4       = 0
userreal1      = 0
userreal2      = 0
userreal3      = 0
userreal4      = 0
grpopts:
  nrdf: 177357
  ref_t: 0
  tau_t: 0
anneal: No
ann_npoints: 0
acc: 0 0 0
nfreeze: N N N
energygrp_flags[ 0]: 0
efield-x:
  n = 0
efield-xt:
  n = 0
efield-y:
  n = 0
efield-yt:
  n = 0
efield-z:

```

```
n = 0
efield-zt:
  n = 0
  bQMMM          = FALSE
  QMconstraints  = 0
  QMMMscheme     = 0
  scalefactor    = 1
qm_opts:
  ngQM           = 0
```

Función de energía total utilizada por AMBER99:

$$E_{total} = \sum_{\text{enlaces}} k_{\ell} (\ell - \ell_0)^2 + \sum_{\text{ángulos}} k_{\theta} (\theta - \theta_0)^2 + \sum_{\text{dihedros}} (V_n/2)(1 + \cos(\omega - \delta)) + \sum_{i-j} [A_{ij}/r_{ij}^{12} - B_{ij}r_{ij}^6 + q_i q_j / Dr]$$

$$+ \sum_{\text{puentes de hidrógeno}} [C_{AH}/r_{AH}^{12} - D_{AH}/r_{AH}^{10}]$$

A y B son parámetros de la ecuación Lennard-Jones
 r_{ij} se refiere a la distancia entre dos átomos o sitios cargados.
 q_i es la carga parcial relativa de la carga de un electrón.

El potencial del modelo de aguas TIP3P esta dado por:

$$E_{ab} = \sum_i^{ona} \sum_j^{onb} \frac{kc q_i q_j}{r_{ij}} + \frac{A}{r_{oo}^{12}} + \frac{B}{r_{oo}^6}$$

En dónde kc es la constante electrostática cuyo valor es de 332.1 Å·kcal/mol
A y B son parámetros de la ecuación Lennard-Jones
 r_{ij} se refiere a la distancia entre dos átomos o sitios cargados.
 q_i es la carga parcial relativa de la carga de un electrón.

Código 1.

```
#!/usr/bin/ipython2.6
import string
import itertools
#Creacion del universo de oligos
product = itertools.product('ACGT', repeat=10)
univ = ""
for i in product:
    print i
    univ = univ + str(i) + '\n'
print univ
f = open('Universooligos.txt','w')
f.write(univ)
f.close()
```

MultiCArG.py

```
#!/usr/bin/ipython2.7
import string
from string import *
import os
import re
FUfin = []
f2 = open("Universooligos2.txt")
datos = f2.read()
lista = string.split( datos, '\n' )
oligoenturno = map( lambda e: string.split( e, ',' ), lista )
numoligo = 0
history1 = open("cantidad.txt", 'w')

for numoligo in range(2000000):
#Secuencia a utilizar
    l = oligoenturno[numoligo]
    #print "\nReescribiendo a formato X3DNA...\n"
    #print l[0]
    #print l[9]
    if l[0] == "A":
        l[0] = 'A-T'
    if l[0] == "C":
        l[0] = 'C-G'
    if l[0] == "G":
        l[0] = 'G-C'
    if l[0] == "T":
        l[0] = 'T-A'
    if l[1] == "A":
        l[1] = 'A-T'
    if l[1] == "C":
        l[1] = 'C-G'
    if l[1] == "G":
        l[1] = 'G-C'
    if l[1] == "T":
        l[1] = 'T-A'
    if l[2] == "A":
        l[2] = 'A-T'
    if l[2] == "C":
        l[2] = 'C-G'
    if l[2] == "G":
        l[2] = 'G-C'
    if l[2] == "T":
        l[2] = 'T-A'
    if l[3] == "A":
        l[3] = 'A-T'
    if l[3] == "C":
        l[3] = 'C-G'
    if l[3] == "G":
        l[3] = 'G-C'
    if l[3] == "T":
        l[3] = 'T-A'
    if l[4] == "A":
        l[4] = 'A-T'
    if l[4] == "C":
        l[4] = 'C-G'
    if l[4] == "G":
        l[4] = 'G-C'
    if l[4] == "T":
        l[4] = 'T-A'
    if l[5] == "A":
        l[5] = 'A-T'
    if l[5] == "C":
        l[5] = 'C-G'
    if l[5] == "G":
        l[5] = 'G-C'
    if l[5] == "T":
        l[5] = 'T-A'
    if l[6] == "A":
        l[6] = 'A-T'
    if l[6] == "C":
        l[6] = 'C-G'
    if l[6] == "G":
        l[6] = 'G-C'
    if l[6] == "T":
        l[6] = 'T-A'
    if l[7] == "A":
        l[7] = 'A-T'
    if l[7] == "C":
        l[7] = 'C-G'
    if l[7] == "G":
        l[7] = 'G-C'
    if l[7] == "T":
        l[7] = 'T-A'
    if l[8] == "A":
        l[8] = 'A-T'
    if l[8] == "C":
        l[8] = 'C-G'
    if l[8] == "G":
```

```

l[8] = 'G-C'
if l[8] == " T":
    l[8] = 'T-A'
if l[9] == " A)":
    l[9] = 'A-T'
if l[9] == " C)":
    l[9] = 'C-G'
if l[9] == " G)":
    l[9] = 'G-C'
if l[9] == " T)":
    l[9] = 'T-A'
#print "\nArchivo en formato X3DNA: \n"
#print l
#Abre el archivo estandar en formato X3DNA y selecciona una porcion de este para ser sustituida.
#print "\nEste es tu archivo estandar en formato X3DNA: \n"
fa = open("bp_step2.par",'r')
cadena = fa.read()
#print cadena
cadena = list(cadena)
cadena[671:674] = l[0]
cadena[772:775] = l[1]
cadena[873:876] = l[2]
cadena[974:977] = l[3]
cadena[1075:1078] = l[4]
cadena[1176:1179] = l[5]
cadena[1277:1280] = l[6]
cadena[1378:1381] = l[7]
cadena[1479:1482] = l[8]
cadena[1580:1583] = l[9]
fa.close()
#Reescribe el archivo bp_step2.par
nuevo = open("bp_step2.par",'w')
nuevo.write("".join(cadena))
nuevo.close()
#print "\nEste archivo esta listo para correrse en X3DNA:\n"
with open("bp_step2.par") as o:
    for line in o:
        print line

try:
    cmd="/home/quetzal/X3DNA/bin/rebuild" + " -atomic+" + bp_step2.par+" CarGrpda.pdb"
    os.system(cmd)
except:
    print "No existe el archivo bp_step2.par"
a = open("CArGoriginal.pdb",'r')
b = a.read()
sus = open("CArGrpda.pdb",'r')
AXC = sus.read()
inicia6 = AXC.find(" A 6")
finaliza6 = AXC.find(" A 7")
inicia7 = AXC.find(" A 7")
finaliza7 = AXC.find(" A 8")
inicia8 = AXC.find(" A 8")
finaliza8 = AXC.find(" A 9")
inicia9 = AXC.find(" A 9")
finaliza9 = AXC.find(" A 10")
inicia10 = AXC.find(" A 10")
finaliza10 = AXC.find(" A 11")
inicia11 = AXC.find(" A 11")
finaliza11 = AXC.find(" A 12")
inicia12 = AXC.find(" A 12")
finaliza12 = AXC.find(" A 13")
inicia13 = AXC.find(" A 13")
finaliza13 = AXC.find(" A 14")
inicia14 = AXC.find(" A 14")
finaliza14 = AXC.find(" A 15")
inicia15 = AXC.find(" A 15")
finaliza15 = AXC.find(" A 16")
inicia26 = AXC.find(" B 26")
finaliza26 = AXC.find(" B 27")
inicia27 = AXC.find(" B 27")
finaliza27 = AXC.find(" B 28")
inicia28 = AXC.find(" B 28")
finaliza28 = AXC.find(" B 29")
inicia29 = AXC.find(" B 29")
finaliza29 = AXC.find(" B 30")
inicia30 = AXC.find(" B 30")
finaliza30 = AXC.find(" B 31")
inicia31 = AXC.find(" B 31")
finaliza31 = AXC.find(" B 32")
inicia32 = AXC.find(" B 32")
finaliza32 = AXC.find(" B 33")
inicia33 = AXC.find(" B 33")
finaliza33 = AXC.find(" B 34")
inicia34 = AXC.find(" B 34")
finaliza34 = AXC.find(" B 35")
inicia35 = AXC.find(" B 35")
finaliza35 = AXC.find(" B 36")

#REEMPLAZANDO LA BASE 35#
p35 = open("p35.pdb",'w')
p35.write(AXC[inicia35-2:finaliza35-2])
p35.close()
p35clon = open("p35.pdb",'r')

```



```

p35sus = p35clon.read()
#print p35sus[0:8]
T35 = open("T35.pdb",'r')
T35 = T35.read()
if p35sus[0:8] == " T B 35":
    p35sus = p35sus.replace(p35sus[0:8], "DT D 235")
    p35sus = p35sus + T35
    b = b.replace(b[87174:87182], "DT D 235")
A35 = open("A35.pdb",'r')
A35 = A35.read()
if p35sus[0:8] == " A B 35":
    p35sus = p35sus.replace(p35sus[0:8], "DA D 235")
    p35sus = p35sus + A35
    b = b.replace(b[87174:87182], "DA D 235")
C35 = open("C35.pdb",'r')
C35 = C35.read()
if p35sus[0:8] == " C B 35":
    p35sus = p35sus.replace(p35sus[0:8], "DC D 235")
    p35sus = p35sus + C35
    b = b.replace(b[87174:87182], "DC D 235")
G35 = open("G35.pdb",'r')
G35 = G35.read()
if p35sus[0:8] == " G B 35":
    p35sus = p35sus.replace(p35sus[0:8], "DG D 235")
    p35sus = p35sus + G35
    b = b.replace(b[87174:87182], "DG D 235")
P35 = open("p35.pdb",'w')
P35.write(p35sus)
reemplazo35 = p35sus[1:2000]
p35clon.close()
P35.close()
#print reemplazo35
#REEMPLAZANDO LA BASE 34#
p34 = open("p34.pdb",'w')
p34.write(AXC[incia34-2:finaliza34-21])
p34.close()
p34clon = open("p34.pdb",'r')
p34sus = p34clon.read()
#print p34sus[0:8]
T34 = open("T34.pdb",'r')
T34 = T34.read()
if p34sus[0:8] == " T B 34":
    p34sus = p34sus.replace(p34sus[0:8], "DT D 234")
    p34sus = p34sus + T34
    b = b.replace(b[84582:84590], "DT D 234")
A34 = open("A34.pdb",'r')
A34 = A34.read()
if p34sus[0:8] == " A B 34":
    p34sus = p34sus.replace(p34sus[0:8], "DA D 234")
    p34sus = p34sus + A34
    b = b.replace(b[84582:84590], "DA D 234")
C34 = open("C34.pdb",'r')
C34 = C34.read()
if p34sus[0:8] == " C B 34":
    p34sus = p34sus.replace(p34sus[0:8], "DC D 234")
    p34sus = p34sus + C34
    b = b.replace(b[84582:84590], "DC D 234")
G34 = open("G34.pdb",'r')
G34 = G34.read()
if p34sus[0:8] == " G B 34":
    p34sus = p34sus.replace(p34sus[0:8], "DG D 234")
    p34sus = p34sus + G34
    b = b.replace(b[84582:84590], "DG D 234")
P34 = open("p34.pdb",'w')
P34.write(p34sus)
reemplazo34 = p34sus[1:2000]
p34clon.close()
P34.close()
#print reemplazo34
#REEMPLAZANDO LA BASE 33#
p33 = open("p33.pdb",'w')
p33.write(AXC[incia33-2:finaliza33-21])
p33.close()
p33clon = open("p33.pdb",'r')
p33sus = p33clon.read()
#print p33sus[0:8]
T33 = open("T33.pdb",'r')
T33 = T33.read()
if p33sus[0:8] == " T B 33":
    p33sus = p33sus.replace(p33sus[0:8], "DT D 233")
    p33sus = p33sus + T33
    b = b.replace(b[81990:81998], "DT D 233")
A33 = open("A33.pdb",'r')
A33 = A33.read()
if p33sus[0:8] == " A B 33":
    p33sus = p33sus.replace(p33sus[0:8], "DA D 233")
    p33sus = p33sus + A33
    b = b.replace(b[81990:81998], "DA D 233")
C33 = open("C33.pdb",'r')
C33 = C33.read()
if p33sus[0:8] == " C B 33":
    p33sus = p33sus.replace(p33sus[0:8], "DC D 233")
    p33sus = p33sus + C33
    b = b.replace(b[81990:81998], "DC D 233")

```

```

G33 = open("G33.pdb",'r')
G33 = G33.read()
if p33sus[0:8] == " G B 33":
    p33sus = p33sus.replace(p33sus[0:8], "DG D 233")
    p33sus = p33sus + G33
    b = b.replace(b[81990:81998], "DG D 233")
P33 = open("p33.pdb",'w')
P33.write(p33sus)
reemplazo33 = p33sus[1:2000]
p33clon.close()
P33.close()
#print reemplazo33
#REEMPLAZANDO LA BASE 32#
p32 = open("p32.pdb",'w')
p32.write(AXC[incia32-2:finaliza32-21])
p32.close()
p32clon = open("p32.pdb",'r')
p32sus = p32clon.read()
#print p32sus[0:8]
T32 = open("T32.pdb",'r')
T32 = T32.read()
if p32sus[0:8] == " T B 32":
    p32sus = p32sus.replace(p32sus[0:8], "DT D 232")
    p32sus = p32sus + T32
    b = b.replace(b[79398:79406], "DT D 232")
A32 = open("A32.pdb",'r')
A32 = A32.read()
if p32sus[0:8] == " A B 32":
    p32sus = p32sus.replace(p32sus[0:8], "DA D 232")
    p32sus = p32sus + A32
    b = b.replace(b[79398:79406], "DA D 232")
C32 = open("C32.pdb",'r')
C32 = C32.read()
if p32sus[0:8] == " C B 32":
    p32sus = p32sus.replace(p32sus[0:8], "DC D 232")
    p32sus = p32sus + C32
    b = b.replace(b[79398:79406], "DC D 232")
G32 = open("G32.pdb",'r')
G32 = G32.read()
if p32sus[0:8] == " G B 32":
    p32sus = p32sus.replace(p32sus[0:8], "DG D 232")
    p32sus = p32sus + G32
    b = b.replace(b[79398:79406], "DG D 232")
P32 = open("p32.pdb",'w')
P32.write(p32sus)
reemplazo32 = p32sus[1:2000]
p32clon.close()
P32.close()
#print reemplazo32
#REEMPLAZANDO LA BASE 31#
p31 = open("p31.pdb",'w')
p31.write(AXC[incia31-2:finaliza31-21])
p31.close()
p31clon = open("p31.pdb",'r')
p31sus = p31clon.read()
#print p31sus[0:8]
T31 = open("T31.pdb",'r')
T31 = T31.read()
if p31sus[0:8] == " T B 31":
    p31sus = p31sus.replace(p31sus[0:8], "DT D 231")
    p31sus = p31sus + T31
    b = b.replace(b[76806:76814], "DT D 231")
A31 = open("A31.pdb",'r')
A31 = A31.read()
if p31sus[0:8] == " A B 31":
    p31sus = p31sus.replace(p31sus[0:8], "DA D 231")
    p31sus = p31sus + A31
    b = b.replace(b[76806:76814], "DA D 231")
C31 = open("C31.pdb",'r')
C31 = C31.read()
if p31sus[0:8] == " C B 31":
    p31sus = p31sus.replace(p31sus[0:8], "DC D 231")
    p31sus = p31sus + C31
    b = b.replace(b[76806:76814], "DC D 231")
G31 = open("G31.pdb",'r')
G31 = G31.read()
if p31sus[0:8] == " G B 31":
    p31sus = p31sus.replace(p31sus[0:8], "DG D 231")
    p31sus = p31sus + G31
    b = b.replace(b[76806:76814], "DG D 231")
P31 = open("p31.pdb",'w')
P31.write(p31sus)
reemplazo31 = p31sus[1:2000]
p31clon.close()
P31.close()
#print reemplazo31
#REEMPLAZANDO LA BASE 30#
p30 = open("p30.pdb",'w')
p30.write(AXC[incia30-2:finaliza30-21])
p30.close()
p30clon = open("p30.pdb",'r')
p30sus = p30clon.read()
#print p30sus[0:8]
T30 = open("T30.pdb",'r')

```

```

T30 = T30.read()
if p30sus[0:8] == " T B 30":
    p30sus = p30sus.replace(p30sus[0:8], "DT D 230")
    p30sus = p30sus + T30
    b = b.replace(b[74214:74222], "DT D 230")
A30 = open("A30.pdb", 'r')
A30 = A30.read()
if p30sus[0:8] == " A B 30":
    p30sus = p30sus.replace(p30sus[0:8], "DA D 230")
    p30sus = p30sus + A30
    b = b.replace(b[74214:74222], "DA D 230")
C30 = open("C30.pdb", 'r')
C30 = C30.read()
if p30sus[0:8] == " C B 30":
    p30sus = p30sus.replace(p30sus[0:8], "DC D 230")
    p30sus = p30sus + C30
    b = b.replace(b[74214:74222], "DC D 230")
G30 = open("G30.pdb", 'r')
G30 = G30.read()
if p30sus[0:8] == " G B 30":
    p30sus = p30sus.replace(p30sus[0:8], "DG D 230")
    p30sus = p30sus + G30
    b = b.replace(b[74214:74222], "DG D 230")
P30 = open("p30.pdb", 'w')
P30.write(p30sus)
reemplazo30 = p30sus[1:2000]
p30clon.close()
P30.close()
#print reemplazo30
#REEMPLAZANDO LA BASE 29#
p29 = open("p29.pdb", 'w')
p29.write(AXC[incia29-2:finaliza29-21])
p29.close()
p29clon = open("p29.pdb", 'r')
p29sus = p29clon.read()
#print p29sus[0:8]
T29 = open("T29.pdb", 'r')
T29 = T29.read()
if p29sus[0:8] == " T B 29":
    p29sus = p29sus.replace(p29sus[0:8], "DT D 229")
    p29sus = p29sus + T29
    b = b.replace(b[71622:71630], "DT D 229")
A29 = open("A29.pdb", 'r')
A29 = A29.read()
if p29sus[0:8] == " A B 29":
    p29sus = p29sus.replace(p29sus[0:8], "DA D 229")
    p29sus = p29sus + A29
    b = b.replace(b[71622:71630], "DA D 229")
C29 = open("C29.pdb", 'r')
C29 = C29.read()
if p29sus[0:8] == " C B 29":
    p29sus = p29sus.replace(p29sus[0:8], "DC D 229")
    p29sus = p29sus + C29
    b = b.replace(b[71622:71630], "DC D 229")
G29 = open("G29.pdb", 'r')
G29 = G29.read()
if p29sus[0:8] == " G B 29":
    p29sus = p29sus.replace(p29sus[0:8], "DG D 229")
    p29sus = p29sus + G29
    b = b.replace(b[71622:71630], "DG D 229")
P29 = open("p29.pdb", 'w')
P29.write(p29sus)
reemplazo29 = p29sus[1:2000]
p29clon.close()
P29.close()
#print reemplazo29
#REEMPLAZANDO LA BASE 28#
p28 = open("p28.pdb", 'w')
p28.write(AXC[incia28-2:finaliza28-21])
p28.close()
p28clon = open("p28.pdb", 'r')
p28sus = p28clon.read()
#print p28sus[0:8]
T28 = open("T28.pdb", 'r')
T28 = T28.read()
if p28sus[0:8] == " T B 28":
    p28sus = p28sus.replace(p28sus[0:8], "DT D 228")
    p28sus = p28sus + T28
    b = b.replace(b[69030:69038], "DT D 228")
A28 = open("A28.pdb", 'r')
A28 = A28.read()
if p28sus[0:8] == " A B 28":
    p28sus = p28sus.replace(p28sus[0:8], "DA D 228")
    p28sus = p28sus + A28
    b = b.replace(b[69030:69038], "DA D 228")
C28 = open("C28.pdb", 'r')
C28 = C28.read()
if p28sus[0:8] == " C B 28":
    p28sus = p28sus.replace(p28sus[0:8], "DC D 228")
    p28sus = p28sus + C28
    b = b.replace(b[69030:69038], "DC D 228")
G28 = open("G28.pdb", 'r')
G28 = G28.read()
if p28sus[0:8] == " G B 28":

```

```

        p28sus = p28sus.replace(p28sus[0:8], "DG D 228")
        p28sus = p28sus + G28
        b = b.replace(b[69030:69038], "DG D 228")
P28 = open("p28.pdb",'w')
P28.write(p28sus)
reemplazo28 = p28sus[1:2000]
p28clon.close()
P28.close()
#print reemplazo28
#REEMPLAZANDO LA BASE 27#
p27 = open("p27.pdb",'w')
p27.write(AXC[incia27-2:finaliza27-21])
p27.close()
p27clon = open("p27.pdb",'r')
p27sus = p27clon.read()
#print p27sus[0:8]
T27 = open("T27.pdb",'r')
T27 = T27.read()
if p27sus[0:8] == " T B 27":
    p27sus = p27sus.replace(p27sus[0:8], "DT D 227")
    p27sus = p27sus + T27
    b = b.replace(b[66438:66446], "DT D 227")
A27 = open("A27.pdb",'r')
A27 = A27.read()
if p27sus[0:8] == " A B 27":
    p27sus = p27sus.replace(p27sus[0:8], "DA D 227")
    p27sus = p27sus + A27
    b = b.replace(b[66438:66446], "DA D 227")
C27 = open("C27.pdb",'r')
C27 = C27.read()
if p27sus[0:8] == " C B 27":
    p27sus = p27sus.replace(p27sus[0:8], "DC D 227")
    p27sus = p27sus + C27
    b = b.replace(b[66438:66446], "DC D 227")
G27 = open("G27.pdb",'r')
G27 = G27.read()
if p27sus[0:8] == " G B 27":
    p27sus = p27sus.replace(p27sus[0:8], "DG D 227")
    p27sus = p27sus + G27
    b = b.replace(b[66438:66446], "DG D 227")
P27 = open("p27.pdb",'w')
P27.write(p27sus)
reemplazo27 = p27sus[1:2000]
p27clon.close()
P27.close()
#print reemplazo27
#REEMPLAZANDO LA BASE 26#
p26 = open("p26.pdb",'w')
p26.write(AXC[incia26-2:finaliza26-21])
p26.close()
p26clon = open("p26.pdb",'r')
p26sus = p26clon.read()
#print p26sus[0:8]
T26 = open("T26.pdb",'r')
T26 = T26.read()
if p26sus[0:8] == " T B 26":
    p26sus = p26sus.replace(p26sus[0:8], "DT D 226")
    p26sus = p26sus + T26
    b = b.replace(b[64008:64016], "DT D 226")
A26 = open("A26.pdb",'r')
A26 = A26.read()
if p26sus[0:8] == " A B 26":
    p26sus = p26sus.replace(p26sus[0:8], "DA D 226")
    p26sus = p26sus + A26
    b = b.replace(b[64008:64016], "DA D 226")
C26 = open("C26.pdb",'r')
C26 = C26.read()
if p26sus[0:8] == " C B 26":
    p26sus = p26sus.replace(p26sus[0:8], "DC D 226")
    p26sus = p26sus + C26
    b = b.replace(b[64008:64016], "DC D 226")
G26 = open("G26.pdb",'r')
G26 = G26.read()
if p26sus[0:8] == " G B 26":
    p26sus = p26sus.replace(p26sus[0:8], "DG D 226")
    p26sus = p26sus + G26
    b = b.replace(b[64008:64016], "DG D 226")
P26 = open("p26.pdb",'w')
P26.write(p26sus)
reemplazo26 = p26sus[1:2000]
p26clon.close()
P26.close()
#print reemplazo26
#REEMPLAZANDO LA BASE 15#
p15 = open("p15.pdb",'w')
p15.write(AXC[incia15-2:finaliza15-21])
p15.close()
p15clon = open("p15.pdb",'r')
p15sus = p15clon.read()
T15 = open("T15.pdb",'r')
t15 = T15.read()
#print p15sus[0:8]
if p15sus[0:8] == " T A 15":
    p15sus = p15sus.replace(p15sus[0:8], "DT C 215")

```

```

        p15sus = p15sus + t15
        b = b.replace(b[35820:35828], "DT C 215")
A15 = open("A15.pdb",'r')
A15 = A15.read()
if p15sus[0:8] == " A A 15":
    p15sus = p15sus.replace(p15sus[0:8], "DA C 215")
    p15sus = p15sus + A15
    b = b.replace(b[35820:35828], "DA C 215")
C15 = open("C15.pdb",'r')
C15 = C15.read()
if p15sus[0:8] == " C A 15":
    p15sus = p15sus.replace(p15sus[0:8], "DC C 215")
    p15sus = p15sus + C15
    b = b.replace(b[35820:35828], "DC C 215")
G15 = open("G15.pdb",'r')
G15 = G15.read()
if p15sus[0:8] == " G A 15":
    p15sus = p15sus.replace(p15sus[0:8], "DG C 215")
    p15sus = p15sus + G15
    b = b.replace(b[35820:35828], "DG C 215")
P15 = open("p15.pdb",'w')
P15.write(p15sus)
reemplazo15 = p15sus[1:2000]
p15clon.close()
P15.close()
#print reemplazo15
#REEMPLAZANDO LA BASE 14#
p14 = open("p14.pdb",'w')
p14.write(AXC[incia14-2:finaliza14-2])
p14.close()
p14clon = open("p14.pdb",'r')
p14sus = p14clon.read()
T14 = open("T14.pdb",'r')
t14 = T14.read()
#print p14sus[0:8]
if p14sus[0:8] == " T A 14":
    p14sus = p14sus.replace(p14sus[0:8], "DT C 214")
    p14sus = p14sus + t14
    b = b.replace(b[33228:33236], "DT C 214")
A14 = open("A14.pdb",'r')
A14 = A14.read()
if p14sus[0:8] == " A A 14":
    p14sus = p14sus.replace(p14sus[0:8], "DA C 214")
    p14sus = p14sus + A14
    b = b.replace(b[33228:33236], "DA C 214")
C14 = open("C14.pdb",'r')
C14 = C14.read()
if p14sus[0:8] == " C A 14":
    p14sus = p14sus.replace(p14sus[0:8], "DC C 214")
    p14sus = p14sus + C14
    b = b.replace(b[33228:33236], "DC C 214")
G14 = open("G14.pdb",'r')
G14 = G14.read()
if p14sus[0:8] == " G A 14":
    p14sus = p14sus.replace(p14sus[0:8], "DG C 214")
    p14sus = p14sus + G14
    b = b.replace(b[33228:33236], "DG C 214")
P14 = open("p14.pdb",'w')
P14.write(p14sus)
reemplazo14 = p14sus[1:2000]
p14clon.close()
P14.close()
#print reemplazo14
#REEMPLAZANDO LA BASE 13#
p13 = open("p13.pdb",'w')
p13.write(AXC[incia13-2:finaliza13-2])
p13.close()
p13clon = open("p13.pdb",'r')
p13sus = p13clon.read()
T13 = open("T13.pdb",'r')
t13 = T13.read()
#print p13sus[0:8]
if p13sus[0:8] == " T A 13":
    p13sus = p13sus.replace(p13sus[0:8], "DT C 213")
    p13sus = p13sus + t13
    b = b.replace(b[30636:30644], "DT C 213")
A13 = open("A13.pdb",'r')
A13 = A13.read()
if p13sus[0:8] == " A A 13":
    p13sus = p13sus.replace(p13sus[0:8], "DA C 213")
    p13sus = p13sus + A13
    b = b.replace(b[30636:30644], "DA C 213")
C13 = open("C13.pdb",'r')
C13 = C13.read()
if p13sus[0:8] == " C A 13":
    p13sus = p13sus.replace(p13sus[0:8], "DC C 213")
    p13sus = p13sus + C13
    b = b.replace(b[30636:30644], "DC C 213")
G13 = open("G13.pdb",'r')
G13 = G13.read()
if p13sus[0:8] == " G A 13":
    p13sus = p13sus.replace(p13sus[0:8], "DG C 213")
    p13sus = p13sus + G13
    b = b.replace(b[30636:30644], "DG C 213")

```

```

P13 = open("p13.pdb", 'w')
P13.write(p13sus)
reemplazo13 = p13sus[1:2000]
p13clon.close()
P13.close()
#print reemplazo13
#REEMPLAZANDO LA BASE 12#
p12 = open("p12.pdb", 'w')
p12.write(AXC[incia12-2:finaliza12-21])
p12.close()
p12clon = open("p12.pdb", 'r')
p12sus = p12clon.read()
T12 = open("T12.pdb", 'r')
t12 = T12.read()
#print p12sus[0:8]
if p12sus[0:8] == " T A  12":
    p12sus = p12sus.replace(p12sus[0:8], "DT C 212")
    p12sus = p12sus + t12
    b = b.replace(b[28044:28052], "DT C 212")
A12 = open("A12.pdb", 'r')
A12 = A12.read()
if p12sus[0:8] == " A A  12":
    p12sus = p12sus.replace(p12sus[0:8], "DA C 212")
    p12sus = p12sus + A12
    b = b.replace(b[28044:28052], "DA C 212")
C12 = open("C12.pdb", 'r')
C12 = C12.read()
if p12sus[0:8] == " C A  12":
    p12sus = p12sus.replace(p12sus[0:8], "DC C 212")
    p12sus = p12sus + C12
    b = b.replace(b[28044:28052], "DC C 212")
G12 = open("G12.pdb", 'r')
G12 = G12.read()
if p12sus[0:8] == " G A  12":
    p12sus = p12sus.replace(p12sus[0:8], "DG C 212")
    p12sus = p12sus + G12
    b = b.replace(b[28044:28052], "DG C 212")
P12 = open("p12.pdb", 'w')
P12.write(p12sus)
reemplazo12 = p12sus[1:2000]
p12clon.close()
P12.close()
#print reemplazo12
#REEMPLAZANDO LA BASE 11#
p11 = open("p11.pdb", 'w')
p11.write(AXC[incia11-2:finaliza11-21])
p11.close()
p11clon = open("p11.pdb", 'r')
p11sus = p11clon.read()
T11 = open("T11.pdb", 'r')
t11 = T11.read()
#print p11sus[0:8]
if p11sus[0:8] == " T A  11":
    p11sus = p11sus.replace(p11sus[0:8], "DT C 211")
    p11sus = p11sus + t11
    b = b.replace(b[25452:25460], "DT C 211")
A11 = open("A11.pdb", 'r')
A11 = A11.read()
if p11sus[0:8] == " A A  11":
    p11sus = p11sus.replace(p11sus[0:8], "DA C 211")
    p11sus = p11sus + A11
    b = b.replace(b[25452:25460], "DA C 211")
C11 = open("C11.pdb", 'r')
C11 = C11.read()
if p11sus[0:8] == " C A  11":
    p11sus = p11sus.replace(p11sus[0:8], "DC C 211")
    p11sus = p11sus + C11
    b = b.replace(b[25452:25460], "DC C 211")
G11 = open("G11.pdb", 'r')
G11 = G11.read()
if p11sus[0:8] == " G A  11":
    p11sus = p11sus.replace(p11sus[0:8], "DG C 211")
    p11sus = p11sus + G11
    b = b.replace(b[25452:25460], "DG C 211")
P11 = open("p11.pdb", 'w')
P11.write(p11sus)
reemplazo11 = p11sus[1:2000]
p11clon.close()
P11.close()
#print reemplazo11
#REEMPLAZANDO LA BASE 10#
p10 = open("p10.pdb", 'w')
p10.write(AXC[incia10-2:finaliza10-21])
p10.close()
p10clon = open("p10.pdb", 'r')
p10sus = p10clon.read()
T10 = open("T10.pdb", 'r')
t10 = T10.read()
#print p10sus[0:8]
if p10sus[0:8] == " T A  10":
    p10sus = p10sus.replace(p10sus[0:8], "DT C 210")
    p10sus = p10sus + t10
    b = b.replace(b[22860:22868], "DT C 210")
A10 = open("A10.pdb", 'r')

```

```

A10 = A10.read()
if p10sus[0:8] == " A A 10":
    p10sus = p10sus.replace(p10sus[0:8], "DA C 210")
    p10sus = p10sus + A10
    b = b.replace(b[22860:22868], "DA C 210")
C10 = open("C10.pdb",'r')
C10 = C10.read()
if p10sus[0:8] == " C A 10":
    p10sus = p10sus.replace(p10sus[0:8], "DC C 210")
    p10sus = p10sus + C10
    b = b.replace(b[22860:22868], "DC C 210")
G10 = open("G10.pdb",'r')
G10 = G10.read()
if p10sus[0:8] == " G A 10":
    p10sus = p10sus.replace(p10sus[0:8], "DG C 210")
    p10sus = p10sus + G10
    b = b.replace(b[22860:22868], "DG C 210")
P10 = open("p10.pdb",'w')
P10.write(p10sus)
reemplazo10 = p10sus[1:2000]
p10clon.close()
P10.close()
#print reemplazo10
#REEMPLAZANDO LA BASE 9#
p9 = open("p9.pdb",'w')
p9.write(AXC[incia9-2:finaliza9-21])
p9.close()
p9clon = open("p9.pdb",'r')
p9sus = p9clon.read()
T9 = open("T9.pdb",'r')
t9 = T9.read()
#print p9sus[0:8]
if p9sus[0:8] == " T A 9":
    p9sus = p9sus.replace(p9sus[0:8], "DT C 209")
    p9sus = p9sus + t9
    b = b.replace(b[20268:20276], "DT C 209")
A9 = open("A9.pdb",'r')
A9 = A9.read()
if p9sus[0:8] == " A A 9":
    p9sus = p9sus.replace(p9sus[0:8], "DA C 209")
    p9sus = p9sus + A9
    b = b.replace(b[20268:20276], "DA C 209")
C9 = open("C9.pdb",'r')
C9 = C9.read()
if p9sus[0:8] == " C A 9":
    p9sus = p9sus.replace(p9sus[0:8], "DC C 209")
    p9sus = p9sus + C9
    b = b.replace(b[20268:20276], "DC C 209")
G9 = open("G9.pdb",'r')
G9 = G9.read()
if p9sus[0:8] == " G A 9":
    p9sus = p9sus.replace(p9sus[0:8], "DG C 209")
    p9sus = p9sus + G9
    b = b.replace(b[20268:20276], "DG C 209")
P9 = open("p9.pdb",'w')
P9.write(p9sus)
reemplazo9 = p9sus[1:2000]
p9clon.close()
P9.close()
#print reemplazo9
#REEMPLAZANDO LA BASE 8#
p8 = open("p8.pdb",'w')
p8.write(AXC[incia8-2:finaliza8-21])
p8.close()
p8clon = open("p8.pdb",'r')
p8sus = p8clon.read()
T8 = open("T8.pdb",'r')
t8 = T8.read()
#print p8sus[0:8]
if p8sus[0:8] == " T A 8":
    p8sus = p8sus.replace(p8sus[0:8], "DT C 208")
    p8sus = p8sus + t8
    b = b.replace(b[17676:17684], "DT C 208")
A8 = open("A8.pdb",'r')
A8 = A8.read()
if p8sus[0:8] == " A A 8":
    p8sus = p8sus.replace(p8sus[0:8], "DA C 208")
    p8sus = p8sus + A8
    b = b.replace(b[17676:17684], "DA C 208")
C8 = open("C8.pdb",'r')
C8 = C8.read()
if p8sus[0:8] == " C A 8":
    p8sus = p8sus.replace(p8sus[0:8], "DC C 208")
    p8sus = p8sus + C8
    b = b.replace(b[17676:17684], "DC C 208")
G8 = open("G8.pdb",'r')
G8 = G8.read()
if p8sus[0:8] == " G A 8":
    p8sus = p8sus.replace(p8sus[0:8], "DG C 208")
    p8sus = p8sus + G8
    b = b.replace(b[17676:17684], "DG C 208")
P8 = open("p8.pdb",'w')
P8.write(p8sus)
reemplazo8 = p8sus[1:2000]

```

```

p8clon.close()
P8.close()
#print reemplazo8
#REEMPLAZANDO LA BASE 7#
p7 = open("p7.pdb",'w')
p7.write(AXC[incia7-2:finaliza7-21])
p7.close()
p7clon = open("p7.pdb",'r')
p7sus = p7clon.read()
T7 = open("T7.pdb",'r')
t7 = T7.read()
#print p7sus[0:8]
if p7sus[0:8] == " T A 7":
    p7sus = p7sus.replace(p7sus[0:8], "DT C 207")
    p7sus = p7sus + t7
    b = b.replace(b[15084:15092], "DT C 207")
A7 = open("A7.pdb",'r')
A7 = A7.read()
if p7sus[0:8] == " A A 7":
    p7sus = p7sus.replace(p7sus[0:8], "DA C 207")
    p7sus = p7sus + A7
    b = b.replace(b[15084:15092], "DA C 207")
C7 = open("C7.pdb",'r')
C7 = C7.read()
if p7sus[0:8] == " C A 7":
    p7sus = p7sus.replace(p7sus[0:8], "DC C 207")
    p7sus = p7sus + C7
    b = b.replace(b[15084:15092], "DC C 207")
G7 = open("G7.pdb",'r')
G7 = G7.read()
if p7sus[0:8] == " G A 7":
    p7sus = p7sus.replace(p7sus[0:8], "DG C 207")
    p7sus = p7sus + G7
    b = b.replace(b[15084:15092], "DG C 207")
P7 = open("p7.pdb",'w')
P7.write(p7sus)
reemplazo7 = p7sus[1:2000]
p7clon.close()
P7.close()
#print reemplazo7
#REEMPLAZANDO LA BASE 6#
p6 = open("p6.pdb",'w')
p6.write(AXC[incia6-2:finaliza6-21])
p6.close()
p6clon = open("p6.pdb",'r')
p6sus = p6clon.read()
p6clon.close()
T6 = open("T6.pdb",'r')
t6 = T6.read()
if p6sus[0:8] == " T A 6":
    p6sus = p6sus.replace(p6sus[0:8], "DT C 206")
    p6sus = p6sus + t6
    b = b.replace(b[12654:12662], "DT C 206")
A6 = open("A6.pdb",'r')
A6 = A6.read()
if p6sus[0:8] == " A A 6":
    p6sus = p6sus.replace(p6sus[0:8], "DA C 206")
    p6sus = p6sus + A6
    b = b.replace(b[12654:12662], "DA C 206")
C6 = open("C6.pdb",'r')
C6 = C6.read()
if p6sus[0:8] == " C A 6":
    p6sus = p6sus.replace(p6sus[0:8], "DC C 206")
    p6sus = p6sus + C6
    b = b.replace(b[12654:12662], "DC C 206")
G6 = open("G6.pdb",'r')
G6 = G6.read()
if p6sus[0:8] == " G A 6":
    p6sus = p6sus.replace(p6sus[0:8], "DG C 206")
    p6sus = p6sus + G6
    b = b.replace(b[12654:12662], "DG C 206")
P6 = open("p6.pdb",'w')
P6.write(p6sus)
reemplazo6 = p6sus[1:2000]
p6clon.close()
P6.close()
#print reemplazo6
#SUSTITUCION EN PDB#
b = list(b)
b[87985:89829] = reemplazo35
b[85393:87156] = reemplazo34
b[82801:84564] = reemplazo33
b[80209:81972] = reemplazo32
b[77617:79380] = reemplazo31
b[75025:76788] = reemplazo30
b[72433:74196] = reemplazo29
b[69841:71604] = reemplazo28
b[67249:69012] = reemplazo27
b[64819:66420] = reemplazo26
b[36631:38475] = reemplazo15
b[34039:35802] = reemplazo14
b[31447:33210] = reemplazo13
b[28855:30618] = reemplazo12
b[26263:28026] = reemplazo11

```



```

b[23671:25434] = reemplazo10
b[21079:22842] = reemplazo9
b[18487:20250] = reemplazo8
b[15895:17658] = reemplazo7
b[13465:15066] = reemplazo6
nuevo = open("CArGmutnuevo.pdb",'w')
nuevo.write("".join(b))
nuevo.close()
a.close()
sus.close()

#####
#OBTENCION DE LA FUERZA DE UNION#
#####

#Formacion del complejo MEF2A_CAR_G
DNA = open("CArGmutnuevo.pdb",'r')
CArG = DNA.read()
protein = open("MEF2A.pdb",'r')
MEF2A = protein.read()
complejo = open("complex.pdb",'w')
complejo.write((MEF2A)+(CArG))
DNA.close()
protein.close()
complejo.close()

#Obtencion de archivos .pqr
try:
    PQRCArG="python" + " pdb2pqr.py.in" + " --ff=amber" + " --fout=amber"+" CArGmutnuevo.pdb"+" CArG.pqr"
    os.system(PQRCArG)
except:
    print "No existe el archivo CArGmutnuevo.pdb"
try:
    PQRcomplex="python" + " pdb2pqr.py.in" + " --ff=amber" + " --fout=amber"+" complex.pdb"+" complex.pqr"
    os.system(PQRcomplex)
except:
    print "No existe el archivo complex.pdb"

#Calculo de FU
try:
    FUCan="apbs" + " FEcalc2propio.in" + " --output-file=history.txt"
    os.system(FUCan)
except:
    print "Hay un error en el calculo de FUCan.pdb"

f = open("history.txt",'r')
htr = f.read()
gE = htr.find("globalEnergy")
gEntera = htr[gE+0:gE+37]
cantidad = htr[gE+13:gE+28]
FU = float(cantidad)
#print htr[gE+30:gE+32]
#print FU
FU = str(FU)
#FUfin = ""
gEnterafin = ""
#FUfin.append(FU)
#FUfin = FUfin + FU
#for i in gEntera:
#    gEnterafin = gEnterafin + i + '\n'
#history = open("cantidad.txt", 'w')
history1.write(FU + '\n')
history2 = open("cantidadentera.txt", 'w')
history2.write(gEnterafin)
f.close()
#history.close
history2.close
#for can1 in range(12):
#    can1 = can1 + 1
#    FUfin[can1]=FU
history1.close

```