



Análisis bioquímico y estructural del factor de transcripción mitocondrial humano A, TFAM, en complejo con la secuencia promotora LSP

Anna Rubio Cosials

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized neither its spreading and availability from a site foreign to the TDX service. Introducing its content in a window or frame foreign to the TDX service is not authorized (framing). This rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.



UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA

DEPARTAMENTO
Biología y Biotecnología Molecular

PROGRAMA DE DOCTORAT
Biotecnología

2012/2013

**ANÁLISIS BIOQUÍMICO Y ESTRUCTURAL DEL FACTOR DE
TRANSCRIPCIÓN MITOCONDRIAL HUMANO A, TFAM, EN
COMPLEJO CON LA SECUENCIA PROMOTORA MITOCONDRIAL
LSP**

Memoria presentada por Anna Rubio Cosials para optar al título de doctor por la Universidad de Barcelona

Directora de tesis, Maria Solà Vilarrubias

Doctoranda, Anna Rubio Cosials

Tutora, Josefa Badia Palacín

Anna Rubio Cosials, 2012

No te quedes inmóvil
al borde del camino
no congeles el júbilo
no quieras con desgana
no te salves ahora

Ni nunca

No te salves

Mario Benedetti, No te salves

ÍNDICE

ABREVIATURAS Y ANGLICISMOS	15-19
RESUMEN	21
OBJETIVOS	21
INTRODUCCIÓN	23-104
A. <u>MITOCONDRIA</u>	25-48
1. Introducción a la mitocondria	
1.1. Historia	
1.2. Origen de la mitocondria: teoría endosimbiosis	
2. Estructura	
2.1. Membrana externa	
2.2. Espacio intermembrana	
2.3. Membrana interna	
2.4. Cristae	
2.5. Matriz	
2.6. Membrana Asociada a RE (MAM)	
3. Organización y distribución de la mitocondria	
4. Función	
4.1. Cadena de transporte electrónico (CTE) y fosforilación oxidativa	
4.1.1. <u>La importancia del equilibrio: CTE y fosforilación oxidativa, consecuencias de su desacople</u>	
4.2. Almacenamiento de iones de Calcio	
4.3. Funciones adicionales	
5. Genoma mitocondrial	
5.1. Características del genoma mitocondrial humano	
5.2. D-loop	
5.2.1. <u>D-loop en mitocondria</u>	
5.2.1.1. D-loop: regiones ricas en citosinas y relación con cáncer	
5.2.1.2. D-loop: CSBII y formación de <i>quadruplex</i>	
5.3. Organización del genoma mitocondrial: Nucleoides	
5.4. Heridabilidad del mtADN	
5.5. Elevada tasa de mutación en mtADN	
5.6. Mecanismo de reparación de mtADN	
6. Mitocondria y apoptosis celular	
7. Mitocondria y enfermedades	
7.1. Clasificación de las enfermedades mitocondriales	
7.1.1. <u>Enfermedades debidas a mutaciones en mtDNA</u>	
7.1.2. <u>Mitocondria y diabetes</u>	
7.1.3. <u>Enfermedades neurodegenerativas</u>	
7.1.3.1. Esclerosis lateral amiotrofica (ELA)	
7.1.3.2. Enfermedad de Alzheimer (AD)	
7.1.4. <u>Mitocondria y cáncer</u>	

7.1.5.	<u>Posible relación con el envejecimiento</u>	
7.1.5.1.	Teoría mitocondrial del envejecimiento	
8.	Mitocondria y efectos secundarios en terapias farmacéuticas	
B.	<u>REGULACIÓN Y TRANSCRIPCIÓN MITOCONDRIAL</u>	49-65
1.	Transcripción mitocondrial	
1.1.	Unidades de la transcripción mitocondrial	
1.2.	Promotores para la cadena H: HSP1 y HSP2	
1.3.	Promotor para la cadena L: LSP	
1.4.	Sitio de inicio de transcripción LSP/HSP	
2.	Maquinaria de la transcripción en mamífero	
2.1.	RNA polimerasa mitocondrial (mtRNAP)	
2.2.	TFB1M y TFB2M	
2.2.1.	<u>Descubrimiento</u>	
2.2.2.	<u>TFB1M y TFB2M interactúan con la maquinaria transcripción</u>	
2.2.3.	<u>TFB1M y TFB2M, y su dualidad funcional</u>	
2.2.4.	<u>Papel en transcripción de TFB2M</u>	
2.3.	mTERF y la terminación de la transcripción	
2.4.	TFAM	
3.	Modelos para el complejo de iniciación de la transcripción mitocondrial	
4.	Biogénesis mitocondrial: coordinación entre genes nucleares y mitocondriales	
5.	Relación entre la formación de encefalopatas para la replicación y la transcripción de LSP	
6.	Transcripción mitocondrial en levadura	
7.	Sistema de traducción en mitocondria	
C.	<u>FACTOR DE TRANSCRIPCIÓN MITOCONDRIAL HUMANO A. TFAM</u>	66-104
1.	Introducción	
2.	Descubrimiento de TFAM	
3.	Isoforma de TFAM y modificaciones post-traduccionales	
3.1.	Isoforma	
3.2.	Modificación pos-traduccionales: acetilación de TFAM	
3.3.	Modificación pos-traduccionales: fosforilación de TFAM	
4.	TFAM y la curvatura del ADN	
5.	TFAM y su unión al ADN	
5.1.	TFAM: unión específica de secuencia (LSP,HSP1)	
5.1.1.	<u>Región de protección para LSP y HSP1 por parte de TFAM</u>	
5.1.1.1.	Afinidad por las regiones de unión LSP, HSP de TFAM	
5.1.1.2.	Patrón de reconocimiento de TFAM en LSP	
5.1.2.	<u>Región de protección de TFAM en las secuencias X e Y</u>	
5.2.	TFAM: unión inespecífica de secuencia	
5.2.1.	<u>Unión a la región D-loop</u>	
5.3.	TFAM: unión a otro tipo de ADN	
5.3.1.	<u>Unión de TFAM a ADN dañado</u>	
5.3.1.1.	Unión a ADN dañado con cisplatino por TFAM	
5.3.1.2.	Unión a ADN oxidado por TFAM	

- 5.4. Unión de TFAM a ADN cruciforme
- 6. HMG-box**
 - 6.1. Clasificación HMGbox
 - 6.2. Características estructurales de dominios HMGbox
 - 6.3. HMGbox inducen curvatura en el ADN
 - 6.4. Secuencias putativas de unión HMGbox
- 7. Funciones de TFAM**
 - 7.1. Mantenimiento del mtADN
 - 7.1.1. Modelo de compactación y coordinación de mtADN
 - 7.1.2. Equilibrio entre transcripción y empaquetamiento del mtADN
 - 7.1.3. Modelo de titración de TFAM para regular el número de mtADN copias
 - 7.2. Inicio de la transcripción y especificidad de secuencia
 - 7.3. TFAM y síntesis de encebadores para la replicación
 - 7.4. TFAM y su posible papel en el mecanismo de reparación del mtADN
- 8. TFAM y el papel de su cola C-terminal**
 - 8.1. Unión específica a ADN e inicio de la transcripción
 - 8.2. Unión a TFB2M
 - 8.3. Estabilización de complejo mtADN/TFAM y desenrollamiento
- 9. Interacción de TFAM con otras proteínas**
 - 9.1. Proteasa Lon
 - 9.1.1. Introducción
 - 9.1.2. Efecto de Lon sobre la cantidad de TFAM y el número de copias mtADN
 - 9.1.3. Degradación de TFAM por Lon: reconocimiento
 - 9.2. Proteína supresora tumoral: p53
 - 9.2.1. Introducción a p53
 - 9.2.2. Interacción p53-TFAM
 - 9.2.3. Relevancia biológica de la interacción
 - 9.3. TFB2M
 - 9.3.1. Introducción TFB1M/TFB2M
 - 9.3.2. Interacción de TFB1M/TFB2M con TFAM
 - 9.3.3. Relevancia biológica de la interacción
 - 9.4. Tiorredoxina 2 (TRX2)
 - 9.4.1. Introducción a Tiorredoxina 2
 - 9.4.2. Interacción de TRX2 con TFAM
 - 9.4.3. Relevancia biológica de la interacción
- 10. TFAM en otros organismos**
 - 10.1. TFAM en *Saccharomyces cerevisiae* (ABF2P)
 - 10.1.1. Caracterización de ABF2p
 - 10.1.2. Unión de ABF2p a ADN
 - 10.1.3. Participación en transcripción de ABF2p
 - 10.1.4. Mecanismo de compactación mtADN en levadura
 - 10.2. TFAM en *Caenorhabditis elegans* (HMG-5 o ce-TFAM)
 - 10.2.1. Caracterización de HMG-5
 - 10.2.2. Unión de HMG-5 a ADN

10.2.3.	<u>HMG-5 y su unión a telómeros</u>	
10.3.	TFAM en <i>Xenopus laevis</i> (xl-TFAM)	
10.3.1.	<u>Caracterización de xl-TFAM</u>	
10.3.2.	<u>Unión de xl-TFAM a ADN</u>	
11.	TFAM y su relación en enfermedades	
11.1.	TFAM y estrés oxidativo	
11.2.	TFAM y diabetes	
11.3.	TFAM y obesidad	
11.4.	TFAM y cáncer	
11.4.1.	<u>TFAM y terapias contra cáncer</u>	
11.4.2.	<u>Cáncer colorectal por inestabilidad en microsatélite</u>	
11.5.	TFAM y enfermedades neurodegenerativas	
11.5.1.	<u>Enfermedad Parkinson (PD) y TFAM</u>	
11.5.2.	<u>Enfermedad Alzheimer (AD) y TFAM</u>	
11.6.	TFAM e insuficiencia cardíaca	
RESULTADOS		105-223
A.	<u>CLONACIÓN, EXPRESIÓN, PURIFICACIÓN Y CARACTERIZACIÓN DE TFAM</u>	107-120
1.	Análisis bioinformático	
2.	Clonación	
2.1.	TFAM-HisTag, MBP-TFAM	
2.2.	Constructos con la cola C-terminal delecionada	
2.3.	Dominios: D1, D2, H1, H2S, H2L	
2.4.	Mutagénesis dirigida: mutantes de unión a ADN, mutantes <i>linker</i>	
3.	Pruebas de expresión y solubilización de la proteína	
4.	Purificación	
5.	Homogeneidad de la proteína	
5.1.	Calibración, posible dimerización TFAM	
5.2.	Caracterización de la degradación en TFAM	
5.3.	Optimización del tampón proteico por Thermofluor y DLS	
B.	<u>FORMACIÓN DEL COMPLEJO PROTEÍNA-ADN</u>	121-126
1.	Diseño de oligos para la cristalización	
2.	Formación del complejo	
2.1.	Optimización de la condiciones del complejo	
2.2.	Ensayos de unión con ADN	
3.	Purificación del complejo	
C.	<u>CRISTALIZACIÓN</u>	127-133
1.	Cristalización TFAM-LSP22	
1.1.	Confirmación de la presencia de ADN en los cristales	
2.	Cristalización TFAM-Half1	
3.	Cristalización TFAMSeMet-LSP22	
4.	Cristalización TFAM-LSP22Br	

D. <u>PROCESAMIENTO DE DATOS Y RESOLUCIÓN DE LA ESTRUCTURA</u>	134-140
1. Difracción y procesamiento de TFAMSeMet-LSP22	
2. Difracción y procesamiento de TFAM-LSP22Br	
E. <u>ANÁLISIS ESTRUCTURAL DE TFAM-LSP22</u>	141-149
1. Curvatura del ADN ofrecida por la proteína en el complejo TFAM-LSP22	
2. Superficie electrostática y conservación general en TFAM	
3. Características del ADN LSP22 en la estructura TFAM-LSP22	
4. Contactos TFAM con ADN	
4.1. Dominios HMGbox y su interacción con la región 1 y 2	
4.2. Linker y su interacción con el ADN	
4.3. Cola C-terminal y su interacción con el ADN	
5. Análisis de la perturbación sobre el ADN y el motivo AAC	
F. <u>ANÁLISIS DE LA FLEXIBILIDAD INTRÍNSECA DEL ADN</u>	150-153
G. <u>CARACTERÍSTICAS BIOQUÍMICAS Y BIOFÍSICAS DE HMGbox DE TFAM</u>	154-164
1. Análisis estructural dominios HMGbox de TFAM	
2. Análisis de unión a ADN de los dominios HMGbox de TFAM	
1.1. Diferente patrón de reconocimiento	
1.2. Diferente afinidad de la secuencia LSP22	
2. Análisis electrostático de los dominios HMGbox	
3. Análisis electrostático de los dominios HMGbox	
4. Representación en forma de rueda de hélices alpha	
5. Análisis estructural entre HMGbox y su núcleo hidrofóbico	
6. Comparativa estructural del dominio HMG2 en su forma unida y no unida	
H. <u>ANÁLISIS ESTRUCTURAL DE LAS DOS ESTRUCTURAS CRISTALOGRÁFICAS DE TFAM</u> ..	165-175
1. Superposición global	
1.1. Superposición de los dominios HMGbox	
1.2. Organización de la cola C-terminal	
2. Análisis estructural del ADN	
3. Contactos TFAM-ADN	
3.1. Contactos HMG1 con ADN	
3.2. Contactos HMG2 con ADN	
3.3. Contactos linker con ADN	
3.4. Contactos cola C-terminal con ADN	
4. Discusión final	
I. <u>ANÁLISIS ESTRUCTURAL DE LA ESTRUCTURA NOVEL BIDOMINIO HMGbox</u>	176-178
1. Interacción con ADN y curvatura final	
2. Orientación dominios HMGbox	
3. Región conectora o linker	
J. <u>INTERACCIONES PROTEÍNA-PROTEÍNA</u>	179-185
1. Empaquetamiento cristalino	
2. Análisis de la superficie de interacción y predicción de ensamblajes	

K. <u>PERMUTACIÓN CIRCULAR PARA EL COMPLEJO TFAM-LSP22</u>	186-189
1. Técnicas de medida de la curvatura del ADN	
2. Ensayo con el complejo TFAM-fragmento ADN wt y mutado	
3. Ensayo con dominios	
4. Limitaciones de la técnica	
L. <u>DISPERSIÓN DE RAYOS X A BAJO ÁNGULO (SAXS)</u>	190-203
1. Descripción de la técnica	
2. Caracterización de TFAM libre en solución mediante SAXS	
2.1. Determinación de la flexibilidad en TFAM	
3. Características del sistema TFAM-LSP22	
3.1. Caracterización de TFAM-LSP22 en solución mediante SAXS	
3.2. Comparación de SAXS con modelo cristalográfico	
3.3. Estudio de la dinámica del complejo TFAM-LSP22 en solución	
M. <u>spFRET PARA EL ESTUDIO DE LA CURVATURA DEL ADN EN SOLUCIÓN</u>	204-223
1. Modelos de ADN para aplicar en estudios spFRET	
2. Primeros experimentos TFAM-30bp	
2.1. Comportamiento ADN	
2.2. TFAM en complejo con C'D' 30bp	
3. Datos para TFAM, D1 y mutante en complejo con ADN 30pb	
3.1. D1, H1	
3.2. Mutantes para unión de segunda HMG2	
4. Datos para TFAM, D1 y mutante en complejo con ADN 50pb	
4.1. TFAM	
4.2. D1, H1	
4.3. Mutantes para unión L182	
5. Modificación del cálculo <i>crossstalk correction</i>	
6. Resultados finales	
6.1. Valores de proximity ratio	
6.1.1. <u>Valores de proximity ratio para TFAM y ADN 30pb/ 50pb</u>	
6.1.2. <u>Valores de proximity ratio dominios y mutantes puntuales</u>	
6.2. Proximity ratio vs E_{FRET}	
6.3. Cálculo de R, distancia entre fluoróforos	
6.4. Confirmación del sistema en forma de U	
6.5. Presencia de conformaciones múltiples	
DISCUSIÓN	225-249
A. <u>TFAM IMPONE UNA CURVATURA DE 180° SOBRE LSP</u>	227-229
1. TFAM provoca una torsión de 180° en LSP	
2. Dinámica del complejo TFAM-LSP	
B. <u>TFAM Y SU RECONOCIMIENTO ESPECÍFICO DE SECUENCIA</u>	229-235
1. Secuencias putativas de unión por dominios HMGbox	

2. Motivo invertido AAC	
2.1. Conservación entre mLSP y hLSP, y intercambio funcional	
2.2. Conservación del motivo AAC entre LSP de hominoides	
2.3. Conservación del motivo AAC entre las secuencias de unión a TFAM	
2.4. Ejemplo de motivo en secuencias de unión de xl-TFAM	
3. Implicación energética AAC y <i>pre-bending</i> inicial	
4. Región central (<i>Nxlinker</i>)	
4.1. Conservación <i>cluster</i> de guaninas central	
C. <u>EL DOGMA DE LA COLA C-TERMINAL</u>	235-240
1. Actividad en reconocimiento específico	
2. Cola C-terminal y flexibilidad intrínseca	
3. Cola C-terminal y cooperatividad	
4. Orientación opuesta al inicio de transcripción	
D. <u>IMPORTANCIA DEL LINKER</u>	240-243
1. La zona del linker en la torsión del ADN	
2. Papel del linker en la unión al ADN y función en la activación de la transcripción	
3. El linker participa en el core hidrofóbico que estabiliza HMG2	
4. Reorganización de los dominios por parte del linker	
5. Longitud del linker y correlación con la secuencia de unión de TFAM	
E. <u>MODELO DE INICIO DE LA TRANSCRIPCIÓN EN LSP</u>	244-245
F. <u>MODELO DE EMPAQUETAMIENTO DEL mtADN</u>	246-248
G. <u>TFAM Y BIOGÉNESIS MITOCONDRIAL: zoom-out</u>	249
CONCLUSIONES	251-254
MATERIALES, MÉTODOS Y FUNDAMENTOS	255-303
<u>CLONACIÓN</u>	256-259
1. Clonación estándar	
2. Clonación independiente de ligación: In-fusion	
3. Mutagénesis dirigida	
<u>PRUEBAS DE EXPRESIÓN</u>	259
<u>SOLUBILIZACIÓN</u>	259-260
<u>PURIFICACIÓN</u>	260-263
1. Columna de afinidad	
1.1. Columna de afinidad por níquel	
1.2. Columna de afinidad por MBP	
1.3. Columna de afinidad por Heparina	
2. Cromatografía de exclusión por peso molecular	

3. Columna de intercambio iónico

IDENTIFICACIÓN DE PROTEÍNAS.....263-264

1. Electroforesis de proteínas en condiciones desnaturalizantes
2. Digestión trípica de proteínas
3. Degradación Edman

CARACTERIZACIÓN DE LA ESTABILIDAD DE PROTEÍNAS EN SOLUCIÓN.....264-265

1. Dynamic Light scattering (DLS)
2. Thermofluor

FORMACIÓN DEL COMPLEJO PROTEÍNA-ADN.....265-267

1. Geles nativos de retardo (EMSA)
 - 1.1. Preparación e la muestra
 - 1.2. Preparación de los geles
 - 1.3. observación de los geles

CRISTALOGRAFIA DE PROTEÍNAS.....268-277

1. **Introducción**
 - 1.1. Cristales y simetría
2. **Cristalización**
 - 2.1. Procesos de optimización
 - 2.2. Congelación de los cristales
3. **Rayos X**
 - 3.1. Difracción de rayos X
4. **Recogida de datos**
5. **Procesamiento de datos**
 - 5.1. Análisis de la calidad de datos
6. **Resolución problema de la fase**
 - 6.1. Método de dispersión anómala múltiple (MAD)
 - 6.2. Método de remplazamiento molecular (MR)
7. **Optimización de la fase**
8. **Modelado, refinamiento y validación de la estructura**
 - 8.1. Refinamiento estructural
 - 8.2. Validación de la estructura
 - 8.2.1. Parámetros estadísticos
 - 8.2.2. Parámetros estereoquímicos

PERMUTACIÓN CIRCULAR.....278-280

9. Descripción de la técnica
10. Realización del ensayo de permutación circular

DISPERSIÓN DE RAYOS X A BAJO ÁNGULO (SAXS).....280-289

1. Dispersión de rayos X a bajo ángulo (SAXS)
2. Análisis de los datos
3. **Combinación de datos cristalográficos y SAXS para modelar estructuras**
 - 3.1. Cálculo de curvas de dispersión teóricas a partir de modelos atómicos: CRY SOL
 - 3.1.1. Validación
 - 3.2. Modelado Rigid Body combinado con datos de SAXS: pre_BUNCH
 - 3.3. Estudio de sistemas flexibles, con múltiples conformaciones: EOM
4. **Estudio de sistema dinámicos en SAXS**
5. **Recogida de datos**

5.1. TFAM en su forma no unida al ADN	
5.2. TFAM en complejo con LSP22	
<u>FRET (FÖRSTER RESONANCE ENERGY TRANSFER)</u>	289-296
1. Introducción a FRET	
1.1. Principio físico	
1.1.1. <u>Limitaciones en FRET</u>	
1.2. Aplicación de FRET en complejos proteína-ADN	
2. Single Molecule FRET (spFRET)	
2.1. Diagrama del sistema	
2.2. Limitaciones del sistema confocal de fluorescencia para spFRET	
2.2.1. <u>Detección de la fluorescencia</u>	
2.2.2. <u>Burst size distribution</u>	
2.2.3. <u>Background o ruido de fondo</u>	
2.2.4. <u>Presencia de eventos múltiples</u>	
2.3. Cálculo de la Eficiencia de FRET en spFRET	
<u>SINGLE PARTICLE FRET (spFRET) PARA TFAM EN COMPLEJO CON LSP</u>	296-303
1. Diseño del experimento	
2. Características de los oligos y purificación HPLC del duplex	
3. Formación del complejo y preparación de la muestra para sp-FRET	
4. Adquisición de los datos	
5. Procesamiento de datos	
5.1. FRETtchen	
5.2. Programa IGOR y ORIGIN	
AGRADECIMIENTOS	305-307
BIBLIOGRAFÍA	307-322

ABREVIATURAS Y ANGLICISMOS

A

ABF2P: homólogo de TFAM en *S. cerevisiae*

ADN: Ácido desoxi-ribonucleico

AFM: *Atomic Force Microscopy*

ANT: transportador de nucleótidos de adenina

ARN: Ácido ribonucleico

AVRNs : *AntiViral Ribonucleoside Analogs*

B

Beamline: línea de difracción de rayos X en sincrotrón

BER: mecanismo de reparación por escisión de base

Burst size: distribución de la suma de fotones detectados

C

Cadena L: cadena ligera o *light*

Cadena H: cadena pesada o *heavy*

CSBs: *conserved sequence blocks*

COX: Citocromo C Oxidasa

Crosslinking: reacción química de entrecruzamiento mediante la reacción residuos de Lysina

Cryobuffers: condiciones crioprotectantes

CTE: Cadena de Transporte Electrónico

CTT'S: constructos de TFAM con la cola C-termina delecionada (CTT1, CTT2, CTT3, CTT4)

D

D1: constructo TFAM para el dominio HMG1, de Ser44-Gln125

D2: constructo TFAM para el dominio HMG2, deLeu149-Arg227

D2C: constructo TFAM para el dominio HMG2 con la cola C-terminal, deLeu149-Cys246

DBDs: DNA Binding Domains

DLS: *Dynamic Light Scattering*

DM: *Density Modification*

D_{max} : distancia máxima intramolecular

DPE: *Distal Promoter Element*

DPN-1: *Dynamin-related Protein 1*

E

E_{FRET} : Eficiencia del fenómeno FRET

EMSA: *Electrophoretic Mobility Shift Assay*

EOM: *Ensemble Optimization Method*

ESRF: *European Synchrotron Facility*

F

FAD: dinucleótido de flavina y adenina

Footprinting: análisis de digestión controlada sobre ADN

FRET: *Fluorescence Resonance Energy Transfer*

G

GSH: glutatión reducido

GSL: glicosfingolípidos

H

H1: constructo TFAM para el dominio HMG1 junto con el *linker* (Ser43-Leu152)

H2L: constructo TFAM para el dominio HMG2 junto con el *linker* y cola C-terminal (Ser115-Cys246)

H2S: constructo TFAM para el dominio HMG2 junto con el *linker* (Ser115-Glu219)

HisTag: cola de Histidinas para permitir la purificación mediante una columna de afinidad catiónica

HisTrap: columna de afinidad por Histidinas

HMGbox: *High Mobility Group domain*

HMGbox SS: *HMGbox sequence specific*, de unión específica

HMGbox NSS: *HMGbox nonsequence specific*, de unión inespecífica

HU: *Heat-Unstable nucleoid*

HSP: *Heavy Strand Promotor*

HSP1: *Heavy Strand Promotor 1*

HSP2: *Heavy Strand Promotor 2*

I

$I(0)$: intensidad de dispersión inicial

IHF: *Integration Host Factor*

L

LEF-1: *Lymphoid enhancer-binding factor 1*

Linker: región conectora

LOAD: *Late-onset Alzheimer's*

Loop: lazo

LSP: *Light Strand promotor*

LSP22Br: secuencia de unión de TFAM a LSP cristalizada en complejo, con Timina remplazada pro 5-Bromo Uracilo

M

MAM: *Membrana Mitochondrial Asociada a RE*

MAD: *Multiple-wavelength Anomalous Dispersion*,

MBP-TFAM: TFAM con la dina MBP en N-terminal. MBP (Maltose Binding Protein)

MD: *Molecular Dynamics*

MELAS: *Mitochondrial Encefalomiopatía, Lactic Acidosis ans stroke-like episodes*

MLS: *Mitochondria Leader Sequence*

MR: *Molecular Replacement* o Reemplazamiento Molecular

MonoS: columna de intercambio de intercambio aniónico

mtADN: ADN mitocondrial

mtRNAP: RNA polimerasa de mitocondria

mtSSB: proteína mitocondrial de unión a DNA de cadena única

mTERF: Factor mitocondrial de terminación de transcripción

Mtfb1p: factor de transcripción mitocondrial en *Saccharomyces cerevisiae*

MTP: poro de transición mitocondrial

MW: peso molecular aparente

N

NAD: Dinucleótido de Adenina y Adenina

NER: mecanismo de reparación por escisión de base

NCR: *Non Coding Regions*

NCS: *non-crystallographic symmetry*

NRF-1: factor nuclear de respiración 1

O

O_H: Origen replicación cadena H

O_L: Origen replicación cadena L

Open complex: burbuja creada para permitir la transcripción por parte de RNAPolimerasa

P

PAC: Plataforma Automatizada de Cristalografía

PIC: complejo de pre-iniciación de la transcripción (*Pre-Initiation Complex*)

proximity ratio (P): valor extraído de los experimento de spFRET, relacionado con E_{FRET}

PPARs: *Peroxisome Proliferator-Activated Receptor*

PPE: *Proximal Promoter Element*

PPR domain : *pentatricopeptide repeat*

R

R: distancia entre fluoróforos

R_{merge}: parámetro estadístico

R_g : radio de giro

RE: retículo endoplasmático

RMN: Resonancia Magnética Nuclear

RNAi: RNA interferencia

RNApol T7: RNA polimerasa del bacteriófago T7

ROP41p: RNAPolimerasa mitocondrial en *Saccharomyces cerevisiae*

ROS: especies reactivas de oxígeno

S

SAD: Single-wavelength Anomalous Diffraction

SAXS : Experimentos de rayos X de bajo ángulo (*Small Angle X-ray Scattering*)

SeMet: Seleno Metionina

Screenings: soluciones de escaneo de condiciones de cristalización

Sliding: mecanismo de deslizamiento de proteínas de unión al ADN

SNPs: *single nucleotide polymorphisms*

spFRET o smFRET: *single-particle* FRET

SPR: *Surface Plasmon Resonance*

SRY: *Sex-determining Region Y*

SRY.B: proteína híbrida entre dominio HMGbox de SRY y dominio HMGbox de HMG1

Stacking: apareamiento entre bases de dos hebras diferentes

STED : *Stimulated Emission Depletion Microscopy*

Sup 75 HL 26/60: gel filtración superdex 75 High Load 26/60

T

TAS : *Termination-associated sequences*

TBP: TATA box-Binding Protein

TEV protease: protease de sitio específico del virus del Tabaco (*Tobacco Etch Virus*)

TLS: *Tranlation Libration Screw-motion*

TFAM: mitocondrial Transcription Factor

TFAM-HisTag: TFAM con la dina HisTag en C-terminal

TFB1M/TFB2M: Factor de Transcripción Mitocondrial ½

TRX2: tiorredoxina mitocondrial

U

UBF: *Upstream Binding Factor*

Unwound o *unwind*: desenrollamiento

V

VDAC: canal dependiente de voltage aniónico

Ve= volumen de elución

Δ

ΔP : Potencial electrostático de membrana

μ

$\mu M/\mu E$: movilidad relativa del complejo vs la movilidad del ADN libre

RESUMEN

En esta tesis se ha conseguido la caracterización bioquímica, biofísica y estructural del factor de transcripción mitocondrial A humano, TFAM.

TFAM presenta múltiples funciones en la biogénesis de la mitocondria y está implicada en transcripción, replicación, empaquetamiento del ADN mitocondrial, reparación del ADN mitocondrial... Para tratar de hallar los mecanismos moleculares se procedió a la resolución de la estructura cristalográfica de TFAM en complejo con su secuencia diana de unión en el promotor LSP (*Light Strand Promotor*).

La estructura de TFAM en complejo con una secuencia de 22 pb de LSP (Rubio-Cosials, Sidow et al.) mostraba cómo TFAM impone una torsión global del ADN de 180°, con ambos dominios HMGbox introduciendo un punto de curvatura mediante la unión al surco menor del ADN. Estudios en solución mediante SAXS (*Small Angle X-ray Scattering*) de TFAM en forma no unida permitieron asignar un alto grado de flexibilidad para la región conectora entre los dominios HMGbox y la cola C-terminal, ambas regiones cargadas positivamente. Esta flexibilidad de la zona conectora permite un gran número de conformaciones diferentes para TFAM en su forma libre, donde los dos dominios HMGbox se encuentran con posiciones relativas diferentes.

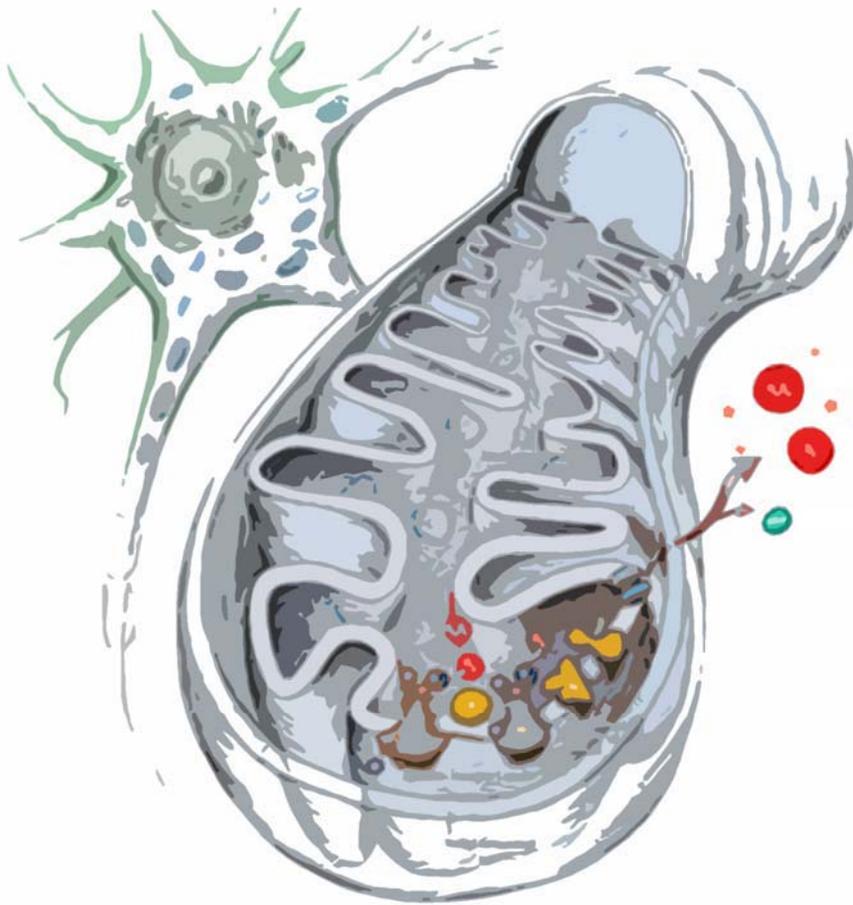
Posteriormente mediante estudios de spFRET (*single-particle FRET*), se confirmó el sistema de curvatura del ADN definido por la estructura cristalográfica con la introducción de una torsión en éste en forma de U. Los estudios de SAXS y spFRET indicaban que esta unión presentaba un comportamiento ligeramente dinámico, posiblemente debido a la flexibilidad intrínseca que muestra la región conectora entre los dos dominios HMGbox en solución.

Los resultados obtenidos permiten entender la multitud de funciones asignadas a TFAM y en concreto, su papel en la activación de la transcripción a partir de LSP.

OBJETIVOS

El objetivo del trabajo realizado durante la tesis doctoral fue caracterizar el factor de transcripción mitocondrial humano A, TFAM, tanto en su forma libre como en la forma unida al ADN, concretamente a su secuencia de unión a la región promotora LSP. La primera etapa del proyecto implicó la obtención de la proteína estable y establecer un protocolo de formación del complejo con el ADN. La segunda fase consistió en la caracterización molecular de la unión de TFAM a LSP. Para ello se aplicaron diversas metodologías: cristalografía de proteínas, dispersión de rayos X a bajo ángulo en solución (SAXS), técnicas basadas en el principio de fluorescencia (FRET), ensayos bioquímicos,.... Todo ello permitió entender los mecanismos moleculares del complejo TFAM-ADN y su relación con las múltiples funciones asociadas a TFAM dentro de la mitocondria.

INTRODUCCIÓN



A. MITOCONDRIA

1. Introducción a la mitocondria

La palabra mitocondria proviene de *mitochondrion* del griego *μίτος* *mitos*, hilo, + *χονδρίον* *chondrion*, granular. La mitocondria tiene un tamaño medio de aproximadamente 0.3-1.0 por 5-10 μm . Puede desplazarse a través del citoplasma celular, y su estructura básica la forman dos membranas diferentes en cuanto a composición y función. La mitocondria es conocida como el proveedor oficial de ATP, fuente de energía para la célula.

1.1. Historia

Las primeras observaciones de estructuras intracelulares que posiblemente representaban mitocondrias fueron publicadas en 1840. La mitocondria fue el primer orgánulo celular relacionado con una enfermedad humana. Esto sucedió en 1962, cuando Luftet describió un paciente con hipermetabolismo el cuál presentaba disfunción mitocondrial.

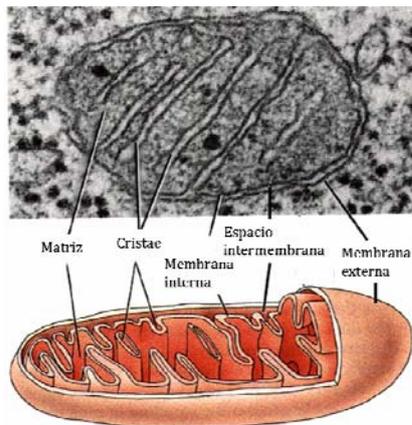
1.2. Origen de la mitocondria: teoría endosimbiosis

Lynn Margulis formuló la teoría de la endosimbiosis que propone que las actuales mitocondrias son descendientes de organismos primitivos con capacidad aeróbica, los cuales fueron “ingeridos” por células protoeucariotas anaeróbicas hace 1.500-2.000 millones de años. Como consecuencia de esta asociación las dos especies se habrían beneficiado: el huésped utilizando energía generada por la actividad aeróbica de la bacteria incorporada, y éste accediendo a nutrientes que conseguiría el huésped.

La teoría de la endosimbiosis se basa en algunas semejanzas entre las bacterias actuales con las mitocondrias y los cloroplastos: ambos orgánulos tienen unas dimensiones parecidas a las bacterias, poseen hebras circulares de ADN en su interior y sus ribosomas son 70S, similares a los de las bacterias. Además, son capaces de replicarse de forma independiente en el interior celular y la doble membrana de estos orgánulos sugiere una incorporación por invaginación de la membrana del protoeucariota, la membrana interna sería de origen procariota y la externa de origen protoeucariota.

2. Estructura

A continuación se describen brevemente los diferentes compartimentos presentes en mitocondria (Fig_A1).



Fig_A1: Imagen microscópica de una mitocondria acompañado de un esquema ilustrativo que muestra la estructura mitocondrial (fuente: *Wikipedia*).

2.1. Membrana externa

La membrana externa mitocondrial engloba todo el orgánulo y está formada por un ratio proteína: fosfolípido similar a la membrana plasmática eucariota. Presenta un gran número de porinas, proteínas que actúan formando canales y permitiendo la difusión de proteínas de igual o menor peso molecular a 5000 Da. La membrana mitocondrial externa se encuentra asociada a la membrana del Retículo Endoplasmático (RE), en una estructura llamada MAM (*Membrana Mitocondrial Asociada a RE*). Esta estructura es importante para la señalización de calcio entre RE y mitocondria, así como en el transporte de lípidos entre RE y mitocondria.

2.2. Espacio intermembrana

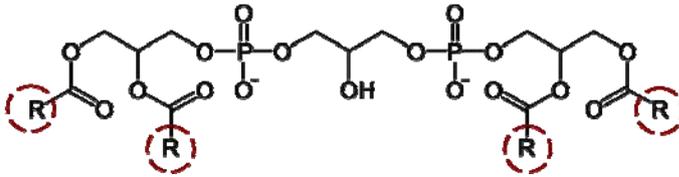
El espacio intermembrana es aquél situado entre la membrana exterior y la interior. Debido a que la membrana exterior es completamente permeable a moléculas pequeñas, la concentración de moléculas pequeñas, como iones y azúcares, es la misma entre el espacio intermembrana y el citosol. Sin embargo, las moléculas mayores deben presentar una secuencia específica de señalización para ser importadas a través de la membrana exterior, con lo que la composición de proteínas en este espacio es diferente de la del citosol. Una de las proteínas situadas en este espacio intermembrana es el CitocromoC.

2.3. Membrana interna

La membrana mitocondrial interna contiene proteínas involucradas en cinco tipos de funciones: fosforilación oxidativa, transporte específico de metabolitos dentro y fuera de la matriz, fusión y fisión de mitocondria e importación de proteínas.

La membrana interna contiene más de 151 polipéptidos diferentes, y tiene un elevado ratio proteína:fosfolípido (en peso más de 3:1, es decir, alrededor de 1 proteína por 15 fosfolípidos). La membrana interna constituye un 50% del total de proteína en la mitocondria.

Esta membrana es rica en un fosfolípido inusual, cardiolipina (Fig_A2). La cardiolipina tiene 4 ácidos grasos (comúnmente los fosfolípidos presentan sólo dos ácidos grasos), y puede ayudar a impermeabilizar la membrana interna. Al contrario que la membrana externa, la membrana interna no contiene porinas, lo que la hace altamente impermeable a todas las moléculas, por lo que todos los iones y moléculas requieren transportadores de membrana específicos para entrar o salir de la matriz.



Fig_A2: Fórmula de la cardiolipina. Posee una estructura dimérica única que contiene unidos cuatro ácidos grasos y dos cargas negativas. Una de las principales funciones de la cardiolipina es que sin su presencia en la membrana mitocondrial interna, una gran cantidad de enzimas de la fosforilación oxidativa no podrían desarrollar su función correspondiente. Variaciones en el contenido o estructura de cardiolipina se han descrito en numerosas alteraciones patológicas (fuente: *Wikipedia*).

2.4. Cristae

La membrana interna se encuentra compartimentada en numerosas crestas o cristae, permitiendo así expandir su superficie ampliando su capacidad para producir ATP. Por ejemplo, en el caso de mitocondria de hígado, el área de la membrana interna es cinco veces mayor que la membrana exterior. La cantidad de cristae está relacionada con la demanda de ATP y, por tanto, depende del tipo de célula.

2.5. Matriz

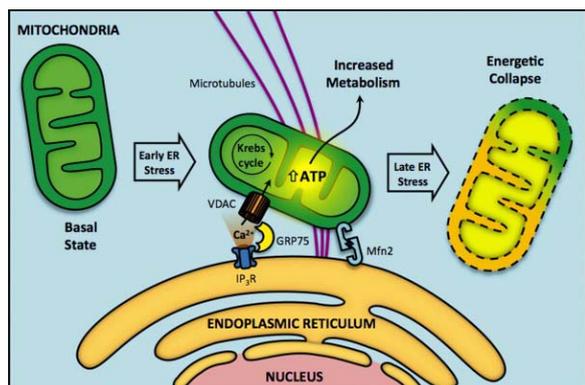
La matriz es el espacio que encierra la membrana interna. Contiene un 66% del total proteico de la mitocondria. La matriz es importante para la producción de ATP con la ayuda de la ATP sintasa, que se encuentra en la membrana interna. La matriz contiene una elevada concentración de enzimas (implicadas en funciones como la oxidación del piruvato y ácidos grasos), ribosomas mitocondriales, tRNA y varias copias de ADN mitocondrial (mtADN).

2.6. Membrana Mitocondrial Asociada a RE (MAM)

MAM es otro elemento estructural conocido por su papel crítico en la fisiología celular y la homeostasis. Se trata de un espacio membranoso situado entre el RE y la membrana exterior de la mitocondria. MAM es imprescindible para RE, ya que presenta las chaperonas de RE (incluidas lectina, calnexina,...) que controlan la homeostasis de RE, pero también juegan un papel en el inicio de la apoptosis medida por estrés en el RE, mediante la señalización por calcio en el RE (Fig_A3). Los roles opuestos de las chaperonas localizadas en MAM sugieren la existencia de mecanismos que regulen el equilibrio entre estado pro-homeostático o pro-apoptótico. Recientemente, la GTPasa Rab32 ha sido identificada como regulador de las propiedades de MAM.

La mitocondria está en contacto estrecho con las vías de secreción. De forma mayoritaria, estos contactos ocurren en MAM. Esta región entre RE y mitocondria ha sido recientemente asignada como mayor eje de señalización celular. La señalización en MAM utiliza la transferencia de calcio entre RE y mitocondria. La importación eficiente de calcio en RE mantiene las chaperonas funcionales y evita la acumulación de proteínas desplegadas en RE, hecho que sucede en situaciones de estrés en RE; pero en la situación inversa, existe un rápido flujo desde RE y se transfiere a la mitocondria conduciendo a la apoptosis por estrés. Por otra parte, el metabolismo mitocondrial se beneficia del incremento de calcio mediante un incremento de producción de NADH a través de piruvato, pero una sobrecarga o exceso de calcio resulta en una producción masiva de especies reactivas de oxígeno y una pérdida del potencial de membrana mitocondrial. Por lo tanto, el transporte de calcio de RE a MAM actúa como un arma de doble filo, sugiriendo que bajo mecanismos de regulación desconocidos puede actuar como pro-homeostático o pro-apoptótico.

Además, MAM está involucrada en la biosíntesis y tráfico de fosfolípidos entre RE y mitocondria. Es en MAM donde se encuentran las enzimas que participan en la biosíntesis de GSL (glicosfingolípidos). GSL son componentes esenciales de las membranas de todas las células eucariotas, conformando un 76% de las membranas.



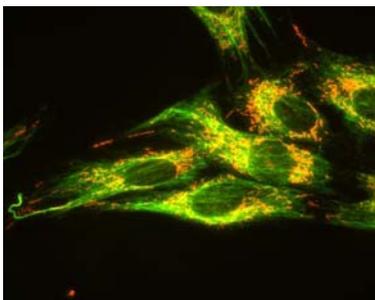
Fig_A3: Regulación del equilibrio homeostasis/apoptosis mediante la estrecha comunicación existente entre MAM y mitocondria. La mitocondria está en contacto estrecho con las vías de secreción. Mayoritariamente, estos contactos ocurren en MAM. Esta región entre RE y mitocondria ha sido recientemente asignada como mayor eje de señalización celular, en la interacción entre MAM y RE se encuentra involucrada la mitofusina Mfn2. La señalización en MAM utiliza la transferencia de calcio entre RE y mitocondria a través del canal VDAC (*Voltage-dependent Anion Channel*). Una importación eficiente de calcio entre RE y mitocondria permite mantener las chaperonas de RE estables. En situaciones de estrés en RE, existe un rápido flujo de calcio conduciendo a una apoptosis por estrés. La mitocondria también puede llegar a apoptosis en situaciones donde existe una mayor producción de ATP, dando lugar a un elevado número de especies ROS. Por lo tanto, el transporte de calcio de RE a MAM actúa como un arma de doble filo, sugiriendo que bajo mecanismos de regulación desconocidos puede actuar como pro-homeostático o pro-apoptótico. (Bui, M. et al. 2010)

3. Organización y distribución de la mitocondria

Las mitocondria se encuentran en casi todo los eucariotas. La variedad en el número y la localización depende del tipo de célula. En los organismos unicelulares normalmente

encontramos una sola mitocondria, mientras que en células de hígado humanas encontramos entre 1.000-2.000 mitocondrias por célula, representado un 1/5 del volumen celular.

Normalmente se encuentran formando un entramado 3D en complejo con el citoesqueleto (Fig_A4). La asociación con el citoesqueleto determina la forma de la mitocondria, lo que puede afectar también la función. Recientemente se ha propuesto que vimentina, uno de los componentes del citoesqueleto, es crítico para la asociación mitocondria-citoesqueleto. Además, la forma general o final de las mitocondrias depende del balance entre fusión y fisión de cada mitocondria. Estos procesos se encuentran controlados por una serie de proteínas, incluida *dynamin-related protein 1* (DPN-1) y de mitofusinas (por ejemplo, mfn-1 y 2).



Fig_A4: Distribución de la mitocondria (en rojo, anticuerpo contra hsp70) en relación con los filamentos intermedios (en verde, anticuerpo contra vimentina) en células de fibroblasto. (De Vos, K. et al. 2007)

La mitocondria también presenta un movimiento a través de la célula, posiblemente gracias a la participación de los motores actina/miosina y dineina/kinesina a lo largo de las redes de microfilamentos y microtúbulos (Iborra, Kimura et al. 2004).

4. Función

Una de las funciones más conocidas de la mitocondria es producir ATP a través de la respiración, y así regular el metabolismo celular. Sin embargo, la mitocondria tiene otras funciones además de producir ATP.

4.1. Cadena de transporte electrónico (CTE) y fosforilación oxidativa

La **Cadena de Transporte Electrónico** (CTE) se considera la principal función mitocondrial, ya que proporciona alrededor de un 90% del requerimiento celular de ATP.

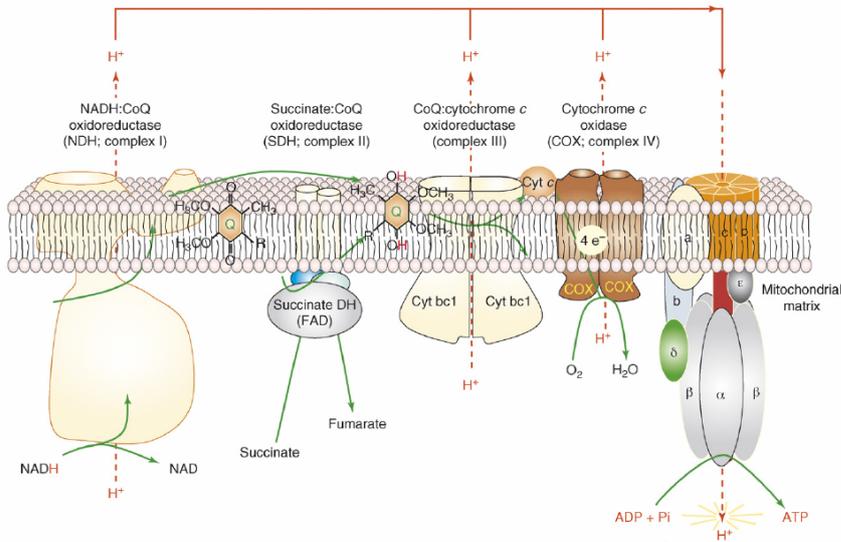
En la mitocondria entran ácidos grasos que son oxidados a acetilCo enzima A (acetilCoA) en el proceso llamado *B-oxidación*. El piruvato (producto final de la glicólisis en el citoplasma celular) también entra a la mitocondria, y se descarboxila por la piruvato descarboxilasa, que incorpora la molécula de la coenzima A y se produce acetilCoA. Posteriormente acetilCoA se oxida en el ciclo de los ácidos tricarbónicos o *ciclo de Krebs*, al final del cual se generan dos moléculas de CO₂ y 4 pares de átomos de hidrógeno, los cuáles se transfieren al dinucleótido de nicotinamida y adenina (NAD⁺) y al dinucleótido de flavina y adenina (FAD) que se reducen a NADH y FADH₂, respectivamente. Estos compuestos tienen una capacidad donadora

de electrones muy elevada y esta energía se va liberando en una serie de reacciones redox que tienen lugar en las crestas mitocondriales, donde se encuentra una serie de complejos enzimáticos y coenzimas formando la cadena de transporte de electrones (CTE) o cadena respiratoria mitocondrial (CRM) (Fig_A5).

Las proteínas que intervienen en CTE están agrupadas en complejos localizados en la membrana de las crestas. Las coenzimas están envueltas en los lípidos y difunden a lo largo de la membrana o a través de la superficie. El complejo I, NADH ubiquinona oxidoreductasa (o NADH deshidrogenasa), es el más grande y está compuesto por un mínimo de 42 polipéptidos. Cataliza la transferencia de electrones desde el NADH a la ubiquinona o coenzima Q (CoQ). El complejo II, succinato CoQ oxidoreductasa, está formado por 4 péptidos y cataliza la oxidación de succinato a fumarato (dentro del ciclo de Krebs), y transfiere los electrones. La citocromo C oxidoreductasa o complejo III está formada aproximadamente por 10 subunidades. El citocromo C actúa como intermediario para la transferencia de electrones entre los complejos III y IV y está localizado en el espacio intermembrana. El complejo IV, citocromo C oxidasa (COX) formado por 13 subunidades, transfiere 4 electrones desde el citocromo C al oxígeno molecular, y se genera una molécula de agua.

Esta transferencia de electrones de un miembro de CTE a otro genera energía que se invierte en la síntesis de ATP por una vía indirecta. En el inicio de la cadena, las moléculas de NADH y FADH₂ transfieren los electrones al siguiente complejo enzimático, y liberan los protones a la matriz mediante mecanismos no conocidos aún. Cuando los complejos enzimáticos transfieren electrones al siguiente componente de la cadena, utilizan energía generada en este proceso para bombear protones al espacio intermembrana traspasando la membrana interna. Esta acumulación de protones en el espacio intermembrana genera un potencial eléctrico que tiende a llevar a los protones de nuevo hacia la matriz. Pero la membrana interna es extremadamente impermeable a los protones y es necesaria la acción del complejo V (F₀-F₁-ATP sintasa) para que estos puedan ser transportados de nuevo al interior mitocondrial. La energía generada con la entrada de protones se utiliza para sintetizar ATP a partir de ADP citoplasmático, importado a través del transportador de nucleótidos de adenina (ANT). El complejo V está constituido por 12-14 polipéptidos, y presenta dos regiones: la F₀, que está ligada a la membrana y F₁, que representa la parte catalítica de la molécula incrustada en la matriz mitocondrial. El proceso de síntesis de ATP es conocido como **fosforilación oxidativa**.

El año 1961, Peter Mitchell propuso el proceso del metabolismo oxidativo por el cual las células consiguen ATP. Para que la fosforilación oxidativa tenga lugar hacen falta dos condiciones: que la membrana interna esté intacta de forma que los protones sólo puedan volver a la matriz por un proceso acoplado a la síntesis de ATP; y que se acumule una concentración elevada de protones en el espacio intermembrana.



Fig_A5. Esquema de CTE que consiste en 5 complejos respiratorios, localizados en la membrana interna de la mitocondria, cada complejo compuesto por múltiples subunidades. CTE finaliza con el proceso de la fosforilación oxidativa que da lugar a la generación de moléculas de ATP. (Wallace, K.2008)

4.1.1. La importancia del equilibrio: CTE y fosforilación oxidativa, consecuencias de su desacople

Como se ha comentado anteriormente, los complejos proteínicos en la membrana interna (NADH deshidrogenas, citocromo C oxidoreductasa y COX) llevan a cabo la transferencia de electrones y el incremento de energía gracias al bombeo de protones (H⁺) en el espacio intermembrana. Este proceso es eficiente, pero un pequeño porcentaje de electrones pueden reducir el oxígeno dando lugar a formas o especies reactivas del oxígeno (ROS), como el superóxido (O₂⁻). Eso puede provocar estrés oxidativo en la mitocondria y contribuir a la disfunción de la mitocondria asociada con el envejecimiento (ver apartado Introducción A.7.1.5).

En situaciones de exceso de calorías y ausencia de ejercicio, existe una ausencia de transporte de ADP citosólico a mitocondria que provoca la imposibilidad de sintetizar moléculas de ATP nuevas. Sin embargo, CTE continúa generando electrones hasta que existe una hiperpolarización del potencial de membrana o electrostático (ΔP) que inhibe el bombeo de protones. Es en este punto donde CTE frena y los transportadores de electrones quedan totalmente reducidos (es decir, cargados de electrones). Los electrones pueden ser transferidos espontáneamente a O₂, generando así un incremento de ROS y llevando a la célula a una situación de estrés oxidativo.

En contraste, en los individuos que realizan ejercicio activamente y generan ADP, la ATP sintasa mantiene ΔP hipopolarizado y los electrones continúan fluyendo a través de la CTE manteniendo el ΔP . De esta forma los transportadores de electrones se mantienen de forma

oxidada y pueden captar electrones. Esto implica una reducción de la generación de ROS, reduciéndose así el estrés oxidativo (Wallace 2005).

4.2. Almacenamiento de iones de calcio

La concentración de calcio en la célula puede regular diversas reacciones, ya que se encuentra ligada a muchos procesos de señalización. La mitocondria puede transitoriamente almacenar calcio, contribuyendo a la homeostasis de la célula. De hecho, la rapidez con la que puede captar el calcio para luego liberarlo rápidamente, hace que la mitocondria actúe como “tampón citosólico” para el calcio, un proceso que puede ir relacionado con la comunicación con RE a través de MAM.

4.3. Funciones adicionales

La mitocondria juega un papel central en muchas otras acciones metabólicas, como la regulación del potencial de membrana, apoptosis programada, regulación de la proliferación celular y síntesis de esteroides.

Algunas de estas funciones adicionales mitocondriales son llevadas a cabo específicamente en algunos tipos de células. Por ejemplo, las mitocondrias de células de hígado pueden detoxificar amonio, un residuo producto del metabolismo proteico. Una mutación en alguno de los genes que regulan estas funciones puede resultar en una enfermedad mitocondrial.

5. Genoma mitocondrial

En 1963 Nass y Nass (Nass and Nass 1963) descubrieron que las mitocondrias contienen su propio DNA (mtDNA). En la actualidad se conoce la secuencia completa del mtDNA de mucho vertebrados e invertebrados, y se sabe que sus funciones esenciales están muy conservadas. Aunque el mtDNA de algunas especies de protozoos ciliados, algas o levaduras es lineal, la gran mayoría de mtDNAs son circulares. El sistema genético mitocondrial más estudiado es el de los humanos.

De acuerdo con la ampliamente aceptada teoría de la endosimbiosis, la mitocondria proviene de una alpha-proteobacteria (Gray, Burger et al. 1999). Con el paso del tiempo, los genes ancestrales bacterianos han sido transferidos al genoma nuclear, algo que se puede evidenciar con la presencia de genes ortólogos en el genoma mitocondrial de algunas especies y en el genoma nuclear de otras especies. La transferencia genética explica porqué todas las proteínas necesarias para la replicación, transcripción y traducción del genoma mitocondrial están codificadas en el núcleo. Sin embargo, no es aún conocido porqué la mitocondria conserva un genoma independiente. Es posible que exista una relación entre regulación de la expresión génica del genoma mitocondrial y regulación del metabolismo de la célula eucariota, a través de la CTE. De hecho es sabido que en cloroplastos la transcripción génica está directamente relacionada con el estado de reducción u oxidación del orgánulo (Pfannschmidt, Nilsson et al. 1999).

5.1. Características del genoma mitocondrial humano

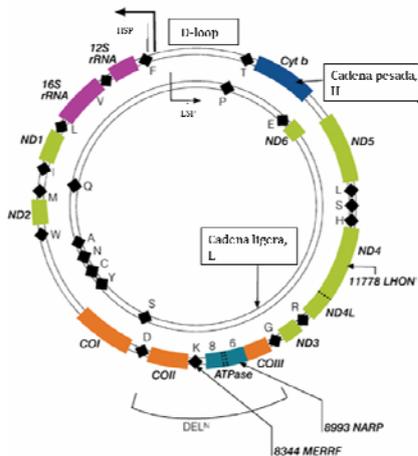
El genoma mitocondrial humano consiste en una molécula de ADN circular (mtADN) de alrededor 16,6kb, donde cada cadena codifica (Anderson, Bankier et al. 1981) (Fig_A6). Las dos cadenas de mtADN pueden ser diferenciadas en función de su diferente contenido en G/C y pueden ser separadas por densidad en gradientes desnaturizantes, dando lugar a cadena pesada o H, y ligera o L.

mtADN codifica para 37 genes: 13 para subunidades de la cadena respiratoria I, III, IV y V; 22 para tRNA mitocondrial (para 20 de los aminoácidos estándar, más un extra gen para Leucina y Serina), y 2 para rRNA. Una mitocondria puede contener entre 2 y 10 copias de este ADN. Cabe destacar que el código genético de mitocondria presenta variaciones respecto al nuclear, presentando un código de uso alternativo.

Como en procariotas, una gran porción del mtADN que codifica con ausencia de regiones repetidas e intrones. Los genes mitocondriales son transcritos de forma policistrónica, y luego son procesados (escindidos y poliadenilados) para dar resultado a mARNs maduros. La mayoría de genes que codifican proteínas son deficitarios de codón de terminación, después del último codón con sentido se encuentra una base T o un par TA, e inmediatamente después empieza otro gen. El codón de terminación se formará después de la transcripción por poliadenilación del extremo 3'.

No todas las proteínas son codificadas en este ADN, la mayoría están codificadas en genes nucleares y las correspondientes proteínas son luego importadas a mitocondria. La transcripción mitocondrial se da a partir de las secuencias promotoras para la cadena H y L (HSP, LSP), que se encuentran adyacentes pero con dirección opuesta. Además, cerca del sitio LSP encontramos la región de inicio de replicación para la cadena H (O_H), mientras que el inicio de replicación para L (O_L) se encuentra en otra posición del genoma.

Se creía que el estado normal de mtADN era una estructura superenrollada poco asociada con proteínas, en comparación con el ADN nuclear (nADN). Las primeras publicaciones sugerían la asociación del mtDNA con la membrana interna a través de la "proteína knob" (Albring, Griffith et al. 1977). Observaciones en levadura demostraron que mtADN se encuentran organizando en estructuras supramoleculares llamadas nucleoides, que contienen diversas copias de mtADN (3-4) y unos 20 polipéptidos diferentes (Miyakawa, Sando et al. 1987). Posteriormente, han sido descritas las mismas estructuras para mtADN humano, asociados a la membrana mitocondrial (Hobbs, Srinivasan et al. 2001). La presencia de estas estructuras conocidas como nucleoides, se ha propuesto como mecanismo de segregación (Kajander, Rovio et al. 2000).



Fig_A6. Representación del genoma mitocondrial, molécula circular donde las dos cadenas son codificantes (L y H). Los rombos representan tARNs, mientras que mARNs y rARNs se encuentran coloreados en forma de cajas. (Kyriakouli, D, et al. 2008)

5.2. D-loop

En biología molecular, un loop desplazado o D-loop es una estructura de ADN donde las dos cadenas de la doble hélice de ADN se encuentran separadas por la acción de una tercera cadena de ADN. La tercera cadena tiene una secuencia nucleotídica complementaria a una de las cadenas principales y se complementa con ella, provocando el desplazamiento de la otra cadena en esa región. Los D-loop se encuentran en determinadas situaciones, incluyendo mecanismos de reparación ADN por recombinación, en telómeros (conocido como T-loop), y en estructuras semi-estables del mtADN.

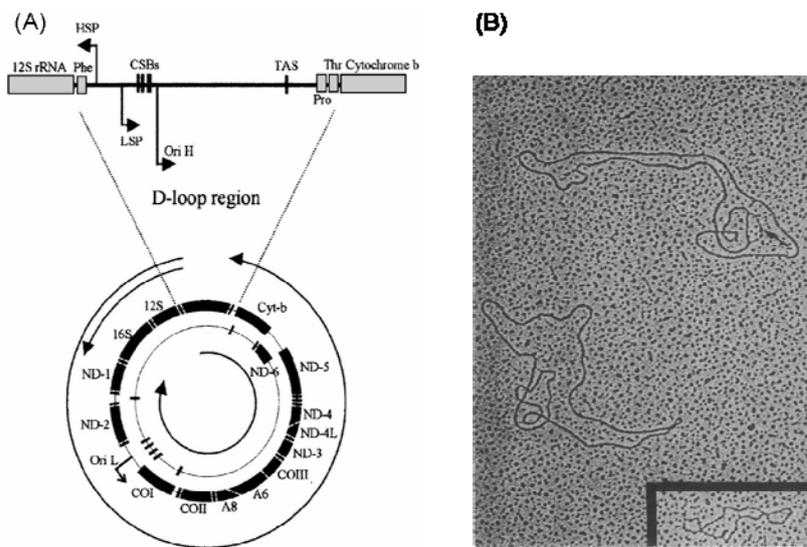
5.2.1. D-loop en mitocondria

En 1971 *Kasamatsu, et al.* descubrieron por microscopía electrónica un fragmento corto de mtADN con tres cadenas al que llamaron '*Displacement loop*' (Fig_A7). Alrededor del 95% de los eventos de replicación de la cadena H finalizan a unos 600 pb río abajo del O_H (Brown, Shine et al. 1978). Los productos nacientes arrestados de la cadena H permanecen estabilizados en la molécula circular parental, formando una estructura de triple cadena caracterizada por el desplazamiento de la cadena parental H. Esta triple estructura se conoce como *displacement-loop* (D-loop) y se encuentra involucrada en muchos casos de cáncer.

La función del D-loop no está aún clara, aunque recientemente se está sugiriendo su participación en la organización de los nucleoides mitocondriales (He, Mao et al. 2007)

La zona que incluye el D-loop (entre tRNAPhe y tRNAPro, de 1.122pb) contiene el origen de replicación de la cadena pesada y los promotores de la transcripción de las dos cadenas. Los segmentos D-loop son variables en tamaño en función de la especie. Esta región también se conoce como la zona de control y alberga regiones llamadas "hipervariables" debido a su elevada variabilidad de secuencia.

La estructura del D-loop para diferentes animales contiene también secuencias conservadas como CSBs (*conserved sequence blocks*), TASS (*Termination-associated sequences*) y estructuras de conformación similar a t-RNA (Shadel and Clayton 1997), que probablemente juegan un rol en la replicación del DNA.

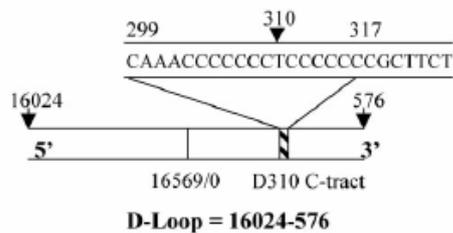


Fig_A7. (A) Esquema del mtADN, con la ampliación de la zona de control o D-loop. En ella se encuentran HSP, LSP, O_H, TAS, CSBs. (B) Imagen microscópica donde aparece el genoma mitocondrial con la presencia del D-loop, señalado con una flecha (Kasamatsu H, et.al. 1971)

5.2.1.1. D-loop: regiones ricas en citosinas y relación con cáncer

El D-loop es una región rica en citosinas. Aunque es sabido que las islas ricas en citosinas tienen una predisposición a sufrir inserciones y/o deleciones, hasta hace 10 años no se empezó a investigar la región del D-loop y la presencia de mutaciones debido a la gran cantidad de citosinas. Recientemente ha sido identificada una región rica en citosinas (*C-tract*), llamada D-310, como *Hot-spot* (punto caliente) para mutaciones puntuales presentes en varios cánceres (Fig_A8) (Mambo, Gao et al. 2003).

Esta región consiste en un conjunto de 12-18 citosinas que son interrumpidas por una base de timina en la posición 310. Un 37% de las lesiones pre-malignas albergan alteraciones en el *C-tract*, inserciones o deleciones de uno o dos nucleótidos. La prevalencia incrementa de hiperplasia benigna a displasia y carcinoma *in situ*, implicando entonces que las mutaciones en *C-tract* pueden ser útiles como marcadores en la progresión maligna de determinados tipos de cáncer, aunque son necesarios estudios clínicos que demuestren esta correlación (Ha, Tong et al. 2002).



Fig_A8. Esquema del D-loop mostrando la región conocida como D-310, rica en citosinas, donde el residuo 310 es una timina (Mambo, *et al.* 2003)

Mambo, *et al.* sugirieron que el D-loop, y en particular la región D310, es extremadamente susceptible a mutaciones posiblemente debido a la combinación de vulnerabilidad al daño por agentes y la ineficiente reparación. Debido a que el área D310 se encuentra en la región CSBII, que presumiblemente controla la replicación y transcripción del mtADN, es muy probable entonces que variaciones en la longitud del *C-tract* de D310 tenga un importante papel en la regulación del metabolismo del mtADN.

5.2.1.2. D-loop: CSBII y formación de *quadruplex*

La zona del D-loop es una de las más divergentes entre los diferentes genomas mitocondriales en vertebrados, pero existen tres pequeños bloques de secuencia altamente conservadas (CSB: CSBI, CSBII y CSBIII) situados entre LSP y O_H, y con la misma extensión. Estas regiones parecen ser importantes para la estabilización de los encebadores ARN en la cadena molde del mtADN (Shadel and Clayton 1997).

La secuencia conservada humana CSBII tiene una longitud de 16pb, donde 15 son GC con todas las guaninas situadas en una cadena H. Se ha demostrado que la terminación específica de secuencia puede ser debida a la formación de estructuras de G-quadruplex que se forman en el ARN rico en guaninas después de la transcripción de CSBII (Wanrooij, Uhler *et al.*).

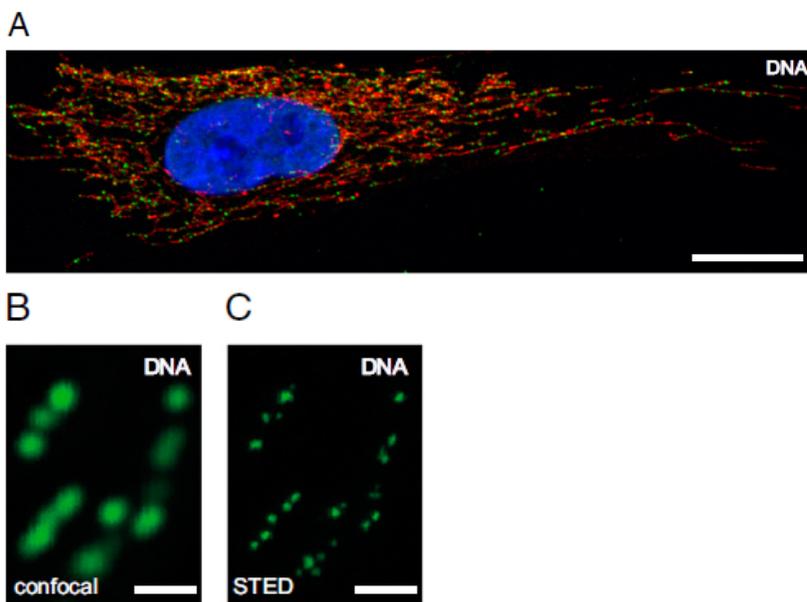
Las estructuras G-quadruplex se encuentran en secuencias ricas en guaninas debido a al *stacking* planar de cuartetos o tétradas de G. Estas estructuras son muy estables debido a los enlaces de hidrógeno entre guaninas entre el cuarteto y además, están estabilizados por iones monovalentes que se mantiene en la cavidad central. Las secuencias ricas en G se encuentran a lo largo del genoma humano en telómeros, ADN ribosomal, regiones de cambio de inmunoglobulinas, minisatélites y en regiones promotoras. Se ha sugerido que más de un 40% de promotores contienen secuencias motivo con predisponibilidad a formar Gquadruplex (Huppert and Balasubramanian 2007). La presencia de estos Gquadruplex en secuencias promotoras implica un mecanismo de control o regulación génica.

5.3. Organización del genoma mitocondrial: Nucleoides

Los nucleoides presentan una dinámica estable, facilitando los procesos de heradibilidad y segregación de estos ensamblados (Kajander, Rovio *et al.* 2000). Se cree que la organización

en nucleoides contribuye a la rápida segregación de mtADN durante la maduración del oócito, proceso conocido como *bottleneck*. Un nivel elevado de fusión mitocondrial parece ser esencial para el mantenimiento del mtADN en nucleoides. Debido a los continuos procesos de fusión y fisión mitocondrial, es necesaria una estructura como los nucleoides que permita el acceso de mtADN a un amplio *pool* de proteínas para que difundan por el ADN y de esa forma permitir los procesos de replicación y transcripción necesarios para la supervivencia de la mitocondria y célula (Bogehagen, Rousseau et al. 2008).

Análisis cuantitativos del tamaño de los nucleoides, así como de la cantidad de mtADN presente en cultivos de células de mamífero sugieren una media de 5-7 mtADN empaquetados en una esfera de diámetro 70 nm (Iborra, Kimura et al. 2004). El empaquetamiento en nucleoides y su densidad sería similar al caso de los nucleoides bacterianos. Además de esta similitud también en ambos casos se encuentran asociados a componentes de membrana, aunque se desconoce exactamente el mecanismo de anclaje. Recientemente, se ha descrito mediante el uso del microscopio STED (*Stimulated Emission Depletion Microscopy*) combinado con biología molecular, nuevas dimensiones para los nucleoides (Fig_A9). STED reveló un tamaño de nucleoide muy uniforme en mamíferos entorno a 85-111 nm, inferior al visto mediante microscopia confocal. Esta uniformidad en el tamaño del nucleoide sugiere la fuerte conservación evolutiva para el mantenimiento del mtADN. Además, a diferencia de otras publicaciones donde se adjudicaba 2-8 copias de mtADN por nucleoide, el uso de STED permite observar una media de 1.4 moléculas de mtADN por nucleoide, lo cual indica que muchos nucleoides sólo contienen una copia de mtADN (Kukat, Wurm et al.).

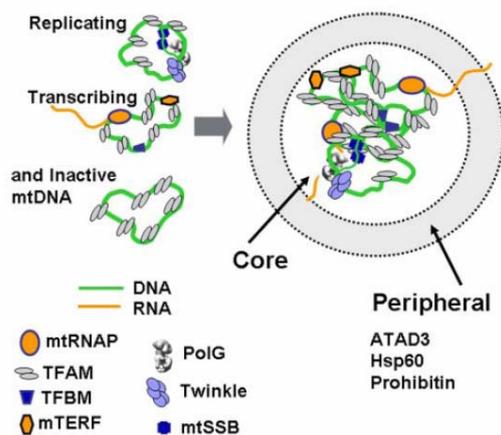


Fig_A9: Los nucleoides presentan un tamaño uniforme en las células. (A) mtADN, en verde, se encuentra compactado en nucleoides en células de fibroblasto humano. En azul vemos la tinción del núcleo y en rojo, la tinción de las redes tubulares de mitocondria. Escala de la barra es 20 μm (B) Imágenes de microscopio confocal,

donde el diámetro de los nucleoides parece ser de unos 300 nm. (C) Imagen de STED donde el diámetro de los nucleoides es de unos 99 nm. Esta técnica permite una mayor resolución y, por lo tanto, un mejor cálculo del tamaño de los nucleoides. Escala de barra en (B,C) de 0.5 μ m. (Kukat, *et al.*)

El empaquetamiento de mtADN en nucleoides es posible gracias a la ayuda de proteínas de unión al ADN, y de forma mayoritaria por el factor de transcripción mitocondrial A, TFAM, y la proteína de unión a cadena sencilla, mtSSB (Kaufman, Durisic *et al.* 2007).

En el trabajo del grupo de Bogenhagen D.F., *et al.* (Bogenhagen, Rousseau *et al.* 2008) e Iborra F.J., *et al.* (Iborra, Kimura *et al.* 2004). se propone un modelo de estructura del nucleoide (Fig_A10), con una zona *core* del nucleoide y una zona periférica. La zona *core* sería claramente el centro de biosíntesis del mtADN, con procesos como la replicación y transcripción. Este *core* se encontraría rodeado por una zona periférica que contiene todas aquellas proteínas que no interaccionan con mtADN eficientemente; pero sí se encuentran en la extracción de nucleoides de forma nativa, como sería el caso de ATAD3 que no contacta directamente con mtADN. Por otra parte, existe una estrecha relación entre nucleoide y ribosomas mitocondriales, reflejándose así una coordinación entre la síntesis de proteínas codificadas en mtADN y un ensamblaje de complejos.



Fig_A10. Estructura del nucleoide y la distribución de proteínas durante la biogénesis del mtADN, como se ha detallado claramente se puede distinguir una zona core y una periférica. (Iborra, Kimura *et al.* 2004).

5.4. Heridabilidad del mtADN

La mitocondria se divide por fisión binaria de forma similar a la división celular bacteriana, pese a que a diferencia de la bacteria, la mitocondria es capaz de fusionarse con otras mitocondrias. La regulación de esta división difiere entre eucariotas. En la mayoría de células eucariotas unicelulares, el crecimiento y la división está ligada al ciclo celular. Por ejemplo, una sola mitocondria puede dividirse de forma sincronizada con el núcleo. El proceso de división y segregación debe ser bien regulado de forma que cada célula hija recibe al menos una mitocondria. En otros eucariotas, por ejemplo en mamíferos, la mitocondria puede replicar su ADN y dividirse en repuesta a necesidades energéticas de la célula, más que estar

sincronizado con el ciclo celular. Cuando las necesidades energéticas de la célula son elevadas, la mitocondria crece y se divide. Por el contrario, si las necesidades son menores, las mitocondrias son destruidas o se convierten inactivas. La comprensión de la dinámica mitocondrial, descrita como el balance entre el proceso de fusión y fisión, ha sido relacionado con alteraciones morfológicas mitocondriales que juegan un papel importante en diversas patologías.

La mitocondria se hereda por vía materna (Giles, Blanc et al. 1980). Es el oócito, con unas 100.000 moléculas de mtADN, el que aporta la inmensa mayoría de mitocondrias en el momento de la fecundación. El mtADN paterno se elimina en la fase temprana de la embriogénesis. Parece que las escasas moléculas de mtDNA aportadas por el espermatozoide son eliminadas al ser reconocidas por señales de ubiquitinación (Sutovsky, Moreno et al. 2000), y en ningún caso parece que participen en el genotipo de la descendencia. La heredabilidad uniparental deja poca oportunidad a la recombinación genética mediante diferentes linajes de la mitocondria. Por esa razón se cree que el mtADN se reproduce mediante fisión binaria.

La práctica ausencia de recombinación de genética en el mtADN hace que sea una fuente de información para los estudios de genética de poblaciones y biología evolutiva, debido a que el mtADN es heredado como una única unidad. De hecho la relación entre mtADN de diferentes individuos puede ser representado con un árbol genealógico.

5.5. Elevada tasa de mutación en mtADN

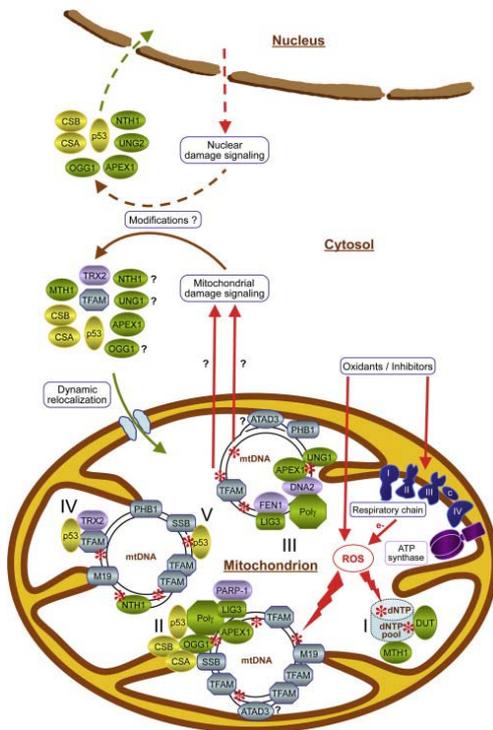
La elevada tasa de mutación del mtADN en comparación con el ADN nuclear (nADN) (Singh, Neckelmann et al. 1987) es probablemente consecuencia del déficit de histonas que protege la doble hélice de ADN. Por otra parte, el mtADN también puede ser dañado por exceso de moléculas ROS subproductos de la fosforilación oxidativa. Este efecto dañino está favorecido por el hecho que el mtADN está en contacto con la membrana mitocondrial interna y, por lo tanto, muy cerca de CTE que genera ROS.

Pese a que se han descrito algunos mecanismos de reparación del mtDNA en respuesta a daños específicos (Croteau, Stierum et al. 1999), hasta este momento no parece existir dentro de la mitocondria un sistema general reparador de mutaciones, lo que puede provocar que ciertas mutaciones se vayan acumulando.

5.6. Mecanismo de reparación de mtADN

Los mecanismos de reparación para ADN fueron primero descritos para nADN, aunque hoy en día existen evidencias que muestran su presencia en orgánulos como mitocondria y cloroplastos. El mecanismo de reparación por escisión de base (BER) parece ser el más bien caracterizado, mientras que mecanismo de reparación por escisión de nucleótido (NER) aún no ha podido ser caracterizado en mitocondria de mamíferos.

Los mecanismos de reparación se encuentran coordinados con otras vías metabólicas del ADN, asociadas a la membrana interna mitocondrial (Fig_A11). Los mecanismos de reparación en mitocondria se encuentran regulados o pendientes de respuesta al estrés. El genoma mitocondrial no codifica para las enzimas que participan en reparación, así que existe un translocación de proteínas codificadas en núcleo, que son transportadas de citosol a mitocondria. Finalmente, cabe destacar que cambios en la fiabilidad y eficiencia de los mecanismos de reparación son la causa de la acumulación de daños en el mtADN, enfermedad y envejecimiento (Boesch, Weber-Lotfi et al.)



Fig_A11. Organización de los procesos de reparación por estrés oxidativo que suceden dentro de la mitocondria. Las principales enzimas de la reparación se encuentran en color verde, mientras que factores que participan en reparación se encuentran en amarillo y lila, en gris se encuentran las principales proteínas de interacción con ADN. (Boesch, et al.)

6. Mitocondria y apoptosis celular

El estudio de la muerte celular programada o apoptosis ha sido foco de interés científico desde hace años. Se trata de un proceso crítico para la existencia de muchos organismos, necesaria para modular el sistema nervioso durante el desarrollo y para mantener un funcionamiento normal del sistema inmunológico. La alteración de los mecanismos normales de apoptosis pueden provocar consecuencias graves: cáncer y enfermedades autoinmunitarias si se carece, e infartos o degeneración nerviosa típica de la enfermedad de Alzheimer si se produce en exceso.

La apoptosis puede ser iniciada por una variedad de agentes que dañan el ADN, por ejemplo radiaciones UV y X, y determinadas drogas terapéuticas. Las proteínas que detectan el daño en el ADN e inician la apoptosis también pueden afectar el ciclo celular: división para que el daño pueda ser reparado; o bien muerte celular, si las alteraciones son irreversibles.

Se conocen como mínimo tres procesos o mecanismos generales que reaccionan la mitocondria con apoptosis (Fig_A12):

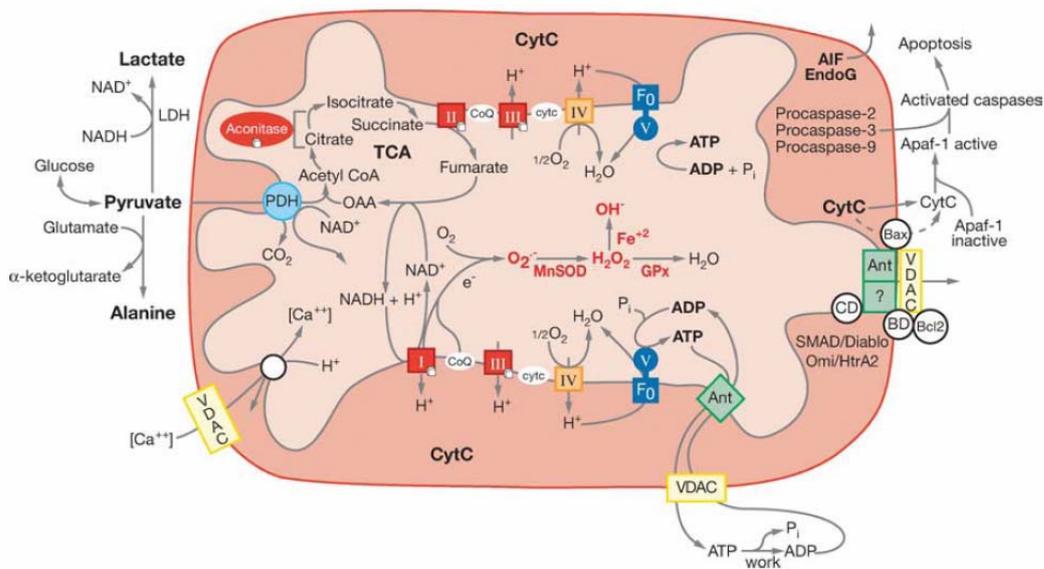
- Interrupción de transporte de electrones
- Alteración del potencial redox celular
- Liberación de factores que promueven la activación de caspasas (por ejemplo, la proteína citocromo C)

Durante la apoptosis, la mitocondria libera citocromo C al citosol (por mecanismos aún no conocidos), es en el citosol donde el citocromo C, juntamente con la molécula Apaf-1 (*Apoptotic Protease Activating Factor*) y la procaspasa-9, forman el "apoptosoma". El apoptosoma activará la caspasa 9, que iniciará la cascada proteolítica que deriva en apoptosis. La salida del citocromo C de la mitocondria parece encaminar la célula hacia la muerte, bien sea por un mecanismo apoptótico mediante la activación de las caspasas por Apaf-1, o por un proceso necrótico más lento causada por el déficit de transporte de electrones, que puede provocar la generación de radicales de oxígeno libre y la disminución de la producción de ATP. El bloqueo de la liberación de citocromo C por parte de Bcl-2 inhibe la apoptosis.

Además del citocromo C, la mitocondria libera otras moléculas, como el factor inductor de apoptosis (AIF) y la procaspasa-3, ambos pro-apoptóticos.

El poro de transición mitocondrial (MTP), un canal no selectivo de estructura parcialmente definida que tiene componentes en las dos membranas mitocondriales, ha estado implicado en la pérdida de calcio y también en la muerte celular. En algunos sistemas experimentales se ha visto que la fase inicial del proceso apoptótico (es decir, el estado que precede a la desintegración nuclear) está precedido por la pérdida del potencial de membrana interna como consecuencia de la apertura de MTP, regulada por Bcl-2 (Marchetti, Hirsch et al. 1996). Esta apertura podría comportar la salida de sustratos metabólicos y sustancias antioxidantes esenciales (ATP, NADH, glutatión reducido o GSH) de la mitocondria. La disminución de GSH puede afectar enormemente la viabilidad mitocondrial ya que es necesario para la reducción del peróxido de hidrógeno y otras especies reactivas de oxígeno (ROS).

Por otra parte, la apertura del MTP en la membrana interna podría provocar que se equilibrasen las concentraciones de protones de la matriz y del espacio intermembrana, con lo que se desacoplaría la CTE. En este caso el volumen mitocondrial aumenta por la entrada de moléculas de agua. Esto puede comportar la rotura de la membrana externa, con la liberación desde el espacio intermembrana de proteínas activadoras de las caspasas (Wallace 2005).



Fig_A12. Representación esquemática de la mitocondria, Donde se pueden diferenciar los mecanismo que conducen a la generación de ROS (acoplados a la CTE) y la vía mitocondrial de activación apoptótica. (Wallace, D. 2005).

7. Mitocondria y enfermedades

El concepto de enfermedad mitocondrial se aplica sólo a aquellas enfermedades que tienen alterado el funcionamiento de la CTE, aunque existen enzimas mitocondriales también implicadas en enfermedades del metabolismo intermediario, como por ejemplo en alteraciones de la β -oxidación o el ciclo de Krebs. Como se ha visto, en la CTE participan polipéptidos codificados por la mitocondria y sintetizados en la matriz mitocondrial, y también muchas subunidades codificadas por el nADN, sintetizadas en el citosol y guiadas hasta la membrana interna de la mitocondria mediante péptidos-señal amino-terminales. Es por ello que las enfermedades mitocondriales pueden tener el origen no únicamente en defectos del genoma mitocondrial, sino también en defectos del genoma nuclear. Por ejemplo, todos los factores para la replicación del mtADN son codificados por el nADN, por lo que no resulta extraño que las mutaciones en nADN afecten de una forma importante el funcionamiento de la mitocondria.

7.1. Clasificación de las enfermedades mitocondriales

El hecho de que las mitocondrias estén ampliamente distribuidas entre los diferentes tejidos y órganos hace que las típicas clasificaciones anatómicas no sean oportunas para el conjunto de las enfermedades mitocondriales. La primera característica que define a las enfermedades mitocondriales es la gran heterogeneidad de las manifestaciones clínicas. Los fenotipos pueden ser diversos y coincidentes: una mutación concreta en el mtADN puede producir fenotipos muy variables, y mutaciones diferentes pueden provocar síntomas similares.

Las enfermedades mitocondriales muestran unos patrones de herencia variables: materna, mendeliana, y una combinación de las dos. Las alteraciones primarias en el mtADN se traducen en una disminución de la producción de energía, sea cuál sea el gen mutado. Los tejidos con mayores necesidades energéticas (cerebro, corazón, músculo...) son los más afectados. Las enfermedades resultantes se transmiten por herencia materna. Otro grupo de enfermedades se encuentra asociado a deleciones múltiples del mtDNA pero se transmiten de forma autosómica, por este motivo se cree que son consecuencia de una comunicación defectuosa entre el núcleo y la mitocondria.

Los desordenes mitocondriales se presentan frecuentemente como patologías neurológicas, pero también se pueden mostrar como miopatías, diabetes, endocrinopatías o variedad de otras manifestaciones sistemáticas. En la mayoría de los casos, las enfermedades son transmitidas de las madres a sus hijos.

En otras enfermedades, debidas a defectos en genes nucleares provocan disfunción de proteínas mitocondriales, éste es el caso de la ataxia de Friederich's y enfermedad de Wilson. Así bien, el ambiente (fenotipo) también puede interactuar con las predisposiciones hereditarias y causar enfermedades mitocondriales. Por otra parte, también existe el efecto del estrés oxidativo y su relación con la disfunción mitocondrial y determinadas enfermedades.

7.1.1. Enfermedades debidas a mutaciones en mtDNA

Las enfermedades debidas a mutaciones puntuales en mtDNA pueden ser divididas en 3 categorías:

- Mutaciones con cambio de pauta de lectura, debido a inserciones o deleciones en el mtDNA, que pueden ser heredadas o espontáneas
- Mutaciones sin efecto en la codificación
- Mutaciones en gener de síntesis de proteínas (rARN y tARN). Posiblemente, unas de las mutaciones en la síntesis de proteínas más comunes es a nivel del nucleótido 3243 (A>G) en el gentRNA^{Leu}(^{UUR}). Esta mutación es interesante debido a la variabilidad que presenta en sus manifestaciones clínicas. Cuando se encuentra en niveles bajos (10-30%) en sangre, el paciente puede manifestar sólo diabetes tipo II con presencia o ausencia de sordera. Entre un 0.5-1% de los casos de diabetes tipo II son debidos a esta mutación en mtADN. En contraste, cuando está presente en más de un 70% del mtADNs, no hay presencia de diabetes tipo II, pero los efectos son mucho más severos incluyendo cardiomiopatías, encefalomiopatías, síndrome de MELAS.

Las manifestaciones clínicas relacionadas con mutaciones en mtDNA afectan a cerebro, corazón, músculo esquelético, hígado, y sistema endocrino, que son los mismos tejidos afectados en el envejecimiento. Síntomas específicos incluirían ceguera, sordera, desordenes

de movimiento, demencias, enfermedades cardiovasculares, debilidad muscular, disfunción renal, y problemas endocrinos como la diabetes.

7.1.2. Mitocondria y diabetes

Estudios en pacientes con diabetes tipo II han revelado problemas en las funciones y biogénesis de mitocondria, generalmente una regulación a la baja. Los pacientes resistentes a insulina, con diabetes tipo II, presentan un déficit o fallo energético en mitocondria. Además, una disminución en los niveles de en el coactivador 1 de *peroxisome-proliferation-activated receptor gamma (PPAR γ)*, (PGC-1), el principal regulador de la biogénesis en mitocondria y la oxidación de ácidos grasos.

Por otro lado, existe una relación causa/efecto muy estrecha entre defectos en mitocondria y problemas con la secreción de insulina por células β en dos modelos de ratón. En el primer modelo, el factor de transcripción mitocondrial A (TFAM) es inactivado en células β -pancreáticas. Esto conduce a un incremento en la concentración de glucosa en sangre y, un progresivo declive en masa de células β por apoptosis (Silva, Kohler et al. 2000). En el segundo modelo, se redujo la afinidad por ATP del canal ATP-dependiente de K, resultando en una severa reducción en sérum de insulina y hiperglicemia (Koster, Marshall et al. 2000). Estos modelos demuestran que la producción de ATP mitocondrial es crítico en el sistema de señalización de las células β para la liberación/secreción de insulina

7.1.3. Enfermedades neurodegenerativas

Se ha especulado que las neuronas pueden ser especialmente vulnerables a la disminución del metabolismo mitocondrial relacionado con el proceso de envejecimiento. El tejido neuronal es postmitótico y, en consecuencia, se pueden acumular mutaciones espontáneas en el mtADN que no serán eliminadas, ya que no existe selección clonal por división celular. La disfunción mitocondrial puede generar radicales libres y daño oxidativo, factores implicados en la patogénesis de las enfermedades neurodegenerativas.

En numerosas enfermedades neurodegenerativas se han observado defectos en la fosforilación oxidativa y una producción incrementada de radicales libres, que no están causados por defectos en mitocondria. En estos casos se cree que la disfunción mitocondrial es probablemente un epifenómeno que en todo caso contribuye a precipitar la muerte celular.

A continuación se presenta un resumen breve de las principales enfermedades neurodegenerativas y del papel que tiene la mitocondria en ellas.

7.1.3.1. Esclerosis lateral amiotrófica (ELA)

La esclerosis lateral amiotrófica (ELA) es una de las enfermedades degenerativas que aparece hacia los 40-50 años, y en un tiempo medio de tres años comporta la parálisis y la muerte del individuo. La mayoría de los casos son esporádicos y de causas desconocidas. Entre un 10-20% de los pacientes son casos familiares y entre estos, en un 10% se han encontrado

mutaciones en el gen de la superóxido dismutasa, gen esencial para la eliminación de radicales O_2^- , los cuales no son especialmente tóxicos pero pueden transformarse en radicales hidroxil (OH^\cdot), mucho más dañinos. De esta manera, se ha sugerido que ELA se produciría en casos donde el equilibrio de especímenes ROS está alterado y ello conduciría al daño oxidativo.

7.1.3.2. Enfermedad de Alzheimer (AD)

Es la demencia más frecuente de la vejez. Un 5% de los casos se transmiten de forma autosómica dominante y son causados por mutaciones en el gen de la proteína amiloide o presenilina. La gran mayoría de pacientes con AD representan casos esporádicos sin que se conozca el defecto genético que los origina. Neuropatológicamente se caracterizan por la presencia de agregados neurofibrilares y de placas amiloides, y por pérdida neuronal. En cerebros de enfermos se ha visto que la actividad COX está reducida. Diversas mutaciones en el mtADN se han asociado con AD, pero no se ha demostrado un papel de éstas en la patogénesis de la enfermedad.

7.1.4. Mitocondria y cáncer

Defectos genéticos en determinados genes nucleares que codifican para proteína mitocondriales han sido directamente relacionados con algunos cáncers hereditarios. Un ejemplo sería el caso de la fumarato hidratasa, que ha sido relacionada con carcinomas renales.

Además también cabe destacar el papel de ROS en cáncer. En ratones *knock-out* para MnSOD (superóxido dismutasa) presentan un incremento del 100% en la incidencia de cáncer y es posible revertir este fenotipo mediante la incorporación de cADN para MnSOD (Van Remmen, Ikeno et al. 2003). Por lo tanto, mutaciones que afecten la fosforilación oxidativa o la CTE conllevaran a un incremento de ROS y, por tanto, estos agentes ROS pueden contribuir al incremento de incidencia o desarrollo de cáncer.

7.1.5. Posible relación con el envejecimiento

Durante CTE existen alguna pérdidas de electrones que dan lugar a la formación de radicales libres altamente reactivos y dañinos (ROS). ROS pueden resultar en un elevado estrés oxidativo para la mitocondria y, en consecuencia, a un elevado porcentaje de mutaciones en el mtADN, dando lugar a anormalidades enzimáticas y desembocando en la alteración de la mitocondria. Como ya se ha explicado anteriormente, la tasa de mutación del mtDNA es 10 veces superior a la del nADN. Mientras que mutaciones en células de la línea germinal provocan enfermedades de herencia materna, la acumulación de mutaciones somáticas en tejidos postmitóticos puede tener como consecuencia el declive progresivo del metabolismo mitocondrial observado en el desarrollo de enfermedades heredadas y también en el envejecimiento. El inicio tardío y el curso progresivo de muchas enfermedades mitocondriales sugiere que el funcionamiento de las mitocondrias empeora con el tiempo.

Diversos trabajos han descrito incrementos estadísticos significativos en la acumulación de mutaciones somáticas en el mtADN con la edad, tanto deleciones como mutaciones puntuales (Goodell and Cortopassi 1998). Lo que no está claro es que el envejecimiento implique la disminución de la actividad del CTE. Existen estudios que detectan reducciones en la actividad de algunos complejos, mientras que otros no encuentran correlación entre edad y capacidad energética del músculo (Barrientos, Volpini et al. 1996). Es posible que estos resultados contrapuestos sean consecuencia de diferencias en cuanto a factores como el grado de ejercicio o el consumo de tabaco u otras sustancias tóxicas en los grupos estudiados por diversos laboratorios.

7.1.5.1. Teoría mitocondrial del envejecimiento

La teoría mitocondrial del envejecimiento propone que si las especies ROS se producen en exceso y no son neutralizadas por mecanismos antioxidantes de los que dispone la célula, pueden reaccionar con el mtADN y dañarlo. Esto afectará el funcionamiento de la propia CTE, y entonces se producirán más radicales libres. Este círculo vicioso puede comportar una disminución en la producción de ATP que alteraría el potencial transmembrana e induciría la vía de la apoptosis celular. El exceso de oxidantes también puede afectar a lípidos de membrana, proteínas y otras macromoléculas.

Las mutaciones somáticas son probablemente consecuencia del daño oxidativo, que se incrementa con la edad tanto en ratas como en humanos. Diversos trabajos sugieren que las mutaciones somáticas se acumulan con la edad en tejidos postmitóticos a causa del daño ocasionado por ROS. La disminución de la fosforilación oxidativa resultante alteraría la capacidad energética del tejido hasta sobrepasar un determinado límite a partir del cuál aparecería el envejecimiento y síntomas o enfermedades degenerativas que normalmente lo acompañan (problemas cardiovasculares, disfunción cerebral, cataratas, cáncer, afectación del sistema inmunológico, etc.).

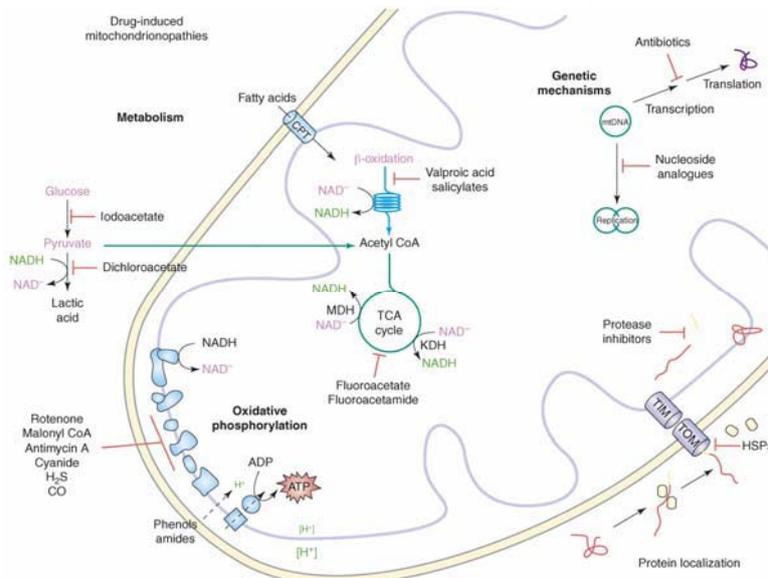
8. Mitocondria y efectos secundarios en terapias farmacéuticas

Recientemente, se ha descubierto que la mitocondria puede resultar en diana involuntaria para diversos tratamientos con fármacos y, posiblemente, es la responsable de algunos de los eventos limitantes de dosis asociados a una amplia variedad de fármacos (Wallace 2008).

Como se ha comentado anteriormente la mitocondria está involucrada en una gran variedad de funciones, cadena de transporte electrónico, ciclo ácidos grasos, oxidación de aminoácidos, ciclo de Krebs, etc. Por lo que la afectación de estas funciones por vía directa o indirecta conlleva a la desregulación de la mitocondria y afecta su biogénesis.

Cada año un porcentaje pequeño de fármacos son retirados del mercado debido a efectos secundarios inesperados durante los test pre-clínicos y los siguientes ensayos clínicos. Recientemente se ha descubierto que una fracción de estos fármacos presenta la mitocondria como diana involuntaria, como se muestra en la Fig_A13. Los mecanismos que generan toxicidad en mitocondria se pueden clasificar en:

- Inhibidores directos de la fosforilación oxidativa
- Sustratos Inhibidores de la oxidación en mitocondria
- Inhibidores de la biogénesis en mitocondria



Fig_A13. Esquema de los mecanismos que generan toxicidad en mitocondria debido al uso de fármacos que interactúan con procesos clave para la viabilidad de mitocondria. (Wallace, K. 2008)

Inhibidores directos de la fosforilación oxidativa

Son compuestos químicos que interfieren directamente en la fosforilación oxidativa incluyendo inhibidores de componentes individuales de CTE y desacopladores de la fosforilación oxidativa. Ejemplos clásicos de inhibidores de CTE son *rotenone*, antimicina A, *cyanide*.

Sustratos Inhibidores de la oxidación en mitocondria

Los sustratos inhibidores de la oxidación en mitocondria son, la mayoría de ellos, miméticos estructurales de los sustratos naturales. Se trata de inhibidores por competición de las respectivas reacciones. Algunos ejemplos incluyen la inhibición de la oxidación de ácidos grasos por *valproic*.

Inhibidores de la biogénesis en mitocondria

La biogénesis de mitocondria es un proceso complejo que involucra un gran número de pasos, entre ellos la replicación, transcripción, transducción....

Un ejemplo claro de estos efectos secundarios sería el efecto que provocan en la RNA polimerasa mitocondrial (mtRNAP) algunos antimicrobianos y el uso de antivirales análogos de ribonucleosidos (AVRNs).

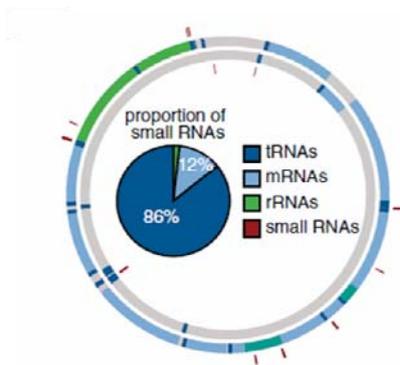
- En la replicación: Un ejemplo sería el de los antivirales análogos de nucleosidos diseñados para interferir con la transcripción inversa del RNA invasor viral. La ADN polimerasa de mitocondria puede adquirir estos análogos en la síntesis de ADN durante la replicación. Esto provoca la truncación y un mtDNA inestable, provocando la delección final del mtADN. La disminución del número de copias de mtADN provocaría de forma directa una reducción de las proteínas sintetizadas en mitocondria que participan en CTE y, por lo tanto, un problema energético.
- En la traducción: La síntesis de proteínas también se encuentra afectada. Algunas terapias antibacteriales implican el uso de oxazolidinonas y cloramfenicol que pueden unirse a las proteínas mitocondriales e inhibirlas.
- En la transcripción: Los ribonucleósidos antivirales (AVRNs) después de entrar en la célula son metabolizados a la forma activa trifosforilada. AVRNs puede actuar como sustrato para la mtRNAP y, en consecuencia, inhibir la transcripción mitocondrial. El uso de AVRNs por mtRNAP y la consiguiente inhibición de la transcripción mitocondrial, ha sido sugerida como causante de la disrupción funcional de la mitocondria y, en consecuencia, de la toxicidad en pacientes que reciben AVRNs durante los ensayos clínicos para tratar infecciones de virus RNA.

B. REGULACIÓN Y TRANSCRIPCIÓN MITOCONDRIAL

1. Transcripción mitocondrial

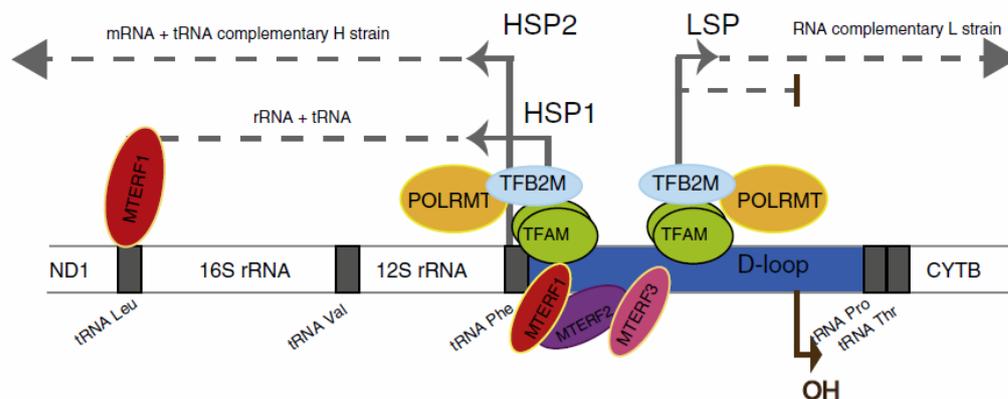
1.1. Unidades de la transcripción mitocondrial

Tal como se ha comentado anteriormente, el mtADN en mamíferos consiste en un ADN circular de doble cadena de 16.5 Kb que codifica solo para 37 genes (Fig_B1): 13 subunidades de la fosforilación oxidativa, 2rRNAs (12S rRNA y 16S rRNA), y 22tRNAs. El genoma mitocondrial codifica proteínas esenciales para los complejos I, III, IV y V de la CTE, pero curiosamente las subunidades del complejo II se encuentran todas codificadas por el genoma nuclear. Los genes del mtADN se encuentran agrupados de forma muy asimétrica entre la cadena H (*heavy*, pesada) y la cadena L (*light*, ligera), llamadas así por su diferencia de densidad en un gradiente de cloruro de cesio. La cadena H presenta la mayoría de los genes, 12 de los 13mRNAs, 2rRNAs y 22tRNAs, mientras que la cadena L codifica sólo para la subunidad ND6 del complejo I y 8 tRNAs (Attardi and Schatz 1988).



Fig_B1. Proporción entre tRNAs, mRNAs y rRNAs en el genoma mitocondrial humano, donde un 86% consiste en tRNAs (Mercer, et al. 20011)

Existen dos regiones no codificantes (NCR) en el mtADN que regulan la transcripción y replicación. La región no codificante mayor es de unos 1200 pb y comúnmente se conoce como el D-loop (Fig_B2), contiene las regiones promotoras de transcripción para las dos cadenas y el origen de replicación para la cadena H (O_H). Además, diversos estudios han constatado la asociación de múltiples proteínas al D-loop, indicando que el D-loop es el principal sitio de regulación de la transcripción. La otra región no codificante, se encuentra localizada entre tRNA Cys y tRNA Asn, y es relativamente pequeña (30pb), contiene el origen de replicación para la cadena L (O_L).



Fig_B2. Esquema del D-loop con los diferentes promotores (LSP, HSP1, HSP2) y sitios de unión a proteínas que se encuentran en mtADN de vertebrados. Las flechas indican la dirección de la transcripción para las cadenas del mtADN. Las formas ovaladas indican las proteínas involucradas en la transcripción, tanto iniciación como terminación de ésta (Peralta, et al. 2011).

1.2. Promotores para la cadena H: HSP1 y HSP2

Para la cadena H se han identificado dos regiones promotoras, HSP1 (*Heavy Strand Promotor 1*) y HSP2 (*Heavy Strand Promotor 2*) (Fig_B2). Existe mucha controversia entorno a la existencia de HSP2, ya que no es fácilmente observable en ensayos *in vitro*. Sin embargo, las evidencias *in vivo* sugieren fuertemente la existencia de un segundo sitio de iniciación (Montoya, Christianson et al. 1982). Los promotores HSP1 y HSP2, definidos mediante los ensayos de protección a *S1 nucleasa*, se encuentran localizados muy cerca, espaciados por 100 pb en la región D-loop y son transcritos en la misma dirección (Martin, Cho et al. 2005). La transcripción iniciada desde el promotor HSP2 da lugar a un transcrito completo de la cadena H que contiene la información para 2 rRNAs, 12 mRNAs y 13 tRNAs, con una frecuencia 20 veces menor al transcrito a partir de HSP1. Mientras que la transcripción desde HSP1 termina de forma prematura río abajo de 16S rRNA, transcribiendo sólo para tRNA Val, tRNA Phe y 2 rRNAs. La transcripción iniciada desde HSP1 parece que finaliza como resultado de la unión específica del factor de terminación de la transcripción MTERF1 (Kruse, Narasimhan et al. 1989). La terminación de los transcritos desde HSP2 ha sido menos estudiada, pero existe una región rica en A/T, llamada H2 cercana a tRNA Phe, que se ha propuesto como región de terminación (Camasamudram, Fang et al. 2003). Recientemente se han identificado varias proteínas que se unen a la región de terminación H2 de ratón, entre ellas ATAD3.

1.3. Promotor para la cadena L: LSP

Los transcritos originados desde el LSP codifican para una proteína (ND6) y una colección de tRNAs. Además, se han detectado eventos de poliadenilación en las posiciones 160-185 nt, río debajo de la región promotora LSP dentro de la zona del D-loop, es decir, existen productos de una transcripción truncada (Bonawitz, Clayton et al. 2006). Estudios de mapeo demuestran

que los transcritos son truncados alrededor de CSBII y éstos sirven como encebador para la iniciación de la replicación de la cadena H a partir de O_H situado cercano a CSBII. Además, esta región presenta una gran protección frente a DNasaI, sugiriendo que la unión de proteínas a esta región puede inducir la terminación de la transcripción (Mercer, Neph et al.). Esta terminación prematura puede ser importante para la formación de encebadores para la replicación, por lo tanto, la identificación de las proteínas que se unen a D-loop puede ser un importante hallazgo.

Los eventos de terminación combinados con la iniciación en el HSP1 conllevan a un mecanismo de control de la cantidad entre rRNA y mRNA.

1.4. Sitio de inicio de transcripción LSP/HSP

Para el promotor HSP1 se definió el sitio de inicio de transcripción a partir del nucleótido 559+/-1 y a para HSP2 en el nucleótido 561. Mientras que para LSP el inicio de la transcripción es en el nt 407+/-1 (Fig_B3). Además han sido identificadas las secuencias necesarias para el inicio de la transcripción. Para el promotor HSP1 se asignó la secuencia situada entre -16 y +7 del inicio de transcripción, mientras que para LSP se definió la región de -28 a +16 (Chang and Clayton 1984).

De cada secuencia necesaria para el inicio de la transcripción es posible definir una región consenso (con la introducción de ciertos espacios) (Fig_B3): 5'-CANACC(G)CC(A)AAAGAPyA-3', donde el inicio de la transcripción ocurre de 6 a 8 nucleótidos después de 3'. Esta región de consenso de inicio de la transcripción se caracteriza por una región rica en citosinas seguida por un conjunto de adeninas.

De forma interesante, la búsqueda de otras regiones homólogas indica la presencia de una secuencia en CSBIII, muy similar a HSP1, que sólo difiere en un nt. En el caso de HSP1 presenta una guanina como inicio de transcripción, mientras que para CSBIII presenta una adenina. Lo cual sugiere que esta posición es esencial para la región de control (Chang and Clayton 1984).

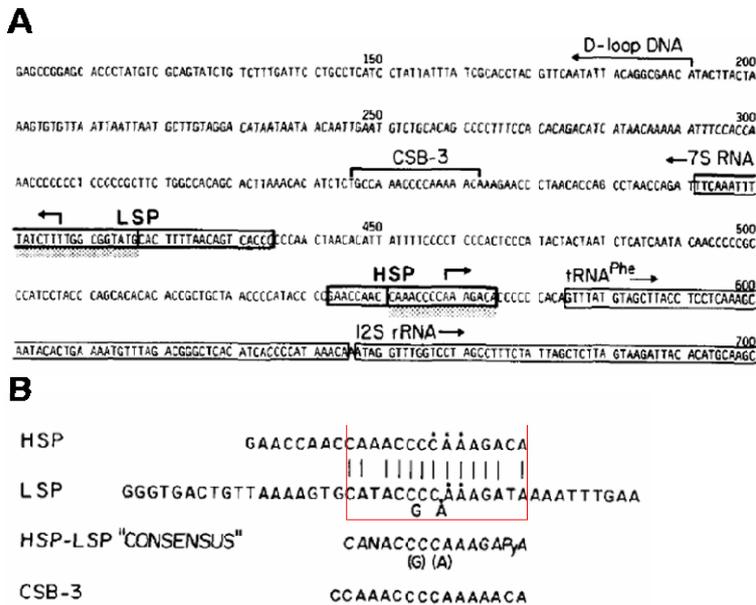


Fig.B3. (A) Secuencia de la región de control D-loop que contiene LSP y HSP, los rectángulos indican regiones imprescindibles para el inicio de la transcripción. Las flechas en LSP y HSP indican la dirección de la transcripción. (B) Secuencia homóloga entre HSP y LSP, así como la región CSB3. (Chang, *et al.* 1984). En esta figura Chang, *et al.* llama HSP al referirse a HSP1.

Un dato observado por Chang, *et al.* es que cuando se delecciona la región de -60 a -16 en HSP1 existe una menor reducción de los transcritos, que para el caso de LSP. Sorprendentemente esta región coincide con la secuencia de unión a TFAM y, ello, se podría relacionar con el hecho que HSP1 no es tan sensible a TFAM.

2. Maquinaria de la transcripción en mamífero

En la transcripción encontramos elementos que actúan de forma cis o trans, los elementos trans han sido descubiertos más lentamente. En un principio se definió que para la transcripción es necesario: RNA polimerasa mitocondrial (**mtRNAP**) (Tiranti, Savoia *et al.* 1997) y por lo menos tres factores de transcripción: factor de transcripción mitocondrial A (**TFAM**) (Fisher and Clayton 1988), factores de transcripción B (**TFB1M&TFB2M**) (Falkenberg, Gaspari *et al.* 2002) para el inicio, y el factor mitocondrial de terminación (**mTERF**) para la terminación (Fernandez-Silva, Martinez-Azorin *et al.* 1997).

2.1. RNA polimerasa mitocondrial (mtRNAP)

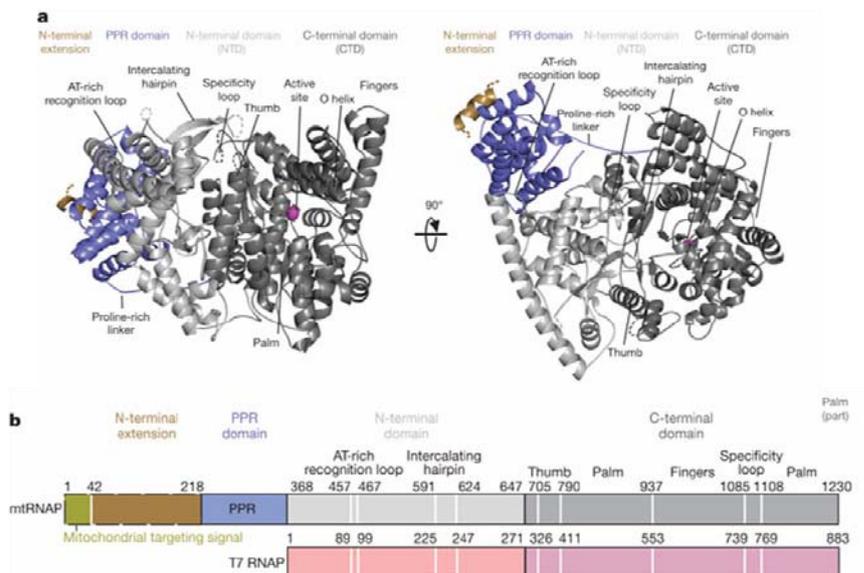
Relativamente relacionada con RNA polimerasa del bacteriófago T7 (RNAPol T7). Sin embargo RNAPol T7 puede iniciar la transcripción por ella sola, mientras que la mtRNAP requiere de los factores TFAM y TFB2M para la unión y apertura del promotor de ADN. TFAM es una abundante proteína que une y dobla la secuencia promotora de ADN de 15-40pb río arriba del sitio de inicio de la transcripción, y estimula la reclutación de la mtRNAP y TFB2 al promotor.

TFB2M asiste a la mtRNAP durante la apertura (*melting*) del promotor y acerca el sitio activo de mtRNAP para interactuar con la primera base del híbrido RNA-DNA.

Recientemente se ha publicado la estructura cristalográfica de la mtRNAP humana a una resolución de 2.5 Å que permite entender las diferencias funcionales entre mtRNAP y RNAP de bacteriófago T7 a través de las diferencias estructurales (Ringel, Sologub et al.) (Fig_B4). La estructura muestra:

- Un dominio catalítico carboxiterminal necesario para catalizar la incorporación de nucleótidos. Este dominio catalítico es similar a la RNAPol T7.
- Un dominio amino-terminal (NTD) que remotamente se parece al dominio de unión al promotor de RNAPol T7.
- Un novedoso dominio “*pentatricopeptide repeat*” (PPR domain) es un dominio helicoidal que conecta con el NTD mediante un linker rico en prolina, prácticamente lineal. Este dominio consiste en nueve α-hélices, de las cuales 4 constan de 2 dominios PPR. Los dominios PPR son típicos de proteínas de plantas y mitocondrias. Las proteínas que presentan dominios PPR se encuentran involucradas en el procesamiento de RNA.
- Una extensión N-terminal flexible

La estructura de mtRNAP tiene la forma de una mano derecha con los subdominios correspondientes a la palma, dedos y pulgar, característicos de la familia polA de las nucleotidil transferasas, donde la palma y dedos corresponden con el dominio C-terminal (CTD), que se encuentra muy conservado. El pulgar presenta el dominio N-terminal y el PPR.



Fig_B4. (A) Estructura cristalográfica de mtRNAP. (B) Representación gráfica de los diferentes dominios presentes en mtRNAP y T7RNAP. Para mtRNAP, en marrón se muestra la región N-terminal flexible, en azul el dominio PPR. En gris claro se muestra el dominio amino-terminal, en forma de pulgar, que contiene el loop de reconocimiento de zonas AT y la horquilla intercalante. En gris oscuro se muestra el dominio C-terminal. (Ringel, *et al.*)

El dominio PPR secuestra o retiene el loop de reconocimiento de zonas ricas en 'AT', que une al ADN en el caso de RNAPol T7, lo que probablemente explicaría la necesidad del factor TFAM durante la unión al promotor. De hecho, la sustitución del residuo conservado Arg en el 'loop de reconocimiento', o la exposición de este loop por delección de la parte N-terminal de mtRNAP, no presenta ningún efecto en la transcripción.

Por otra parte, el dominio CTD y la horquilla intercalante, que provocan la apertura del ADN en RNAPol T7, están reposicionados, explicando así la necesidad de TFB2M durante la apertura del promotor.

Ringel, *et al.* proponen un modelo donde TFB2M se insertaría entre la horquilla intercalante para su reestructuración. La reestructuración de la horquilla intercalante permitiría su acción en la apertura del promotor. Además, TFB2M se situaría cercano al sitio de activación en CTD para estabilizar el sustrato de la transcripción mediante la unión de la cadena molde y el NTP llegado. Es así como la pérdida de funciones evolutiva, como unión y apertura del ADN de mtRNAP, implica la acción de factores de iniciación tanto en *trans* como *cis*, para permitir que la regulación del genoma mitocondrial se adapte a las funciones de la mitocondria y los cambios en el ambiente.

Además de su papel en transcripción, mtRNAP sirve como primasa para la replicación del ADN mitocondrial.

2.2. TFB1M y TFB2M

2.2.1. Descubrimiento

Estudios en modelos de sistemas *Saccharomyces cerevisiae* y *Xenopus laevis* revelaron el requerimiento o la necesidad de otro factor de transcripción mitocondrial (MTF1p), distinto de TFAM, sugiriendo la existencia de un segundo factor de transcripción mitocondrial en vertebrados. Fue en este momento cuando se describieron en humanos dos factores de transcripción homólogos a sc-mtTFB/MTF1p, llamado TFB1M y TFB2M, relacionados ambos con la familia de *RNA metiltransferasas* (Seidel-Rogol, McCulloch et al. 2003).

Para este factor existen dos isoformas humanas, llamadas TFB1M y TFB2M, identificadas de forma independiente. Análisis filogenéticos de esta familia de proteínas indican que la presencia de estos dos homólogos de TFB en *metazoos* es el resultado de un evento de duplicación génica ocurrido en las primeras etapas de la evolución en eucariotas, antes de la divergencia en *fungi* y *metazoos* (Cotney and Shadel 2006).

En levadura, la RNA polimerasa mitocondrial (ROP41p) se encuentra asociada con sc-mTFB de 43 kDa. La estructura primaria de sc-mtTFB recuerda a σ -factors de procariotas, pero la

estructura cristalográfica revela cierta homología con rRNA metiltransferasas. Aunque individualmente ni RPO41p ni MTF1p pueden interactuar con promotores mitocondriales de levadura, juntos llevan a cabo reconocimiento específico del promotor (Diffley and Stillman 1991).

De forma interesante, al igual que el factor de levadura, los dos TFBMs humanos están relacionados con rRNA metiltransferasas. TFB1M puede unir S-adenosil-metionina y metilar rRNA. Sin embargo, esta actividad no es necesaria para la activación de la transcripción llevada a cabo por este factor.

2.2.2. TFB1M/TFB2M interactúan con la maquinaria transcripción

Inicialmente se demostró que tanto TFB1M como TFB2M son capaces de interactuar directamente con TFAM, ya que coimmunoprecipitaban en extractos de células HeLa y se observó interacción directa *in vitro*. Además se ha identificado que era la cola C-terminal de TFAM la que interactúa con TFB1M y TFB2M (McCulloch and Shadel 2003).

También se ha descrito que tanto TFB1M como TFB2M pueden interactuar directamente con mtRNAP formando un heterodímero estable mediante co-purificación (McCulloch and Shadel 2003).

El grupo de Falkenberg, *et al.* describieron que TFB2M presenta al menos un orden de magnitud mayor en promover la transcripción que TFB1M (Falkenberg, Gaspari et al. 2002). Además, observaron que al eliminar TFB2M de los extractos mitocondriales existía una pérdida del nivel de transcripción.

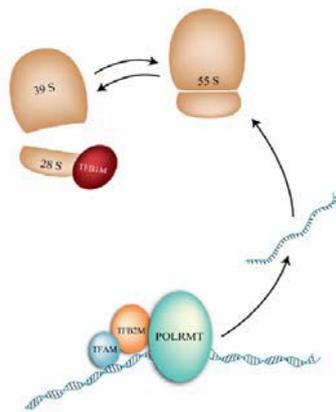
2.2.3. TFB1M/TFB2M y su dualidad funcional

En un principio tanto TFB1M como TFB2M fueron descritas como isoformas con dualidad funcional, con capacidad para participar en la transcripción y actuar como metiltransferasas (McCulloch and Shadel 2003). Sobre todo en el caso de TFB1M, en el que se constató la actividad rRNA metiltransferasa y su capacidad para metilar los residuos adenina adyacentes en la estructura *stem-loop* conservada en moléculas rRNA del 16S bacteriano y 12S mitocondriales humano (Seidel-Rogol, McCulloch et al. 2003) (Fig_B5). Por lo tanto, en un determinado momento se planteó la cuestión de por qué estas proteínas, y sobre todo para el caso de TFB1M, presentaban la dualidad funcional participando tanto en la transcripción como en la traducción de proteínas (Shoubridge 2009).

1. En cultivo de células de *Drosophila*, knockdown de RNA interferencia (RNAi) para la isoforma TFB2M resultaban en una disminución de la transcripción y del número de copias mtDNA, mientras que RNAi para la isoforma TFB1M no presentaba ningún efecto en la transcripción ni replicación, pero sí implicaba una reducción de la traducción (Matushina Y, et al. 2005). Cabe destacar que tanto la sobreexpresión de TFAM como TFB2M incrementaba el número de copias de mtDNA, mientras que la sobreexpresión de TFB1 no afectaba.

2. Ratones knockout para TFB1M mostraban que una pérdida de TFB1M anula la traducción mitocondrial (Metodiev, Lesko et al. 2009). Estos resultados demostraban que las dos isoformas de TFBM presentaban funciones independientes y diferentes *in vivo*: TFB1M actúa como dimetiltransferasa modificando 12S rRNA, mientras que TFB2M presenta un papel en el inicio de la transcripción.

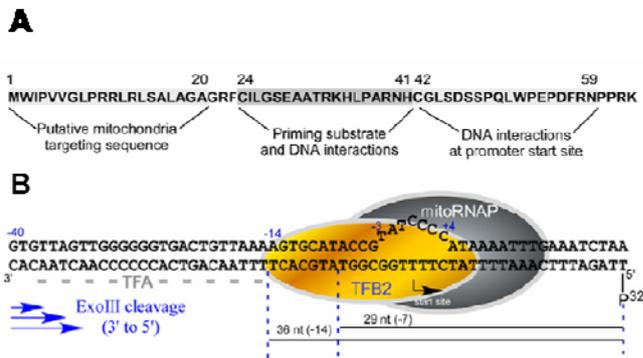
Esto sugirió que TFB1M presentana una pérdida en la capacidad de activar la transcripción y que las dos isoformas de TFB no funcionaban de la misma forma.



Fig_B5. Representación de la función de TFB1M en el ensamblaje de los ribosomas mitocondriales. El RNA producto de la transcripción del mtADN, es posteriormente traducido mediante los ribosomas mitocondriales 55S. Préviamente es necesaria la acción metiltransferasa por parte de TFB1 a la subunidad pequeña 28S (Metodiev, et al.)

2.2.4. Papel en transcripción de TFB2M

Recientemente, se ha demostrado que TFB2M lleva a cabo una serie de interacciones con el ADN cerca del inicio de transcripción y facilita la apertura del promotor, aunque no es esencial para el reconocimiento del mismo (Ringel, Sologub et al.) (Fig_B6). Además, TFB2M es capaz de interactuar con el sustrato de la transcripción, sugiriendo así que TFB2M actúa como componente transitorio en el sitio de catálisis del complejo de iniciación de la transcripción (IC). Mapeando TFB2M se ha detectado una región de N-terminal (residuo 1-62) que se encuentra involucrada tanto en el proceso de apertura como en la interacción con el nuevo producto de la transcripción. Sin embargo, cabe destacar que TFB2M por el sólo es incapaz de unirse al promotor, ya que no se ha detectado ningún producto mediante footprinting (Sologub, Litonin et al. 2009).

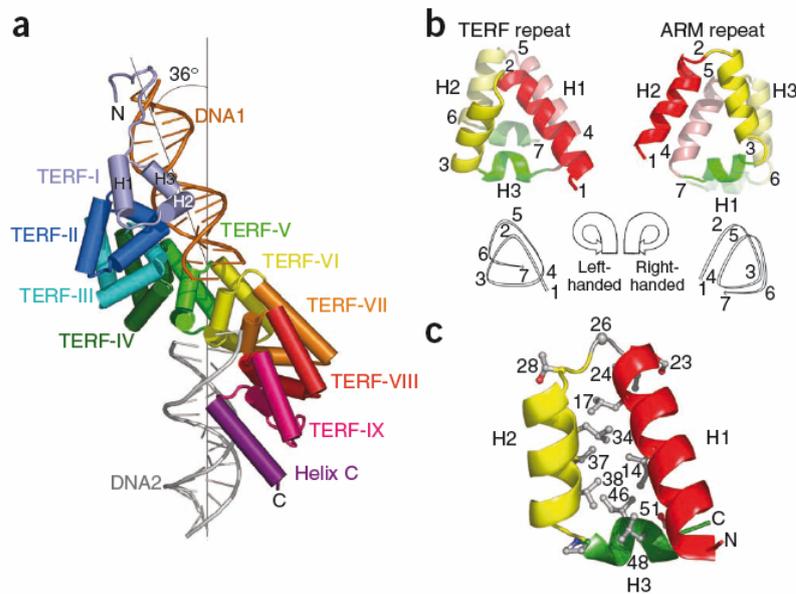


Fig_B6. (A) Secuencia proteica de TFB2M y su participación en la interacción con mtRNAP, función de apertura y interacciones con ADN. (B) Representación de las regiones de protección de TFAM (-40 a -14), mtRNAP (-7 a +15), TFB2M/mtRNAP (-14 a +15) (Ringel, et al.).

2.3. mTERF y la terminación de la transcripción

Los eventos específicos de terminación o fin de la transcripción juegan un papel importante en la dirección de los niveles de transcritos en la mitocondria. La elevada relación entre la síntesis de rRNA y mRNA ha sido atribuida a la mayor frecuencia de transcripción de la cadena H a partir del sitio HSP1. HSP1 dirige la transcripción a partir de su sitio dentro del gen tRNA Phe. La mayoría de los transcritos iniciados en HSP1 terminan río abajo del 16 rRNA, en una secuencia terminadora fuertemente bidireccional localizada dentro del gen tRNA Leu(UUR). Este evento de terminación requiere la participación del factor de terminación de la transcripción (mTERF). La actividad de promover el fin de la transcripción de mTERF es bidireccional, y dado que no existen genes codificados río abajo del sitio de unión de mTERF, puede entonces también actuar en el fin de la transcripción de la cadena ligera.

mTERF es una proteína de 39 kDa, que contiene tres dominios cremallera de leucinas y une a la secuencia de 28 pb situada río abajo del gen que codifica para 16S rRNA, en el gen para tRNA Leu, donde se conoce que finaliza la transcripción. La estructura cristalográfica de mTERF en complejo con ADN revela que esta proteína está formado por 9 repeticiones helicoidales que están reorganizadas de forma triangular en sentido contrario a las agujas del reloj, dando lugar a un núcleo hidrofóbico central. Este tipo de organización permite definir un nuevo dominio llamado dominio *Zurdo* (Jimenez-Menendez, Fernandez-Millan et al.) (Fig_B7). Además tanto la estructura cristalográfica como experimentos de SAXS permiten concluir que mTERF genera una curvatura en el ADN de 35°, lo cual permite entender el mecanismo in vivo de esta proteína.



Fig_B7. (A) Estructura cristalografía del factor de terminación mTERF, con las nueve repeticiones. La región N-terminal está contactando con el DNA1, mientras que la región C-terminal está en contacto con el DNA2. (B) Representación de las nueve repeticiones helicoidales que están reorganizadas de forma triangular en el sentido contrario a las agujas del reloj, definiendo el dominio Zurdo. (C) Representación de los residuos mutados en el repeat VII. (Jiménez-Menendez, *et al.*)

La proteína recombinante humana mTERF es activa como monómero y no parece ser necesaria su fosforilación para que sea activa (Asin-Cayuela, Schwend et al. 2005). Interesantemente. Durante la caracterización inicial de esta proteína, se observó que mTERF era capaz de estimular el inicio de la transcripción a partir de HSP1, indicando la existencia de un nexo de comunicación entre mTERF y la maquinaria de inicio de la transcripción, es decir, entre la terminación y el inicio de la transcripción (Kruse, Narasimhan et al. 1989; Asin-Cayuela, Helm et al. 2004). Además, mTERF puede unir simultáneamente el sitio de iniciación en HSP1 y el de terminación, resultando en un *looping* de rRNA. Se ha propuesto que estas interacciones pueden promover la reiniciación de la transcripción en HSP1 resultando en un elevado ratio de síntesis de rRNA, por tanto, un aumento del ratio rRNA/mRNA. Y este hecho es consistente con el conocimiento que la diferencia en ratio entre rRNA y mRNA es debido a la diferencia en transcripción entre HSP1 y HSP2.

2.4. TFAM

Los promotores LSP y HSP comparten río arriba una secuencia de unión para TFAM (Fisher and Clayton 1988). TFAM fue primero caracterizado como una proteína HMGbox (*High Mobility Group*) que estimula la transcripción mediante la unión específica a sus sitios de unión río arriba en LSP y HSP.

Estructuralmente TFAM consiste en dos motivos HMGbox en tándem y una cola C-terminal. TFAM así como otras proteínas HMGB puede doblar y desenrollar el DNA, propiedades potencialmente ligadas a su habilidad para estimular la transcripción después de la unión al DNA. En la unión específica y la promoción del inicio de la transcripción parece jugar un papel importante la cola C-terminal de TFAM (Dairaghi, Shadel et al. 1995).

Además de su unión específica de secuencia, TFAM puede unir de forma no específica a otros sitios de ADN, desarrollando así su función como proteína empaquetadora del mtADN en estructuras compactas llamadas nucleoides.

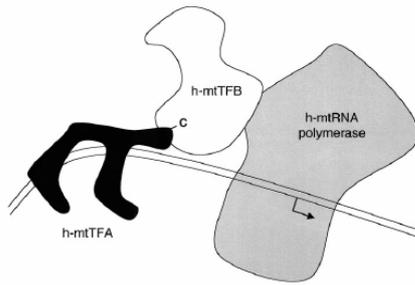
Estudios con ratones knockout para TFAM mostraban una letalidad embrionaria y una ausencia de forma en el mtADN confirmando así el papel de TFAM en el mantenimiento del mtADN en mamíferos (Larsson, Wang et al. 1998). Además, los niveles de TFAM correlacionan con la abundancia de mtADN, ya que se observó que la sobreexpresión de TFAM provocaba un incremento del número de copias de mtADN (Ekstrand, Falkenberg et al. 2004). En concreto, la cantidad de mtADN cuantificada en tejidos somáticos y embrionarios era proporcional a la cantidad de TFAM expresada en ellos, sugiriendo que TFAM podría funcionar como factor limitante en el número de copias de mtADN (Kanki, Ohgaki et al. 2004).

3. Modelos para el complejo de iniciación de la transcripción mitocondrial

El modelo de inicio de la transcripción (o PIC, *pre-initiation complex*) y el papel de las proteínas involucradas está en continua discusión debido a la gran investigación en mitocondria y la gran cantidad de información disponible cambiante en breve espacio de tiempo.

A continuación se realiza un breve resumen de alguno de los modelos más aceptados y su evolución con el tiempo.

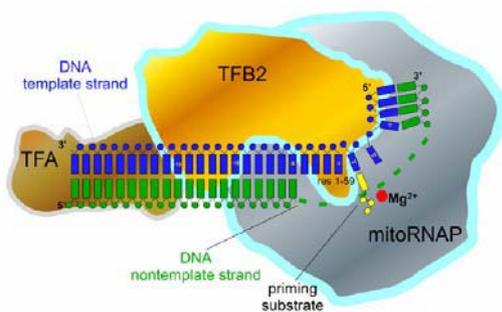
Inicialmente, se describió un modelo basado en las interacciones conocidas para las proteínas descritas como esenciales en el inicio de la transcripción: TFAM, TFB1M y TFB2M, mtRNAP (McCulloch and Shadel 2003) (Fig_B8). Donde mtRNAP realiza el reconocimiento y unión a la secuencia promotora con la ayuda de TFAM. TFAM permitiría la interacción con el ADN ya sea gracias a su reconocimiento específico de secuencia o mediante la generación de una conformación estructural determinada debido a su capacidad para curvar el ADN. De acuerdo con este modelo, TFB2M presentaría una función de “adaptador” que facilitaría la accesibilidad de mtRNAP a la secuencia promotora, gracias a la interacción de TFB2M con la cola C-terminal de TFAM.



Fig_B8. Modelo de inicio de la transcripción mitocondrial en base a las interacciones entre las diferentes proteínas (McCulloch&Shadel_2003), donde la cola C-terminal interacciona con TFB2M que actúa como puente o cemento entre TFAM y mtRNAP.

Posteriormente se propuso el modelo de la burbuja de transcripción o “*open complex*”, donde la región promotora de -35 a +10 pb se encuentra cubierta por el complejo de inicio de la transcripción formado por TFAM, TFB2M y mtRNAP, existiendo una apertura o burbuja de 3 o 4 pb río abajo del sitio de inicio de transcripción (Sologub, Litonin et al. 2009) (Fig_B9). Sologub, *et al.* sugerían un mecanismo de inicio de la transcripción gobernado por mtRNAP asistida por dos factores de iniciación de la transcripción, TFAM y TFB2M. Donde mtRNAP posee funciones de reconocimiento de la secuencia promotora, pero depende de TFAM y TFB2M para la apertura del promotor (Sologub, Litonin et al. 2009).

En este modelo, TFAM se une al mtADN, protegiendo la región río arriba del sitio de inicio de la transcripción en LSP entre 14-35 pb, y asiste al ensamblaje del complejo de iniciación TFB2M/mtRNAP y/o facilitaría la apertura inicial del promotor. El principal papel de TFB2M es la apertura del promotor y estabilizar el complejo abierto mediante la unión simultánea del sustrato para la transcripción y el RNA nascente. Además, parece ser que el desplazamiento de TFB2M puede ser el evento que gobierne la transición del estado de iniciación de la transcripción al estado de elongación.



Fig_B9. Modelo de complejo para el inicio de la transcripción, donde TFB2 favorece la apertura del ADN y favorece la participación de mtRNA. (Sologub, *et al.*)

Al mismo tiempo se planteó un modelo de inicio de la transcripción donde el complejo de inicio de la transcripción mitocondrial está regulado por un sistema de dos componentes *in vitro*, siendo estos dos componentes mtRNAP y TFB2M, llevando a cabo la iniciación

específica a partir de los promotores mitocondriales, LSP y HSP1 (Shutt, Bestwick et al.). Es decir, TFAM sería prescindible para el inicio de transcripción. Aunque existe cierta dependencia entre la cantidad de TFB2M y su actividad, cuando menor es la cantidad de TFB2M mayor es la dependencia de la acción de TFAM. Este modelo de dos componentes, donde TFAM resulta prescindible, difiere totalmente con las anteriores descripciones del PIC basadas en un sistema de tres componentes. Cabe destacar que en sistema *in vitro* existen muchas variables a tener en cuenta que pueden afectar los resultados observados, como la concentración de proteínas, la concentración de sal en el sistema,... Recientemente se ha descrito que concentraciones de sal bajas pueden favorecer la formación o estabilización de la burbuja de transcripción y crear artefactos en el sistema (Shutt, Bestwick et al.).

Otro punto de controversia es que Shutt, *et al.* observaron que la utilización del promotor (LSP o HSP1) varía en función de la cantidad de TFAM presente. Detectaron que la cantidad de TFAM necesaria para activar el promotor HSP1, es mayor que para LSP. Además la estimulación de LSP por TFAM era significativamente mayor que para HSP1. Esta diferencia entre transcripción de LSP y HSP1 en función de la cantidad de TFAM es debida a los diferentes roles de los transcritos originados desde HSP1 y LSP. La transcripción a partir de LSP además de dar lugar a un mRNA (ND6) y tRNAs, también produce los encefaladores necesarios para iniciar la replicación del mtADN a partir de O_H, mientras que la transcripción a partir de HSP1 genera 2 rRNAs necesarios para la biogénesis de los ribosomas mitocondriales y la traducción.

Por lo tanto según este modelo, el PIC se trataría de un sistema de dos componentes homólogo al de levadura que ayudaría a entender la regulación diferencial de los promotores del mtADN mediante la participación de TFAM, que actuaría como activador transcripcional en lugar de ser un componente obligado del complejo de iniciación de la transcripción: LSP *in vitro* no es intrínsecamente más activo que HSP1, pero es más sensible a la estimulación por TFAM que HSP1. Es por ello que los autores proponen un mecanismo de iniciación similar en los dos promotores, pero TFAM los activa por mecanismos distintos. Además, se especula que la necesidad de mayores cantidades de TFAM en HSP1 podría indicar un mayor orden o estructuración del complejo TFAM/mtADN así como involucrar secuencias de la región interpromotora (Shutt, Bestwick et al.).

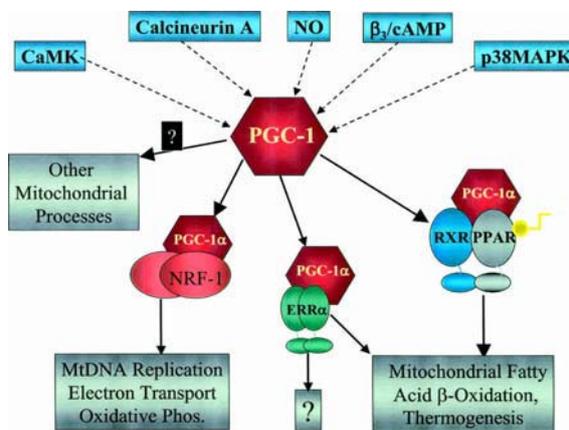
4. Biogénesis mitocondrial: coordinación entre genes nucleares y mitocondriales

La regulación de la biogénesis mitocondrial involucra la acción coordinada de genes nucleares y mitocondriales, ya que la actividad de la mitocondria no es independiente y requiere de la expresión de determinados genes nucleares para sintetizar enzimas catalíticas, para formar parte de los complejos de CTE, para sintetizar proteínas que participen en la importación y ensamblado de complejos, para proteínas involucradas en procesos de expresión y replicación del mtADN.

El factor nuclear de respiración 1 (NRF-1) ha sido relacionado con la biogénesis mitocondrial. Los genes diana para NRF-1 han sido identificados mediante la caracterización funcional de los sitios de unión de NRF-1 dentro de sus promotores. Muchos de los genes identificados como diana son subunidades de los cinco complejos de CTE. No obstante la red regulatoria controlada por NRF-1 se extiende desde los genes que codifican para subunidades CTE a otro tipo de genes. Por ejemplo, genes involucrados en el ensamblaje del aparato respiratorio o constituyentes de la transcripción y replicación del mtADN, o bien componentes de la importación de proteína, etc. TFAM, proteína molecular codificada en núcleo, es una de las dianas génicas para NRF-1 viéndose así reflejado el papel de NRF-1 como integrador en interacciones o la relación núcleo-mitocondria.

NRF-1 es una proteína de 68 kDa que presenta un dominio C-terminal de activación de la transcripción. Esta proteína se une como homodímero a los sitios palindrómicos específicos del ADN. La fosforilación de una serina en el dominio N-terminal incrementa la afinidad a ADN y, por lo tanto, sus funciones de activación trans.

Se ha descrito el coactivador transcripcional PGC-1 como coordinador de la regulación de la biogénesis mitocondrial a través de núcleo (Fig_B10). Este coactivador transcripcional lleva a cabo múltiples funciones como modificación de la cromatina a través de modificaciones post-traduccionales, interacción con RNA polimerasa II, procesamiento de mRNA o reclutamiento de otros coactivadores transcripcionales, como NRF-1. Además PGC-1 es el único que exhibe un enriquecimiento en determinados tejidos (como músculo esquelético o tejido adiposo) y es altamente inducible (la expresión de este activador se ve incrementada en situaciones de demanda de ATP o calor) (Kelly and Scarpulla 2004).



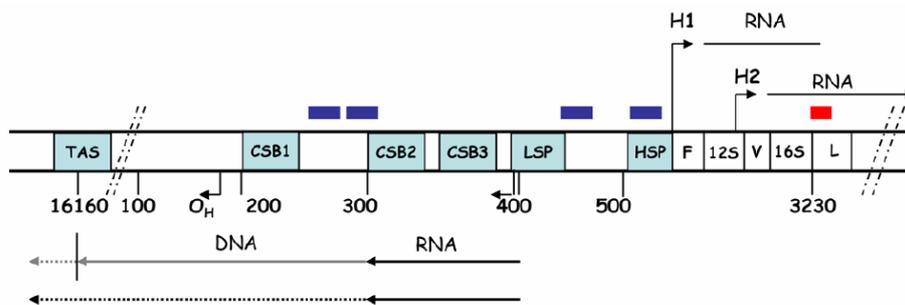
Fig_B10: Representación esquemática del papel regulador de PGC-1 en la célula y sus diferentes interacción con factores de transcripción, represores, supresores de tumores, etc. (Nelly, et al. 2004)

5. Relación entre la formación de encebadores para la replicación y la transcripción de LSP

Aunque la zona del D-loop es una de las más divergentes entre los diferentes genomas mitocondriales, en vertebrados existen pequeños bloques de secuencias altamente

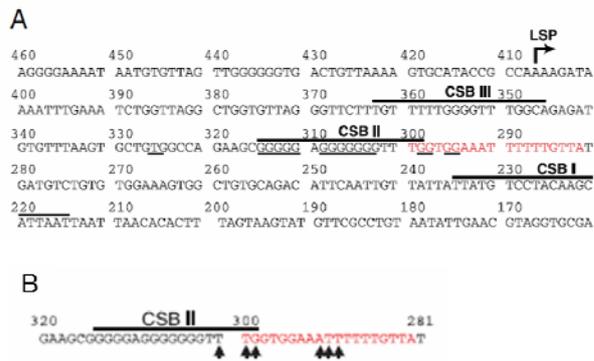
conservadas situados entre LSP y O_H. Estos tres elementos son llamados “*conserved sequence blocs*” (CSBI, CSBII y CSBIII) y presentan diferente extensión (Fig_B11). Estas regiones parecen ser importantes para la estabilización de los encebadores RNA en la cadena molde del mtDNA (Shadel and Clayton 1997). En general, la mayoría de los transcritos finalizan la transcripción hasta el final de su ADN molde. No obstante, existen evidencias de un fenómeno de terminación de la transcripción en CSBII *in vitro*, correspondiente con la posición 300-282 en el genoma mitocondrial, dando lugar a transcritos prematuros de unos 100pb aproximadamente (Pham, Farge et al. 2006).

La secuencia conservada humana CSBII tiene una longitud de 16pb, donde 15 son GC con todas las guaninas situadas en una cadena. El RNA sintetizado a partir de LSP contiene las guaninas, y este hecho contribuye a favorecer la estabilidad de la estructura híbrida ARN-ADN en CSBII (Xu and Clayton 1996). La estabilización del híbrido ARN/ADN inmediatamente río arriba del O_H puede incrementar la probabilidad de iniciación de la replicación. Esto puede ser beneficioso en casos donde es necesaria una rápida amplificación del mtDNA (Bonawitz, Clayton et al. 2006).



Fig_B11. Representación esquemática del D-loop (marcado con líneas diagonales discontinuas), donde se representa las regiones conservadas: CSBI, CSBII, CSBIII; las secuencias promotoras para la cadena L y H: LSP y HSP. Los rectángulos azul oscuro hacen referencia a las zonas descritas de interacción TFAM con ADN. El rectángulo rojo representa el sitio de unión para mTERF. En la parte inferior se observa la síntesis de RNA continua, así como el bloqueo en CSB2 para dar lugar a los encebadores. (Farge, et al. 2006)

Recientemente se ha demostrado que el proceso de terminación prematura de la transcripción en la cadena L es un proceso dependiente de secuencia (Pham, Farge et al. 2006). Mientras que mutaciones en CSBIII provocan una leve disminución de la terminación de la transcripción prematura, los mutantes en CSBII casi eliminan la formación de productos prematuros de la transcripción, en concreto los mutantes de la región 300-304. La mutación en CSBI no provoca ningún efecto.



Fig_B12. (A) Representación de las regiones CSB en la secuencia de mtADN. Los residuos en rojo muestran la zona donde ocurre la terminación. Las guanina subrayadas son las susceptibles de formar G-quadruplex. (B) Las flechas indican los residuos cuya mutación provoca una reducción de la terminación prematura de la transcripción. (Farge, et al. 2006)

Además se ha demostrado que esta terminación específica de secuencia es debida a la formación de estructuras de G-quadruplex que se forman en el ARN después de la transcripción de CSBII (Wanrooij, Uhler et al.). También se demostró que la secuencia de poli-dT río abajo de CSBII es importante para la terminación prematura de la transcripción. Este mecanismo es reminiscente del sistema de terminación en procariotas independiente de Rho, con la excepción que las estructuras de G-quadruplex reemplazan la formación de un *hairpin loop* (una horquilla) en el mRNA bacteriano durante la terminación de la transcripción.

6. Transcripción mitocondrial en levadura

El mtADN de *S.cerevisiae* (86 Kb) tiene una organización compleja con abundantes zonas no codificantes y genes interrumpidos por intrones, mientras que el mtADN mamífero es mucho más pequeño (16 Kb) y contiene muy pocas regiones no codificantes y ausencia de intrones. Además, la iniciación de la transcripción en *S.cerevisiae* ocurre a partir de múltiples promotores, mientras que en mtADN de mamíferos sólo existen tres secuencias promotoras descritas (Gaspari, Falkenberg et al. 2004).

La maquinaria básica para la transcripción en *S.cerevisiae* consiste únicamente en dos factores o componentes: RPO41 o *sc*-mtRNAP, y el factor de transcripción MTF1p, también conocido como *sc*-mtTFB. En este caso la transcripción es totalmente independiente de productos codificados en núcleo. En contraste, la transcripción en mamíferos se ha definido como un sistema de tres componentes: mtRNAP, TFB2M y TFAM, aunque recientemente existe mucha controversia en este campo. Es interesante destacar que el homólogo de TFAM en levadura, ABF2p, no presenta actividad en la transcripción pero tiene un importante papel en el empaquetamiento del mtADN.

En levadura, el heterodímero RPO41/MTF1p se une a una región del promotor de 9nt (región consenso ATATAAGTA) desde donde se inicia la transcripción (Mangus, Jang et al. 1994). MTF1p permanece asociada a RPO41 durante la iniciación y en los primeros estadios de la

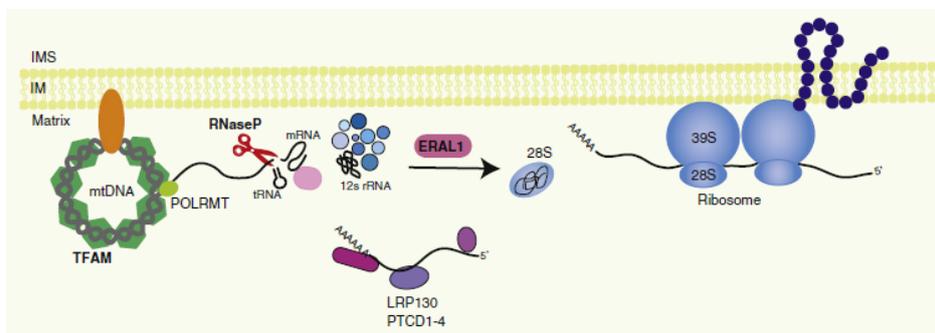
elongación, pero es liberada cuando empieza la etapa de elongación. RPO41p es una RNA polimerasa de subunidad, que comparte similitud de secuencia con las polimerasas T7 y T3 de bacteriófagos. Aunque RPO41 puede iniciar la transcripción no selectivamente a partir de ADN sintéticos, la especificidad del factor MTF1p es necesaria para el reconocimiento de la secuencia promotora y la iniciación selectiva (Winkley, Keller et al. 1985). De forma interesante, RPO41p mediante su interacción con los promotores mitocondriales y los nucleósidos trifosfatos, puede actuar como sensor para la disponibilidad de ATP. La abundancia de transcritos mitocondriales está relacionada con la respiración y la síntesis de ATP, y cuando es necesario un requerimiento energético elevado (es decir, elevada concentración de ATP) existe una activación del inicio de la transcripción de la subunidades de la cadena respiratoria.

7. Sistema de traducción en mitocondria

La síntesis de las 13 proteínas codificadas en mtADN es llevada a cabo por un mecanismo de traducción especializado situado en la matriz mitocondrial (Fig_B13). Los componentes de la maquinaria mitocondrial de la traducción (factores de traducción, factores de elongación y mitorribosomas) son parecidos a la maquinaria de bacteria y difieren de los encontrados en el citosol celular.

El mitorribosoma mamífero es una partícula 55S, compuesta por una subunidad pequeña, 28S, y una mayor, 39S, y contiene muchas más proteínas y RNAs que el ribosoma bacteriano 70S.

La transcripción a partir de LSP y HSP da lugar a productos policistrónicos, donde los tRNAs son escindidos para procesar los transcritos y dar lugar a mRNA y rRNA. Este proceso de puntuación es llevado a cabo mediante la actividad endonucleasa de RNasa P. La proteína ERAL1, localizada en la matriz mitocondrial, se une a rRNA de la subunidad pequeña 28S, siendo un principal constituyente de la subunidad pequeña del ribosoma mitocondrial y permite su ensamblaje. Se ha podido observar que muchas de las proteínas que participan en el metabolismo del RNA presentan interacciones con el factor de transcripción y empaquetamiento TFAM, existiendo un vínculo entre transcripción, traducción y nucleoides, todo ello localizado en la matriz mitocondrial (Uchiumi, Ohgaki et al.).



Fig_B13. Mecanismo de traducción, procesamiento de RNA y ensamblaje de mitorribosomas, localizado en la matriz mitocondrial. (Uchiumi, et al.)

C. FACTOR DE TRANSCRIPCIÓN MITOCONDRIAL HUMANO A, TFAM

El factor de transcripción mitocondrial humano A, TFAM, es la proteína objeto de nuestro estudio. A continuación se detallará alguna de las características de esta proteína.

1. Introducción

El primer factor de transcripción identificado en células de mamífero fue el factor de transcripción mitocondrial A humano, TFAM, (Fisher and Clayton 1988). El aislamiento del gen que codifica para TFAM reveló que es un miembro de la familia de proteínas *high-mobility-group* (HMGB) (Parisi and Clayton 1991), que confieren a TFAM las propiedades para curvar y cubrir el ADN. Los dominios HMGB presentados una estructura ampliamente conservada donde tres hélices alpha se encuentran plegadas en forma de L.

TFAM se encuentra involucrado en varios procesos relacionados con el ADN como transcripción, replicación y mantenimiento del ADN mitocondrial (Thomas and Travers 2001). Relacionado con su capacidad para activar la transcripción, TFAM puede unir de forma específica el ADN río arriba de LSP y HSP1, como se pudo comprobar mediante ensayos de *footprint* (Fisher, Topper et al. 1987). En estos promotores, el sitio de unión para TFAM se encuentra situado río arriba del sitio de inicio (-12 a -39pb). Parece ser que la distancia de 10pb entre la región definida para el inicio de la transcripción y la región de unión para TFAM es crítica para la transcripción (Chang, Hixson et al. 1986). El segundo sitio de inicio de transcripción, HSP2, presenta una secuencia de 15 pb similar a la conservada en los otros promotores, pero aparentemente no presenta sitio de unión para TFAM río arriba.

TFAM es una proteína de 246 aminoácidos (24.4 kDa) codificada y sintetizada en el núcleo, que es posteriormente translocada a mitocondria gracias a la presencia de una región N-terminal, llamada MLS (*mitochondrial leader sequence*) que presenta las propiedades necesarias para importar la proteína hasta la matriz de la mitocondria, incluyendo cargas positivas uniformemente distribuidas así como potenciales hélices anfipáticas. (Fig_C1).

El análisis de la secuencia peptídica revela dos regiones conservadas que fueron descritas como dos dominios HMGB en tándem, y aunque el segundo dominio es más divergente, éstos presentan un 26% de identidad y 36% de similitud global cada una (permitiendo espacios)(Parisi and Clayton 1991). El 70% de la proteína está compuesto por los dominios HMGB, lo cual hace pensar en la relevancia funcional de estos dominios para TFAM. Como más adelante se detallará, existen dos tipos principales de proteínas HMGB en función del número de HMGB y su capacidad de unión al ADN de forma específica o inespecífica. Fue Parisi *et al.*, quienes observaron la relación existente entre los dominios HMGB de TFAM y HMGB del factor humano nuclear de transcripción, UBF (*Upstream Binding Factor*), proteína humana conocida por ser importante para la transcripción en el núcleo.

Aunque el tamaño de UBF humano (hUBF) y TFAM sea dispar (97 kDa y 25 kDa, respectivamente), su similitud en cuanto a secuencia primaria de aminoácidos es indudable. La primera HMGbox de TFAM es mucho más similar a la primera HMGbox de hUBF. Sin embargo, la segunda HMGbox de TFAM es similar a la segunda de hUBF, siendo ambas mucho más cortas y divergentes que otro tipo de HMGbox (Parisi and Clayton 1991).

A partir de la estructura primaria de la proteína se predijo la existencia de una secuencia linker entre los dos dominios HMGbox, con un número de residuos y carga positiva similar a la cola C-terminal. La cola C-terminal ha sido descrita como crítica para el reconocimiento específico de secuencia, así como para una eficiente activación de la transcripción (Dairaghi, Shadel et al. 1995).

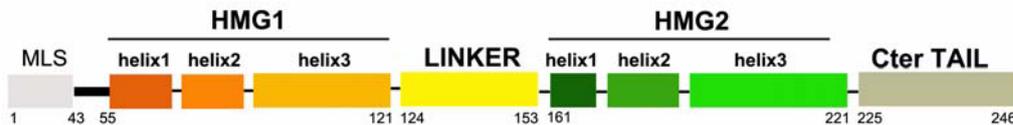


Fig. C1. Representación de los diferentes dominios de TFAM. Naranja para el primer dominio HMGbox (HMG1) con sus tres hélices y en verde para el segundo dominio HMGbox (HMG2). Entre los dos dominios se encuentra un linker (en amarillo), y finalmente un cola C-terminal (en gris). En la región N-terminal se encuentra MLS, para translocar la proteína a mitocondria. La delimitación de los dominios se ha hecho en base a la estructura tridimensional (Rubio-Cosials, et al.).

2. Descubrimiento de TFAM

TFAM fue primero definida como factor de transcripción mitocondrial 1 (Mtf1), necesario para el reconocimiento *in vitro* de los dos principales promotores en mitocondria humana. Posteriormente se descubrió que presentaba la capacidad de unir secuencias de ADN de forma inespecífica (Fisher and Clayton 1988). Además, se detectó un patrón de migración con la presencia de múltiples bandas al aumentar el ratio de TFAM por ADN, lo cual representaba complejos TFAM/ADN con diferente estequiometría.

Inicialmente, la investigación entorno TFAM se focalizó en el papel como factor de transcripción mitocondrial. La comparación de las secuencias para LSP y HSP sugiere que TFAM pueda funcionar de forma independiente a la orientación (Fisher, Topper et al. 1987). Mediante ensayos de afinidad se observó que el complejo TFAM/LSP era más estable que para el caso TFAM/HSP, cuando se realizaban ensayos de competición con una secuencia random. Además para la realización de experimentos de transcripción *in vitro* era necesaria una menor cantidad de TFAM para estimular LSP que para HSP, y esto seguramente tiene mucho que ver con la diferencia funcional de ambos promotores *in vivo*. Los resultados *in vitro* sugerían ya entonces que la concentración de TFAM podía regular la selección del promotor donde la cantidad de TFAM es importante para regular transcripción/regulación del mtADN.

No obstante, en 1987 Fisher, *et al.* no asociaban una implicación funcional a la capacidad de TFAM de unir inespecíficamente a ADN, sino que la discusión sobre este aspecto se limitaba a reflexionar sobre la capacidad de algunas proteínas de unión específica a unir ADN de forma inespecífica para regular la unión de determinados activadores o represores a unirse a determinadas regiones del genoma.

3. Isoforma de TFAM y modificaciones post-traduccionales

3.1. Isoforma

El gen de TFAM (ID 7019) se expande unas 10 Kb y consiste en 7 exones y cinco intrones. Se ha identificado una isoforma alternativa de TFAM *in vivo* que corresponde a la delección del exon5 conocida como TFAM- Δ exon5 (Tominaga, Hayashi et al. 1993). TFAM- Δ exon5 es producto de una pequeña delección en una pequeña región de 32 aminoácidos de la HMGbox2 de TFAM. Aunque no se conoce función *in vivo* para TFAM- Δ exon5, en experimentos *in vitro* se pudo constatar que esta isoforma es incapaz de llevar a cabo la transcripción (Dairaghi, Shadel et al. 1995), pero presenta una débil protección a *DNaseI* en la mitad del sitio LSP más proximal al sitio de inicio de la transcripción. TFAM- Δ exon5 retiene la habilidad de unión inespecífica a ADN.

Esta isoforma es similar a la isoforma existente de UBF, factor de transcripción para RNA polimerasa I (O'Mahoney, Brandon et al. 1991). Como TFAM, UBF presenta múltiples HMGbox y una cola C-terminal. Sin embargo, al contrario de TFAM donde la cola es básica, en el caso de UBF la cola es ácida y se cree que interacciona con HMGbox1 de UBF modulando su capacidad de unión a ADN, en vez de estabilizar su unión. Existe una forma alternativa de UBF, UBF2, presente *in vivo* y donde la segunda HMGbox ha resultado parcialmente deleccionada. La isoforma UBF2 no presenta habilidad para activar la transcripción *in vivo*. Tanto TFAM como UBF presentan una pérdida de la función transcripcional cuando se delecciona uno de los exones por *splicing* alternativo, esta pérdida parece afectar la capacidad de regular la unión a ADN. Aunque el papel de UBF2 y TFAM- Δ exon 5 parece ser desconocido *in vivo*, es posible que éstas formas cortas de proteína muestren funciones de unión en otros sitios del genoma nuclear o mitocondrial.

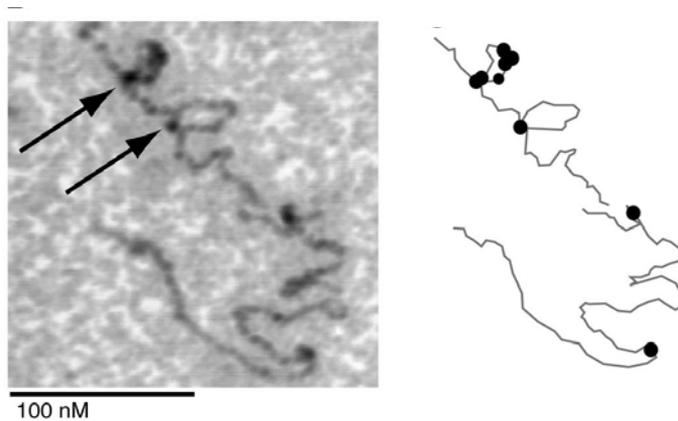
3.2. Modificación pos-traduccionales: acetilación de TFAM

La acetilación de los grupos ϵ -amino de lysina es una importante modificación reversible que permite regular la expresión génica. En particular en núcleo, la acetilación de histonas activa la expresión génica probablemente debido a la neutralización de parte de la carga positiva de las histonas implicando una reducción de las interacciones histona-ADN e histona-histona. Ello provoca la desestabilización de la estructura del nucleosoma y permite la accesibilidad de factores de transcripción al genoma. Aunque la acetilación fue definida para las histonas existe una creciente lista de proteína-NO histonas, las cuales juegan un papel en la regulación de la transcripción y la proliferación celular. En particular HMG1 y otras proteínas HMGbox son acetiladas.

ADN (Fisher, Lisowsky et al. 1992). Mediante estudios con ADN circular relajado y topoisomerasa, se demostró que TFAM era capaz de introducir vueltas de superhélice o superenrollamiento negativo. Esta capacidad de introducir superenrollamiento negativo parece estar ampliamente ligado con el hecho que TFAM es un activador transcripcional, que frecuentemente provocan perturbaciones conformacionales en la estructura del ADN como la introducción de vueltas de hélice negativas que promueven un desenrollamiento local y la apertura del ADN.

Además mediante ensayos de permutación circular Fisher *et al.*, constataron la alteración conformacional que TFAM induce en los promotores HSP y LSP después de su unión específica, mediante la introducción puntos de curvatura en el mtADN. También observaron que el doblamiento del ADN no requería una unión estable y sucedía también en secuencias control inespecíficas. Por lo tanto, Fisher *et al.* fueron los primeros en disociar la unión específica del doblamiento del ADN.

Mediante AFM bajo condiciones de alto vacío se realizó la medida del número de eventos de unión individuales, donde la unión de TFAM inducía doblamientos del ADN de $100 \pm 20^\circ$ en una corta distancia (Kaufman, Durisic et al. 2007) (Fig_C3). Estas mediadas difieren de lo observado en el caso de ABF2p mediante AFM en modo *tapping* (Friddle, Klare et al. 2004) que eran de unos 78° . Una dato a destacar era la presencia de proteína en la base de estructuras tipo loop o Ω .



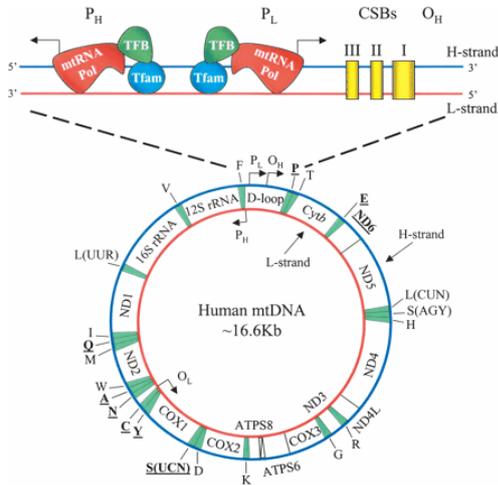
Fig_C3. Imágenes de AFM en condiciones de elevado vacío, donde se puede observar la compactación del ADN y sus especies intermedias. La imagen muestra la formación de loops a baja ocupación de TFAM, las flechas indican la presencia de TFAM en la base de las estructuras tipo Ω . En el panel derecho se muestra un esquema del la unión de TFAM al ADN, los puntos representan a TFAM (Kaufman, et al.).

5. TFAM y su unión al ADN

TFAM presenta la propiedad de unir ADN de forma específica e inespecífica. Esta capacidad de unir tanto a sitios específicos del ADN como a regiones inespecíficas la hace única dentro de la familia de proteínas HMGB. Además, TFAM presenta la capacidad de unir a otros tipos de ADN, como ADN dañado con cisplatino o horquillas de recombinación.

5.1. TFAM: unión específica de secuencia (LSP,HSP1)

Como se ha comentado anteriormente, TFAM participa en el inicio de transcripción mediante su unión específica a regiones río arriba de las zonas promotoras para la transcripción, LSP y HSP1 (Fig_C4).



Fig_C4. Esquema del genoma mitocondrial humano con la ampliación de la zona llamada D-loop, donde se encuentra las regiones promotoras de la transcripción.

5.1.1. Región de protección para LSP y HSP1 por parte de TFAM

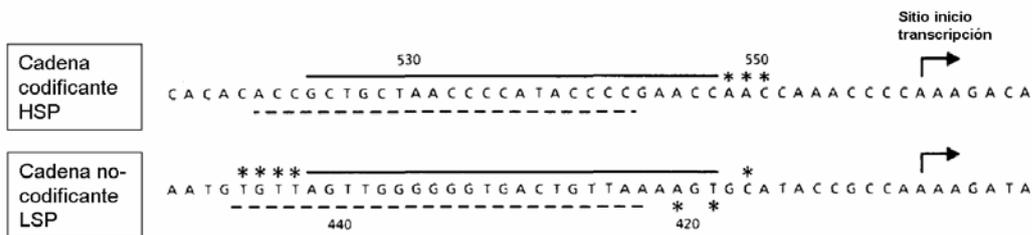
Mediante análisis por *footprinting* se identificaron dos regiones de unión de TFAM a mtADN en las zonas proximales a los promotores (LSP y HSP1). Los dos sitios de unión (HSP-proximal y LSP-proximal) se encuentran situados entre la posición -12 a -39pb, río arriba del sitio de inicio de transcripción, pero con orientación opuesta a la dirección de transcripción (Fig_C5). Ello sugiere que TFAM podría funcionar de forma bidireccional (Fisher, Topper et al. 1987).

- HSP1

TFAM protege débilmente la región río arriba de -13 a -35pb (posición 559 nt) en la cadena no-codificante. La debilidad de la interacción TFAM-HSP (comparada con la interacción TFAM-LSP) se ve reflejada en los análisis de protección mediante *DNasa I*, donde es necesario un mayor ratio de TFAM para la protección.

- LSP

Aproximadamente 23pb son protegidos fuertemente en cada cadena, con una marcada asimetría entre la cadena no-codificante, donde la protección completa sucede de -12 a -35 del sitio de inicio, mientras que en la cadena codificante, la protección completa sucede de -16 a -39pb.



Fig_C5. Regiones de protección frente a ensayos de digestión con DNasa para las secuencias HSP y LSP. Los asteriscos muestran los sitios de corte para DNasa. La línea sólida muestra la región de protección para codificante y la línea discontinua para la cadena no-codificante. En la figura HSP hace referencia a HSP1. (Fisher, *et al.*)

Por lo tanto, las dos regiones proximales de unión que presenta TFAM (HSP1, LSP) están situadas exactamente a la misma distancia respecto al sitio de iniciación y su longitud es aproximadamente unos 23pb. Sin embargo, existen dos diferencias a destacar. La afinidad de TFAM por su sitio proximal a LSP es mayor que para HSP1 (la protección de LSP sucede con ratios de TFAM 4 veces menor). Además, al contrario del caso para LSP donde la protección se da en las dos cadenas, para HSP1 existe menor protección para la cadena codificante de HSP.

Las dos regiones de unión para TFAM son análogas en posición y función transcripcional, además presentan cierta homología de secuencia lo cuál sugiere que en algún momento una duplicación, parcial o completa, de un promotor bidireccional permitió la separación de la transcripción de ambas cadenas, y permitió así la regulación diferencial de las dos unidades de transcripción representadas por la cadena L y H. Por lo tanto, la duplicación de esta secuencia podría haber permitido la “especialización” de promotores específicos de cadena en el mtADN (Fisher, Topper *et al.* 1987).

5.1.1.1. Afinidad por las regiones de unión LSP, HSP1 de TFAM

In vitro, TFAM presenta una mayor afinidad para unirse a LSP que a HSP1 (Fisher and Clayton 1988). Mediante ensayos de afinidad se observó que el complejo TFAM/LSP era más estable que para el caso TFAM/HSP1, cuando se realizaban ensayos de competición con una secuencia random. Además para la realización de experimentos de transcripción *in vitro* era necesaria una menor cantidad de TFAM para estimular LSP que para HSP, y esto seguramente tiene mucho que ver con la diferencia funcional de ambos promotores *in vivo*. Por lo tanto, TFAM podría regular la selección del promotor y en consecuencia la estrecha relación entre transcripción y regulación del mtADN. Esta diferencia en la activación entre HSP1 y LSP fue asociada a la necesidad de una perturbación estructural mayor en HSP1 (Chang and Clayton 1984).

Recientemente mediante ensayos de *Surface Plasmon Resonance* (SPR), se confirmó que TFAM presenta una mayor afinidad por su región de unión a LSP comparado con la afinidad por HSP1. La diferencia de afinidad entre estas dos secuencias es de un orden de 10 veces. Kaufman *et al.*, definían la presencia de un dímero sobre la secuencia de 45 nt y un tipo de

unión cooperativa; tanto la presencia del dímero como el efecto cooperativo de TFAM sobre ADN serán discutidos más adelante (Kaufman, Durisic et al. 2007).

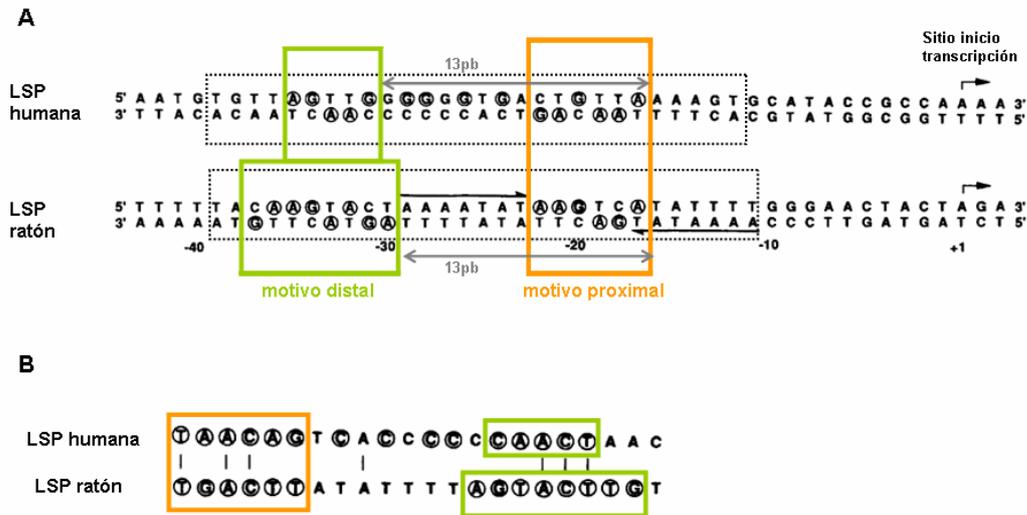
5.1.1.2. Patrón de reconocimiento de TFAM en LSP

Fisher *et al*, mapearon los contactos entre TFAM y pares de bases específicos de LSP que estabilizan el complejo proteína-ADN, tanto para humano como para ratón, usando el sistema de metilación por interferencia (que consiste en metilar purinas, en el átomo de N7 guanina y en N3 de adenina, y observar la incapacidad de TFAM para unir a los fragmentos metilados). Estos experimento revelaron que tanto hTFAM (humano) como mTFAM (ratón) contactan purinas de ambas cadenas del ADN tanto en el surco mayor como menor en la regiones definidas mediante los ensayos de protección usando *DNaseI* (Fisher, Parisi et al. 1989).

Estos experimentos muestran que hTFAM realiza extensivos contactos con el surco mayor a lo largo de todo el ADN, lo que hace pensar que se trata de un tipo de reconocimiento a lo largo de toda la secuencia, dando lugar a una forma de complejo proteína-ADN extendida. Sin embargo, en el caso de mTFAM el patrón de reconocimiento es mucho más simple, consistiendo en dos regiones de contacto situadas en los extremos de las secuencia de unión, separadas por una vuelta de ADN (~10 pb). Sin embargo, debe tenerse en cuenta que entre estos dos dominios de mTFAM no existen pares GC, lo cual es prerequisite para el mapeado de los contactos en el surco mayor mediante metilación por interferencia. Por lo tanto, estos autores concluyeron que mTFAM permanecería a lo largo del surco mayor en una forma extendida del complejo proteína-ADN, y algo similar a la que existiría para hTFAM.

En el patrón de reconocimiento por hTFAM y mTFAM es posible diferenciar dos motivos de unión en función del sitio de inicio de la transcripción; uno más proximal y más conservado, y otro más distal (Fig_C7). (Fisher, Parisi et al. 1989).

Finalmente, estos autores observaron que ambos factores contactan de forma extensiva con el surco mayor y menor en aproximadamente las mismas posiciones en relación al sitio de inicio de la transcripción, realizando contactos en más de una cara de la doble hélice y con TFAM uniéndose a lo largo del ADN de forma extendida. El hecho de que el tipo de reconocimiento o unión a ADN para hTFAM y mTFAM fuese similar, mientras que la secuencia sea divergente, sugirió que esta divergencia no implicaba grandes cambios estructurales en los motivos de ADN (Fisher, Parisi et al. 1989).



Fig_C7: (A) Regiones promotoras para hLSP y mLSP alineadas en función del sitio de inicio de transcripción. Las cajas punteadas indican las regiones de protección mediante *footprint* análisis. Los residuos de contacto con elADN presentan un círculo. Las cajas verdes y naranja destacan los dos motivos de contacto definidos dentro de la región de protección de TFAM. (B) Alineamiento de secuencia entre las zonas de unión de TFAM a mLSP y Hlsp, donde se muestran los motivos de contacto con cajas de colores y con líneas verticales, la conservación entre las secuencias. (Fisher, et al. 1987)

5.1.2. Región de protección de TFAM en las secuencias X e Y

Además de las regiones de unión localizadas en HSP1 y LSP, fueron identificados dos sitios de unión adicionales para TFAM mediante análisis de *footprinting* mediante *DNasa I*. La protección llevada a cabo por TFAM se extendía aproximadamente desde la posición 230 a 290, esta región considerablemente mayor a las zonas proximales de HSP1 y LSP puede ser dividida en dos subzonas que presentan cierta homología a LSP y HSP1 (Fisher, Topper et al. 1987) (Fig_C8). Por lo tanto, existen cuatro regiones de unión específica para TFAM situadas dentro D-loop: **H** para la zona proximal de HSP1 (523-551), **L** para la zona proximal de LSP (418-445), **X** de la posición 233 a 260 e **Y** de la posición 276 a 303.

Mirando con más detalles la localización de las secuencias X e Y podemos observar que éstas se encuentran parcialmente incluidas dentro de CSBII y CSBI, respectivamente (Fig_C8). Aunque no existe información funcional en relación a estas secuencias X y Y, recientemente algunos experimentos indican la relación entre CSBII (donde la secuencia Y se encuentra parcialmente localizada) y la terminación prematura de la transcripción para LSP, que da lugar a la formación de encebadores para la replicación de HSP desde O_H (situado también en D-loop) (comentado anteriormente en el capítulo transcripción mitocondrial).

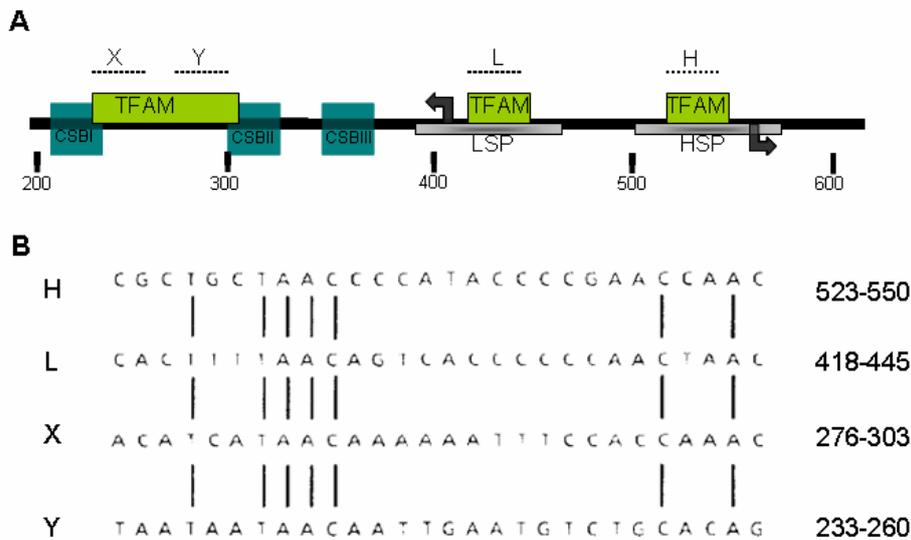


Fig.C8. Esquema de los sitios de unión para TFAM dentro del D-loop observados mediante ensayos de protección con *DNasal*. (A) Ubicación relativa de los sitios de unión de TFAM dentro del D-loop con las regiones de inicio de transcripción para la cadena H y L, así como las regiones conservadas CSBII y CSBIII. (B) Comparación de secuencia de las diferentes regiones de unión a TFAM. (Fisher, et al. 1987)

La comparación de los sitios de unión de ortólogos para TFAM permite establecer similitud entre secuencias: L: $94\pm 2\%$, H: $85\pm 7\%$, X: $79\pm 15\%$ e Y: $61\pm 14\%$. Los sitios H y L no presentan deleciones ni varianza entre las diferentes secuencias de ortólogos; mientras que los sitios X e Y presentan deleciones y una menor similitud (Foran, Hixson et al. 1988). El análisis de estos resultados permite sugerir que los sitios H y L probablemente presentan una función similar en chimpancés, gorilas y humanos; mientras que las secuencias X e Y, si presentasen función, sería bajo unas condiciones restrictivas mucho más relajadas que para H y L.

5.2. TFAM: unión inespecífica de secuencia

Inicialmente se demostró que TFAM podía unir de forma específica a mtADN de -12 a -19 pb respecto al sitio de inicio de la transcripción para HSP1 y LSP. Posteriormente se demostró que TFAM puede contactar con el ADN de forma no específica de secuencia (Fisher, Parisi et al. 1989), así como la capacidad de curvar y compactar el ADN (Fisher, Lisowsky et al. 1992).

La capacidad de TFAM para unir ADN doble cadena de forma inespecífica con una elevada afinidad no podía ser detectada mediante ensayos de protección con *DNasal*. Estudios de AFM demostraron que TFAM une aproximadamente cada 35pb, independientemente de la secuencia, por lo tanto, ello sugería su relevancia en el empaquetamiento del mtADN (Kaufman, Durisic et al. 2007).

Además, diversas técnicas permitieron concluir que TFAM unía de forma cooperativa al ADN. Mediante EMSA, la movilidad electroforética del complejo presenta un doble retardo al aumentar la concentración de proteína. Mediante imágenes de AFM, la cuantificación de TFAM

en forma unida a ADN generaba una curva sigmoideal, correspondiente a una primera unión de TFAM a ADN, una segunda fase donde la unión se da de forma exponencial y finalmente una fase plateau; experimentos adicionales de SPR dilucidaron la existencia de cooperatividad positiva en respuesta a la concentración de proteína (Kaufman, Durisic et al. 2007). Kaufman *et al*, sugirió dos hipótesis para explicar la unión cooperativa de TFAM. La cooperatividad podía ser debida gracias a la interacción entre proteínas unidas al ADN, o debida a cambios en la estructura del ADN proximal.

5.2.1. Unión a la región D-loop

Además de su unión inespecífica por todo el mtADN, TFAM presenta una flexibilidad en el reconocimiento de secuencias de mtADN en el D-loop, dependiendo de las conetraciones de saturación. Mediante análisis *footprint in organello* demostraron que existía un gran nivel de interacción de TFAM dentro de la región de control D-loop, en una extensión de alrededor 500 pb. Comparando estos resultados con estudios de protección *in vitro* mediante *DNasal* indicaron que este tipo de interacciones proteína-ADN eran debidas a la unión no especifica de secuencia realizada por TFAM en la región control del mtADN (Ghivizzani, Madsen et al. 1994).

Además de las regiones de unión inespecífica, Ghivizzani *et al*, definieron regiones de baja afinidad para la unión de TFAM a D-loop con un cierto nivel de protección a metilación del ADN (nunca 100%) y protección a *DNasal*. Estas regiones se encuentran situadas adyacentes a CSBI, CSBIII; y presentan extensiones de 70pb y 30pb respectivamente. Cabe destacar que se observaron *in organello* numerosos sitios de hipermetilación en la cadena L situados entre HSP y CSBI. Ello sugiere alteraciones específicas en la conformación de ADN inducido por la unión de TFAM a ADN, ya que ésta presenta capacidad para doblar y distorsionar el ADN uniéndose al surco menor y mayor del ADN.

Este modelo de unión en fase o "*Phased TFAM DNA binding*" observado cerca de los elemento CSB *in vitro* ha sido propuesto como consecuencia de la organización de TFAM en mtADN para facilitar la interacción de ortos elementos en trans y cis (Ghivizzani, Madsen et al. 1994).

5.3. TFAM: unión a otro tipo de ADN

TFAM es capaz de unirse a estructuras como ADN cruciformes y reconocer ADN dañado con cisplatino; seguramente esta capacidad viene conferida gracias a que forma parte de la familia de proteínas HMGbox (Yoshida, Izumi et al. 2002).

5.3.1. Unión de TFAM a ADN dañado

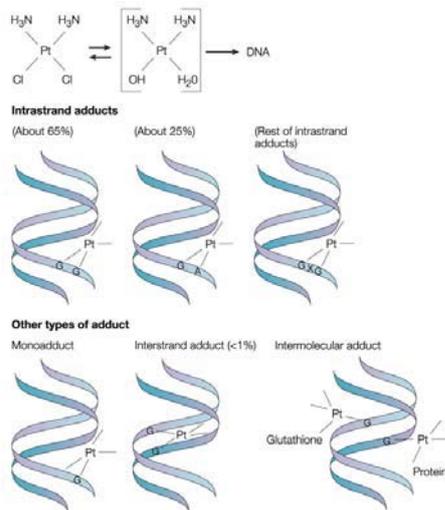
El mtADN es mucho más susceptible a daños que el nADN, debido a la falta de un empaquetamiento supraestructural como la cromatina. Esto sugiere que los mecanismos de reparación en mitocondria deberían ser más efectivos que los existentes en núcleo, ya que el daño es mayor. Han sido identificados diversos genes específicos de mitocondria involucrados en reparación, como la glicosilasa mitocondrial homóloga a MutY de *E.coli* (mtMYH) y la

glicosilasa 8-oxoguanina mitocondrial (mtOGG1), ambas presentan actividad glicosilasa con reconocimiento para ciertos tipos de ADN's dañados (Boesch, Weber-Lotfi et al.).

La mitocondria actúa como un centro fundamental de decisión en muchos de los tipos de respuesta apoptótica. Por lo tanto, en la mitocondria existe una dualidad de respuesta entre apoptosis y mecanismos de reparación del ADN.

5.3.1.1. Unión a ADN dañado con cisplatino por TFAM

Cisplatino es un potente agente anticancerígeno usado para tratar muchos tipos de tumores sólidos. El efecto citotóxico del cisplatino se cree que es debido principalmente por su interacción con el ADN, creando entrecruzamientos en el ADN que pueden ser en cis o trans. El entrecruzamiento que ocurre con mayor frecuencia se da entre purinas vecinas, pudiendo ocurrir dentro de la misma cadena (intracadena o en cis) o entre cadenas distintas (intercadena o en trans) (Fig_C9).



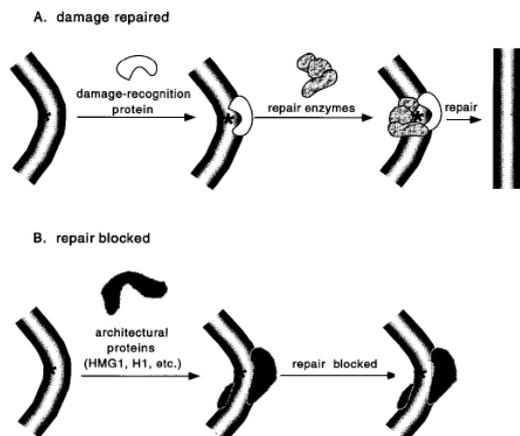
Fig_C9. El compuesto cisplatino es capaz de provocar la muerte celular debido a su interacción con ADN, por ello es usado como un potente anticancerígeno. Una vez ha tenido lugar la hidrólisis del cisplatino, éste puede interactuar con el ADN. Existen muchos tipos de complejos ADN-cisplatino. Dentro los complejo intracadena uno de los más frecuentes ocurre entre guanina adyacentes (65%) o entre una guanina y adenina (25%). También puede suceder entre dos bases situadas en la misma cadena, pero distanciadas. Estos aductos causan una perturbación en el ADN provocando la activación de los mecanismos de reparación. Cisplatino puede también crear aductos entre bases de distintas cadenas, llamados intercadena. Aunque este tipo de alteraciones son bastante poco frecuentes (sobre un 1%), se cree que son bastante citotóxicas. (Fuente: Wikipedia).

Existen ciertas proteínas que son capaces de unirse al ADN dañado por cisplatino, como por ejemplo YB-1 o proteínas que forman parte de la familia HMGB, por ejemplo, HMG1 o HMG2 (proteínas nucleares). TFAM muestra dos veces más afinidad por ADN cisplatinado que HMG1 y HMG2 (Yoshida, Izumi et al. 2002). Mediante electroforesis ha sido posible constatar que TFAM presenta una mayor afinidad por ADN cisplatinado que por ADN dañado por oxidación. Esta mayor afinidad se encuentra relacionada con el grado de daño en el ADN, ya que el ADN

cisplatinado presenta 4-5 puntos de entrecruzamiento por el platino por oligonucleótido, mientras que el ADN oxidativamente dañado en el estudio presenta un 8-oxo-dG por oligonucleótido (Yoshida, Izumi et al. 2002).

Las proteínas que reconocen ADN cisplatinado de forma preferente pueden ser clasificadas en dos categorías (Fig_C10). Por una parte, proteínas que participan en mecanismos de reparación del ADN y reconocen de forma específica los sitios dañados, es decir, funcionando como primer paso en el proceso de reparación. Por otra parte existen proteínas que pertenecen a la clase de proteínas “arquitecturales”, es decir, proteínas que se unen a este tipo de ADN por similitud estructural con sus sitios de unión naturales. Dentro de esta segunda clase encontraríamos las proteínas perteneciente a la familia HMGbox, y concretamente, TFAM. Se ha descrito que estas proteínas inducen una curvatura en el ADN posterior al provocado por la modificación con cisplatino. El hecho que TFAM pueda unir de forma preferente a ADN dañado por oxidación proporciona información sobre como funciona la mitocondria (Zlatanova, Yaneva et al. 1998).

Estas dos clases de proteínas podrían estar involucradas en procesos de citotoxicidad, aunque en papeles opuestos. Las proteínas que reconocen de forma específica el ADN cisplatinado son clave para alterar la actividad anticancer de la terapia con cisplatino.



Fig_C10. Representación esquemática de los dos modos de acción de las proteínas que reconocen de forma preferente ADN cisplatinado. (A) Proteínas que reconocen específicamente los lugares con ADN cisplatinado y permiten la actuación de los mecanismos de reparación. (B) La unión de proteínas arquitecturales como TFAM interfieren en los mecanismos de reparación bloqueando la unión de los sistemas de reparación. (Zlatanova, et al. 1998).

5.3.1.2. Unión a ADN oxidado por TFAM

Las células eucariotas producen radicales libres o especies reactivas de oxígeno (ROS) debido a pequeños desacoples de la CTE. La acumulación de ROS provoca daños de tipo oxidativo en el mtADN que están relacionados con procesos de envejecimiento, cáncer y otro tipo de enfermedades humanas. Una de las formas oxidadas de ADN más comunes se da en la base

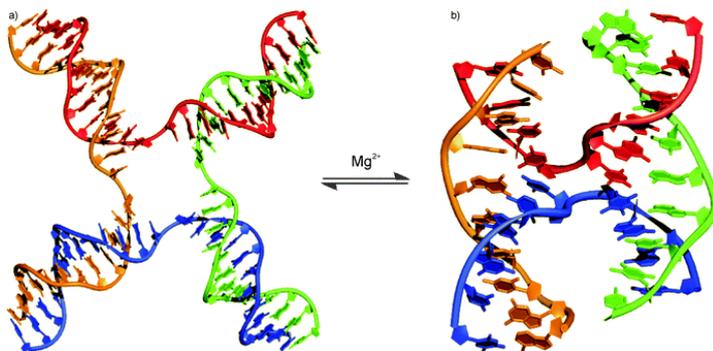
guanina, 8-oxo-7,8-dihydroguanina (8-oxo-dG). Este 8-oxo-dG es una de las alteraciones más abundantes en el ADN, provocadas por la exposición a ROS (Kasai, Crain et al. 1986).

TFAM es capaz de reconocer de forma preferente ADN oxidado. La afinidad de unión de TFAM a ADN oxidado es mayor que para la glicosilasa mitocondrial mtMYH. (Yoshida, Izumi et al. 2002).

5.3.2. Unión de TFAM a ADN cruciforme

Las estructuras tipo Holliday o *Four-way Junction* son estructuras cruciformes debido a la interacción móvil de cuatro cadenas de ADN, ampliamente conservadas de bacterias a mamíferos. Estas estructuras son llamadas así, ya que fue Robin Holliday quien las descubrió en 1964, observó estas estructuras mientras estudiaba los procesos de recombinación homóloga en levadura.

Mediante estudios con rayo X y RMN ha sido posible estudiar la estructura tridimensional de *four-way junction*. Existen dos posibles formas (Fig_C11); la forma extendida con las hebras en posición antiparalela, o la forma con las hebras en posición paralela o forma de X



Fig_C11. Estructuras de ADN que representan four-way junction (A) es la forma extendida, no estable, y (B) es la forma planar X. En presencia de la suficiente cantidad de cationes divalentes, adquiere la forma X. PDB códigos; 1C7Y y 3IGT. (Khuu, et al. 2009).

Algunas proteínas HMGB presentan una gran afinidad de unión para *four-way junction* que mimetizan a las estructuras Holliday u horquillas de recombinación, que son intermediarios putativos de la recombinación. También se sabe que la recombinación del ADN juega un papel crítico en la replicación del mtADN y su reparación. Es sabido que en levadura una buena función de la recombinación es esencial para mantener el mtADN. Recientemente se ha detectado actividad recombinatoria en extractos de mitocondria de rata, pese a que la existencia de recombinación en mitocondria de mamíferos es aún un hecho de controversia y discusión. Además se han descrito estructuras tipo Holliday en el D-loop, cercano al origen de replicación. Por lo tanto, parece ser que este tipo de estructuras intermediarias de la recombinación formarían un papel importante en el mantenimiento del mtADN en mamífero.

TFAM une a estructuras tipo four-way junction con aproximadamente 10 veces más afinidad que a su correspondiente ADN linear y prácticamente con la misma afinidad que a sus secuencia de unión en LSP de 30 nt. TFAM se une como monómero a estas estructuras y ambos dominios HMGbox presentan capacidad para unirse (Ohno, Umeda et al. 2000).

Ello hace plantear la posibilidad que este tipo de estructuras sean fisiológicamente importantes en la biogénesis mitocondrial. Particularmente en levadura, las reacciones de recombinación son importantes en el mantenimiento del mtADN y la proteína ABF2p, homóloga de TFAM, estabiliza las horquillas de recombinación o estructuras Holliday intermediarias *in vivo*.

Por lo tanto, la unión de TFAM a intermediarios de la recombinación como son las estructuras cruciformes podría estar involucrado en los procesos de recombinación implicados en el mantenimiento del mtADN.

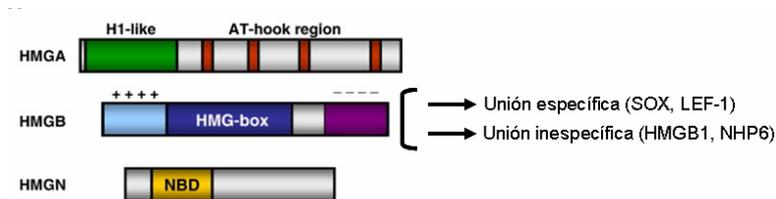
6. HMGbox

TFAM presenta dos dominios HMGbox. La principal característica de un dominio HMGbox es su interacción con el surco menor del ADN y su capacidad para distorsionar la estructura del ADN de forma pronunciada después de la unión (Grosschedl_1994).

Las proteínas HMGbox son, después de las histonas, la segunda familia de proteínas cromosomales más abundantes. En contraste con las histonas, se encuentran en cantidades variables y se cree que unen a un 10% de nucleosomas. Hace más de 50 años fueron identificadas como contaminantes de preparaciones de histona H1.

6.1. Clasificación HMGbox

Dentro de la superfamilia de HMG existen tres diferentes clases estructurales, en función de su masa molecular, característica de unión a ADN y secuencia aminoacídica en los motivos HMG (Fig_C12): *HMG-nucleosome-binding family* (HMGN), *HMG-AT-hook family* (HMGA) y *HMG-box family* (HMGB). Dentro de la familia HMGB podemos encontrar dos tipos de proteínas: HMGB de unión inespecífica y HMGB de unión específica.



Fig_C12. Representación esquemática de los diferentes grupos y subgrupos dentro de la familia de proteínas HMG.

En general, las proteínas HMGbox de unión inespecífica a ADN presentan dos dominios HMGbox y una cola C-terminal ácida (ej: HMGB1), mientras que las proteínas HMGbox de

unión específica presentan un tipo de unión específica con un solo dominio HMGbox y una cola C-terminal no acídica (ej: como los factores determinantes de sexo SRY y SOX).

También existen proteínas tipo HMGB que presentan más de dos dominios HMGbox, sería el caso de la proteína UBF. UBF contiene seis dominios HMGbox, que interactúan con el ADN en forma de dímero. Cada dominio individual de UBF presenta propiedades funcionales diferentes para el reconocimiento del ADN e interacción proteína-proteína (Kuhn, Voit et al. 1994).

Otra diferencia notoria entre HMGB de unión específica y unión no específica, es su diferente distribución. Normalmente las proteínas de unión específica se encuentran restringidas a determinados tipos celulares, mientras que las no específicas son ubicuas. Esta restricción en la distribución puede tener una importante implicación reguladora, que ayude a las proteínas a ensamblarse en complejos nucleoproteicos sólo en tipos celulares específicos (Weir, Kraulis et al. 1993).

Otra característica fascinante de los dominios HMGbox es que presenta capacidad para reconocer estructuras irregulares del ADN, como estructuras cruciformes de ADN o ADN dañado por cisplatino (Bianchi, Beltrame et al. 1989).

El caso de TFAM es más diverso y difícil de clasificar en alguno de los dos subgrupos definidos, ya que contiene dos dominios HMGbox y una cola C-terminal básica. Además TFAM presenta la habilidad de unir de forma específica e inespecífica el ADN.

6.2. Características estructurales de dominios HMGbox

Los dominios HMGbox suelen constar con unos 80 aminoácidos y presentan tres hélices alpha (I, II, III), que constituyen el 80% de los aminoácidos del dominio.

Las tres hélices se encuentran relacionadas formando un plegamiento en forma de L, típico y conservado para todos los dominios HMGbox (Fig_C13). Las hélices I y II se encuentran en estrecho contacto formando un ángulo de 80° con la tercera hélice, más larga que las anteriores. Existen una serie de residuos aromáticos en el dominio HMGbox que permite la ordenación de la tres hélice en forma de L (frecuentemente se encuentran conservados). En la región N-terminal que tiene forma extendida y estabiliza la hélice III, se encuentra un motivo de aminoácidos conservados (Lys-Arg-Pro) (Grosschedl, Giese et al. 1994).

Es interesante notar que los residuos básicos en la superficie concava de las hélices I y II participan en la unión con el surco menor del ADN (Giese, Pagel et al. 1994).

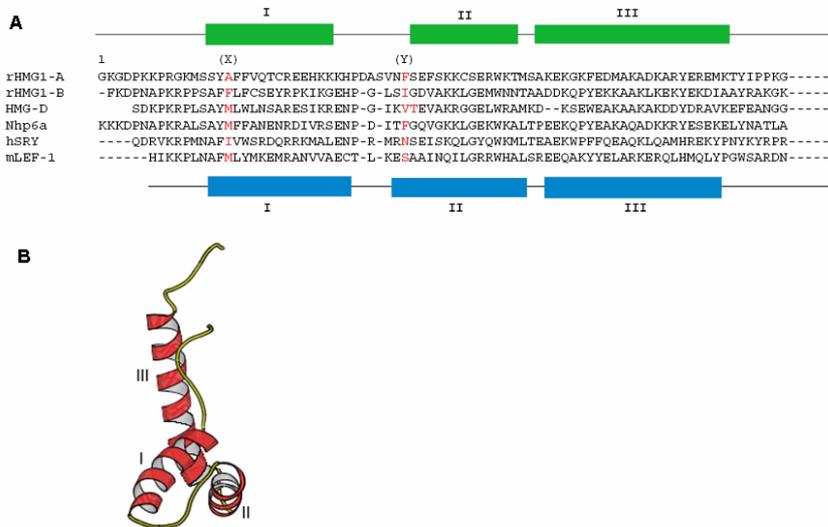


Fig. C13. (A) Alineamiento secuencial de diferentes HMGB de unión específica e inespecífica. Los lugares X e Y son las posiciones definidas para la intercalación de residuos hidrofóbicos en HMGBbox de unión no específica. (B) Plegamiento típico para los dominios HMGBbox, que consiste en tres hélices alpha.

6.3. HMGBbox inducen curvatura en el ADN

Es ampliamente sabido que HMGBbox presentan la capacidad de doblar el B-ADN lineal, mediante la intercalación de residuos hidrofóbicos en el surco menor del ADN.

La curvatura inducida por las proteínas HMGB sobre el ADN fue previamente demostrada mediante ensayos de recircularización con ligasa por la proteína procariota HMG1 (Paull, Haykinson et al. 1993). La determinación de la estructura tridimensional de complejos de dominios HMGBbox con ADN muestran que la inducción de la curvatura es debida a la intercalación de residuos hidrofóbicos voluminosos de los dominios HMGBbox entre sucesivas bases en el surco menor, acompañado por un desenrollamiento parcial (por la apertura del surco menor y curvatura hacia el lado del surco mayor). Los residuos intercalantes de los dominios HMGBbox se encuentran rodeados por residuos conservados que forman contactos electrostáticos con el ADN.

Generalmente, la curvatura del ADN en proteínas de unión inespecífica involucra múltiples residuos hidrofóbicos que intercalan dos pasos diferentes del ADN, mientras que las de unión específica sólo presentan un punto de intercalación en el ADN mediado por un único residuo (Klass, Murphy et al. 2003) (Fig. C13). La mutación de los residuos intercalantes provoca una pérdida o disminución de la curvatura y, por lo tanto, una disfunción.

6.4. Secuencias putativas de unión HMGBbox

Han sido descritas regiones consenso de ADN para el reconocimiento y unión por parte de las HMGB de unión específica. Estas regiones consenso han sido definidas a través del estudio de los sitios de unión para SRY, LEF-1, SOX; y su alineamiento (Laudet, Stehelin et al. 1993). Pese

al bajo nivel de homología ha sido posible definir un motivo de unión en el surco menor del ADN (Fig_C14).



Fig_C14. Secuencia consenso para los dominios HMGbox de unión específica. (Laudet, et al. 1993)

7. Funciones de TFAM

A continuación se describen algunas de las principales funciones de TFAM en mitocondria:

- Mantenimiento del mtADN
- Inicio de la transcripción
- Síntesis de encebadores para la replicación
- Regulación de la biogénesis del mtADN, control entre transcripción y número de copias de mtADN
- Implicación en mecanismos de reparación del mtADN

7.1. Mantenimiento del mtADN

Como se ha descrito en la introducción del genoma mitocondrial, el mantenimiento de la integridad del mtADN es importante para permitir un buen funcionamiento celular y evitar condiciones patológicas. TFAM juega un papel crucial en el mantenimiento del mtADN, siendo uno de los principales componentes.

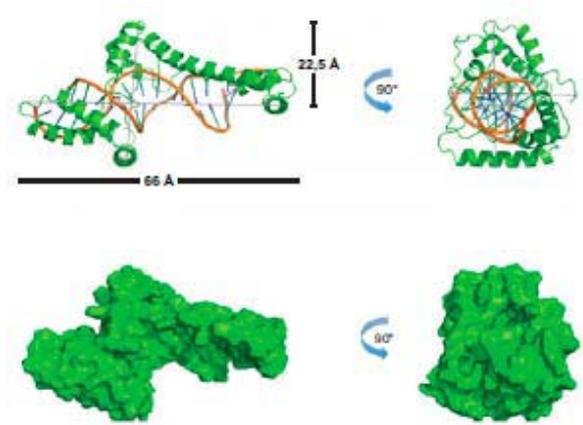
La propiedad de unión inespecífica para TFAM junto con el hecho que se encuentra en abundancia sugiere que juega un papel importante en la estabilización y el mantenimiento de mtADN. ABF2p, proteína homóloga a TFAM en levadura, tiene una función en el empaquetamiento del ADN. Aunque son muy divergentes evolutivamente, ABF2p no presenta linker entre las dos HMGbox ni una cola C-terminal, existe una intercambiabilidad de fenotipos por lo que respecta al empaquetamiento (Diffley and Stillman 1991).

Otros experimento han demostrado la función de empaquetamiento del mtADN por parte de TFAM. La anulación o disrupción homocigota del gen *tfam* es letal, en parte debido a que ello provoca una disminución del número de copias de mtADN, resultando en una pérdida de la capacidad para llevar a cabo la fosforilación oxidativa (Matsushima, Matsumura et al. 2003). En células heterocigotas, la expresión de TFAM se encuentra reducida sobre un 50% y la cantidad de mtADN disminuye hasta la mitad.

El mtADN de mamíferos se encuentra empaquetado en complejos nucleoproteicos llamados nucleoides, que pueden ser visualizados como pequeños cuerpos esféricos situados en la matriz (Nass 1969). Cada nucleoide contiene entre 1 y 10 genomas, dependiendo del tipo celular. Una de las proteínas que se encuentra en mayor cantidad en estas estructuras

nucleoproteínica es TFAM, tal y como se observó mediante estudios inmunocitoquímicos (Garrido, Griparic et al. 2003).

Recientemente, Kukat *et al.*, fueron capaces de cuantificar la cantidad de TFAM por nucleóide gracias al uso del microscopio STED en combinación con biología molecular (Kukat, Wurm et al.). Concluían que existen aproximadamente 1000 moléculas de TFAM por molécula de mtADN, es decir, una molécula de TFAM por 16.6pb de mtADN en fibroblastos humanos. Estos autores ofrecían un modelo estructural de cómo TFAM unía a mtADN mediante Pyhre & I-Tasser (Fig_C15). El modelo estructural propuesto en cuanto a volumen presentaba una correlación con el cálculo anterior de 16pb por cada TFAM.



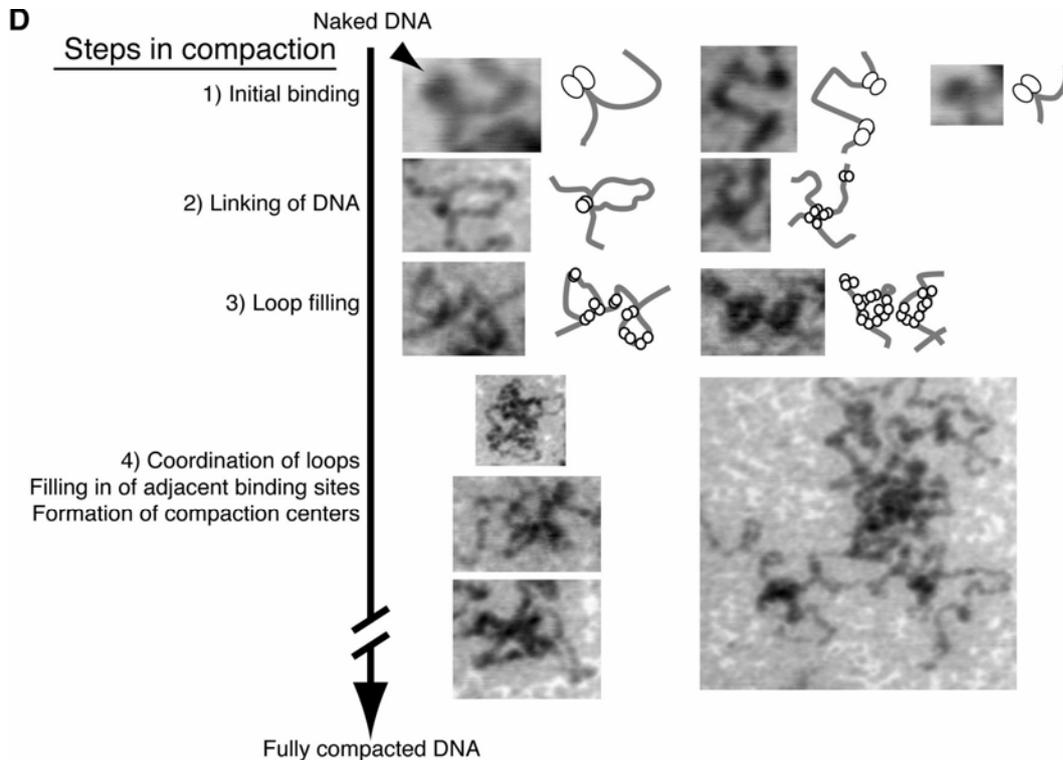
Fig_C15. Predicción de la estructura de TFAM en complejo con ADN mediante Phyre & I-Tasser. En la figura de abajo se representa la superficie de TFAM en forma unida a ADN, donde la superficie accesible a solvente se muestra en color verde. En base al volumen correspondiente los autores calcularon el número de moléculas de ADN por nucleóide. (Kukat, *et al.*).

7.1.1. Modelo de compactación y coordinación de mtADN

Existe un modelo propuesto para la compactación de mtADN por parte de TFAM en base a las observaciones mediante AFM (Kaufman, Durisic et al. 2007) (Fig_C16). En él, el proceso de compactación propuesto empieza con la unión de un dímero de TFAM y el correspondiente doblamiento del ADN. Posteriormente se produce la formación de estructuras tipo *loop* o X, y múltiples moléculas de ADN pueden ser unidas. Cuanto más TFAM se une a los sitios proximales, el tamaño del contorno del ADN decrece rápidamente, lo cuál es fácilmente compatible con el modelo de unión cooperativa al ADN. Si la unión de TFAM al ADN sigue incrementando se produce la formación de estructuras de tipo globular, donde los complejos TFAM/ADN se encuentran altamente compactados. Estas estructuras compactas serían parecidas a los nucleóides.

En la teoría propuesta por Kaufman, la formación de los *loops* forma parte de los primeros pasos hacia la compactación. *In vivo* se han observado estructuras similares en fracciones enriquecidas con nucleóides de mitocondria de hígado de rata (Van Tuyle and McPherson 1979). Además es sabido que las proteínas con dominios HMGbox presentan la capacidad de

afectar la estructura del ADN induciendo superenrollamiento, plegamiento y formación de bucles de ADN.



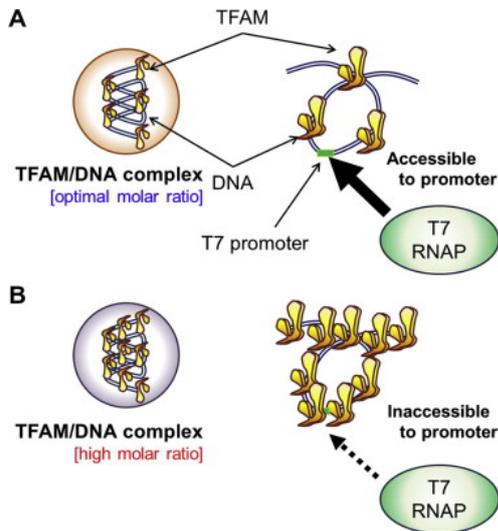
Fig_C16. Imágenes AFM, en condiciones de ultra vacío del proceso de compactación del ADN por parte de TFAM. Modelo de compactación propuesto por Kaufman et al. El primer paso sería la unión de TFAM, seguidamente se daría la formación de estructuras tipo Ω o X, que permitirían la unión de diversas moléculas de ADN. La unión de más moléculas de TFAM permitiría la compactación fina del ADN. Barra de escala de 100 nm (Kaufman et al.)

7.1.2. Equilibrio entre transcripción y empaquetamiento del mtADN

Cómo la dinámica de los nucleoides está regulada no es sabido aún. Se ha especulado que la dinámica del nucleoide mitocondrial requiere la regulación o la existencia de una elevada afinidad de TFAM con el ADN para promover la transcripción o replicación del mtADN, similar a los mecanismos que existen en la dinámica del genoma nuclear. Recientemente, se ha descrito la cantidad de complejo TFAM/ADN que permite la existencia de eventos de transcripción junto con el empaquetamiento del mtADN.

Furukawa *et al.*, han descrito cómo la forma en que el mtADN se encuentra empaquetado por el efecto de TFAM tiene un impacto en la activación de la transcripción o su inhibición (Fig_C17). Una molécula de TFAM unida cada 20pb permite la activación de la transcripción *in vitro*, mientras que el mismo ensayo con protamina no permitía la transcripción, posiblemente debido a que no ofrecía una conformación de ADN adecuada. Teniendo en cuenta esta estequiometría, es posible especular que la transcripción es activa cuando existe una óptima conformación del ADN, es decir, cuando TFAM se une a ADN en un apropiado

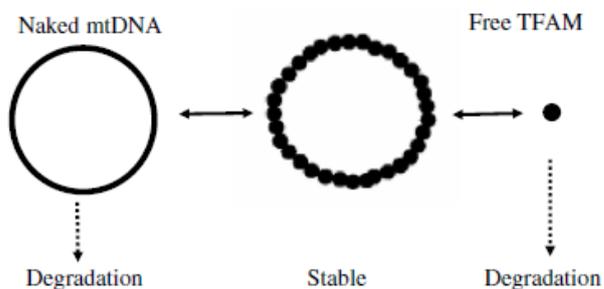
intervalo. No obstante, en relaciones molares elevadas, la transcripción del ADN empaquetado con TFAM resulta inhibida (Furukawa,2012).



Fig_C17. Ilustración esquemática del empaquetamiento del mtADN con TFAM y su actividad transcripcional. TFAM ofrece una conformación activa de ADN para los procesos de transcripción, cuando el ADN se encuentra doblado o superenrollado, mientras que cuando existe un sobre empaquetamiento del ADN por exceso de TFAM existe una inhibición de la transcripción. (Furkawa, et al.)

7.1.3. Modelo de titración de TFAM para regular el número de copias de mtADN

Mediante estudios de depleción y sobreexpresión de TFAM es posible concluir que la cantidad de TFAM es el principal determinante de la concentración de mtADN. Y el efecto inverso, cuando la cantidad de mtADN es disminuida mediante el tratamiento con bromuro de etidio, la cantidad de TFAM disminuye dependiendo de la cantidad de mtADN (Seidel-Rogol and Shadel 2002). Por lo tanto, TFAM es un factor limitante de dosis para mtADN. Kang *et al.* postularon que TFAM titraba la cantidad de mtADN en función de su unión a mtADN (Kang, Kim et al. 2007).



Fig_C18. Titración del mtADN por TFAM. TFAM se muestra como un criculo cerrado. El ADN descubierto y TFAM en su forma no unida son inestables en mitocondria y, por ello, rápidamente degradados (Kang, et al.2007).

7.2. Inicio de la transcripción y especificidad de secuencia

Durante una serie de años ha sido ampliamente discutido el papel de reconocimiento de secuencia específica por TFAM y su papel en el inicio de la transcripción. Análisis mutacionales de TFAM permitieron observar que en todos los mutantes creados para TFAM, la actividad transcripcional parecía estar correlacionada con la unión específica a ADN, sugiriendo que estas dos funciones pueden ser inseparables. Además revelaron que la cola C-terminal es importante para el reconocimiento específico del ADN y esencial para la activación de la transcripción (Dairaghi, Shadel et al. 1995).

Anteriormente, 20 años atrás, estaba ampliamente aceptado que TFAM era el principal componente necesario para el inicio de la transcripción a partir de los dos promotores HSP1 y LSP, localizados en la región del D-loop del mtADN (Chang and Clayton 1984) (Fisher and Clayton 1985) (ver Introducción, apartado B.1). La unión de TFAM específica de secuencia a las regiones situadas río arriba de los sitios de inicio de transcripción, de -12 a -39 pb del sitio de inicio, se creía que era imprescindible para la activación de la transcripción (Hixson and Clayton 1985) (Topper and Clayton 1989).

Recientemente existe una amplia discusión entorno al modelo inicio de la transcripción, como se ha descrito en la introducción, transcripción mitocondrial (ver Introducción, apartado B.3)

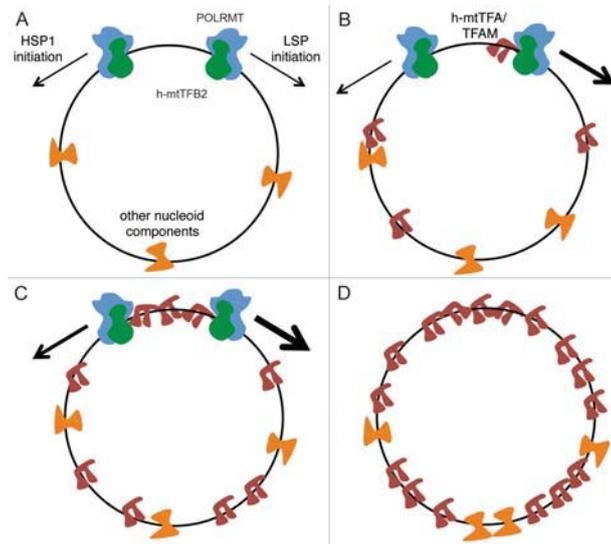
Uno de los posibles roles de TFAM en el modelo de iniciación de la transcripción puede ser el de introducir perturbaciones estructurales específicas en el mtADN, lo que puede facilitar el inicio de la transcripción (Gaspari, Falkenberg et al. 2004). Este modelo podría explicar porqué parece ser crucial la distancia entre el sitio de unión de TFAM y el sitio de inicio de la transcripción. (Dairaghi, Shadel et al. 1995). Además, también ha sido demostrado que TFAM es capaz de interactuar con TFB2M, hecho que podría conllevar a la reclutación de mtRNAP.

7.3. TFAM y síntesis de encebadores para la replicación

Como se ha comentado anteriormente la utilización del promotor varía en función de la cantidad de TFAM presente.

La cantidad de TFAM necesaria para activar el promotor HSP1, es mayor que para LSP. Además la máxima cantidad de estimulación de LSP por TFAM es significativamente mayor que para HSP1. Esta diferencia en activación de LSP y HSP1 en función de la cantidad de TFAM se relaciona con los diferentes roles de los transcritos originados desde HSP1 y LSP. La transcripción a partir de LSP además de dar lugar a un mRNA (ND6) y tRNA, también produce los encebadores necesarios para iniciar la replicación del mtADN a partir de O_H. Mientras que la transcripción a partir de HSP1 genera 2rRNAs necesarios para la biogénesis de los ribosomas mitocondriales y la traducción. Entonces, cambios en la concentración de TFAM pueden ser responsables de provocar la transcripción para dar lugar a replicación o para la expresión génica dependiendo de qué promotor está activo, LSP o HSP1 (Shutt, Bestwick et al.).

Dado que la concentración de TFAM por nucleoide no es constante, es posible definir un sistema donde la concentración de TFAM actúa como activador transcripcional para la regulación entre expresión génica, replicación y empaquetamiento del mtADN (Shutt, Bestwick et al.) (Fig_C19).



Fig_C19. Modelo de regulación de la regulación del mtADN, donde de A a D existe un aumento en la concentración de TFAM. (a) Situación la que con concentraciones de TFAM muy limitantes, HSP1 y LSP son activos dando lugar a un cierto nivel de expresión génica y replicación en nucleoides que tienen poca cantidad de TFAM asociada. (b) Un incremento de la cantidad de TFAM puede desplazar el sistema hacia un mayor replicación debido a la activación preferencial de LSP, mientras que se mantiene el nivel de transcripción de HSP1. (c) Un mayor nivel de TFAM asociada a nucleoides implicaría un máximo nivel de transcripción a partir de ambos promotores y, por tanto, una mayor expresión génica. (d) En concentraciones de TFAM mucho mayores sería de esperar una inhibición de la transcripción a partir de los dos promotores y daría lugar a nucleoides latentes en cuanto a transcripción y replicación. (Shutt, et al. 2011)

Existen evidencias científicas que corroboran este modelo de regulación de la biogénesis en mitocondria dependiente de la concentración de TFAM. Se ha observado que nucleoides con menor cantidad de TFAM son más activos en replicación (Wai, Teoli et al. 2008). Además, estudios en levadura demuestran que la composición proteica de ABF2p de los nucleoides varía en función de las necesidades en la expresión del mtADN en respuesta a cambios metabólicos (Kucej, Kucejova et al. 2008).

Actualmente aún quedan preguntas por responder como qué mecanismo regula la cantidad de TFAM en los nucleoides o si existe alguna modificación post-traduccional que regule la actividad de TFAM.

7.4. TFAM y su posible papel en el mecanismo de reparación del mtADN

El mtADN se encuentra situado en la membrana interna de la mitocondria, donde son generadas la mayoría de las especies ROS, siendo altamente susceptible al daño oxidativo. Un

posible mecanismo de protección del mtADN es el empaquetamiento en los nucleoides, asociados a la membrana.

Además la mitocondria presenta una robusta actividad de mecanismos BER, pero no mecanismos NER. Ha sido reportado que TFAM es capaz de inhibir mecanismos BER *in vitro*, aunque este mecanismo no ha sido descrito *in vivo* en mitocondria. El efecto inhibitorio de TFAM sobre los mecanismos BER es dosis-dependiente. Ello sugiere que TFAM modula la reparación del ADN mediante mecanismos de competición por unión a ADN (Canugovi, Maynard et al.)

Este efecto inhibitorio de los mecanismos NER puede influenciar en las terapias quimioterapéuticas. Las proteínas con dominios HMGbox como HMG1 y TFAM inhiben de forma específica la reparación de los entrecruzamientos 1,2-intracadena inducidos por el cisplatino por los mecanismos BER/NER (Huang, Zamble et al. 1994). El cisplatino es un medicamento usado para quimioterapia de varios tipos de cáncer. Esta inhibición por parte de HMG1 y TFAM permite sugerir que la diferente sensibilidad de tumores a cisplatino es mediada por proteínas con HMGbox como TFAM y HMG1 (Huang, Zamble et al. 1994). Estos resultados sugieren que el tipo y los niveles de proteínas HMGbox en un determinado tumor pueden influenciar la respuesta del cáncer a la quimioterapia con cisplatino y esto puede proveer bases racionales para la síntesis de nuevas drogas anticancerígenas derivadas del cisplatino.

8. TFAM y la función de su cola C-terminal

8.1. Unión específica a ADN e inicio de la transcripción

Experimentos de mutagénesis e intercambio de dominios evidenciaban que, mientras que los dominios HMGbox son necesarios para las propiedades de unión inespecífica, la cola C-terminal es absolutamente necesaria para la unión específica a ADN y su actividad en el inicio de la transcripción (Dairaghi, Shadel et al. 1995).

El papel crítico de la cola C-terminal humana de TFAM en transcripción fue confirmado mediante la construcción de proteínas quiméricas que intercambiaban regiones similares (dominios HMGbox) entre hTFAM y ABF2p. ABF2p era incapaz de activar la transcripción a partir de hLSP, mientras que la adición de la cola C-terminal confería la especificidad de reconocimiento para LSP y su actividad transcripcional. En todos los mutantes creados para hTFAM, la actividad transcripcional parecía estar correlacionada con la unión específica a ADN, sugiriendo que estas dos funciones pueden ser inseparables.

Ello relacionaba inmediatamente a la cola C-terminal como región crítica para la unión específica de secuencia y la activación de la transcripción en LSP. No obstante, la cola C-terminal es necesaria pero no es suficiente para esta función, ya que mutaciones de TFAM con la cola C-terminal pero con HMGbox1 delecionada o la isoforma Δ exon5 son incapaces de activar la transcripción (Dairaghi, Shadel et al. 1995).

8.2. Unión a TFB2M

Como se ha descrito anteriormente existe una interacción física entre la cola C-terminal y TFB2M (McCulloch and Shadel 2003) (ver Introducción, apartado C.2.2). Para testar la hipótesis que la cola C-terminal de TFAM es necesaria para la iniciación de la transcripción debido a su interacción con TFB2M, se han usado mutantes de la cola C-terminal de TFAM para realizar experimentos de unión con TFB2M (Dairaghi, Shadel et al. 1995). TFAM, con una delección de 5 aa en la cola C-terminal es capaz de interactuar con TFB2M, mientras que la pérdida de 10 aminoácidos o total de la cola C-terminal provocaba una severa o pérdida total de la interacción *in vitro* con TFB2M. Por lo tanto, parece que la cola C-terminal de TFAM sería un punto importante de interacción física y funcional con TFB2M.

8.3. Estabilización de complejo mtADN/TFAM y desenrollamiento

Existen evidencia científica que describen que la cola C-terminal dirige o fortalece la unión del complejo TFAM/ADN, posiblemente debido a que la cola C-terminal presenta 8 aminoácidos básicos (4 Arg y 4 Lys), de 25 aminoácidos totales. De forma que posiblemente la cola C-terminal contribuye a la estabilización del complejo TFAM/mtADN mediante interacciones electrostáticas (Ohgaki, Kanki et al. 2007). Mediante ensayos de afinidad con TFAM y TFAM- Δ C (construcción con la cola C-terminal delecionada según la predicción de estructura secundaria) con tres secuencias de oligonucleótidos diferentes (LSP, TAS y COXI). Ohaki *et al*, observaron que TFAM presentaba una constante de afinidad 3 ordenes menor que TFAM- Δ C. Sorprendentemente, la ausencia de cola C-terminal incrementa la disociación, sugiriendo que la cola C-terminal es importante para la estabilización de la asociación.

Además, ha sido descrito que el D-loop puede ser resuelto mediante desenrollamiento de estados de ADN supergirados y TFAM puede desenrollar el ADN. Ohgaki *et al*, compararon la capacidad para llevar a cabo el desenrollamiento de la proteína recombinante TFAM y TFAM- Δ C. Mientras que TFAM es capaz de llevar a cabo un superenrollamiento negativo del ADN en presencia de topoisomerasa, se requiere ocho veces más cantidad de TFAM- Δ C para llevar a cabo el mismo desenrollamiento. Aunque es posible que esta menor capacidad de desenrollamiento se encuentre relacionada con la pérdida de afinidad de tres ordenes de magnitud entre TFAM y TFAM- Δ C para unir a diferentes secuencias de ADN (Ohgaki, Kanki et al. 2007).

9. Interacción de TFAM con otras proteínas

Se ha definido el contacto de TFAM con diversas proteínas. En este apartado se describen algunos de los contactos con proteínas más caracterizados.

9.1. Proteasa Lon

9.1.1. Introducción

Lon es la principal proteasa en la matriz mitocondrial de eucariotas, y se encuentra ampliamente conservada entre especies. Lon es miembro de la superfamilia de ATPasa asociada a diversas actividades celulares (AAA+ATPasa), y forma una estructura de anillo homo-oligomérica. Lon contribuye al control de la calidad de proteína en mitocondria mediante la degradación preferencial de proteína oxidadas (debido a su modificación) o proteínas desplegadas antes de que suceda su agregación.

Además de su función proteolítica, Lon mitocondrial presenta capacidad de unir ADN *in vitro*, y se ha mostrado que interacciona con mtADN en cultivos celulares (Lu, Yadav et al. 2007). También ha sido demostrado que Lon es un componente de los nucleoides mitocondriales, junto con TFAM (Kucej and Butow 2007).

9.1.2. Efecto de Lon sobre la cantidad de TFAM y el números de copias mtADN

El modelo de ratón knockdown de RNAi para Lon resulta en un incremento de la abundancia de TFAM y el número de copias de mtADN. De forma inversa, una sobreexpresión de Lon reduce el nivel de TFAM y, por lo tanto, el número de copias de mtADN, ya que como se ha descrito anteriormente existe una relación dependiente de dosis entre la cantidad de TFAM y el número de mtADN. Por otra parte, la inducción de la depleción de mtADN en células knockdown para Lon no resulta en una degradación de TFAM, provocando un aumento del ratio TFAM: mtADN (Matushima_Goto_2010_PNAS).

Matushima *et al.*, concluyen que la proteasa Lon regula la transcripción mitocondrial mediante la estabilización del ratio TFAM: mtADN vía la degradación selectiva de TFAM.

9.1.3. Degradación de TFAM por Lon: reconocimiento

Se ha descrito que la degradación de TFAM es estimulada cuando existe una depleción del mtADN (Matushima_Goto_2010_PNAS). Por lo tanto, TFAM en su forma no unida a ADN es susceptible de ser degradada. Aunque el mecanismo por el cual Lon reconoce las proteínas diana aún no es demasiado conocido, recientemente ha sido publicado que Lon proteasa de *E.coli* reconoce zonas ricas en residuos aromáticos que en la estructura nativa de la proteína se encuentran escondidos en forma de núcleos hidrofóbicos, pero que son accesibles en proteínas desplegadas (Gur and Sauer 2009). Curiosamente, los dominios HMGbox de TFAM presentan 4 residuos aromáticos conservados en el núcleo hidrofóbico, y estos se encuentran parcialmente ocultos cuando TFAM se encuentra unida a ADN. Otra posible explicación es que la forma no unida a ADN de TFAM sea expuesta al estrés oxidativo, mientras que la forma unida no lo sea. Y finalmente, la tercera hipótesis estaría relacionada con la capacidad de Lon para unir a ADN y su localización en nucleoides. Sería posible que Lon detecte las grandes superestructuras de compactación que se forman cuando la cantidad de TFAM es mucho mayor, de forma que Lon proteasa reconoce este exceso y lo degrada.

9.2. Proteína supresora tumoral: p53

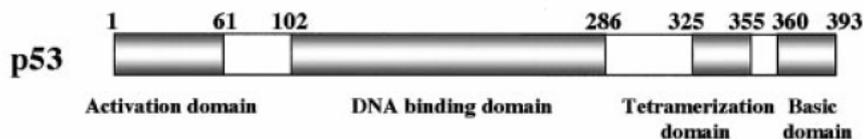
9.2.1. Introducción a p53

El gen de p53, también llamado el "guardián del genoma", se encuentra en el brazo corto del cromosoma 17 (17p13) y codifica un factor de transcripción nuclear de 43.7 kDa.

P53 resulta esencial para inducir la respuesta de la célula ante el daño del ADN, deteniendo el ciclo celular en caso de mutación. p53 es un supresor tumoral que desempeña un papel importante en apoptosis y control del ciclo celular. Está involucrado en multitud de funciones celulares como el arresto del ciclo celular, apoptosis, senescencia, daño del ADN y reparación. Pertenece a una familia de factores de transcripción, a la cual pertenecen también p63 y p73. Estas tres proteínas colaboran en una compleja red de interacciones que aún no se conoce en su totalidad. Sin embargo, p53 es ubicuo (se expresa en todos los tejidos), mientras que p63 y p73 presentan especificidad tisular.

La proteína humana p53 tiene 393 aminoácidos y presenta diversos dominios (Fig_C20) (Wong, Rajagopalan et al. 2009):

- Un dominio ácido de activación de la transcripción (TAD), también conocido como dominio de activación 1 (AD1), que activa factores de transcripción: residuos 1-42, y el dominio de activación 2 (AD2) importante para su función de activación apoptótica : residuos 55-75
- Un dominio rico en prolina importante para la actividad apoptótica; residuos 64-92
- Un dominio central de unión al ADN (DBD), contiene un átomo de zinc y diversos residuos de arginina (residuos 102-292)
- Un dominio de oligomerización (residuos 316-325). La tetramerización es esencial para la actividad de p53 *in vivo*
- Un dominio -Cterminal que se encuentra involucrado en la regulación negativa de la unión a ADN mediante el dominio central (residuos: 356-393)



Fig_C20. Dominios estructurales y funcionales presentes en p53 (Wong, et al.)

9.2.2. Interacción p53-TFAM

p53 intrínsecamente carece de estructura terciaria, aunque pre-existen hélices amfipáticas en TAD1 y TAD2.

Estudios mediante RMN mostraron que TFAM se une primeramente al dominio N-terminal de transactivación de p53, con una constante de afinidad K_d : $1.95 \pm 0.19 \mu M$. El dominio regulador C-terminal de p53 ofrece un segundo lugar de unión para TFAM. Las hélices I y II de los dominios HMGbox de TFAM constituyen la región de unión para p53. El complejo HMGbox-p53 se encuentra estabilizado por las interacciones electrostáticas de las cadenas laterales negativas de p53 y las cargas positivas de HMGbox, así como de interacciones de tipo hidrofóbico entre la superficie cóncava de p53 en N-terminal y el núcleo hidrofóbico de HMGbox (Wong, Rajagopalan et al. 2009).

La interacción de p53-TFAM *in vivo* se puede ver incrementada ante un tratamiento genotóxico (por ejemplo, cisplatino). p53 resulta fosforilada en diferentes residuos de Serina de TAD1 y TAD2, confiriendo una mayor carga negativa y, por lo tanto, favoreciendo el tipo de interacciones electrostáticas entre TFAM-p53 (Nemajerova, Erster et al. 2005).

9.2.3. Relevancia biológica de la interacción

Debido a que TFAM y p53 unen de forma preferente a ADN distorsionado, es posible que la interacción entre TFAM y p53 esté implicada en mecanismo de reparación del ADN dañado.

Existen diversas evidencias que sugieren que TFAM es un sensor de ADN dañado, ya que puede unir con una elevada afinidad a ADN dañado oxidativamente y aductos de cisplatino. Su interacción con el supresor tumoral p53 se cree que puede jugar un papel en los mecanismos de reparación de ADN, ya que esta interacción puede evitar la inhibición de los mecanismos de reparación BER por parte de TFAM (Introducción, ver apartado C.5.3).

Además, es sabido que p53 es capaz de translocarse a mitocondria como respuesta a determinadas situaciones celulares, aunque aún no es conocido el mecanismo por el cual se da la translocación (p53 no presenta MTS). Existe la teoría que mediante tratamiento genotóxico es posible que éste quede unido al transportador mitocondrial VDAC (canal dependiente de voltaje aniónico), mediante interacción con sus cisteínas y metioninas, y provoque una alteración de la permeabilidad de membrana mitocondrial (Cullen, Yang et al. 2007). Además se ha demostrado que la sobreexpresión de TFAM y el bloqueo de su translocación pueden resultar en una acumulación nuclear de la proteína (Pastukh, Shokolenko et al. 2007). TFAM nuclear fue descrita como posible proteína que confiere citoprotección contra drogas quimioterapéuticas, por ejemplo: etopósido, camptotecina y cisplatino. La localización de p53 mitocondrial y la acumulación nuclear de TFAM dan pie a pensar que existe un tráfico bidireccional de proteína después de daño en el ADN: p53 pasaría de núcleo a mitocondria, mientras que TFAM pasaría de mitocondria a núcleo).

9.3. TFB2M

9.3.1. Introducción TFB1M y TFB2M

Pese a que en un principio se creyó en la dualidad funcional de estas proteínas, posteriormente diversos estudios sugerían que las dos isoformas de TFB no tenían la misma

función. TFB1M actúa como dimetiltransferasa modificando 12S rRNA, mientras que TFB2M presenta un papel en el inicio de la transcripción (McCulloch, Seidel-Rogol et al. 2002).

9.3.2. Interacción de TFB1M y TFB2M con TFAM

Se ha demostrado que tanto TFB1M y TFB2M son capaces de interactuar directamente con TFAM, ya que es posible observar su coimmunoprecipitación en extractos de HeLa cells y su interacción directa *in vitro*. Además se ha identificado que es la cola C-terminal de TFAM la que interactúa con TFB1M/TFB2M (McCulloch and Shadel 2003).

9.3.3. Relevancia biológica de la interacción TFAM con TFB2M

En LSP, TFB2M participa en el PIC, permitiendo la formación de la burbuja de transcripción, es decir, ayuda a la apertura de la región promotora. Además, estabiliza el producto inicial de la transcripción (RNA).

Mediante ensayos *footprinting* se ha podido constatar que TFB2M no presenta capacidad de unión al ADN por sí sólo, por lo que la implicación funcional de esta interacción física entre TFAM y TFB2M puede ser permitir la llegada de TFB2M a PIC por parte del reconocimiento específico de la secuencia de TFAM.

9.4. Tiorredoxina 2

9.4.1. Introducción

Las tiorredoxinas son proteínas que actúan como antioxidantes, facilitando la reducción de otras proteínas a través de un intercambio tiol-disulfuro en la cisteína. Las tiorredoxinas se encuentran en casi todos los organismos de forma ubicua y son esenciales para la vida en los mamíferos. Es una enzima oxidoreductasa de 12 kDa que contiene un sitio activo ditiol-disulfuro. De entre los posibles sustratos *in vitro* encontramos: ribonucleasas, coriagonadotropinas, factores de coagulación, receptores de glucocorticoides e insulina.

Las tiorredoxinas se caracterizan a nivel de secuencia de aminoácidos por la presencia de dos cisteínas vecinas formando un motivo CXXC. Estas dos cisteínas son la clave para la capacidad de la tiorredoxina de reducir a otras proteínas. La tiorredoxina también posee una estructura terciaria característica denominada plegamiento tiorredoxina. Las tiorredoxinas son mantenidas en estado reducido por la flavoenzima tiorredoxina reductasa, en una reacción dependiente de NADPH. Las tiorredoxinas actúan como donantes de electrones para peroxidasas y ribonucleótido reductasa.

La tiorredoxina 2 (TRX2) de mamífero es una isoforma mitocondrial y se encuentra ampliamente conservada. En mamíferos, TRX2 se encuentra codificada en núcleo y es transportada a mitocondria mediante una secuencia de localización situada en su extremo N-terminal. Se ha demostrado que TRX2 interactúa con componentes de la CTE, juega un papel en el control del potencial de membrana y participa en la vía de señalización de la apoptosis

mitocondrial. En 2005 fue resuelta la estructura de TRX2 humana en su forma oxidada y reducida (Fig_C21). Aunque el plegamiento es muy similar entre tiorredoxina citosólica (TRX1) y tiorredoxina mitocondrial (TRX2), existen algunas diferencias en la transición de forma oxidada a reducida entre TRX1 y TRX2. Aunque existe dimerización en las dos formas, en el caso de TXN2 no existe Cys involucrada en la dimerización y ésta se da mediante contactos hidrofóbicos (Smeets, Evrard et al. 2005)



Fig_C21. Estructura cristalográfica de TRX2. (A) Estructura de la forma oxidada para TRX2, donde es posible diferenciar tres hélices alpha y cuatro hojas beta correspondientes al típico plegamiento de TRX. (B) Dímero formado por dos moléculas de TRX2 en estado oxidado.

Recientemente, se ha demostrado que la viabilidad de algunas células cancerígenas son más dependiente del sistema TRX que las células normales, sugiriendo que el sistema mitocondrial TRX es una diana terapéutica para cáncer (Zhang, Zheng et al.).

9.4.2. Interacción TRX2 con TFAM

La tiorredoxina 2 (TRX2) interacciona directamente con TFAM *in vivo* y aumenta la capacidad de unión a ADN dañado. Mediante ensayos de *pulldown* se ha visto que es HMGB1 de TFAM quien interacciona directamente con TRX2, mientras que el dominio HMGB2 no interacciona. Además, se ha reportado que la interacción de TRX2 con TFAM aumenta la afinidad de TFAM por ADN dañado de forma oxidativa o cisplatinado (Kidani, Izumi et al. 2009).

9.4.3. Relevancia biológica de la interacción

TRX2 interacciona directamente con TFAM. Los niveles de expresión para TFAM y TRX2 se ven incrementados en líneas celulares resistentes a cisplatino, y la sobreexpresión de estas proteínas hace que las células sean resistentes a este tipo de drogas. Tanto TFAM como TRX2 presentan capacidad para unir a ADN cisplatinado y dañado oxidativamente y, a su vez, presentan capacidad para interactuar con p53. Ello sugiere que la interacción TFAM/TRX2 tiene relevancia en la vía de la apoptosis mitocondrial y mecanismos ROS (Kidani, Izumi et al. 2009).

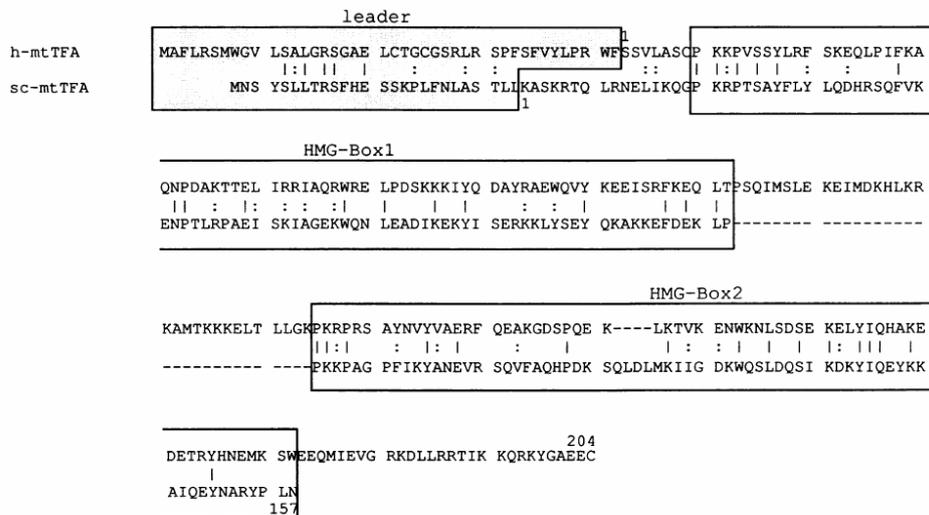
10. TFAM en otros organismos

10.1. TFAM en *Saccharomyces cerevisiae* (ABF2p)

El homólogo de TFAM para levadura, ABF2p, es una proteína abundante compuesta por dos dominios tándem HMGbox y un pequeño linker de ~6 aa. ABF2-p fue originalmente definida como *ARS-binding factor 2*, con capacidad para unirse al origen de replicación del ADN nuclear, aunque más tarde se encontró que estaba localizada en mitocondria (Diffley and Stillman 1991). La disrupción o truncamiento del gen *abf2* provoca una pérdida de la competencia respiratoria y mtADN, cuando las células crecen en glucosa. La expresión de h-TFAM en la cepa de levadura deficitaria para *abf2* mejora el fenotipo, por lo que existe una homología funcional entre h-TFAM y ABF2p.

10.1.1. Caracterización de ABF2p

Mediante un alineamiento de las secuencias para hTFAM y ABF2p, es posible observar que existe un cierto grado de divergencia (Fig_C22). Mientras que ambas presentan dos dominios HMGbox en tándem, ABF2p es deficitaria para la cola C-terminal y linker. Esta divergencia en secuencia hace pensar que existe una divergencia evolutiva que implica cambios funcionales entre las dos proteínas (con un 21% de identidad y 33% de similitud aminoacídica) (Parisi, Xu et al. 1993).



Fig_C22. Alineamiento de secuencia entre *sc*-TFAM y h-TFAM, con cajas se indican la secuencia de translocación a mitocondria y los dominios HMGbox. (Parisi, et al.)

10.1.2. Unión de ABF2p a ADN

ABF2p une a ADN de forma inespecífica, ya que puede unirse a múltiples secuencias de ADN con una afinidad similar. Además exhibe cierta especificidad, por ejemplo, se une específicamente a la secuencia de ADN llamada ARS1, secuencia de replicación autónoma. Se

trata de una región regulatoria situada en el origen de replicación del ADN nuclear de levadura. También se ha visto unión específica para la secuencia adyacente al promotor donde une la *sc-mtRNAP* (Diffley and Stillman 1991).

La diferencia en el tipo de unión a ADN entre ABF2p y TFAM, y la diferencia en la estructura secundaria invitan a pensar que los sistemas han evolucionado para adaptarse. En levadura encontramos un mtADN de 75 Kb con múltiples promotores de transcripción y orígenes de replicación, sugiriendo que es necesaria la presencia de una proteína que empaquete el ADN maximizando la accesibilidad de las principales regiones de control a la maquinaria. En contraste, en vertebrados existe un pequeño genoma mitocondrial con menos zonas de regulación y, por lo tanto, requiere menos en términos de empaquetamiento (Parisi, Xu et al. 1993).

10.1.3. Participación en transcripción de ABF2p

A diferencia de TFAM en mamíferos, ABF2p no es necesaria para iniciar la transcripción mitocondrial en levadura, aunque ABF2p puede estimular la transcripción *in vitro* de forma leve.

Como se ha comentado anteriormente ABF2p presenta función en el mantenimiento del mtADN y una leve capacidad de activar la transcripción, mientras que TFAM presenta una dualidad funcional participando tanto en la transcripción como en el mantenimiento del mtADN. Esta evolución de TFAM en un potente activador de la transcripción no parece ser debida a cambios en las propiedades bioquímicas de HMGbox, sino posiblemente debido a la incorporación de una cola C-terminal y una región conectora entre los dos dominios HMGbox en tándem. Por lo tanto, ABF2p parece tener la función de mantener el mtADN y empaquetarlo en las estructuras conocidas como nucleoides, ya que es posible reemplazar funcionalmente AB2p por HU, histone-like protein de *E.coli* (Megraw and Chae 1993).

Además existen resultados que sugieren que ABF2p se encuentra involucrada en la replicación y segregación del mtADN, quizás mediante la formación o estabilización de intermediarios de la recombinación.

10.1.4. Mecanismo de compactación de mtADN en levadura

Mediante el uso de técnicas de AFM a elevada resolución fue posible observar que ABF2p induce pronunciadas distorsiones en la conformación del ADN. Las imágenes de AFM indican curvaturas de unos $\pm 78^\circ$, que coincide con otros ángulos reportados para proteínas de la familia HMG, por ejemplo, SRY dobla unos 70° , SOX curva 83° . Sin embargo, en el caso de ABF2p contiene dos dominios HMGbox a diferencia de las otras proteínas que doblan $\pm 80^\circ$. Cuando se incrementa la cantidad o cobertura de proteína es posible observar un notable colapso de las moléculas de ADN en un denso complejo molecular. Ello sugiere que ABF2p es capaz de compactar el ADN mediante la simple introducción de curvas o doblamientos en el esqueleto de ADN (Friddle, Klare et al. 2004) (Fig_C23).

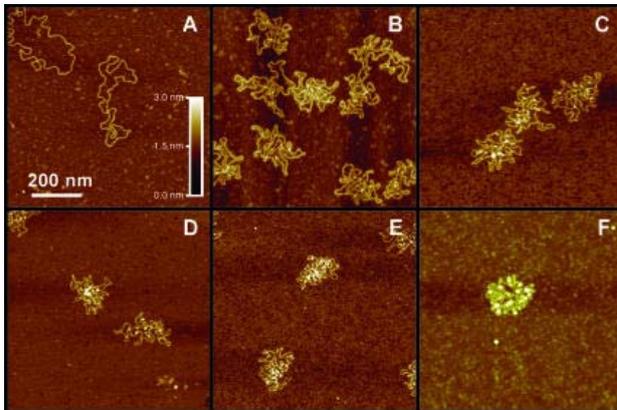


Fig. C23. Imágenes de AFM, donde se puede visualizar el efecto sobre la molécula de ADN linearizada pBR322 de concentraciones crecientes de ABF2p. (A) No existe ABF2p. (B) ratio 1 ABF2p: 20 ADN. (C) ratio 1:4. (D) ratio (1:2). (E) ratio 1:1. (F) imagen ampliada a partir de E. (Friddle, et al. 2004)

10.2. TFAM en *Caenorhabditis elegans* (HMG-5 o *ce*-TFAM)

HMG-5 de *Caenorhabditis elegans*, contiene dos dominios HMGbox y presenta una similitud de secuencia con hTFAM. HMG-5 tiene capacidad de unirse a ADN de forma inespecífica, ya que puede unirse tanto a mtADN humano, como mtADN de *C.elegans* y ADN de un plásmido cualquiera. Además, HMG-5 puede interactuar con ella misma. HMG-5 presenta la capacidad de actuar como factor de empaquetamiento (Sumitani, Kasashima et al.).

10.2.1. Caracterización de HMG-5

HMG-5 presenta una secuencia con una identidad del 23.3% y una similitud de 42.3% con hTFAM. HMG-5 presenta los dos dominios en tandem HMGbox y entre ellos un linker de unos 27 aminoácidos. Cabe destacar la casi total ausencia de la cola C-terminal, una cola de 10 aminoácidos al contrario de hTFAM donde son 25 residuos (Sumitani, Kasashima et al.).

Al igual que TFAM de mamífero puede interactuar formando homodímeros, lo mismo se ha descrito para HMG-5.

10.2.2. Unión de HMG-5 a ADN

HMG-5 se encuentra en la fracción nucleoide y se ha visto que une a mtADN de mamífero. HMG-5 fue originalmente identificada como una proteína de unión ADN telomérico de doble cadena (Im and Lee 2003). HMG-5 se expresa en el núcleo de oócitos y células embrionarias, pero no en adultos. HMG-5 se encuentra principalmente localizada en los extremos de los cromosomas, indicando que HMG-5 se une a telómeros *in vivo*. Estas observaciones sugieren que HMG-5 puede participar en los primeros estadios de la embriogénesis interactuando con los telómeros (Im and Lee 2003).

HMG-5 no sólo une secuencias teloméricas de *C.elegans* compuestas por TTAGGC, sino que también une a secuencias teloméricas humanas y de plantas, mtADN y el plásmido pUC19, haciendo gala de su capacidad de unir ADN de forma inespecífica.

10.3. TFAM en *Xenopus laevis* (xl-TFAM)

La transcripción mitocondrial ha sido extensamente estudiada tanto en sistema de mamíferos como en anfibios. La transcripción en *Xenopus laevis* puede suplir el espacio ausente de conocimiento entre la transcripción en levadura y mamífero (Antoshechkin and Bogenhagen 1995).

xl-TFAM es capaz de regular la eficiencia de utilización de los promotores mitocondriales *in vitro*, aunque no es capaz de reemplazar xl-TFB. Antoshechkin *et al*, observaron que xl-TFAM no era un componente esencial para la maquinaria de transcripción basal, aunque puede estimular significativamente la transcripción específica de promotor. Cabe destacar el hecho que los autores también observan una diferencia en el grado de estimulación de la transcripción por xl-TFAM dependiendo del promotor.

10.3.1. Caracterización de xl-TFAM

xl-TFAM presenta dos dominios HMGbox en tándem, con un alto grado de similitud o homología con h-TFAM. Al contrario del caso de ABF2p que presenta solo unos pocos residuos entre el tándem de HMGbox, las proteínas de vertebrado (xl-TFAM y h-TFAM) presentan también dos regiones altamente conservadas: una de ellas es la región situada entre las dos cajas HMGbox y la segunda hace referencia a la cola C-terminal. Experimentos previos sugieren que estas dos zonas altamente conservadas, linker y cola C-terminal, están involucradas en el reconocimiento específico de ADN y la estimulación de la transcripción.

10.3.2. Unión de xl-TFAM a ADN

xl-TFAM puede unir de forma inespecífica al mtADN, realizando su función de empaquetamiento del genoma mitocondrial, así como unir de forma específica.

Estudios de footprinting con *DNaseI* permitieron identificar los sitios de unión de xl-TFAM dentro de la zona de control cerca de los principales promotores de la transcripción. xl-TFAM cubre cinco regiones del D-loop. Tres de ellas son adyacentes a los promotores mitocondriales y las otras dos se encuentran cercanas a CSBII. Estos resultados de footprinting son similares a los de h-TFAM. La transcripción bidireccional en *Xenopus laevis*, probablemente debido a la unión de xl-TFAM como tetrámero, que permite interacciones simétricas con otros componentes de la maquinaria de transcripción en los promotores adyacentes (Antoshechkin, Bogenhagen et al. 1997).

11. TFAM y su relación en enfermedades

11.1. TFAM y estrés oxidativo

Modelos *in vitro* e *in vivo* han demostrado de forma directa e indirecta, mediante la protección frente a estrés oxidativo y apoyo de funciones mitocondriales respiratorias, la participación de TFAM en enfermedades o procesos relacionados con el estrés oxidativo y el envejecimiento. La inactivación condicional y selectiva de alelos *tfam* en ratones transgénicos hace que estos ratones desarrollen un fenotipo similar a la disfunción progresiva motora que sucede en Parkinson y muerte de neuronas dopaminérgicas. Este tipo de ratones muestran una reducción de la expresión de mtADN y una deficiencia de la CTE (Ekstrand, Terzioglu et al. 2007).

Por otra parte, han sido generados ratones transgénicos que sobreexpresan TFAM. Estos ratones presentan un aumento en el número de copias de mtADN en miocardio y protección del corazón frente a disfunciones cardíacas de origen mitocondrial (Matsusaka, Ikeuchi et al. 2005). Los ratones transgénicos para TFAM exhiben un mejoramiento respecto al envejecimiento, que correlacionan con la buena actividad de los complejos I y IV en cerebro, y reducción de la peroxidación de lípidos con la edad (Hayashi, Yoshida et al. 2008).

Además, se ha demostrado que la sobreexpresión de TFAM en mioblastos de rata acelera la recuperación de mtADN después del daño con peróxido (Noack, Bednarek et al. 2006).

La construcción de estos ratones transgénicos para TFAM y sus efectos demuestran que TFAM mediante su papel de mantenimiento de la integridad de mtADN, ofrece una protección frente al estrés oxidativo.

Finalmente, cabe destacar que la terapia génica basada en la transfección de TFAM provoca un marcado mejoramiento de las funciones mitocondriales con la edad (Keeney, Quigley et al. 2009).

11.2. TFAM y diabetes

La expresión génica de TFAM resulta alterada por las condiciones celulares y puede ser directamente regulada por diferentes factores de transcripción como los factores nucleares respiratorios (NRF-1 y NRF-2). Recientemente, se ha demostrado el papel regulatorio de TFAM en el control del número de copias, su requerimiento para la biogénesis mitocondrial y desarrollo embrionario en ratones knockout. Sin embargo, aún no es demasiado conocida la regulación de la expresión de TFAM.

La metilación en dinucleótidos CpG inactiva la expresión de muchos genes. El promotor de *tfam* consta de 2378 pb y contiene 67 CpG dinucleótidos, llamados islas CpG. Los sitios CpG son conocidos como sitios de metilación en el genoma, imprimiendo y regulando la expresión génica. Se ha predicho diferentes sitios de unión para NRF-1 en el promotor de TFAM y en numerosos genes mitocondriales que codifican para elementos de CTE, como citocromo C. Diversos resultados han sugerido que la metilación de los sitios de unión para NRF-1 provocan el silenciamiento de la actividad del promotor TFAM y disminuyen el número de

copias de mtADN. Además se ha podido demostrar que la expresión de TFAM queda afectada por diversos estímulos exógenos como glucosa (Choi, Kim et al. 2001).

Los autores sugieren que una exposición a glucosa, creciente y suspendida en el tiempo, provoca una metilación de los sitios reconocidos por NRF-1. Ello provoca una disminución de la expresión de TFAM y, por lo tanto, una reducción del número de copias del mtADN. Y relacionan esta reducción con el número de copias de mtADN con el hecho que células con disminución de mtADN o depleción muestran problemas en el metabolismo de la glucosa provocando un estatus diabético (Choi, Kim et al. 2001).

11.3. TFAM y obesidad

Como se ha comentado anteriormente, TFAM es una proteína codificada en el núcleo. La disminución de la expresión de proteínas mitocondriales codificadas en núcleo ha sido relacionada con el desarrollo de obesidad en ratón, sugiriendo que variantes genéticas de TFAM pueden influenciar la biogénesis mitocondrial y, en consecuencia, afectar el metabolismo energético y estimular la deposición de cuerpos grasos. Han sido identificados dos polimorfismos (SNPs, *single nucleotide polymorphisms*) en la región promotora bovina de TFAM. La presencia de estos SNPs puede conllevar a la ganancia/pérdida de seis sitios putativos de unión de factores de transcripción ligados al metabolismo energético (por ejemplo, NRF-1). Ello sugiere que el gen *tfam* juega un papel importante en el metabolismo lipídico y quizás es un importante gen candidato para determinar la obesidad en mamíferos (Zhihua,2005).

11.4. TFAM y cáncer

11.4.1. TFAM y terapias contra cáncer

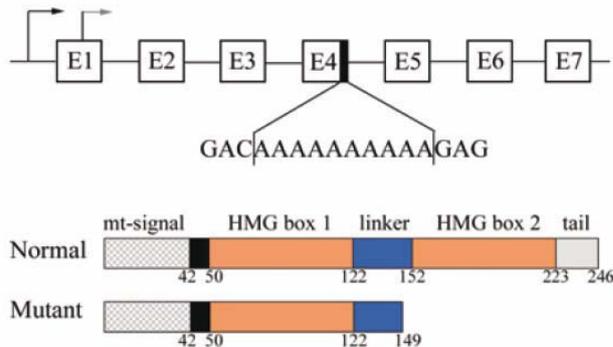
TFAM ha sido localizada en núcleo de células cancerígenas después del tratamiento con ciertas drogas. Aunque aún no está claro qué papel tiene la localización nuclear de TFAM, podría ser que TFAM ofrece algún tipo de citoprotección frente a tratamientos quimioterapéuticos. Pastukh, *et al.* definían este cambio de localización como un posible efecto de “*distribución eclíptica*”, fenómeno en el que existe una distribución irregular de una determinada proteína entre dos o más compartimentos celulares, donde la acumulación de la proteína en un compartimento impide su detección en otro (Pastukh, Shokolenko et al. 2007).

11.4.2. Cáncer colorectal por inestabilidad en microsatélite

El cáncer colorectal (CRC) se puede mostrar por inestabilidad genómica de dos tipos: inestabilidad cromosómica e inestabilidad en microsatélites. La inestabilidad por microsatélites se caracteriza por inactivación de proteínas involucradas en mecanismos de reparación del ADN. Como consecuencia, los tumores por inestabilidad en microsatélites presentan una acumulación de alteraciones genéticas, tanto en las repeticiones de microsatélites codificantes como en las no codificantes.

Una de las regiones microsatélites potencial diana de mutaciones en TFAM es la repetición poli(A)₁₀, en el exón 4 de *tfam*. La mutación en este punto implica la interrupción de la pauta de lectura. La mayoría de veces se muestra en forma de una deleción heterocigota de 1 base

en la región poli(A)10. Esta mutación en la pauta de lectura del gen para TFAM da lugar a una proteína truncada en el codón 149 (Leu149), dando lugar a una delección del segundo dominio HMGbox2 y la cola C-terminal (Fig_C24).



Fig_C24. Esquema de los exones del gen *tfam*, donde en el exón 4 encontramos poli(A)10 susceptible de mutación. Debajo aparecen los dominios para TFAM salvaje y para la forma truncada de TFAM. (Guo, *et al*)

La presencia de esta forma truncada para TFAM en células de cáncer colorectal provoca una disminución del nivel de proteína salvaje tanto *in vivo* como *in vitro*. Esta disminución en la cantidad de TFAM provoca una disminución de mtADN y, por lo tanto, una inestabilidad mitocondrial. Además esta forma truncada de TFAM promueve el crecimiento de células de cáncer colorectal, es decir, promueve la tumorigénesis de este tipo de células. A diferencia de lo que sucede en TFAM salvaje, que precisamente inhibe la tumorigénesis. También se ha visto que esta forma truncada confiere resistencia a este tipo de células para la apoptosis inducida por tratamiento con cisplatino (Guo, Zheng *et al.*).

11.5. TFAM y enfermedades neurodegenerativas

Las disfunciones mitocondriales han sido relacionadas con enfermedades neurodegenerativas como Parkinson (PD) y Alzheimer (AD). Análisis *post-mortem* sugieren una reducción del metabolismo del oxígeno y glucosa tanto en PD como AD. Por lo tanto, una reducción de las funciones mitocondriales (por ejemplo a través de la pérdida de integridad de mtADN debido a la disfunción de TFAM) puede conllevar a una situación neurodegenerativa progresiva y crónica como sucede en PD y AD.

Algunas variantes genéticas para mtADN (haplogrupos de mtADN) han sido relacionadas en la modulación del riesgo a sufrir estas enfermedades. Por ejemplo, se ha detectado el efecto protector del haplogrupo J y K frente a Parkinson, mientras que haplogrupos como H han sido asociados con un incremento al riesgo de sufrir Alzheimer.

Debido a los resultados derivados del diseño de ratones transgénicos para TFAM, se promovió el estudio de las variantes o polimorfismos presentes en el gen *tfam* y su relación con enfermedades. El gen humano *tfam* contiene siete exones y seis intrones y cubre 11.7Kb de la región cromosómica 10q21.1. Dentro del gen de TFAM encontramos diferentes variantes o polimorfismos situados en intrones y exones. Las variantes (SNPs) más citadas y relacionadas con enfermedades para TFAM son rs1937 (exón1, posición 167, G>C, Ser12Thr) y rs2306604 (intrón 4, IVS4+113 A>G) (Belin, Bjork *et al.* 2007). En rs1937 normalmente existe una G que

da lugar al residuo Ser12, pero también existe la variante donde existe una C que da lugar al residuo Thr (Reyes, Mezzina et al. 2002). Mientras que en rs2306604 puede existir un cambio en el intrón 4 de A a G, però ello no provoca ninguna modificación directa en la proteína codificada. Aunque no existen implicaciones funcionales significativas de estos polimorfismos, han sido relacionados estadísticamente con AD y PD.

11.5.1. Enfermedad Parkinson (PD) y TFAM

Mediante estudios de genotipado y fenotipo de poblaciones control y poblaciones con Parkinson, ha sido posible detectar que la variante intrónica rs2306604 modula el riesgo a sufrir Parkinson. Los análisis de genotipado revelaron que el genotipo rs2306604 (G/G vs. A/A+A/G) eran más frecuentes en pacientes con Parkinson que en control (Gaweda-Walerych, Safranow et al.)

11.5.2. Enfermedad Alzheimer (AD) y TFAM

Diversos estudios han demostrado la relación entre AD y los cromosomas 9, 10 y 12. Interesantemente la región 10q, que coincide con el gen para TFAM, ha sido ampliamente estudiada debido a su relación con AD.

Se ha descrito una leve asociación de la variante rs2306604 (G) con la enfermedad de Alzheimer. De hecho se ha sugerido que el alelo A puede potenciar el riesgo a sufrir Alzheimer (Belin, Bjork et al. 2007).

Además, se ha detectado relación entre el polimorfismo rs1937 y LOAD (Late-onset Alzheimer's). La presencia del alelo C tiene un papel protector frente al desarrollo de LOAD, cuando se compara con la presencia del principal alelo G. Ello puede ser interpretado también a la inversa, creyendo que la presencia del alelo G presenta una mayor riesgo a sufrir LOAD (Zhang, Yu et al.) (Alvarez, Corao et al. 2008).

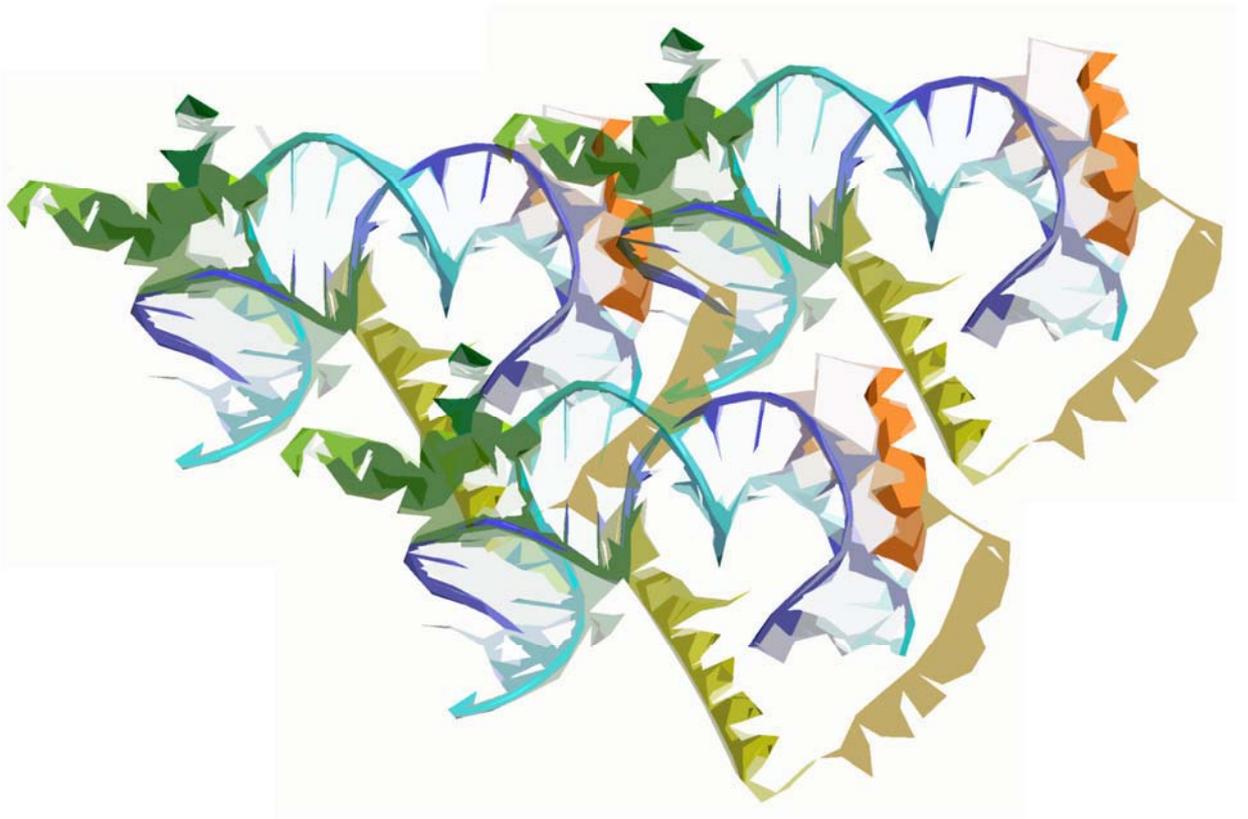
11.6. TFAM e insuficiencia cardíaca

La insuficiencia cardíaca es una afección compleja caracterizada por la incapacidad del corazón para responder a un gasto cardíaco acorde con la demanda metabólica del resto de los tejidos. Esta enfermedad representa el estado terminal de la cardiopatía isquémica, valvular o hipertensiva. Recientes hallazgos indican que cambios en la morfología mitocondrial podrían ser relevantes en la fisiopatología cardiovascular, pues influyen en la capacidad metabólica celular (Parra, Verdejo et al.). Tejidos que presentan una alta demanda energética, como el corazón y el músculo esquelético, tienden a presentar una red mitocondrial fusionada, mientras que aquellos con baja demanda energética, como el hígado, presentan una red más fisionada.

El ratón con la delección del factor mitocondrial TFAM desencadena una miocardiopatía embrionaria letal, mientras que en los ratones heterocigotos se desarrolla miocardiopatía dilatada severa, y son viables sólo hasta los 20 días de desarrollo (Larsson, Wang et al. 1998). Además mediante estudios genéticos de pacientes con hipertrofia cardíaca, ha sido posible identificar 2 variantes genéticas para TFAM, que no están presentes en la población control. Se trata de un cambio situado en la región no codificante -91 (5'UTR) (C>A), es una región

que contiene sitios de unión para factores de transcripción como NRF. La segunda variante consiste en una mutación sin sentido en el exón 4, A105T. Este aminoácido se encuentra en el primer dominio HMGbox, concretamente en la Hélice 3, posiblemente el cambio de Ala por Thr puede afectar la unión de TFAM a mtADN. Estas dos variantes tan solo se identificaron en uno de los pacientes y se trataba de un caso esporádico de hipertrofia cardíaca, con lo cual es necesario un estudio con un mayor número de pacientes, para evitar limitaciones (Alonso-Montes, Castro et al. 2008).

RESULTADOS



A. RESULTADOS: CLONACIÓN, EXPRESIÓN, PURIFICACIÓN y CARACTERIZACIÓN DE TFAM

En este apartado se detallará el trabajo realizado de clonación, expresión, purificación y caracterización de TFAM y diversos constructos.

1. Análisis bioinformático

La predicción de la estructura secundaria de TFAM en su forma madura (Ser43-Cys246), cuya secuencia señal para mitocondria ha sido procesada mediante el programa *Psipred* (McGuffin, Bryson et al. 2000) permite observar que se trata de una proteína con alto contenido en hélices alpha (Fig_A1), ya que como se comentó en la Introducción, TFAM presenta dos dominios HMGbox separados por un *linker*. Estos dominios HMGbox, normalmente formados por unos 80 residuos, consisten en una región N-terminal elongada seguida de tres hélices alpha, dos pequeñas y la última larga, que se empaquetan en forma de bota o L. Además de los dominios HMGbox de TFAM, encontramos un *linker* (~25 residuos) y una cola C-terminal (~27 residuos).

La predicción de la estructura secundaria es importante para abordar la clonación de la proteína, debido a que es importante mantener las regiones con conformación secundaria constantes, para evitar una posible desestructuración de la proteína en el momento de expresarla.



Fig_A1. Predicción de la estructura secundaria de TFAM en base al programa *PsiPred*, donde se indica el tipo de conformación predicha y la confianza de la predicción para cada residuo. También se encuentran indicadas las hélices de los dominios HMGbox. (Lobley, Sadowski et al. 2009).

2. Clonación

2.1. TFAM-HisTag, MBP-TFAM

La proteína madura (43-246 aminoácidos, sin la señal de translocación a mitocondria) fue clonada mediante las enzimas de restricción *Nco I* y *Xho I* (*Biolabs*) en el vector de expresión PET28b (+) (*Novagen*) a partir del cDNA de TFAM cedido por el Dr. H.Jacobs (*FinMIT*, Tampere, Finlandia). La clonación en este vector permite introducir una cola de poli-Histidinas (HisTag) en el extremo C-terminal, formada por 6 Histidinas consecutivas permitiendo así la purificación de TFAM mediante una columna de afinidad cargada con cationes.

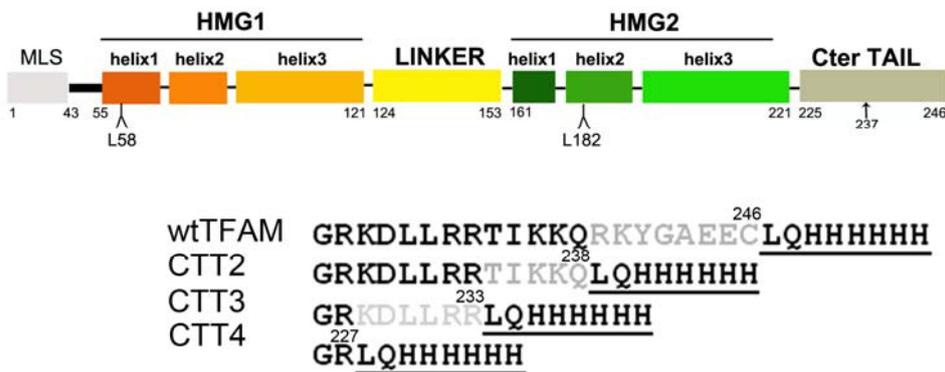
Así mismo, también se clonó la forma madura de TFAM dentro del vector pETM40 usando el mismo tipo de enzimas de restricción. Este vector permite la expresión de la proteína con una fusión a la proteína de unión a maltosa (MBP) en el extremo N-terminal seguida de un sitio de corte por digestión con TEV.

2.2. Constructos con la cola C-terminal deleccionada

Mediante programas de predicción secundaria como *PsiPred*, se observó que la cola C-terminal no presentaba en su totalidad estructura secundaria predecible, lo cuál indicaba que

la cola C-terminal podía dificultar la cristalización de la proteína. Además como se ha comentado en la Introducción (apartado C.8), esta cola C-terminal está involucrada en múltiples funciones: especificidad de secuencia, interacción con TFBM, introducción de supercoling negativo, etc. Por ello, se decidió generar una batería de mutantes con la cola C-terminal delecionada en diferentes puntos. Estos mutantes fueron generados a partir del clon TFAM-HisTag y contenían la HisTag en C-terminal.

En la Fig_A2 se indican los diferentes mutantes, llamados CTT2 (Ser44-Glu238), CTT3 (Ser44-Arg233) y CTT4 (Ser44-Arg227).



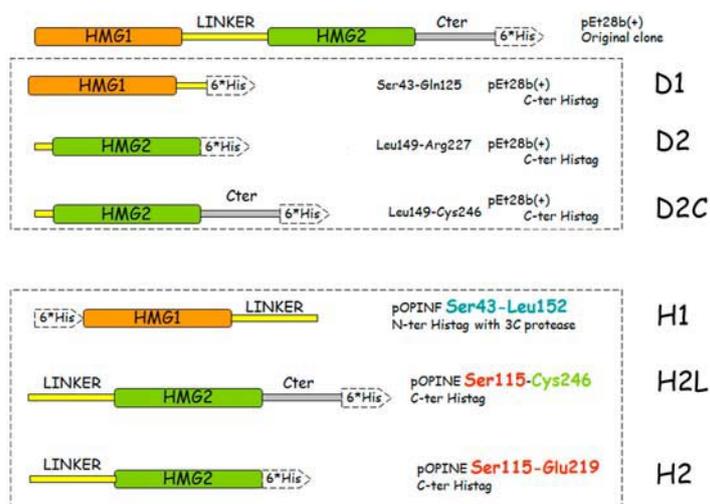
Fig_A2. Representación con cajas de la secuencia aminoacídica de TFAM, en la parte inferior se indica secuencia de la cola C-terminal para TFAM y los diferentes constructos. En gris claro se indican los residuos a seleccionar en el siguiente constructo inferior. Los números indican el residuo final de cada constructo más la cola HisTag.

2.3. Dominios: D1, D2, H1, H2S, H2L

Antes de conocer la estructura cristalográfica de la proteína, se realizó la clonación de los dominios HMGbox de TFAM mediante el uso de programas de predicción secundaria como *Psipred* y el alineamiento con otros dominios HMGbox con estructura resuelta (SRY, LEF-1, NHP6A, HMG1...). Se consiguieron tres constructos diferentes, a partir del mutante TFAM-HisTag (Fig_A3):

- Dominio1 (D1): Ser44-Gln125
- Dominio2 (D2): Leu149-Arg227
- Dominio2 + cola C-terminal (D2C): Leu149-Cys246

Además, una vez resuelta la estructura de TFAM en complejo con el ADN se diseñaron constructos de los dominios con el *linker*, ya que la estructura dejaba entrever que este *linker* era funcionalmente importante para la interacción con el ADN. El dominio HMGbox 1 (HMG1) más *linker* fue clonado en el vector de expresión pOPINF, que añade una Histag en el extremo C-terminal y fue llamado H1. Para el dominio HMGbox2 (HMG2) más *linker*, llamado H2S (H2short), y linker+HMG2+cola C-terminal, llamado H2L (H2long), fue usado el vector pOPINE añadiendo la Histag en el extremo C-terminal. El constructo H1 dentro del sistema pOPINF presenta posibilidad de corte de la diana mediante digestión con proteasa 3C.



Fig_A3: Representación esquemática de los diferentes constructos de dominios realizados, donde HMG1 y HMG2 hace referencia a los dominios HMGbox de TFAM.

2.4. Mutagénesis dirigida: mutantes de unión a ADN y mutantes en el linker

A partir de la estructura de TFAM en complejo con LSP22, fue posible dilucidar los contactos que realizan los dominios HMGbox (HMG1, HMG2) con la bases del ADN. Para HMG1, los contactos que contactan el ADN son: Tyr57, L58, Thr77, Thr78 y Ile81, para HMG2: Pro178, Gln179, Tyr162, Asn163, Val166 y Leu182 (ver más adelante en la descripción de la estructura)

Con la finalidad de realizar ensayos de afinidad al ADN y curvatura del ADN, se llevó a cabo mutagénesis dirigida usando el constructo TFAM-Histag de cada uno de estos residuos, que se mutaron a Ala. También se realizó un doble mutante (Leu58/Ala-Leu182/Ala)

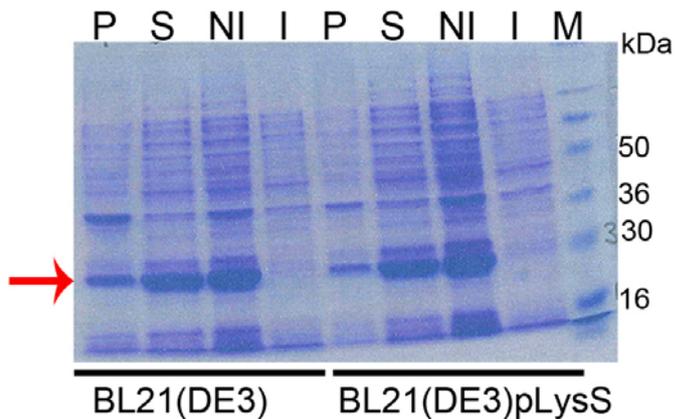
Además, debido a la posible dimerización de TFAM se construyeron mutantes del linker: M127A, E130A, E130R. Tanto los dobles mutantes como los mutantes para el linker no han sido caracterizados en su expresión, purificación, etc...

3. Pruebas de expresión y solubilización de la proteína

Para obtener un buen rendimiento de expresión de la proteína se realizaron pruebas de expresión, que presentan múltiples variables para optimizar el proceso. Se realizaron estos ensayos con cuatro tipo de cepas: BL21(DE3), BL21(DE3)pLysS, Rossetta2, Rossetta pLysS; a dos temperaturas diferentes (16°C y 37°C) con una agitación de 250 rpm. La concentración de inductor era de 1mM IPTG, tiempos de expresión después de la inducción de 3.5 h para 37°C y 24 h para 16°C, en medio LB.

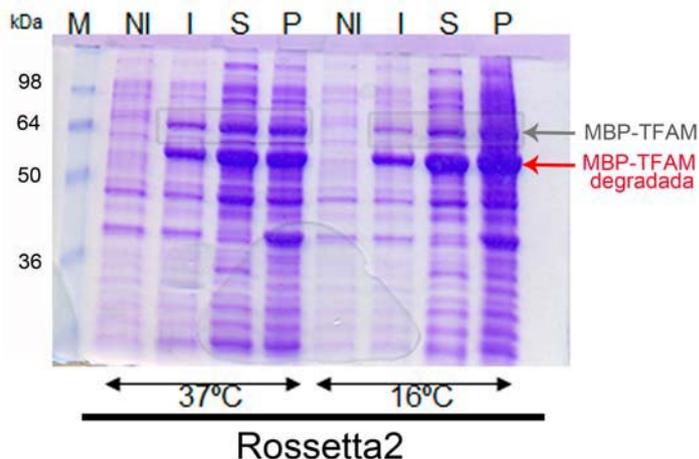
Para el constructo TFAM-Histag, las condiciones de expresión eran similares en todos los ensayos, así que finalmente se escogió la cepa BL21(DE3)pLys y una temperatura de inducción de 37°C (Fig_A4). Estas mismas condiciones se aplicaron para la expresión de los

constructos con la cola C-terminal recortada (CTT), los dominios (D1, D2, D2C) y los mutantes de unión a ADN. Sólo en el caso de los mutantes Leu58/Ala y Tryr57/Ala mostraron degradación durante la expresión, por ello se realizó una expresión de 24 h a 16°C para tratar de evitar esta degradación. Por otra parte, la expresión de los constructos H2L y H2S no se expresaron en ninguna de las cepas usadas, lo cuál parece indicar que se trata de constructos muy poco estables.



Fig_A4. Gel SDS-PAGE al 15% donde se muestra la expresión de TFAM-HisTag en BL21(DE3) y BL21(DE3)pLysS a 37°C con las siguientes muestras (leer de derecha a izquierda): M indica marcador, NI fracción no inducida, I fracción inducida, S fracción soluble y P fracción insoluble. La flecha roja indica la banda correspondiente a la proteína TFAM-HisTag.

Para la expresión del constructo MBP-TFAM, la cepa con mayor expresión de la proteína era Rossetta2, independientemente de la temperatura. Durante la expresión de este constructo existe una degradación constante dando lugar a una forma truncada de la proteína (~60 kDa) (Fig_A5).



Fig_A5. Gel SDS-PAGE al 12% donde se muestra la expresión de MBP-TFAM en Rossetta2 a dos temperaturas diferentes (37°C y 16°C) con las siguientes muestras: M indica marcador, NI fracción no inducida, I fracción

inducida, S fracción soluble y P fracción insoluble. La flecha gris indica la banda correspondiente a la proteína MBP-TFAM, mientras que la flecha roja indica la forma truncada o degradada de MBP-TFAM. El buffer de lisis era 750mM NaCl, 50mM HEPES pH7.5, 1mM DTT, 1mM PMSF, DNasa.

El siguiente paso después de la expresión, es la solubilización de la proteína y para ello es clave la elección del tampón de solubilización en función de las características físico-químicas de la proteína. Mediante el servidor *ProtParam* (Bairoch, Apweiler et al. 2005) es posible conocer algunos de los parámetros fundamentales para decidir el tampón de solubilización como son el pI teórico de la proteína que determinará el pH del tampón o el número de Cys que determinará la presencia de agentes reductores (Tabla_A1). Además, es sabido que las proteínas de unión al ADN, como TFAM, requieren una gran concentración de sal para mimetizar la carga negativa del ADN.

Para determinar el tampón de lisis se consideraron los parámetros fisicoquímicos de los constructos de proteína (Tabla_A1), se procedió a probar diferentes tampones, concentración de sal, agentes reductores, etc. Después de una número de pruebas y optimizaciones para determinar el tampón de solubilización se escogió: 750 mM NaCl, 50 mM HEPES pH 7.5, 1 mM DTT.

constructos	Num aa	PM (Da)	pI	Cys
TFAM-His	214	25652.4	9.65	2
CTT2	206	24715.4	9.72	1
CTT3	201	24116.6	9.65	1
CTT4	195	23334.7	9.55	1
D1	93	11113.6	9.76	1
D2C	109	13138	9.16	1

Tabla_A1. Tabla resumen con algunas de las propiedades físico-químicas de los diferentes constructos para TFAM obtenidas mediante el servicio *ProtParam*.

4. Purificación

La purificación de TFAM en su forma madura y soluble (25.6 kDa), así como de los constructos con la cola C-terminal (CTT'S) y los dominios con o sin extensión de linker (D1, D2, D2C, H1), todos ellos con presencia de HisTag, presenta dos etapas consecutivas: una primera etapa que consistía en una columna de afinidad, cargada con un agente quelante que une a níquel y que atrapa a las colas de histidinas, y la segunda en una cromatografía de exclusión molecular.

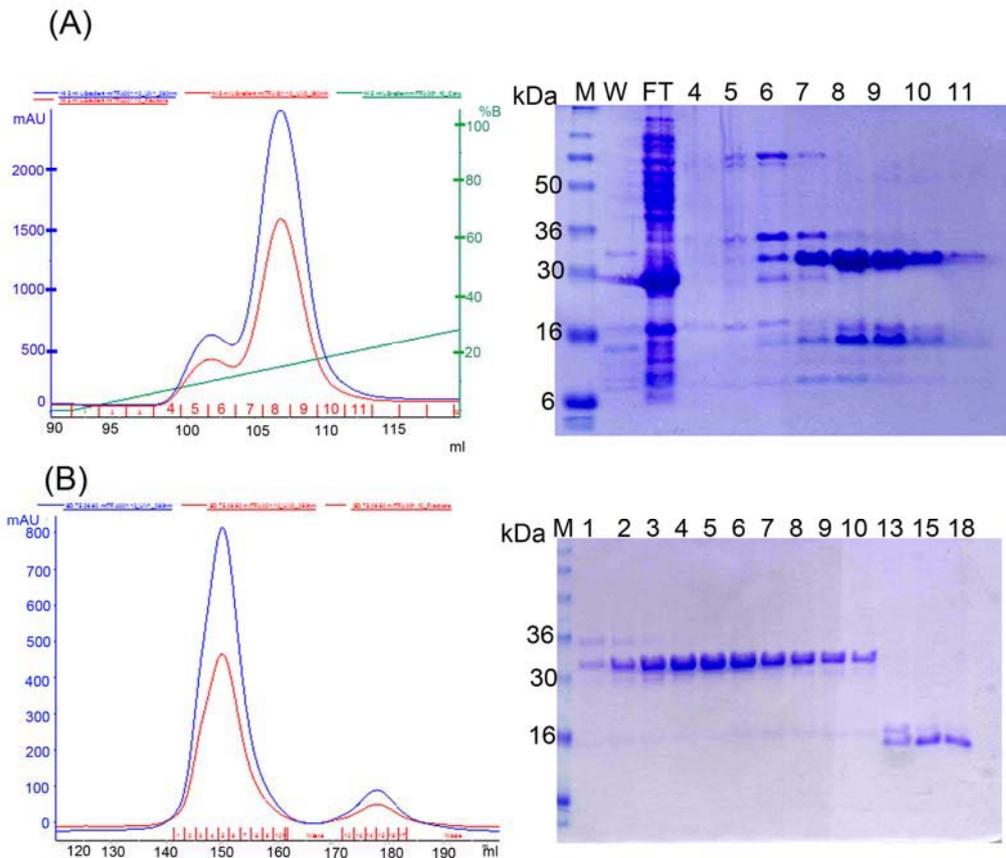
En la columna de afinidad (HisTrap5ml) para la purificación de TFAM (Fig_A6 (A)) se aplicaba un gradiente de imidazol lineal, donde alrededor de 12% existía un primer pico, debido a la presencia de contaminantes con unión inespecífica a la columna cargada de níquel. A medida que aumentaba la concentración de tampón B (1 M imidazol, 750 mM NaCl, 50 mM HEPES pH 7.5, 1 mM DTT) se eluye la proteína interés (~20 % B). Las fracciones se cargaban en un gel

SDS-PAGE (15%) para confirmar la presencia de la proteína, y posteriormente las fracciones elegidas eran cargadas en la gel filtración (Superdex 75 HL 26/60)

Tal y como se observa en los cromatogramas de ambas columnas de afinidad y de exclusión molecular para TFAM(Fig_A6), la línea de absorbancia 280nm está por encima de la de 260, esto es indicativo de la ausencia de oligonucleótidos en la muestra, y esto es especialmente importante en la purificación de proteínas de unión a ADN. Las fracciones más puras se juntaron y concentraron para proceder al siguiente paso de purificación.

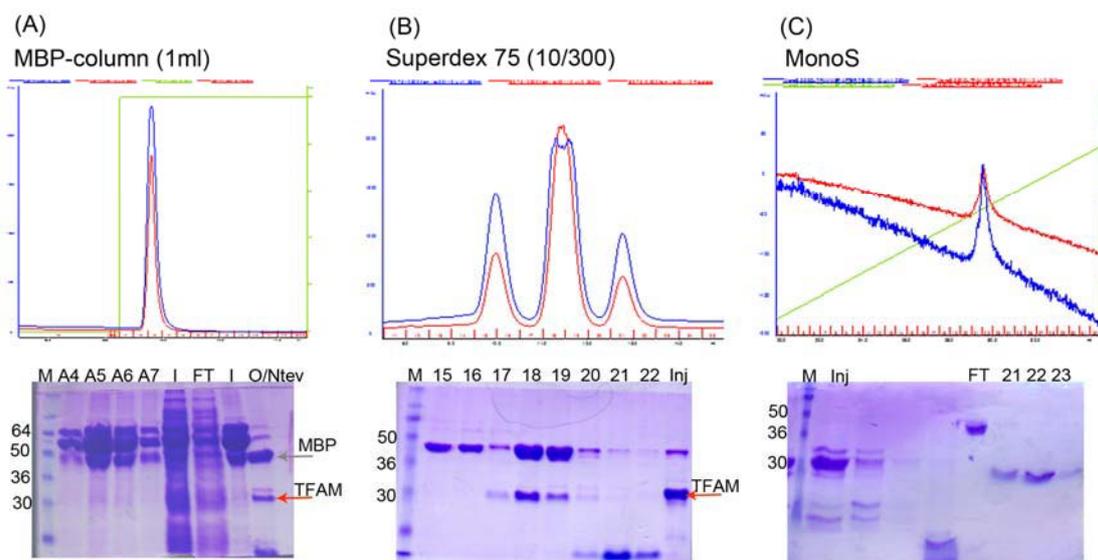
La purificación por gel filtración o cromatografía de exclusión por peso molecular muestra que la proteína eluye en un solo pico moderadamente ancho con un ligero hombro al principio. El análisis de las fracciones por SDS-PAGE indicó una alta pureza de las fracciones eluídas, así como la presencia de una banda superior para las fracciones primeras que podrían explicar dicho hombro, correspondiente a contaminantes. La anchura del pico era compatible con la presencia de distintos estados oligoméricos (dímeros, monómeros). La ausencia de un pico en el volumen de exclusión de la columna de gel filtración indicó ausencia de proteína agregada y confirmó que el tampón de purificación era apropiado. No obstante se observó un patrón de degradación correspondiente a un segundo pico en el cromatograma de la gel filtración, por ello se decidió realizar la purificación a 4°C. El análisis de las muestras por electroforesis y *protein mass fingerprinting* se identificó como producto de degradación de TFAM.

El rendimiento final de la expresión y la purificación para TFAM-HisTag es aproximadamente de unos 15 mg por 1 L de cultivo. Debido a que el rendimiento de la purificación para el constructo TFAM-HisTag era mucho mayor que MBP-TFAM, se decidió abordar la cristalización empleando este constructo.



Fig_A6. (A) Cromatograma de elución de TFAM de una columna HisTrap 5ml. Eje vertical izquierdo, densidad óptica de las fracciones eluyentes; eje derecho, porcentaje del tampón B. La pendiente verde simboliza el incremento del porcentaje de tampón B. El cromatograma indica las fracciones eluyentes y su respectiva absorbancia a 280nm (en azul) y a 260nm (en rojo). En general, para una proteína pura la relación 260/280 es de 0.7. Gel SDS-PAGE 15% de las fracciones eluidas en la columna de afinidad. En vertical se indican los pesos moleculares del marcador utilizado. Carriles: M: marcador; W: fracción de lavado de la columna; FT: fracción de carga de columna o "flowthrough"; 4-11: fracciones eluidas en A. (B) Cromatograma Superdex 75 HL 26/60 TFAM. La representación del gráfico sigue los mismos criterios que en (A), pero con ausencia del gradiente de imidazol. Gel SDS-PAGE 15% de las fracciones de la gel filtración. Obsérvese la pureza de las fracciones (>95%).

La purificación del constructo MBP-TFAM implicaba cuatro pasos secuenciales (Fig_A7): columna de afinidad para MBP, digestión con proteasa TEV, gel filtración de la muestra digerida y finalmente una columna de intercambio catiónico o columna de heparina. Pese a que normalmente las proteínas fusionadas con MBP presentan un alto rendimiento de expresión y purificación, ya que MBP permite una mayor solubilización de la proteína, y el tipo de columna permite una elución de la proteína en alta concentración, el rendimiento final (~ 2 mg por 1L cultivo) era muy bajo comparado con TFAM-Histag. Posiblemente debido a los consecutivos pasos de diálisis, tanto para digestión con proteasa TEV como para la inyección en MonoS o columna de Heparina, que pueden provocar la desestabilización de la proteína.



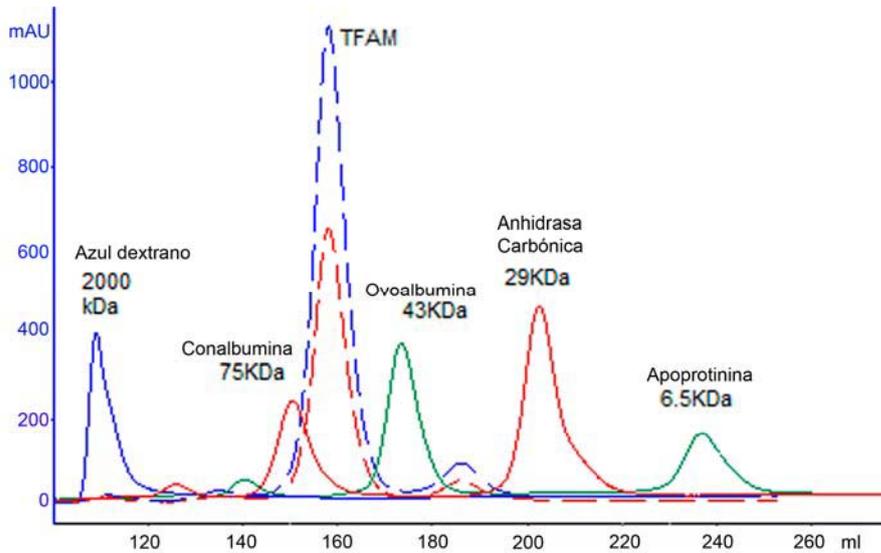
Fig_A7. Purificación del constructo MBP-TFAM. (A) Perfil de elución de la columna de afinidad por MBP, mediante un el cambio de tampón con las mismas condiciones pero conteniendo 10mM maltosa era posible eluir la proteína interés. Posteriormente se procedía a la diálisis de la muestra hasta 300mM de NaCl para la digestión O/N con TEV a 20°C, ver fracción O/Ntev. (B) Una vez digerida la muestra se cargaba en una columna de exclusión por peso molecular, donde las fracciones 17-19 contenían TFAM, pero con presencia también de MBP. (C) Mediante un último paso en una columna de intercambio iónico se conseguía TFAM pura, debido a que se aprovechaba el hecho que se trata de una proteína de unión al ADN (cargada positivamente) con lo cual permanecía unida a la columna hasta alcanzar un porcentaje tal de sal que permita eluirla (ver fracciones 21-23).

5. Homogeneidad de la proteína

Tanto para lo ensayos de cristalización como para el estudio mediante SAXS es importante que la proteína sea lo más estable y homogénea posible. Por ello una vez purificada se realizaron diferentes ensayos biofísicos para caracterizar el estado de la muestra, antes de la formación del complejo con ADN.

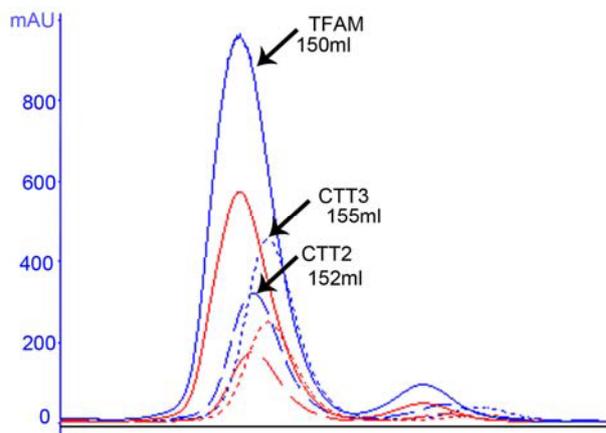
5.1. Calibración, posible dimerización TFAM

Mediante la calibración de la columna de gel filtración (Sup75 HL 26/60) se compara el volumen de elución (V_e) de la proteína con el V_e de las proteínas globulares estándar (por ejemplo: ovalbumina, conalbumina, aprotinina,...) siempre en presencia del mismo tampón que para TFAM (750 mM NaCl, 50 mM HEPES pH7.5, 5 mM DTT). La representación lineal de $K_{av} ((V_e - V_0)/(V_t - V_0))$ vs log del peso molecular de los marcadores o patrones, permite extraer el peso molecular de la proteína diana. Se obtuvo que el peso molecular experimental de TFAM-Histag era ~46 kDa, valor muy cercano al peso molecular teórico del dímero ($2 \times 25.6 \text{ kDa} = 51.2 \text{ kDa}$) (Fig_A8).



Fig_A8. Perfil cromatográfico de la gel filtración (Sup75 HL 26/60) durante la calibración de la columna en las condiciones de tampón para TFAM, donde se indican los volúmenes de elución (V_e) para las diferentes proteínas globulares usadas como patrones y para TFAM.

Cuando se analiza la secuencia aminoacídica de TFAM se observan dos cisteínas, una situada en la parte N-terminal (Cys49) y otra en el último residuo de la proteína (Cys246). Se creyó que estas cisteínas eran las posibles causantes de la dimerización, por lo que comparamos el V_e de TFAM con los constructos con la cola C-terminal delecionada (CTT), que destacan por la ausencia de Cys246 (Fig_A9). La diferencia en el V_e no parece implicar estados oligoméricos distintos entre TFAM y CTT'S, lo que sugiere que el residuo Cys final no participa en la posible dimerización de TFAM.



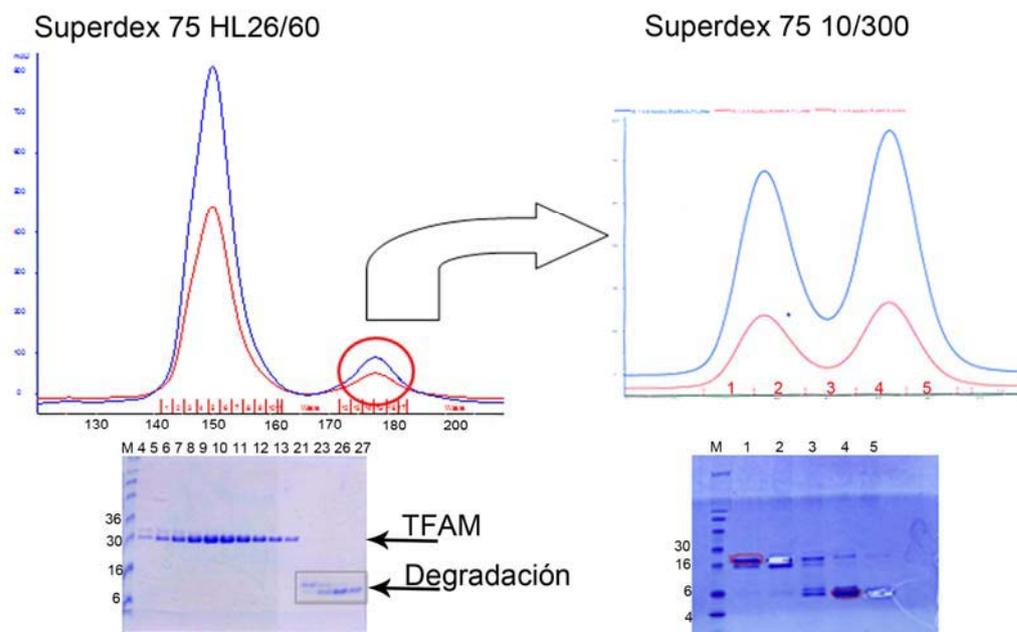
Fig_A9. Perfil cromatográfico de TFAM y los constructos CTT2/CTT3, donde se indican los diferentes V_e en presencia del mismo tampón de purificación.

5.2. Caracterización de la degradación en TFAM

Como se ha comentado anteriormente durante la purificación del constructo TFAM-Histag, existe una degradación constante de TFAM por el extremo N-terminal, ya que esta degradación se arrastra ya durante la columna de afinidad (Fig_A10). En realidad, esta degradación está formada por una doble degradación de ~ 16 kDa y ~ 6 kDa, que pueden

ser separadas mediante el uso de una segunda gel filtración, tal y como se indica en la Fig_A9.

Mediante secuenciación N-terminal de la muestra de 16kDa fue posible obtener la secuencia aminoacídica del inicio de los productos de degradación. El producto de degradación de 16 kDa va de Thr144 a Cys246, por lo tanto incluye parte del linker.



Fig_A10. Procedimiento para la caracterización de la degradación en TFAM-Histag. Cuando se realiza la purificación de TFAM-Histag es posible observar un segundo pico en el cromatograma de la gel filtración, que mediante gel SDS-PAGE al 15% se constata que se trata de dos degradaciones. La muestra con las fracciones eluidas (21-27) son concentradas y cargadas en una segunda gel filtración, que permite la separación de las dos degradaciones presentes (~16 kDa y ~6 kDa) y su posterior caracterización mediante secuenciación N-terminal y digestión proteica de las bandas recortada en membrana PVDF y gel SDS-PAGE.

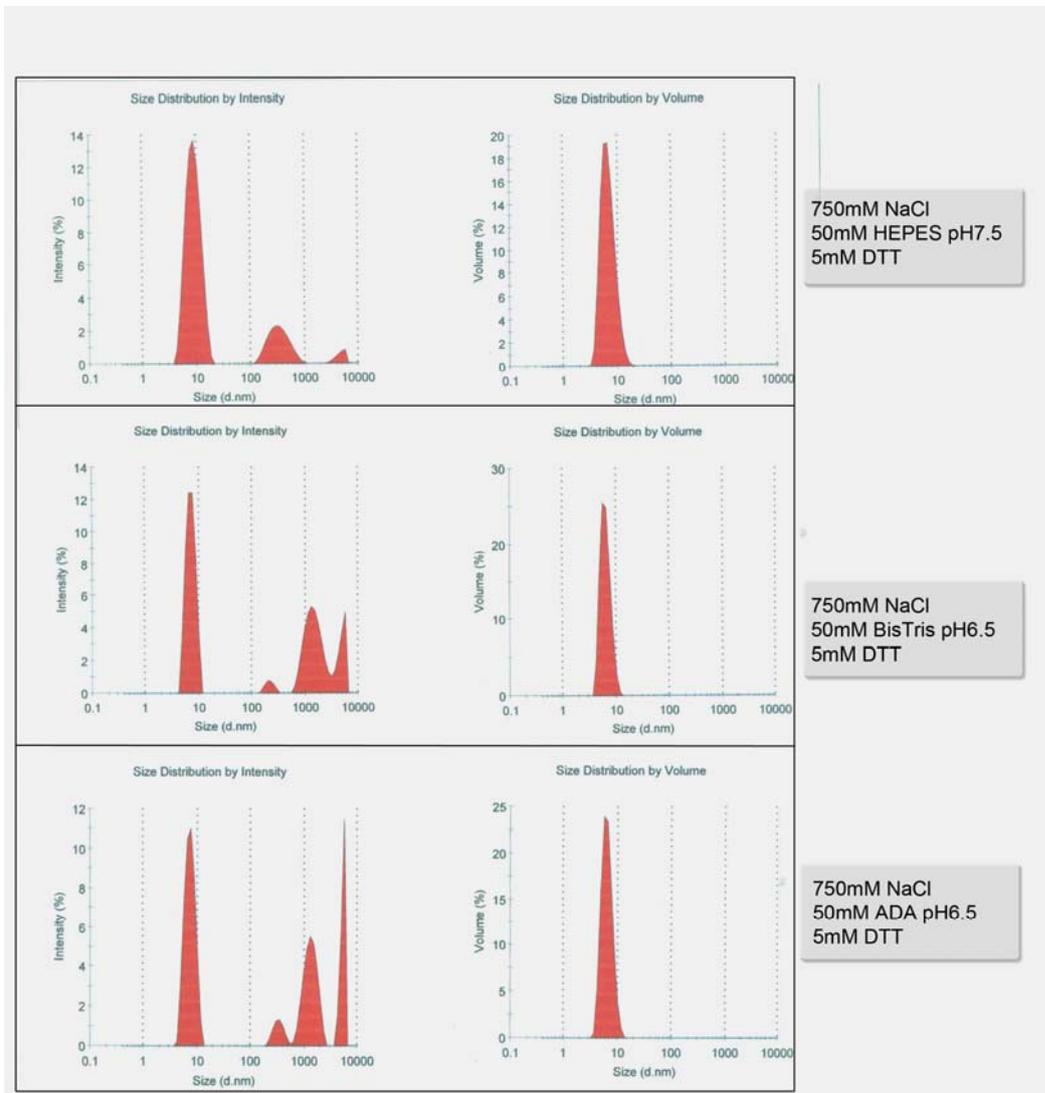
5.3. Optimización del tampón proteico por ThermoFluor y DLS

Se realizaron ensayos de estabilidad térmica de la proteína en función de la composición del tampón, técnica conocida como *ThermoFluor*. Para este ensayo se usa un colorante sensible a la hidrofobicidad del entorno (*SyPro Orange*) y se analiza la variación en la fluorescencia en función del incremento de temperatura. Debido a que TFAM es una proteína con poca presencia de núcleos hidrofóbicos, este ensayo no resultó óptimo para la caracterización de la proteína en solución.

Es por ello que se utilizó la técnica conocida como *Dynamic Light Scattering (DLS)*, que mediante el análisis de la dispersión que se produce después de que un haz laser de luz monocromática traspase una muestra, dado que el movimiento de las moléculas es función del tamaño de las moléculas, el resultado es una función de correlación que permite calcular el coeficiente de difusión, la distribución del radio hidrodinámico (perfil de distribución en función del volumen e intensidad de la partícula en solución) y la polidispersidad del sistema. La polidispersidad es la desviación estándar de la muestra.

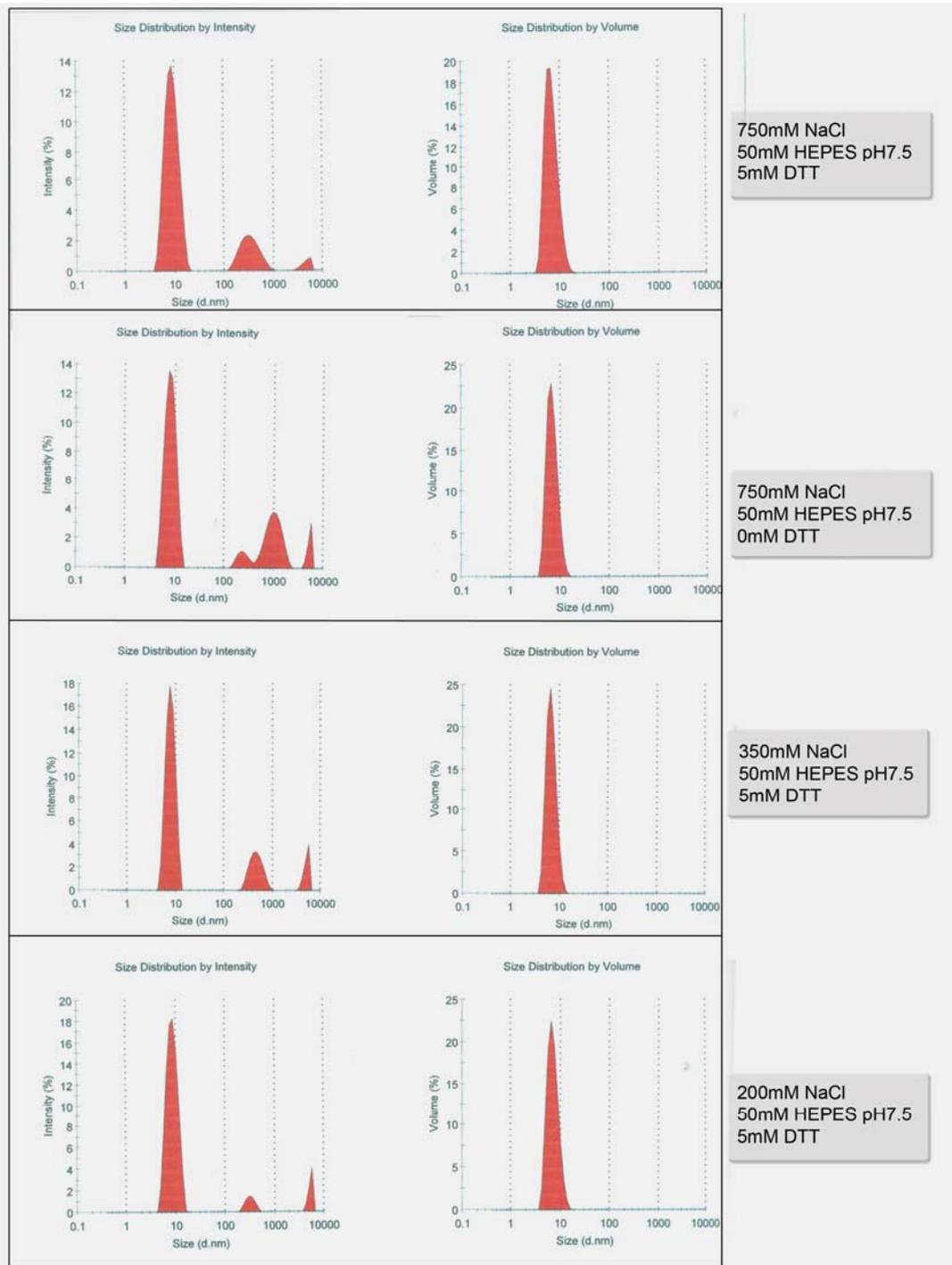
Según este parámetro se pueden distinguir: muestra monodispersa si la polidispersidad es inferior a ~20% y muestra polidispersa si la polidispersidad es superior al 30%.

Se realizaron mediadas de DLS a una concentración de proteína de 13mg/ml, donde las características del tampón proteico eran modificadas en cuanto a concentración de NaCl y pH de la solución. Para determinar la calidad de la muestra se examinaba el valor numérico de la polidispersidad y tamaño de la partícula, así como el perfil de distribución en función de la intensidad y volumen. Este principio físico permitió establecer las condiciones de tampón correctas para realizar experimentos de SAXS con la proteína en la forma no unida, ya que las medidas de SAXS requieren reducir la concentración de NaCl para no añadir ruido a la medida y la presencia de agente reductor para evitar daño por radiación (Tabla_A2) (Fig_A11 Fig_A12). Fue posible observar que era necesario añadir el agente reductor, DTT, y que era posible disminuir la concentración de NaCl hasta 200 mM. Además el tampón en presencia de HEPES a pH7.5 ofrecía una mayor estabilidad a la muestra.



Fig_A11. Perfiles de intensidad y volumen de la muestra de TFAM en presencia de diferentes tampones, claramente el mejor tampón es el usado para la purificación. Partículas muy grandes dan lugar a una fuerte

intensidad (gráfico de la izquierda), a pesar de que el volumen que ocupan (gráfico de la derecha) prácticamente no sea detectable.



Fig_A12. Perfiles de intensidad y volumen de la muestra analizada mediante DLS, para seleccionar la concentración ideal de NaCl del tampón para los ensayos de SAXS.

Curiosamente, a pesar de que la alta concentración de NaCl (750 mM, 1M) optimiza la extracción de proteína tras la lisis, la reducción a 200 mM da lugar a una menor presencia

de agregados una vez la proteína ha sido purificada. No obstante, la polidispersidad es menor a 750 Mm de NaCl (ver Tabla_A2).

Tampón diálisis	Tamaño promedio (nm)	Polidispersidad
750mMNaCl, 50mM HEPES pH 7.5, 5mM DTT	10.33	0.207
750mMNaCl, 50mM HEPES pH 7.5, 0mM DTT	99.80	0.353
350mMNaCl, 50mM HEPES pH 7.5, 5mM DTT	41.90	0.299
200mMNaCl, 50mM HEPES pH 7.5, 5mM DTT	12.97	0.269

Tampón diálisis	Tamaño promedio (nm)	Polidispersidad
750mMNaCl, 50mM HEPES pH 7.5, 5mM DTT	10.33	0.207
750mMNaCl, 50mM Bis-Tris pH 6.5, 5mM DTT	492.5	0.954
750mMNaCl, 50mM ADA pH 6.5, 5mM DTT	124.4	0.641

Tabla_A2. Valores numéricos del tamaño y polidispersidad de la muestra en composiciones de tampón diferentes, analizados mediante DLS. Tabla superior, variación en la combinación de NaCl y DTT. La tabla inferior, variación en el tipo de tampón.

B. RESULTADOS: FORMACIÓN DEL COMPLEJO PROTEÍNA-ADN

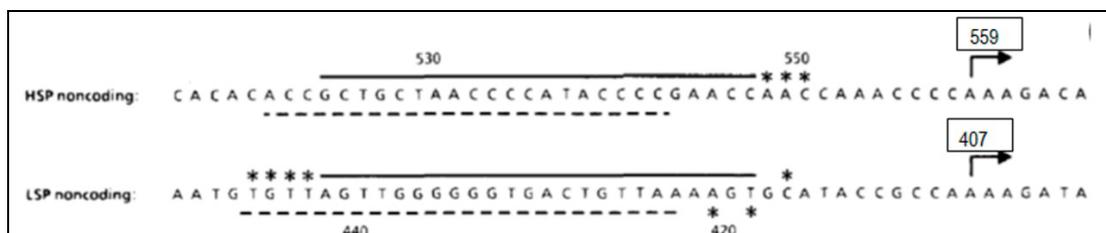
Como se ha comentado durante la Introducción, en mamíferos la transcripción ocurre a través de tres regiones promotoras: LSP, HSP1 y HSP2. Se han caracterizado sitios de unión para TFAM en LSP y HSP1. Debido a que la afinidad de TFAM es superior para LSP y que la implicación funcional en la transcripción parece ser mayor para LSP, se escogió la secuencia de unión LSP para la cristalización con TFAM.

La cristalización de TFAM se ensayó sobretodo con el constructo TFAM-Histag, aunque también fueron usados los constructos CTT's y el constructo MBP-TFAM después de la digestión de su diana de proteólisis para separar MBP y TFAM.

1. Diseño de oligos para la cristalización

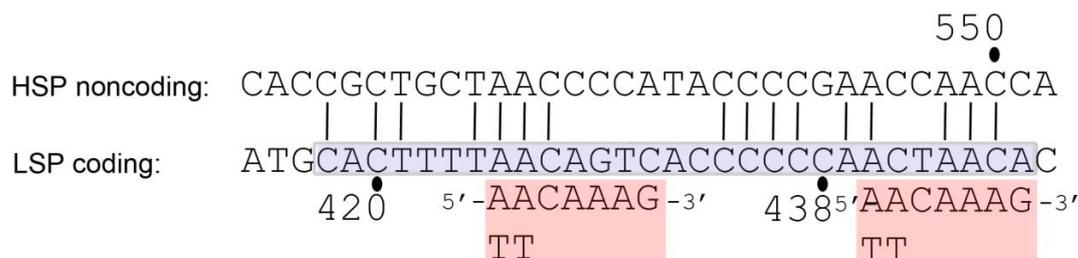
Para la cristalización de TFAM en complejo con su secuencia de unión a LSP fue necesario realizar el diseño de los oligos de cristalización en base a la región de protección de TFAM analizada por *footprinting* (Fisher, Topper et al. 1987) y la secuencia consenso definida para los dominios HMGbox de unión específica de secuencia (Laudet, Stehelin et al. 1993).

Los análisis de *footprinting* muestran una región de protección de ~27 pb para LSP y HSP, donde las cadenas muestran un desplazamiento en la región de protección (Fig1, donde la línea sólida indica la región de protección para la cadena no-codificante y el interlineado para la cadena codificante).



Fig_B1. Análisis *footprinting* con DNasa I para TFAM y sus regiones de unión en LSP y HSP1. Los asteriscos indican los sitios de corte por TFAM en la cadena codificante y no codificante. Las líneas indican la zona de protección ofrecida por TFAM en la cadena no codificante, mientras que el interlineado indica la región de protección para la cadena codificante. (Dairaghi&Shadel, 1995).

Mediante la comparación de las dos secuencias en HSP y LSP, definidas mediante *footprinting* por TFAM, y el alineamiento de las secuencias consenso para los dominios HMGbox; se estableció la secuencia madre de 29 pb a partir de la cual se diseñaron los diferentes oligos con longitud y extremos variables (Fig_B2).



Fig_B2. Alineamiento secuencial de las dos regiones (HSP y LSP) donde TFAM ofrece protección frente a la digestión con DNasaI, con una línea vertical se indican las bases que coinciden entre las dos secuencias. En cajas rojas se indica la secuencia general consenso para los dominios HMGbox de unión específica. La zona sombreada en gris hace referencia a la secuencia madre establecida para iniciar el diseño de los oligos para la cristalización.

A partir de esta secuencia madre de 29 pb se realizó un diseño amplio de oligos para la cristalización dejando extremos cohesivos y no-cohesivos, con longitudes variables, “moviéndonos” a lo largo de la secuencia madre de 29 pb (ver Fig_B3)

bp	sequence
33bp	ATGCACITTTAACAGTCACCCCCCAACTAAC TACGTGAAAATTGTCAGTGGGGGGTTGATTGTG
28bp	CACTTTTAAACAGTCACCCCCCAACTAAC CGTGAAAATTGTCAGTGGGGGGTTGATT TCGTGAAAATTGTCAGTGGGGGGTTGATT
27bp	ACTTTTAAACAGTCACCCCCCAACTAAC GTGAAAATTGTCAGTGGGGGGTTGATT
26bp	CACTTTTAAACAGTCACCCCCCAACTAA TGTGAAAATTGTCAGTGGGGGGTTGATT
25bp	ACTTTTAAACAGTCACCCCCCAACTAA TTGAAAATTGTCAGTGGGGGGTTGATT TTTAAAATTGTCAGTGGGGGGTTGA
	TAACAGTCACCCCCCAACTAAC TTGTCAGTGGGGGGTTGATTGA
	TTAACAGTCACCCCCCAACTAAC ATTGTCAGTGGGGGGTTGATTGA
22bp	TAACAGTCACCCCCCAACTAAC ATTGTCAGTGGGGGGTTGATTG
	__AACAGTCACCCCCCAACTAAC __TGTCAGTGGGGGGTTGATTGT
21bp	__AACAGTCACCCCCCAACTAAC __TTGTCAGTGGGGGGTTGATTG
15bp	CACTTTTAAACAGTCA GTGAAAATTGTCAGT
	CACCCCCCAACTAAC
16bp	GTGGGGGGTTGATTG

Fig3. Esquema de los diferentes oligos empleados para la cristalización, en azul se encuentra señalada la secuencia del oligo de 22 pb cristalizado en complejo junto con TFAM.

2. Formación del complejo

Uno de los puntos clave para la cristalización de la proteína en complejo con ADN es la optimización de las condiciones químicas como el tampón para la formación del complejo

(concentración de NaCl, pH, etc.) y el ratio proteína : ADN óptimo; para ello fueron usados geles de retardo nativo (EMSA: *Electrophoretic Mobility Shift Assay*).

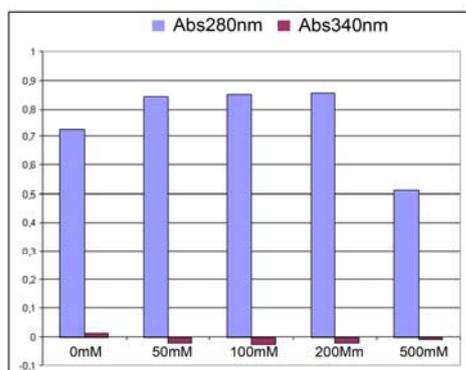
Cabe destacar que el método establecido para la formación del complejo fue siempre mediante diálisis escalonada en gradiente decreciente de NaCl.

2.1. Optimización de las condiciones del complejo

La composición del tampón para la formación del complejo puede ayudar a estabilizar la interacción proteína/ADN. Uno de los factores claves en la formación de complejos proteína/ADN es la concentración sal, ya que mucha de las interacciones que se dan son de tipo electrostático con lo que variaciones en la concentración de sal pueden promover o desestabilizar la formación del complejo. En este caso se realizaron complejos a diferentes concentraciones de sal que fueron analizados mediante un espectrofotómetro y geles EMSA. La medición de la absorbancia a 340 nm permite observar la existencia de agregación de la muestra (Tabla_B1, Fig_B4), experimentos que permitieron delimitar la concentración de NaCl óptima entre 50-200mM NaCl. La visualización mediante electroforesis de las muestras a diferentes condiciones permitió establecer cuál era la concentración de NaCl indicada para desplazar todo el ADN hacia la formación del complejo.

mM NaCl	Abs 280nm	Abs340nm
0	0,723	0,012
50	0,845	-0,02
100	0,852	-0,028
200	0,856	-0,02
500	0,51	-0,009

Tabla_B1. Experimento de formación del complejo a diferentes concentraciones de NaCl. A una concentración constante de proteína y ADN se realizaron diálisis a diferentes concentraciones de NaCl y luego realizó la medida de la Absorbancia a 280nm (donde absorben los residuos aromático de la proteína) y a 340nm (longitud de onda indicador de efectos de agregación).



Fig_B4. Experimento de formación del complejo a diferentes concentraciones de NaCl. Mediante la observación de la relación entre absorbancia a 280 nm y 340 nm, es posible caracterizar la concentración de NaCl óptima

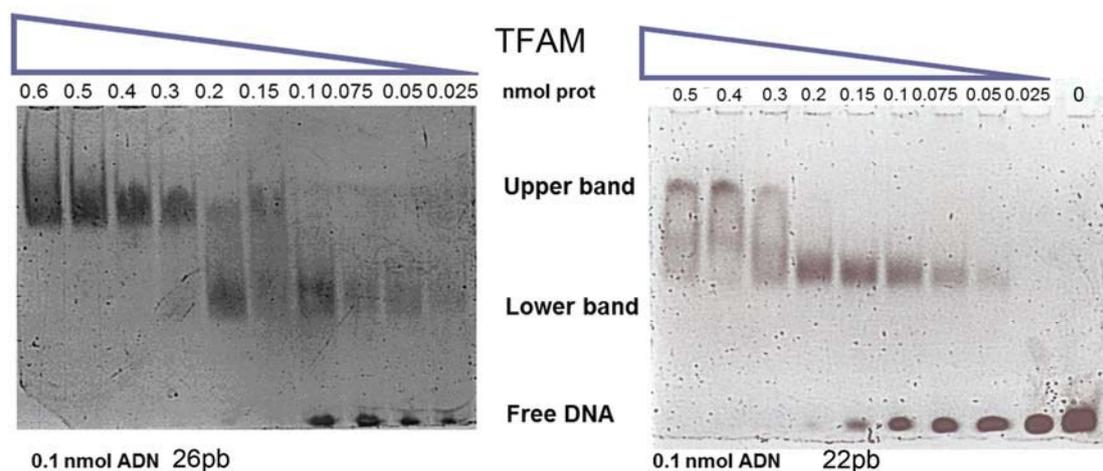
para formar el complejo. Esta concentración se encuentra oscilante entre 50-200mM, finalmente el complejo se formaba a 20 mM donde presentaba una menor cantidad de ADN libre.

Finalmente, el tampón óptimo empleado para la formación del complejo presentaba las siguientes características: 20 mM NaCl, 50 mM HEPES pH 7.5, 5 mM DTT, donde la longitud del ADN no afectaba a la composición del tampón

2.2. Ensayos de unión con ADN

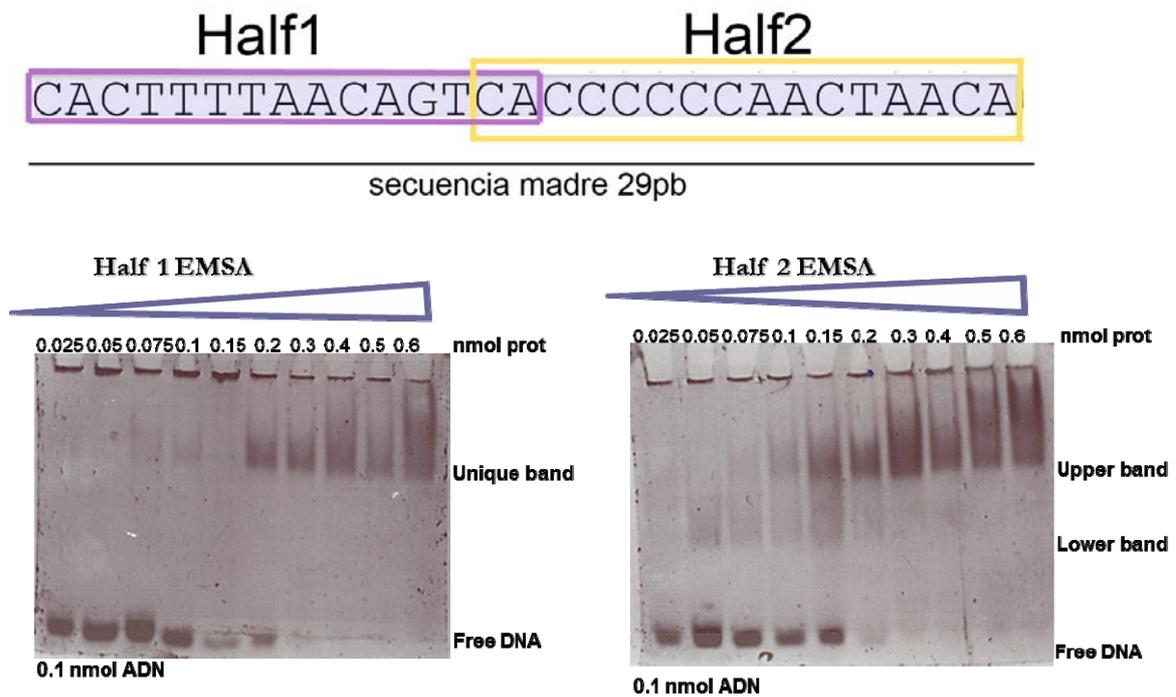
Los ensayos de retardo en la movilidad electroforética (*Electrophoretic Mobility Shift Assays*, EMSA) o más comúnmente llamados geles nativos de retardo, se basan en la observación de que complejos estables proteína/ADN migran más lentamente que los fragmentos de ADN libre en geles de poliacrilamida en condiciones nativas bajo la acción de un campo eléctrico. Se usaron los ensayos EMSA para conocer el comportamiento de la muestra cuando se utiliza una concentración constante de ADN (0.1nmol) frente a concentraciones crecientes de proteína de 0.025 nmol a 0.6 nmol (Fig_B5). El análisis de los resultados indicó que TFAM-Histag formaba diversos tipos de complejos con el ADN, ya que a medida que aumentaba la concentración de proteína aparecían bandas más retardadas en el gel. La observación de geles EMSA con diferente ratio proteína:ADN permitía dilucidar la existencia de dos bandas de complejo (*coger band*, *upper band*). La presencia de la banada *upper band* se daba en complejos con ADN de longitudes de 29 pb hasta 20 pb. La existencia de esta doble banda sugería la posibilidad de diferentes estados oligoméricos dentro de la muestra.

Para los ensayos de cristalización se trabajó con relaciones 2:1 de proteína:ADN, ya que era cuando existía una desaparición de ADN libre.



Fig_B5. EMSA nativo del complejo TFAM-LSP26 y LSP22 (según Fig_B3). A una concentración constante de ADN se añadió cantidades crecientes de proteína. Obsérvese que en ratios donde la presencia de proteína es inferior a la de ADN existe una banda inicial (indicada como *lower band*), que desaparece al incrementar la presencia de proteína en el complejo, transformándose en una banda más retardada (indicada como *upper band*). Además existe una zona confusa o intermedia donde se encuentran las dos bandas. A muy baja concentración de proteína, aparecen bandas correspondientes a ADN libre en la parte inferior del gel.

A raíz de la estructura resuelta por Stott *et al* (Stott, Tang et al. 2006), en la que mediante RMN evidenciaban la estructura de un bidominio HMGbox sintético, que contenía el dominio HMGbox de SRY y el dominio B de HMG1, en complejo con un oligo de 16pb, por ello, se decidió dividir la secuencia madre de 29 pb en dos secuencias, una de 15pb, llamada Half1, y otra de 16pb, llamada Half2. Al realizar los ensayos de unión con los oligos Half1 y otra de Half2, en ambos casos la desaparición total del ADN libre sucedía en el ratio 2 proteína: 1 DNA. Pero a diferencia de lo visto anteriormente para otros ADN's, en el caso del complejo con Half1 no existía una doble banda de retardo y TFAM parece tener una afinidad ligeramente inferior, donde el primer retardo se da a 0.075 nmol frente a 0.05nmol para el oligo Half2 (Fig_B6).



Fig_B6. EMSA nativo del complejo TFAM-Half1 (a la izquierda) y TFAM-Half2 (a la derecha). Mientras que en Half2 a concentraciones crecientes de proteína existe un primer retardo a 0.05 nmol y un segundo retardo a 0.15 nmol. En el caso de TFAM-Half1 no existe doble banda y el primer retardo sucede a 0.075.

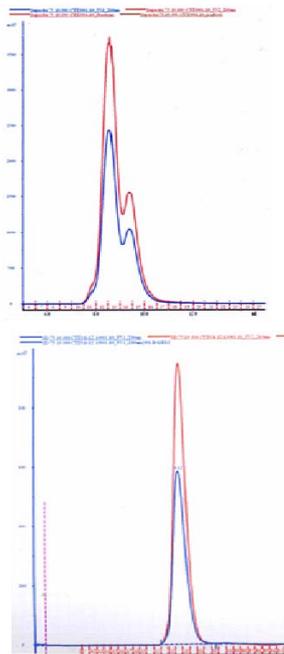
3. Purificación del complejo

Una vez formado el complejo proteína (TFAM, CTT3 ó CTT2)/ADN con ratio 2:1 para los diferentes oligos de cristalización usados (Fig_B3), se prosiguió a la purificación de éste mediante cromatografía de exclusión molecular.

La gel filtración mostraba para todos los complejos proteína-ADN posibles, perfiles cromatográficos con presencia de un doble pico, lo cual indicaba la heterogeneidad de la muestra cargada (fig_B7). Posteriormente, la heterogeneidad era ratificada mediante EMSA donde se observaban diversas bandas de retardo. Además esta heterogeneidad parecía estar en equilibrio, ya que si se volvían a purificar las fracciones eluidas de la primera gel filtración era posible observar la doble banda en el gel nativo EMSA.

La etapa de la purificación del complejo no lograba disminuir la heterogeneidad de la muestra por lo que se creó un punto de inflexión en la estrategia de purificación a partir del cual se decidió eliminar la etapa de purificación del complejo TFAM-ADN (que normalmente se introduce para pulir la calidad de la muestra), ya que no se llegaba a una pureza mayor, y además introducía un tiempo de experimento más largo, diluía la muestra y el rendimiento era menor.

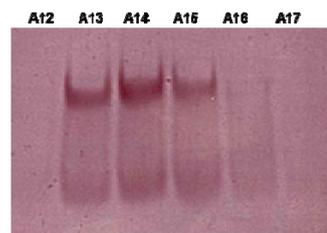
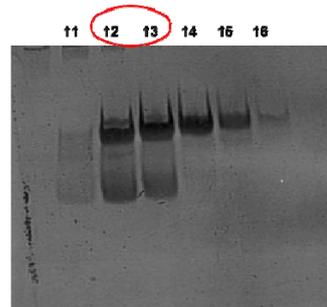
Sup 75 10/30



TFAM + DNA22



EMSA gel



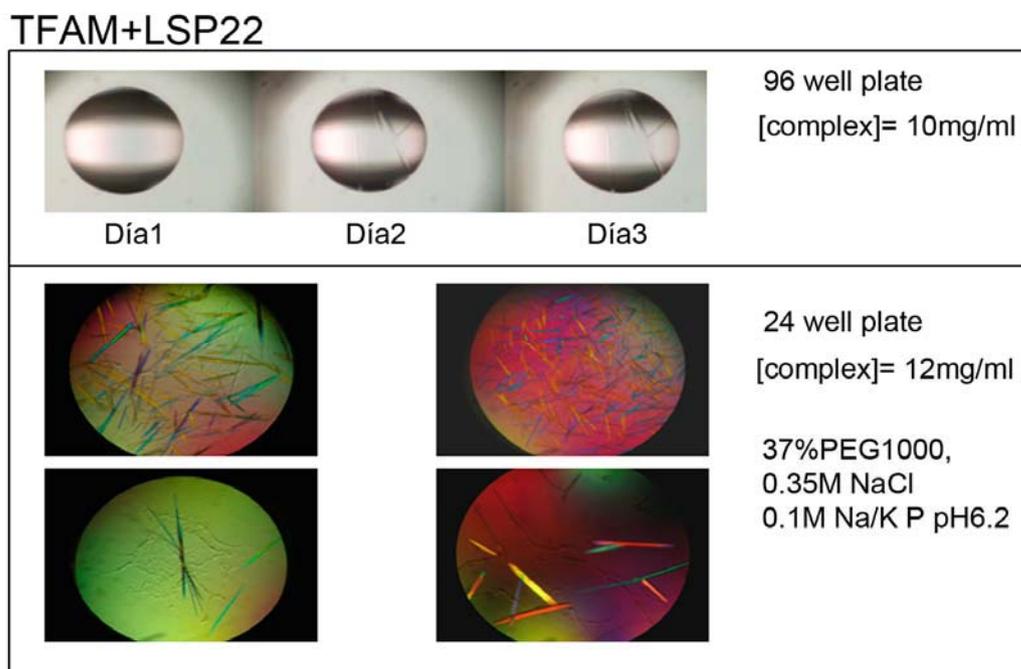
Fig_B6. (A) Gel filtración (Superdex 75 10/30) del complejo TFAM/ADN 22. En el cromatograma (superior izquierda) se observaron 2 picos cuyas fracciones fueron analizadas por EMSA, gel nativo 5.5% archilamida teñido con *SyBr-green* (B) Gel filtración (Superdex 75 10/300) de las fracciones (12-13) del gel anterior (indicadas con círculo rojo) que da lugar a un pico único. Cuando se analizan las fracciones por EMSA es posible observar que la muestra sigue siendo heterogénea. Obsérvese la relación de OD a 280 nm (en azul) por debajo de 260 nm (en rojo), indicativa de la presencia de ADN.

C. RESULTADOS: CRISTALIZACIÓN

La criba inicial de condiciones de cristalización se realizó con los *screenings* comerciales de *Hampton Research* en placas de 96 pocillos y dispensación líquida automática mediante un robot. Las placas se guardaban a dos temperaturas distintas, 4 y 20°C. En ellas se testaron los constructos TFAM, CTT3 y CTT2, solos o en complejo con los diferentes oligos de longitud variable.

1. Cristalización del complejo TFAM-LSP22

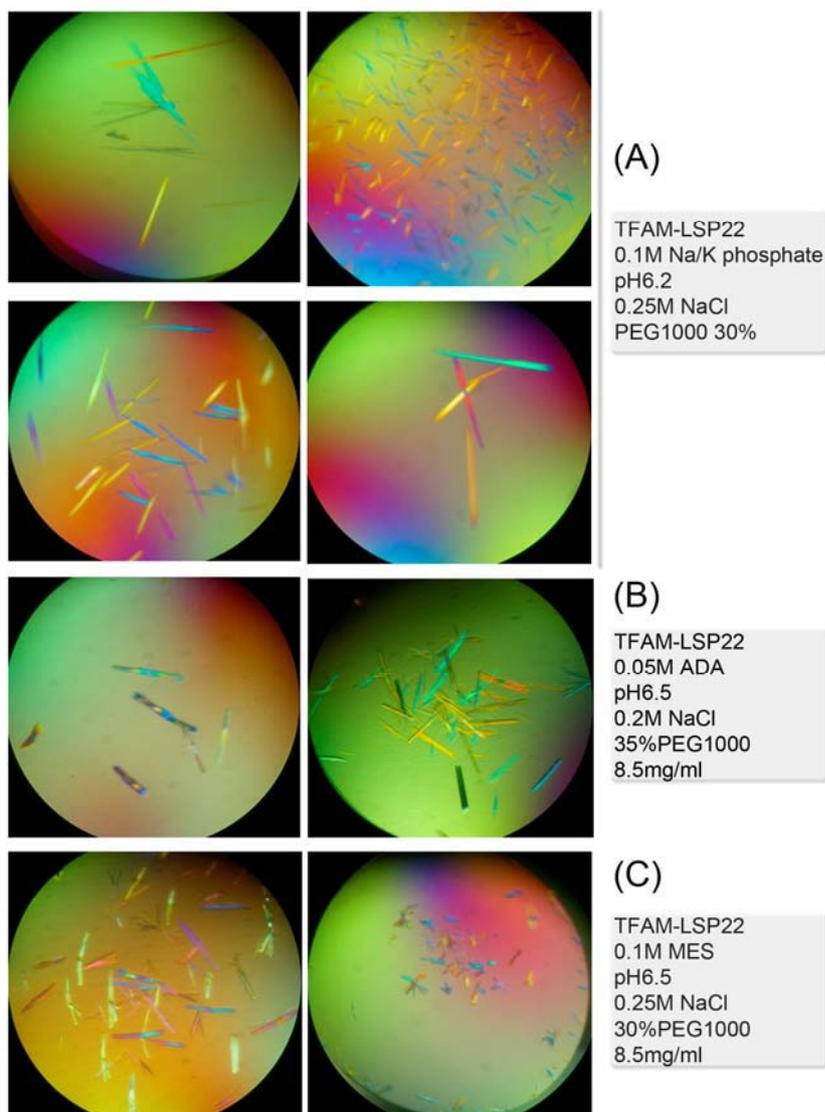
Se obtuvieron cristales con el complejo TFAM-LSP22 (con extremos romos) en la siguiente condición a 20°C: 20%PEG1000, 0.2M NaCl, 0.1M Na/K phosphate pH 6.2, a una concentración 10 mg/ml del complejo y gotas de 200 nl (Fig_C1). Eran cristales con forma de aguja y no del todo tridimensionales, de poco grosor y formados por finas capas. Una estrategia utilizada para incrementar el tamaño de los cristales fue escalar hacia placas de 24 pocillos con gotas de 2 µl y variar las condiciones de cristalización (concentración de precipitante, pH, volumen de la gota, temperatura, etc.); de esta forma se obtuvieron unos cristales más grandes. Estos cristales fueron congelados para evitar el daño por radiación y la desecación del cristal, es importante que durante el proceso de congelación no se formen cristales de agua que podrían afectar la ordenación del cristal y para ello normalmente se buscan soluciones químicas de congelación (*cryobuffers* o soluciones crioprotectoras) que favorezcan la vitrificación del tampón en el que está inmerso el cristal. En el caso de los cristales TFAM-LSP22 no hizo falta buscar soluciones de crioprotección, ya que las condiciones de cristalización ya eran por sí mismas crioprotectoras.



Fig_C1. Imágenes de los primeros cristales obtenidos con el complejo TFAM-LSP22. En la parte superior aparecen los cristales obtenidos en placas de 96 pocillos mediante la dispensación automática de gotas usando los robots de la Plataforma Automatizada de Cristalografía (PAC) del PCB, donde se muestra la evolución de la

gota en 3 días. En la parte inferior se muestran algunas imágenes de los cristales obtenidos después del proceso de escalado a placas de gotas de 24 pocillos.

Estos cristales fueron difractados en la línea ID14-4 en la Instalación Europea de Radiación de Sincrotrón (ESFR, Grenoble). Los primeros datos obtenidos mostraban una difracción aún débil ($\sim 4 \text{ \AA}$), pero sirvieron para determinar el grupo espacial y la celdilla unidad. A partir de este momento se realizaron todo tipo de ensayos de optimización de estos cristales (Fig_C2) (*macroseeding*, *microseeding*, ensayos de deshidratación química, uso de aditivos (TCEP, sarcosine), empleo de aceites para disminuir la velocidad del proceso de vapor de difusión, etc.), así como jugar alrededor del oligo empleado. Ello implica jugar con el oligo de 22pb (llamado LSP22) alterando su longitud y tipo de extremos de forma conservadora. Pese a la batería de oligos creados alrededor del oligo LSP22, tan sólo se obtuvieron cristales con uno de los oligo variantes, y estos cristales difractaban a una mayor resolución que TFAM-LSP22 (Fig_C3). El resto dieron lugar a la formación de núcleos cristalinos, pero no cristales.



Fig_C2. Cristalización del complejo TFAM-LSP22, variando el tampón de la solución de cristalización y su pH (indicados en A, B y C). Es en el caso de la cristalización en las condiciones indicadas en (A) donde los cristales crecen y difractan mejor.



Fig_C3. Esquema de los diferentes oligos variante del oligo LSP22 y su cristalización en la condición: 37%PEG1000, 0.35M NaCl, 0.1M Na/K fosfato pH6.2. En azul se muestra la secuencia del oligo LSP22. Excepto el último oligo con extremos romos, los demás presentan extremos cohesivos. Tan solo con un de los oligos variante se obtuvieron cristales, para los otros casos sólo se obtuvieron estados de fase cercanos a la nucleación.

Mediante los procesos de optimización aplicados a los cristales TFAM-LSP22 fue posible aumentar la resolución hasta 3.2 Å en ID23-1, pertenecientes al grupo espacial ortorrómbico primitivo (P222) y con los siguientes parámetros de celdilla: a=56.12, b=118.41, c=120.05; $\alpha=\beta=\gamma=90^\circ$. Cuando se escalaron los datos, las estadísticas indicaban baja completitud (88.6 Å) y multiplicidad (4.1) de los datos, y factores de error elevados (R_{merge}) (R_{merge} en la última capa de 61 %). Pese a ello se probó resolver la estructura mediante reemplazo molecular con el programa *Phaser* usando la estructura resuelta del segundo dominio HMGbox de TFAM (Gangelhoff, Mungalachetty et al. 2009), (código de acceso en pdb: 3FGH). Debido a que la calidad de los datos y la resolución no eran suficientes, no fue posible encontrar una solución válida. Además, el porcentaje de identidad era reducido. Los valores de *z score* eran de 4, cuando se esperarían valores por encima de 7 o 8).

1.1. Confirmación de la presencia de ADN en los cristales

Para ratificar que los cristales eran producto de la cristalización de la proteína con el ADN se realizó un ensayo simple, que consistía en cargar algunos de los cristales obtenidos en gel nativo para un ensayo EMSA y teñidos con SyBr, que detecta el ADN. Mediante este

pequeño ensayo se pudo constatar que los cristales estaban compuestos por proteína y ADN, y que en el caso de la proteína esta parecía no estar degradada (Fig_C4).

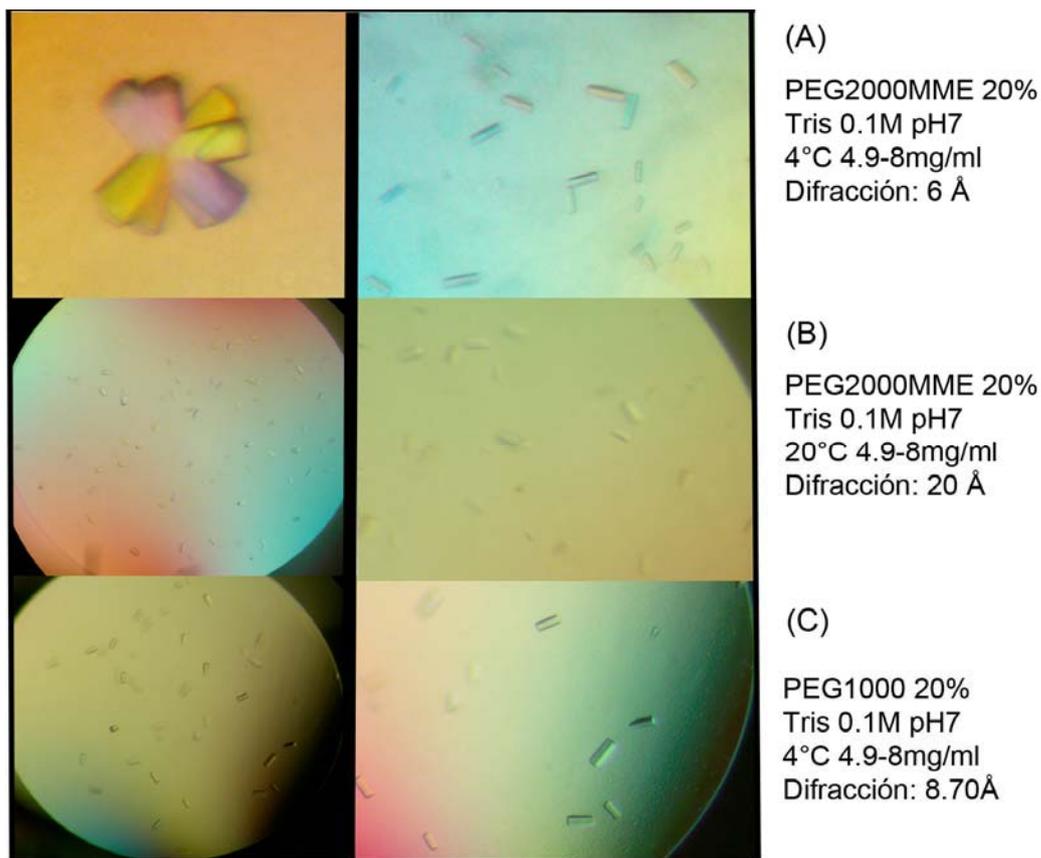


Fig_C4. Gel nativo EMSA al 5% para TFAM-LSP22. Muestras **xtall**: mezcla de cristales para TFAM-LSP22 disueltos en el tampón del complejo; **cmplx**: complejo TFAM-LSP22 en solución; **LSP22**: oligo de 22pb usado para la formación del complejo y posterior cristalización. Mediante este ensayo se ratifica que los cristales presentan la misma composición que el complejo en solución. Es interesante notar que LSP22 no migra como una banda única, hecho que se caracterizó posteriormente.

2. Cristalización TFAM-Half1

Durante la optimización de los cristales TFAM-LSP22, también se obtuvieron cristales para la secuencia de oligo Half1 de 15pb en diferentes condiciones de cristalización y temperatura (Fig_C5). Estos cristales presentaban un aspecto más robusto, con caras definidas y crecimiento más tridimensional. Pese a la gran cantidad de cristales en diferentes condiciones, la resolución de la difracción de estos era pobre, no llegando a más de $\sim 6 \text{ \AA}$ (difracción en ID14.4).

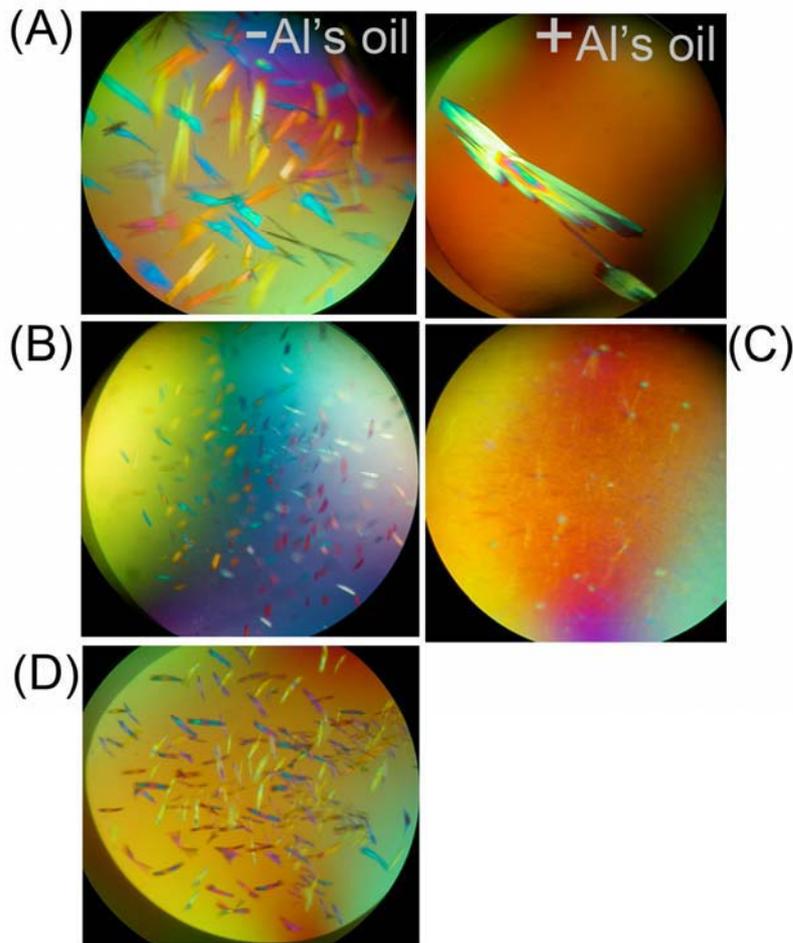
De la misma forma que para el complejo TFAM-LSP22, en este caso se llevaron a cabo procesos de optimización y diseño de nuevos oligos variante de Half1. Pese a ello, no se consiguió mejorar la resolución de la difracción.



Fig_C5. Cristalización TFAM-Half1. (A), (B) y (C) muestran los cristales para diferentes condiciones de cristalización y temperatura, también se muestran los valores de difracción obtenido en la línea ID14.4 en Junio 2009.

3. Cristalización TFAMSeMet-LSP22

Con el objetivo de resolver el problema de la fase de forma experimental se decidió derivatizar la proteína con seleno metionina (SeMet) durante la expresión mediante el uso de un medio preparado para ello. De forma sorprendente, la cristalización de TFAM-SeMet en complejo con el oligo LSP22 se daba en condiciones de cristalización mucho más extensas en gota pequeña, que luego se escalaron a gota grande y se optimizaron (Fig_C6). Además, los cristales pese a presentar la misma forma de aguja, resultaban más robustos. Finalmente fue alrededor de las misma condición que los cristales TFAM-LSP22 (37%PEG1000, 0.35M NaCl, 0.1M Na/K Phosphate pH 6.2) donde se obtuvieron los cristales con mayor resolución (~ 2.6 Å). Los datos de estos cristales permitieron trazar la totalidad de la estructura, incluyendo proteína y ADN.



Fig_C6. Optimización de los cristales obtenidos para TFAMSeMet-LSP22 complejo a una concentración de 8mg/ml. (A) Condición: PEG1000 30%, NaCl 0.25 M, Na/K phosphate 0.1 M pH 6.2, similar a la condición de los cristales nativos. Mediante el uso de *Al's oil* fue posible disminuir la nucleación y favorecer el crecimiento. (B) Condición: PEG 3350 20%, NH₄F 0.2M. (C) Condición: PEG1000 30%, NaCl 0.25M, 0.1M HEPES pH7.5. (D) Condición: PEG1000 30%, NaCl 0.25M, 0.1M MES pH6.5.

4. Cristalización TFAM-LSP22Br

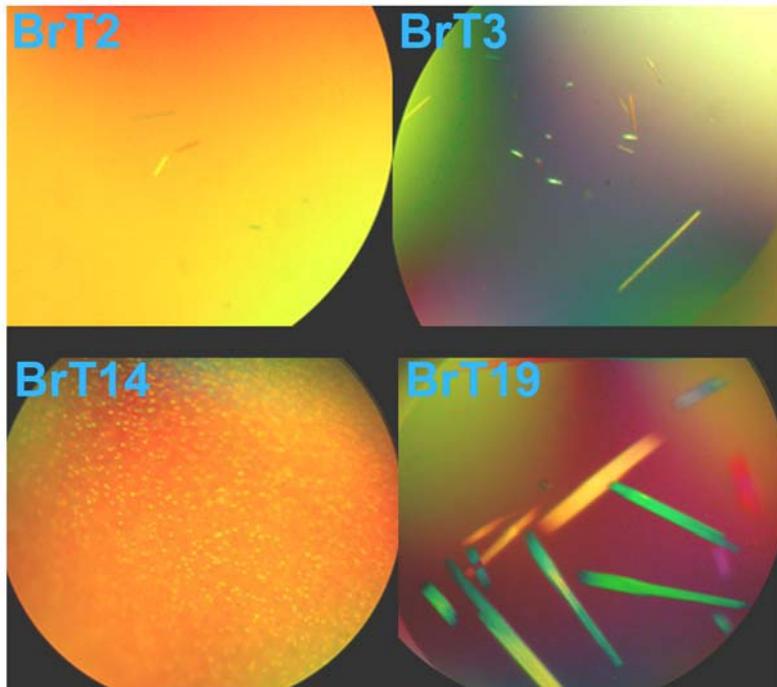
Una vez trazada la cadena polipeptídica de TFAM dentro de la densidad electrónica, se buscó confirmar el orden y orientación de la secuencia nucleotídica del ADN. Por ello se decidió usar ADN derivatizados comercialmente, concretamente substituir una timina de la secuencia original del oligo LSP22 por 5-BromoUracilo. Para ello se escogieron las timinas que no estuvieran implicadas en interacciones proteína-ADN y que presentasen un menor *Bfactor* o factor de temperatura (indicados en azul a continuación).

CadenaC 5'-T1 A2 A3 C4 A5 G6 T7 C8 A9 C10 C11 C12 C13 C14 C15 A16 A17 C18 T19 A20 A21 C22-3'

CadenaD 3'-A22 T21 T20 G19 T18 C17 A16 G15 T14 G13 G12 G11 G10 G9 G8 T7 T6 G5 A4 T3 T2 G1-5'

Los ensayos de cristalización con estos oligos se realizaron alrededor de la condición del complejo nativo. Cabe destacar que la manipulación de estos complejos implicaba trabajar bajo condiciones con baja intensidad de luz, ya que el Bromo (Br) es sensible a la luz debido a su fácil excitación. Los mejores cristales se obtuvieron en el caso donde T19 de la

cadena C era derivatizada a Br-Uracilo y estos cristales difractaban hasta 2.1 Å (datos medidos en ID23.2, ESRF) (Fig_C7).



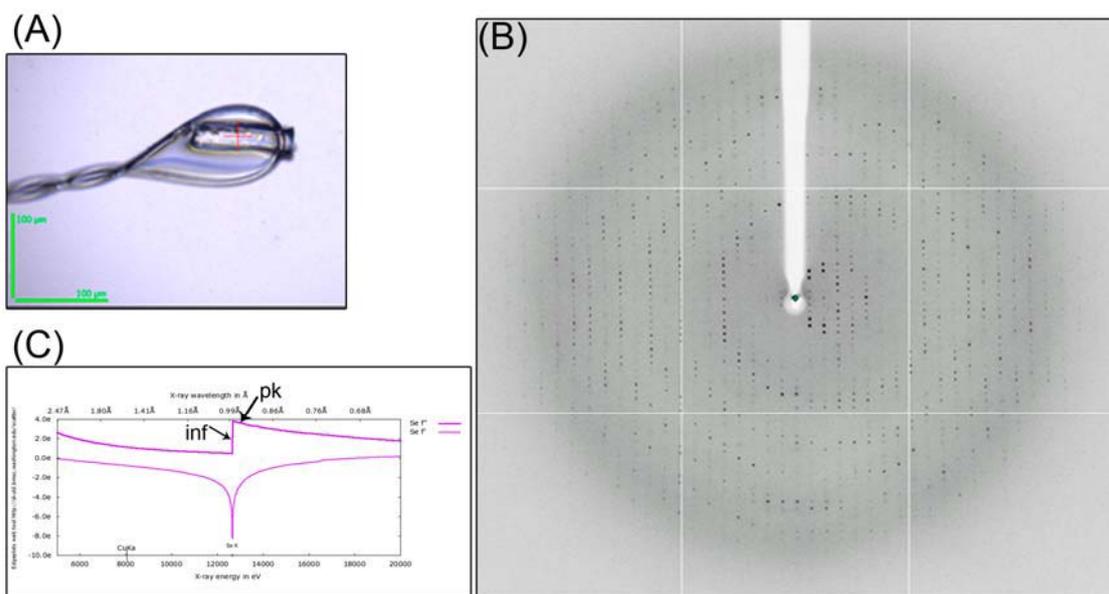
Fig_C7. Cristalización del complejo TFAM-LSP22BrU. Las condiciones de cristalización eran las mismas en todos los casos: 37%PEG1000, 0.35 M NaCl, 0.1 M Na/K Phosphate pH 6.2 a 8mg/ml. Mientras que para BrT14 no fue posible conseguir cristales, en el caso de BrT2 y BrT3 éstos eran lábiles mostrando un mal empaquetamiento del cristal. Los cristales obtenidos con BrT19 presentaban una morfología de aguja y eran robustos.

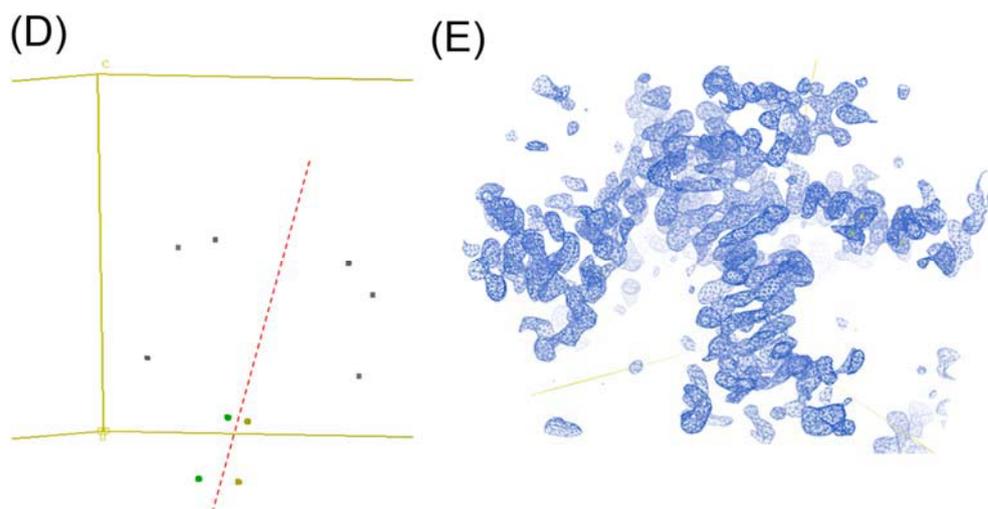
D. RESULTADOS: PROCESAMIENTO DE DATOS Y RESOLUCIÓN DE LA ESTRUCTURA

1. Difracción y procesamiento de los datos de difracción del cristal TFAMSeMet-LSP22

Debido a que no era factible aplicar técnicas de remplazamiento molecular para resolver la estructura nativa de TFAM en complejo con LSP22, se decidió resolver el problema de la fase experimentalmente mediante el uso de proteína derivatizada con SeMet en complejo con el ADN y utilizando el método de dispersión anómala a múltiples longitudes de onda (*Multiple-wavelength Anomalous Dispersion, MAD*).

El mejor conjunto de datos pertenece a un cristal difractado en la ID14-4 del ESRF mediante el uso de un goniómetro cuya orientación kappa es modificable, lo que permite una colección de datos mucho más completa con una precisión ~ 4 microns. Para ello se uso el minikappa a 60° , permitiendo una recogida con todas las reflexiones posibles y elevada redundancia, pero además favoreciendo la presencia de los pares de *Friedel* en la misma imagen (Fig_D1). Para la recogida del set de datos primero se realizó un escaneo de fluorescencia para escoger la longitud de onda óptima para la absorción de Se (fueron colectados datos a dos longitudes de onda distintas, inflexión: ip= 12.6665 KeV y pico: pk= 12.669 KeV).





Fig_D1. Difracción de TFAM SeMet-LSP22. (A) Imagen del cristal difractado dentro del lazo de nylon que permite mantener el cristal en *cryobuffer*. (B) Ejemplo de una imagen del patrón de difracción del cristal TFAM SeMet-LSP22 en ID14.4 ESRF. El patrón de difracción se expande hasta una resolución de 2.6 Å y muestra una ligera anisotropía. (C) Escaneo de fluorescencia para determinar la absorción de Se en el cristal. En ella se indica la energía del punto de inflexión (inf)= 12,666 5Kev ($f' = -9.33$, $f'' = 2.89$), y del pico de absorción (pk)= 12,669 Kev ($f' = -8.25$, $f'' = 5.28$). (D) Posición de los átomos de Se encontrados mediante SHELX. La línea roja indica el eje binario que relaciona los átomos de Se de ambas moléculas. (E) Mapa de densidad electrónica después de llevar a cabo una corrección de anisotropía de los datos y una corrección del solvente. Es posible observar la continuidad en el mapa de densidad, donde es posible diferenciar una hebra de ADN en el eje vertical.

El cristal era ortorrómbico primitivo y contenía 2 complejos proteína/ADN en la unidad asimétrica, con un porcentaje de solvente del 65.83 % y un coeficiente de Matthews de 3.21 Å³/Da. Los datos fueron indexados e integrados con el programa XDS (Kabsch) aplicando una corrección del daño por radiación. El cristal pertenecía al grupo ortorrómbico primitivo, P222, con los siguientes parámetros de celdilla a=56.12; b=118.41; c=120.05; $\alpha = \beta = \gamma = 90^\circ$. Posteriormente estos datos fueron escalados mediante el programa SCALA (Evans 2006), que mediante el cálculo de las extinciones sistemáticas permitía dilucidar que existía un eje helicoidal (l), otro no helicoidal (h) y uno de los ejes (k) no se podía caracterizar. Por lo tanto, dentro del grupo ortorrómbico primitivo podía tratarse de cristales con simetría P222₁ o P2₁2₁2.

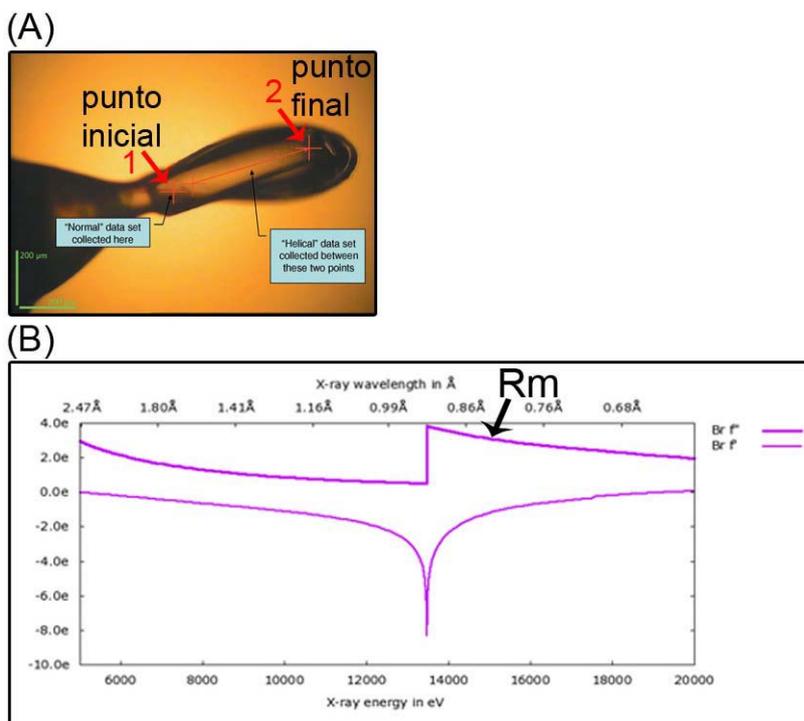
La estructura fue resuelta mediante el método de difracción anómala (*single-wavelength anomalous diffraction, SAD*) con el programa SHELXD (Sheldrick 2008), ya que los datos fueron recogidos a la longitud de onda del pico. Mediante SHELXD fue posible identificar 10 sitios correspondientes a 5 de los seis residuos de SeMet que presentaba TFAM derivatizada (la primera metionina hace referencia al inicio de traducción y normalmente es eliminada); por lo tanto, se ratificó la existencia de dos proteínas por celdilla. La inspección visual de la posición de los átomos pesados (Se) evidencia dos *clusters* de cinco átomos cada uno relacionados por una simetría binaria, es decir, que existe una simetría no cristalográfica (Fig_D1 (D)). El faseado consecutivo con SHELXE reveló que el grupo espacial correcto era P2₁2₁2, presentando una *pseudos-free correlation coefficient* significativamente diferente (debería ser mayor) respecto P222₁. Además mediante SHELXE se aplicó una corrección anisotrópica al mapa de densidad electrónica para poder

realizar el trazado de la cadena polipeptídica y secuencia nucleotídica mediante el uso de COOT (Emsley and Cowtan 2004). También se aplicó una corrección del solvente del mapa de densidad electrónica mediante el programa DM (*Density Modification*) (Cowtan and Main 1996) (Fig_D1 (E)).

Seguidamente, se procedió al refinado con REFMAC5 (Murshudov, Vagin et al. 1997), que incluía refinado mediante TLS (*Translation Libration Screw-motion*), restricciones como NCS (*non-crystallographic symmetry*) en regiones específicas (residuo 51-56; 104-108; 128-131; 143-147; 158-160), y fue necesario construir una librería para establecer los pares de bases Watson y Crick para el ADN. El proceso de refinado se alternó entre uso del programa REFMAC5 y construcción a mano hasta llegar a un modelo final a una resolución de 2.75 Å con los siguientes valores de los indicadores $R_{\text{factor}} = 0.24$ Å y $R_{\text{free}} = 0.30$. El modelo final incluía los residuos Ser43-Gln238 y Ser43-Thr234 para ambas moléculas de proteína, y la secuencia entera del oligo LSP22. La representación de Ramachandran presentaba el % de residuos en regiones favorecidas, que era del 94.04% y debería ser > 98%. Además, la estructura presentaba un 8.6% de rotámeros pobres, parámetro que debería ser < 1%.

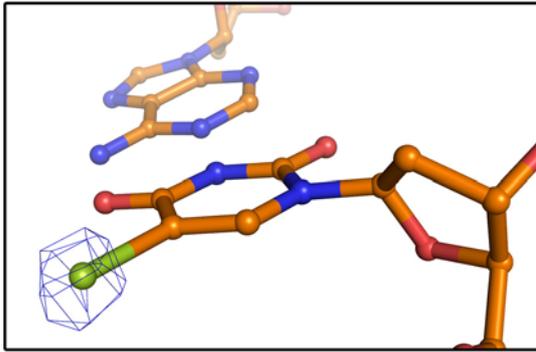
2. Difracción y procesamiento de TFAM-LSP22Br

Para confirmar la orientación de la secuencia nucleotídica en la estructura fueron diseñados oligos LSP22 con bases brominadas en determinadas posiciones (tal y como se ha comentado en el apartado de cristalización). Los cristales que mejor difractaron fueron aquellos con la secuencia mutada en la T19 de la cadena C (llamado BrT19), hasta 2.1 Å (Fig_D2). La recogida de datos se realizó en ID23.1 de ESRF, línea que presenta una longitud de onda fija ($\lambda = 0.872600$ Å, $E = 14.290$ KeV). Esta longitud de onda se encuentra situada en el patrón de absorción de Br más a la derecha, es decir, se realizó una recogida de los datos de difracción en un punto remoto. Para evitar daño por radiación en el cristal y la pérdida de la señal del bromo optamos por una fuerte atenuación de la energía y una recogida de datos de forma helicoidal, mediante el uso de la interfaz *helical collection* dentro de MxCube en el ESRF (Gabadinho, Beteva et al.).



Fig_D2. Estrategia de recogida para los cristales TFAM-LSP22Br. (A) Imagen del cristal colocado dentro del lazo, en este caso se aplicó una recogida de tipo helicoidal, que mediante la selección de un punto inicial (1) y un punto final (2) permite la recogida desplazada a lo largo del cristal, con un poder de difracción similar y estable (*Flot_2010*). (B) Escaneo de fluorescencia para determinar la energía de absorción de átomo de bromo en el cristal. Debido a que la línea utilizada era de longitud de onda fija, ello implicó hacer una recogida en un punto desplazado a la izquierda del pico de absorción $wv = 0.872600 \text{ \AA}$, $Kev = 14.290$.

Debido a que estos cristales difractaban mejor, el conjunto de datos fue usado para llevar a cabo un remplazamiento molecular usando como modelo el pdb obtenido a través de TFAMSeMet-LSP22. El proceso de refinado implicó varios ciclos con el programa REFMAC (incluyendo inicialmente pasos de curepo rígido, TLS, NCS). Los últimos pasos de refinado se llevaron a cabo mediante el programa AUTOBUSTER (Smart, Womack et al.), que realiza un refinado mediante una novedosa restricción basada en LSSR (*Local Structure Similarity Restraints*) que permite explotar la información de estructuras relacionadas y facilitar el uso de NCS. Dentro del programa AUTOBUSTER fueron usadas las aplicaciones Gelly, que permiten fijar restricciones de geometría para evitar violaciones relacionadas con la construcción a mano (útil para el ADN), TLS y *form factor correction* (permite reunir los datos a una longitud de onda donde el factor de dispersión es significativamente diferente). El mapa de densidad de los datos con fases para TFAM-LSP22Br con el modelo parcial TFAMSeMet-LSP22 mostraba un pico de señal anómala ($>5\sigma$) en la posición esperada para el bromo, confirmado la secuencia de ADN (Fig_D3).



Fig_D3. Mapa de densidad anómala (Fo-Fc) donde se revela una señal $> 5\sigma$ correspondiente al Br de la base de bromoUracil que substituye T19 de la cadena C del oligo LSP22.

La estructura final incluía los residuos Ser44-Lys237 para la cadena B y Ser44-Arg233 para la cadena B, así como la secuencia nucleotídica completa para el oligo LSP22. La superimposición de la estructura obtenida mediante TFAMSeMet-LSP22 y TFAM-LSP22Br no revela grandes cambios, de forma que la presencia del ADN brominado no implica cambios estructurales en el conjunto global de la estructura. Por lo tanto, finalmente se obtuvo la estructura de TFAM-LSP22Br con unas estadísticas de refinado óptimas para la resolución final de 2.45 Å, $R_{\text{factor}} = 18.2 / R_{\text{free}} = 22.8$ (Tabla_D1) y un mapa de densidad electrónica de suficiente calidad como para diferenciar el *stacking* entre las bases pirimidínicas y purínicas del ADN (Fig_D4). El análisis de la geometría final mediante el servidor *MolProbity* (Chen, Arendall et al.) permitió observar que el 98.4% de los residuos se encuentran en zona favorables (superior al 94.04 % de la estructura TFAM SeMet-LSP22)

TFAM-Br in complex to LSP-22

Data collection:	
Space group	P2 ₁ 2 ₁ 2
Cell constants	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	113.9, 117.2, 56.53
$\alpha = \beta = \gamma$ (°)	90
Wavelength (Å)	0.87260
Resolution range (outer shell) (Å)	40.84 – 2.45 (2.58–2.45)
<i>R</i> _{sym} (outer shell) (%) [*]	8.0 (46.7)
Average intensity (outer shell) ($\langle I \rangle / \sigma(\langle I \rangle)$)	17.5 (4.2)
Completeness (outer shell) (%)	99.8 (100)
Redundancy	7 (7.4)
Refinement:	
Resolution (Å)	40.84 – 2.45
No. reflections	28,483
<i>R</i> _{work} / <i>R</i> _{free} (%) [‡]	18.2 / 22.8
No. atoms:	
Protein	3238
DNA	1797
Water	185
Ligands	28
<i>B</i> -factors [†] :	
Protein	42.5
Ligand/ion	40.8
Water+ligands/ion	40.2
R.m.s deviations:	
Bond lengths (Å)	0.012
Bond angles (°)	1.4
Main-chain conformational angle analysis [§]	
Nr of residues in favored regions / outliers / all residues	376 / 0 / 4

^{*} $R_{sym} = \sum I - \langle I \rangle / \sum I$, where *I* is the observed intensity and $\langle I \rangle$ is the statistically weighted absolute intensity of multiple measurements of symmetry related reflections.

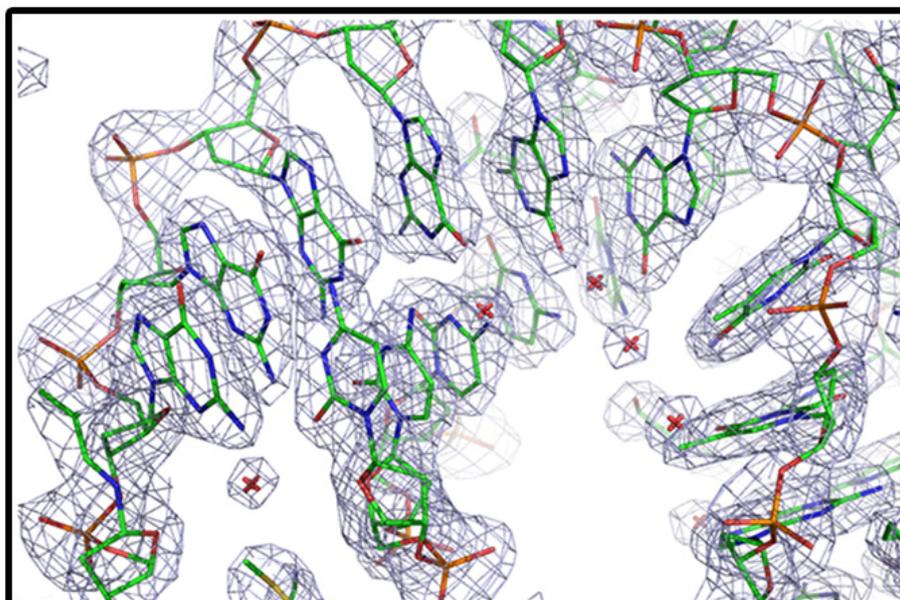
[‡] MPE = Mean Phase Error.

[‡] $R = \sum |F_o - k|F_c| / \sum |F_o|$, *R* from the working set and *R*_{free} from the test set.

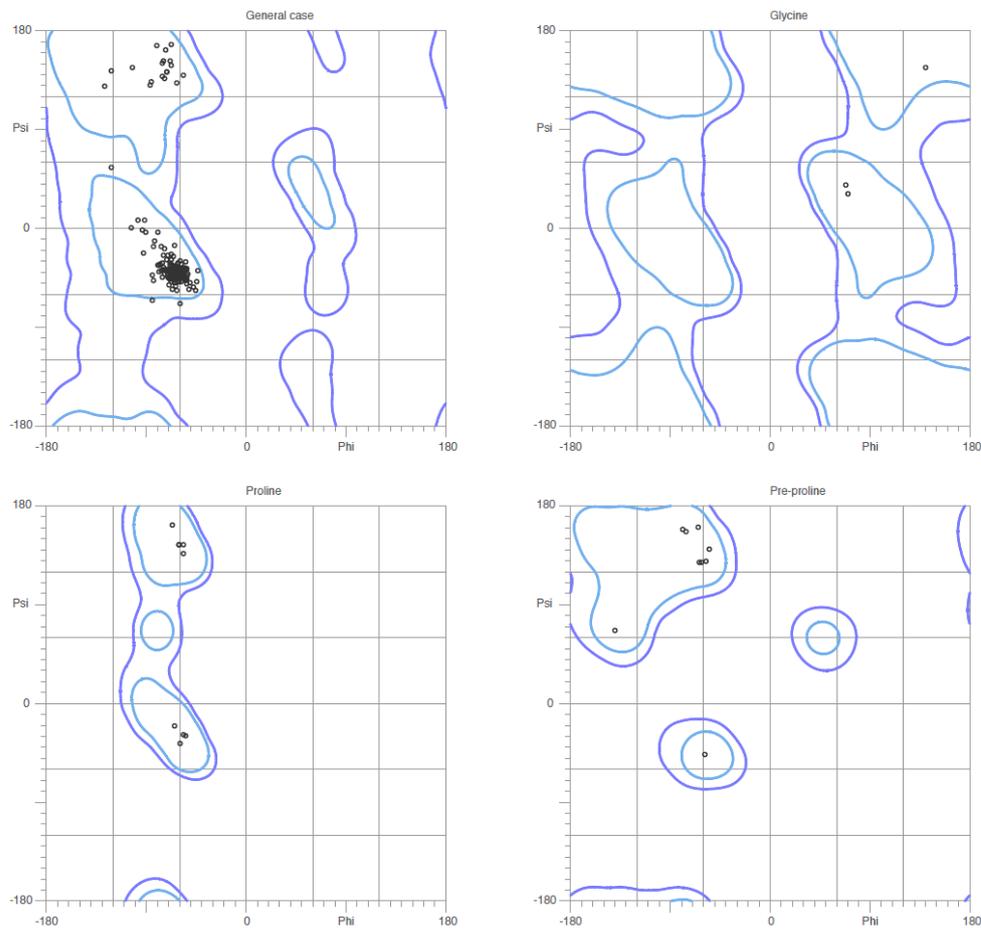
[†] *B* factors as after Refmac5/Phenix refinement including TLS.

[§] According to MOLPROBITY.

Tabla_D1. Estadísticas del procesado y refinado para TFAM-LSP22Br.



Fig_D4. Mapa de densidad electrónica definido para la estructura TFAM-LSP22. El mapa de densidad electrónica se encuentra focalizado en una región del ADN, donde se muestra el *stacking* presente entre las bases del ADN, además es posible distinguir las moléculas de agua.



Fig_D5. Representación de Ramachandran para la estructura TFAM-LSP22Br. 376 de los residuos se encuentra en regiones Ramachandran favorecidas y todos los residuos (380) están en regiones permitidas.

E. RESULTADOS: ANÁLISIS ESTRUCTURAL DE TFAM-LSP22

La proteína madura TFAM no cristalizó sin la presencia de ADN, los mejores cristales fueron obtenidos con el oligonucleótido de la región de protección de TFAM para *Light Strand Promoter* (LSP) de una extensión de 22 pb.

La determinación estructural mediante métodos experimentales, que incluyen datos de rayos X de cristales con proteína derivatizada con selenio-metionina (SeMet-TFAM-LSP22) o ADN derivatizado con Bromo-Uracilo (TFAM-LSP22Br), permitieron obtener un mapa de densidad electrónica suficiente para asignar la secuencia de TFAM y ADN. Un análisis detallado de las interacciones proteína-ADN muestran que ambos cristales no presentan una diferencia substancial de la estructura de proteína nativa.

1. Curvatura del ADN ofrecida por la proteína en el complejo TFAM-LSP22

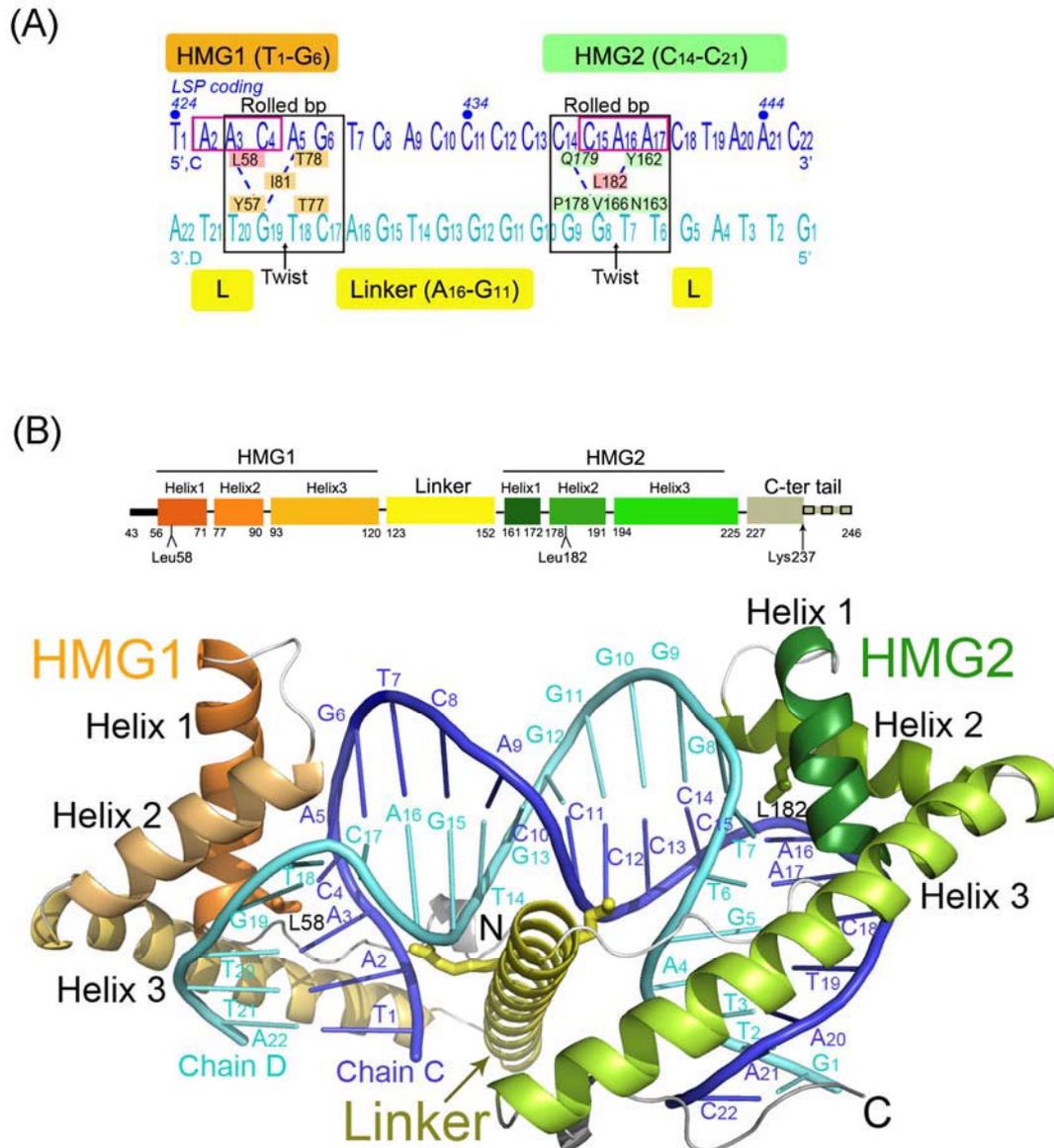
TFAM es una proteína α -modular, que comprende dos dominios HMGbox (HMG1 y HMG2), cada uno de ellos se extiende unos 75 residuos (residuos 44-120 y 125-225, respectivamente) (Fig_E1). Estos se encuentran conectados por un linker en conformación helicoidal (residuos 124-152) y, finalmente, se halla una cola C-terminal en conformación prácticamente extendida (residuos 226-237). Ambas HMGbox consisten en 3 hélices- α plegadas en la típica y conservada disposición de L que presentan los dominios HMGbox. La superficie interna de la L contacta con el surco menor del ADN mediante interacciones polares o no-polares, con residuos específicos que parcialmente o de forma total se intercalan entre las bases. Los dos dominios HMGbox presentan una orientación o disposición de cabeza con cabeza, donde el brazo corto de la L, que incluye las hélices 1 y 2, están orientados hacia la región central de la estructura global y relacionada por un eje de simetría binaria. A diferencia de la orientación cola-cola para el constructo sintético diseñado mediante ingeniería genética SRY.B, que presenta un complejo estable con el ADN (fusión de HMGbox de unión específica con HMGbox de unión inespecífica), donde los dos brazos cortos de la disposición L se encuentran mirando hacia el exterior del eje central (Stott, Tang et al. 2006) (detallado más adelante).

TFAM HMG1 contacta la región 1 de LSP, que comprende la secuencia T₁A₂A₃C₄A₅C₆, mientras que HMG2 contacta la región 2 de LSP, con secuencia C₁₄C₁₅A₁₆A₁₇C₁₈T₁₉A₂₀A₂₁. En estas regiones la proteína allana y ensancha el surco menor del ADN mediante la separación del esqueleto de fosfatos de ambas cadenas, imponiendo una curvatura final del ADN de 180°. Dentro de los dominios HMGbox cabe destacar la existencia de dos residuos que se intercalan en el ADN, Leu58 para HMG1 y Leu182 para HMG2, en los puntos de mayor curvatura en las regiones definidas.

Entre estos dos dominios HMGbox, que se encuentran interaccionando con los surcos menores del ADN, existe un linker helicoidal potencialmente básico alojado en el surco menor de la zona del ADN situada entre la región 1 y región 2, con conformación típica de B-DNA, exactamente en el surco menor (Fig_E1(B)). La orientación del linker es perpendicular al ADN y al eje de simetría que relaciona los dos dominios HMGbox. El linker destaca a simple vista por el hecho de que se encuentra estabilizando la curvatura

inducida sobre el ADN por los dos dominios HMGbox mediante interacciones no específicas con los extremos del ADN que sobresalen de la estructura (Fig_1E(A)).

Finalmente, encontramos una cola C-terminal posterior a HMG2. En la estructura final fue imposible trazar toda la cola C-terminal, faltando 9 residuos finales, es decir, se trazó hasta Lys237 (Fig_1E(B)). Esta cola C-terminal, que al igual que el linker presenta un potencial extremadamente básico, se ha relacionado con la habilidad de TFAM para la unión específica al ADN. Sin embargo, en la presente estructura esta cola C-terminal no muestra una interacción específica con las bases del ADN.



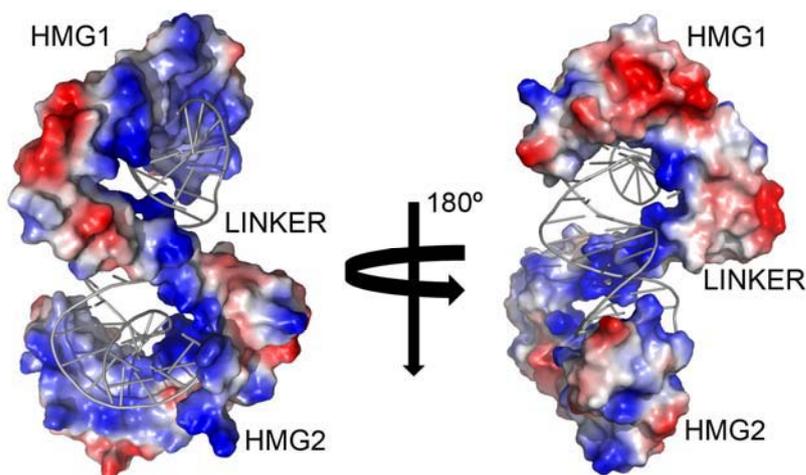
Fig_E1. Complejo TFAM-LSP22. (A) Secuencia de LSP cristalizada (en azul, más intenso para la cadena C o rica en citosinas, y azul cielo para la cadena D o rica en guaninas). Las cajas naranja, verde y amarilla simbolizan los dominios HMG1, HMG2 y el linker (L) respectivamente, que contactan con el ADN. Los residuos de TFAM que intercalan el ADN se muestran con fondo rosado, mientras que los residuos que realizan contactos con las bases se muestran siguiendo el código de colores de los dominios. Las cajas rojas indican el motivo AAC

inverso y las cajas negras las regiones donde existe una fuerte desviación en los valores de *twist* y *roll*. Las líneas discontinuas representan los puentes hidrógeno tipo enlace bifurcados (B) Representación de los dominios de TFAM a lo largo de la secuencia, mostrando los residuos que se intercalan y el último residuo trazado en la estructura. Debajo se muestra una representación de la estructura 3D en forma de lazos. Los dominios HMG1 y HMG2 con sus respectivas hélices se encuentran coloreados en naranja y verde, respectivamente. El linker se representa en amarillo, mientras que las regiones flanqueantes, los extremos N-terminal y C-terminal, en gris. Los residuos que se intercalan entre el ADN (Leu58 y Leu182) se muestran en forma de barras.

El modo en el que el ADN se encuentra curvado en la estructura TFAM-LSP22 muestra similitudes con lo que sucede en la familia de proteínas HU, que son proteínas de unión al surco menor del ADN con una funcionalidad de tipo arquitectural en los nucleoides de procariotas (Rice, Yang et al. 1996; Mouw and Rice 2007). Dentro de esta familia de proteínas HU encontramos proteínas como IHF (*Integration Host Factor*), que induce una curvatura en el ADN de hasta 180°, dando lugar a un forma final de U al ADN. Esta proteína forma dímeros, y une al ADN mediante la intercalación al surco menor de una hoja β o " *β -ribbon arm*". La interfaz de dimerización entre las dos subunidades es rica en residuos positivos y sirve para neutralizar la repulsión existente entre las cargas negativas del esqueleto fosfato.

2. Superficie electrostática y conservación general en TFAM

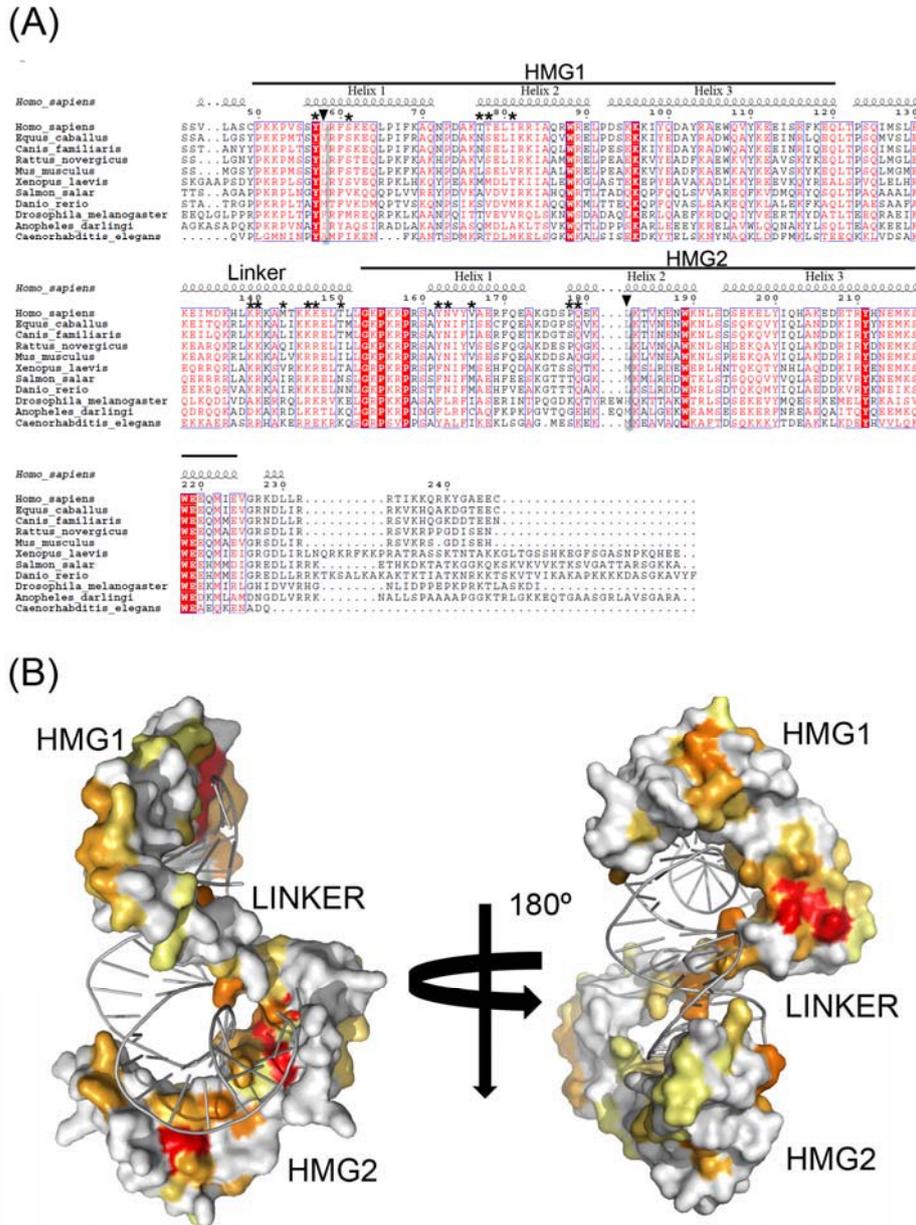
Al analizar el potencial electrostático y representarlo sobre una superficie de Connolly de la estructura de TFAM en complejo con ADN, observamos que se trata de una proteína con un elevado potencial básico. En la superficie se distinguen tres regiones cargadas positivamente, que coinciden con las regiones de máxima interacción con el ADN: cara interna de HMG1, cara interna de HMG2 y la región del linker (Fig_E2).



Fig_E2. Representación de la superficie electrostática de TFAM en complejo con ADN, en dos orientaciones diferentes relacionadas por 180°. En azul se muestran las regiones cuyo potencial electrostático está cargado positivamente y en rojo, cargado negativamente.

Al realizar una alineamiento de TFAM con diferentes homólogos en organismos multicelulares (Fig_E3 (A)), es posible observar que existe un elevado grado de conservación entre las diferentes TFAM, excepto en la zona de la cola C-terminal. En el caso de la Cola C-terminal existen diferentes longitudes e incluso ausencia de cola C-

terminal, como es el caso del homólogo de TFAM en levadura (ABF2p). Al representar la conservación existente en TFAM, con una degradación de colores de rojo a blanco en función del grado de conservación de más a menos (Fig_E3 (B)), es posible observar que existe una distribución de residuos conservados en regiones de la estructura. Concretamente, cabe destacar que las regiones determinadas, donde existe interacción proteína-ADN hay un gran número de residuos conservados, así como en el linker.

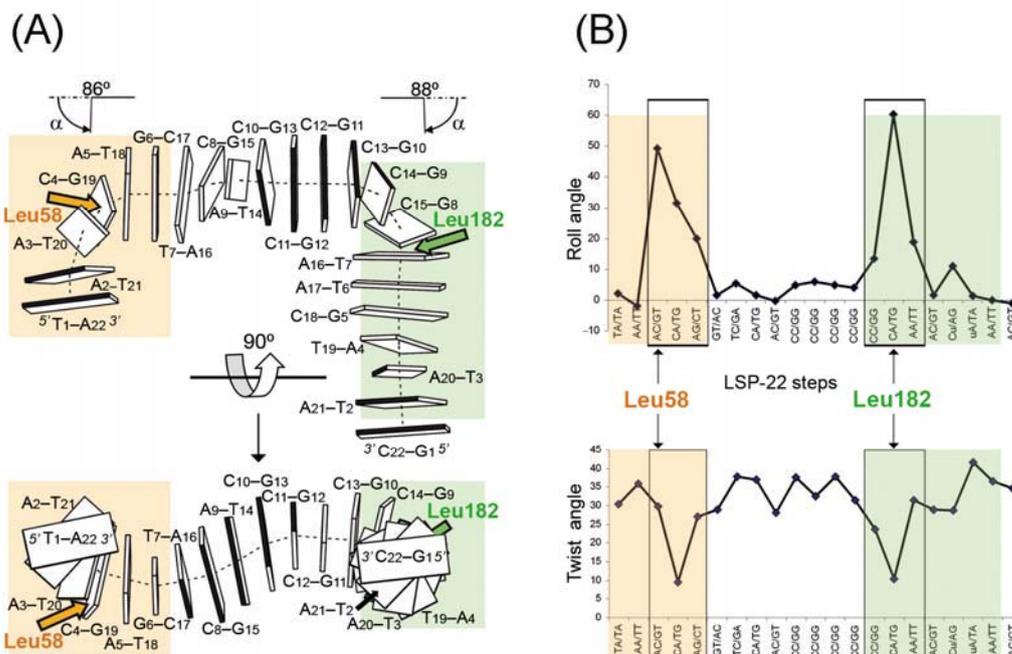


Fig_E3. Conservación entre TFAM homólogas. (A) Alineamiento entre diferentes TFAM de organismos pluricelulares (*metazoos*), donde se indican los dominios en tándem HMGbox, el linker y por último la cola C-terminal. El grado de conservación o identidad entre residuos se indica en color rojo, donde los residuos con fondo rojo están conservados al 100%. Los triángulos indican los residuos de intercalación y con estrellas, residuos destacados por su interacción con el ADN. (B) Representación de TFAM en complejo con LSP22, donde la coloración representa el grado de conservación de los residuos: 0-30%= blanco, 31-50%= amarillo, 51-69% = amarillo intenso, 70-89% = naranja; 90-99% = naranja fuerte; 100% = rojo.

3. Características del ADN LSP22 en la estructura TFAM-LSP22

En la estructura, el ADN sufre dos curvaturas muy marcadas de $\sim 90^\circ$, cada una de ellas provocada por una HMGbox, dando como resultado una curvatura global de 180° y quedando en forma de U (Fig_E4 (A)). Dentro de las regiones 1 y 2 podemos definir los puntos de mayor curvatura que consiste en la secuencia $A_3C_4A_5G_6$ para la región1 y $C_{14}C_{15}A_{16}A_{17}$ en la región 2. En un B-ADN canónico, la distancia entre los fosfatos distanciados por cuatro bases en el surco menor es de 11.7 Å. En el caso de la región1, la distancia entre T_7 (cadena C) y T_{20} (cadena D) es de 22.4 Å, mientras que para la región2 la distancia es de 21.9 Å, entre A_{17} y G_{10} (Fig_E4 (A)). Ello indica el ensanchamiento que la proteína causa en el surco menor.

En estas dos regiones existe un valor extremadamente positivo de *roll* (con máximos de 50° y 60°) debido a la fuerte curvatura inducida, que hace volcar el surco menor hacia el surco mayor, mientras que el valor de *twist* disminuye (con un mínimo de 10°) en el par de bases central (C_4A_5/T_8G_9 , $C_{15}A_{16}/T_7G_8$) con lo que el surco menor se encuentra doblado y abierto. En la cara interna de cada doblamiento, es decir, en el surco mayor de las regiones de LSP 1 y 2, los pares de bases con mayor valor de *roll* realizan interacciones Watson-Crick regulares, así como enlaces de hidrógeno bifurcados entre la misma cadena (Coll, Frederick et al. 1987) (Fig_E1 (B), Fig_E4 (A), (B)). Además existen moléculas de agua coordinadas por los átomos de las bases, que estabilizan la torsión. En la cara externa de la curvatura, es decir, en la parte del surco menor que interacciona con HMGbox, los átomos de oxígeno del esqueleto fosfato son estabilizados mediante contactos polares y cargas positivas de los residuos que conforman las tres hélices- α .



Fig_E4. Características de LSP22 en el complejo TFAM-LSP22, que presenta una forma de U, con las regiones 1 y 2 en naranja y verde, que simbolizan las zonas de contacto con los dominios HMG1 y HMG2, respectivamente.

(A) Representación en pastilla de los pares de bases que conforman el ADN LSP22, donde los puntos de intercalación son indicados con una flecha. (B) Representación gráfica de los parámetros *roll* y *twist*.

4. Contactos TFAM con ADN

4.1. Dominios HMGbox y su interacción con la región 1 y 2

Como se ha detallado anteriormente los dominios HMGbox de TFAM interactúan con las bases del ADN. En HMG1, Leu58 de la hélice1 se intercala entre A₃ y C₄ de la cadena C (rica en citosinas) dificultando su *stacking* y contribuyendo fuertemente al incremento de *roll* en este paso. Esta distorsión se encuentra estabilizada por un residuo cercano altamente conservado, Tyr57 de la hélice 1, que parcialmente se intercala entre las bases T₂₀ y G₁₉ de la cadena D, y realiza contactos del tipo puente de hidrógeno con G₁₉. La hélice 2 de HMG1 también contribuye a la distorsión del ADN, mediante la intercalación parcial de Thr77, Thr78 e Ile81 (Fig_E5 (A)).

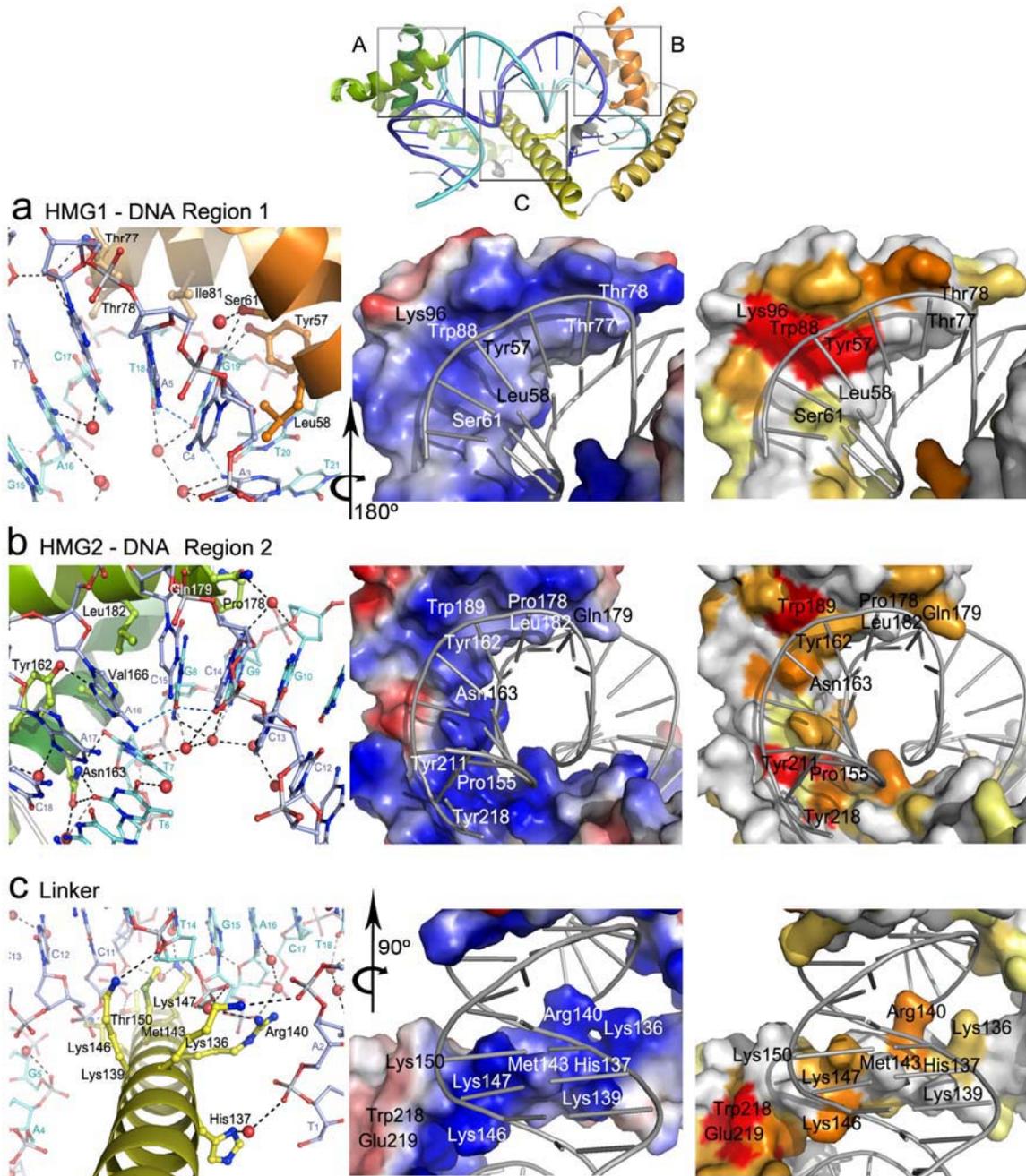
Por lo que se refiere a HMG2, éste presenta un patrón de interacción diferente con la región2. En la posición equivalente a HMG1-Leu58, un residuo polar, Asn163, no se intercala pero realiza puente de hidrógeno con T₆ y T₇ de la cadena D. La cadena lateral del residuo precedente, Tyr162, altamente conservado e incluido en el core hidrofóbico de HMG2, muestra una orientación que permite formar puente de hidrógeno con A₁₆ de la cadena rica en citosina (cadena C), similar al residuo topológicamente equivalente en HMG1, Tyr57. La mayor distorsión dentro de la región 2 está provocada por un residuo hidrofóbico de la hélice 2 de HMG2, Leu182, que se intercala entre C₁₅-G₈ y A₁₆-T₇. A diferencia de su homólogo topológicamente en HMG1, Ile81, que sólo intercala parcialmente. La intercalación de Leu182 se encuentra compensada por Val166, situada en la hélice 1 de HMG2, que semi-intercala entre T₇ y G₈ de la cadena D. Además existen diversos residuos que estabilizan esta distorsión de la región2, situados en el loop formado entre hélice 1 y hélice2 descrito como γ -turn (Ser177-Glu180) (Fig_E5 (B)).

4.2. Linker y su interacción con el ADN

En la estructura, el linker se encuentra contactando HMG1 y HMG2, y compensando la repulsión del esqueleto fosfato del ADN, que se encuentra cercano debido a la torsión provocada por la unión de HMGbox al ADN. En particular, un grupo de residuos positivos situados en el extremo C-terminal del Linker que realizan tres tipos de contacto con el ADN (Fig_E4(C)). Dos de estos contactos son realizados por los residuos altamente conservados Arg140 y Lys146. Arg140 contacta con el surco mayor de la región 1, y Lys 146 realiza el mismo contacto con el surco mayor de la región 2. Estos residuos realizan contactos con los fosfatos del ADN mediante sus cadenas laterales cargadas positivamente, permitiendo estabilizar la curvatura inducida por HMGbox en la otra cara del ADN.

El tercer contacto realizado por el linker es con el surco menor del ADN situado entre la región 1 y la región 2 (regiones de contacto de HMGbox), que comprende las bases desde A₁₆ a G₁₃ (cadena D) y desde A₉ a C₁₁ (cadena C). Lys139, Lys147 y otros residuos cercanos conservados realizan contactos de tipo electrostático, que permiten estabilizar la curvatura en forma de U, sin realizar ningún reconocimiento específico de secuencia. Es decir, el linker yace en el surco menor y estabiliza los dos puntos de curvatura. Mediante su paso perpendicular a través del ADN, estabiliza los contactos de los dos dominios HMGbox al otro lado del ADN y, por lo tanto, contribuye a facilitar la curvatura en forma de

U. En cambio en la estructura de la proteína bidominio artificial SRY.B, el pequeño linker básico permanece estirado a lo largo del surco menor estabilizando la estructura global, pero no realiza ningún contacto que ayude a la estabilización del doblamiento, que es de 101° (Stott, Tang et al. 2006).



Fig_E4: Representación de las interacciones presentes en HMG1 (a), HMG2 (b) y linker (c) con el ADN, siguiendo la pauta de colores original. En el panel de la izquierda se muestra con detalle las interacciones de los residuos proteicos con la secuencia del ADN, los puentes de hidrógenos o enlaces salinos se muestran con un interlineado negro. El panel central muestra los mismos residuos del anterior panel en la superficie de potencial electrostático. Finalmente en el panel de la derecha se muestra la conservación de TFAM con los residuos de interacción con ADN enumerados.

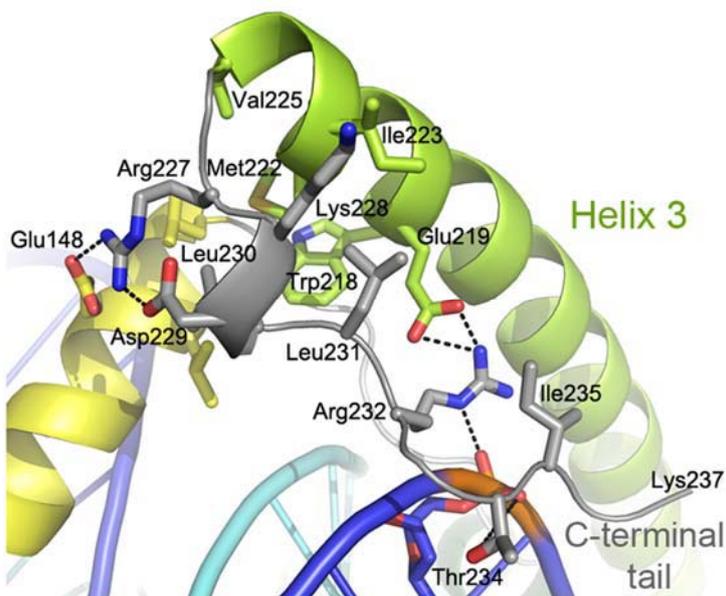
4.3. Cola C-terminal y su interacción con el ADN

La cola C-terminal ha sido descrita como necesaria para el reconocimiento específico del ADN (Dairaghi, Shadel et al. 1995; Gangelhoff, Mungalachetty et al. 2009) y para su interacción con el factor de transcripción B2 para formar el complejo de iniciación de la transcripción (McCulloch and Shadel 2003). En mamíferos, la cola C-terminal está conservada hasta el residuo Lys236 (Fig_E4). Esta cola se encuentra empaquetada de forma antiparalela hacia la hélice3 de HMG2 hasta el residuo Ile235, que se encuentra realizando un contacto tipo hidrofóbico con el residuo Glu219 de la Hélice 3 de HMG2 (Fig_E5). En concreto cabe destacar 4 residuos de la cola C-terminal, Leu230, Leu231, Arg232, Ile235, que realizan contactos de tipo van der Waals con residuos del linker, con la hélice 3 de HMG2 o con ella misma:

- Leu230: linker (Leu152, Leu149), hélice 3 (Met222)
- Leu231: hélice 3 (Met222, Ile223, Glu219)
- Arg232: cola C-terminal (Thr234, leu230)
- Ile235: hélice3 (Glu219)

Además, existen interacciones de tipo electrostático entre Arg232 con Glu219 (localizado en hélice 3), Asp229 con Arg227 (localizado en cola C-terminal) y Arg227 con Glu148 (localizado en *linker*).

Por lo que se refiere a lo contactos que realiza la cola C-terminal con el ADN, tan solo existen 3 contactos con el esqueleto fosfato del ADN, por parte de los residuos Arg232 (con A21 de la cadena C), Arg233 (con C22 de la cadena C), y Thr234 (con C22 de la cadena C) (Fig_E5). Por lo tanto, en la estructura TFAM-LSP22 parece que la cola C-terminal no realiza un reconocimiento específico de secuencia del ADN, ya que no existe interacción con las bases del ADN. Aunque el último residuo trazado es Lys237, que se encuentra expuesto al solvente, los últimos nueve residuos no presentes en la estructura no parecen dar lugar a una zona especialmente cargada positivamente (Q₂₃₈RKYGAEEC₂₄₆).

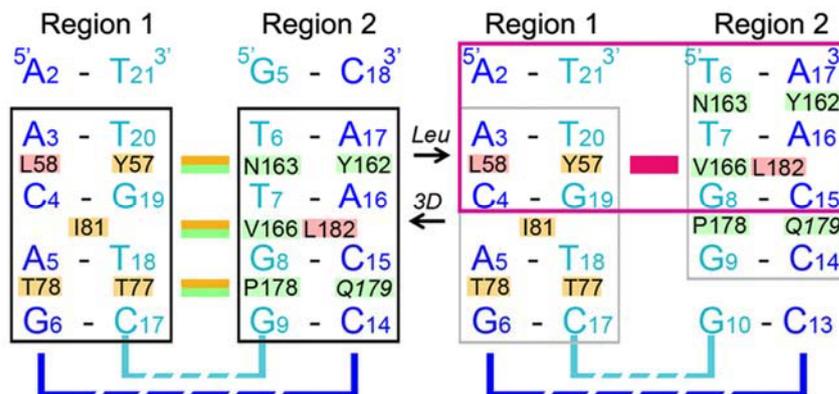


Fig_E5. Representación de la Cola C-terminal (en gris) empaquetada de forma antiparalela contra la hélice 3 de HMG2 (en verde) y que realiza ciertos contactos con el linker (en amarillo). Los residuos que forman parte del

núcleo hidrofóbico o implicados en interacciones con el ADN se muestran en forma de barras. Los oxígenos se muestran en rojo y los nitrógenos en azul; los puentes de hidrógeno o enlaces iónicos se muestran con interlineado. La posición de último residuo trazado, Lys237, finaliza la representación.

5. El motivo simétrico AAC de LSP

Como se ha explicado anteriormente, HMG1 y HMG2 intercalan residuos de diferentes hélice en el ADN: HMG1 intercala Leu58 de la hélice1 entre los pares de base $A_3'C_4A_5G_6$ de la secuencia, mientras que HMG2 utiliza el residuo Leu182, de la hélice2, para intercalarse entre los pares de base $C_{14}C_{15}'A_{16}A_{17}$ (el acento' indica el paso intercalado). Por lo tanto, cada leucina provoca la disrupción del *stacking* en diferentes pasos de su respectiva región. La superimposición de los dos dominios HMGbox (representado esquemáticamente Fig_E6) muestra una coincidencia espacial de los aminoácidos que contactan el ADN (por ejemplo, Tyr57 estructuralmente se alinea con Tyr162). Sin embargo, existe un desplazamiento para los residuos que se intercalan, Leu58 y Leu182, y los pares de bases intercalados. Si se realiza un alineamiento en base a los residuos que se intercalan, entonces observamos que la secuencia $A_2A_3'C_4$ se alinea con $C_{15}'A_{16}A_{17}$. Si invertimos esta última secuencia, encontramos que la secuencia $A_{17}A_{16}'C_{15}$ encaja perfectamente con $A_2A_3'C_4$. Por lo tanto, es posible definir un motivo invertido, $AA'C$ -10pb- $C'AA$, que sigue una simetría similar a la simetría binaria HMGbox.

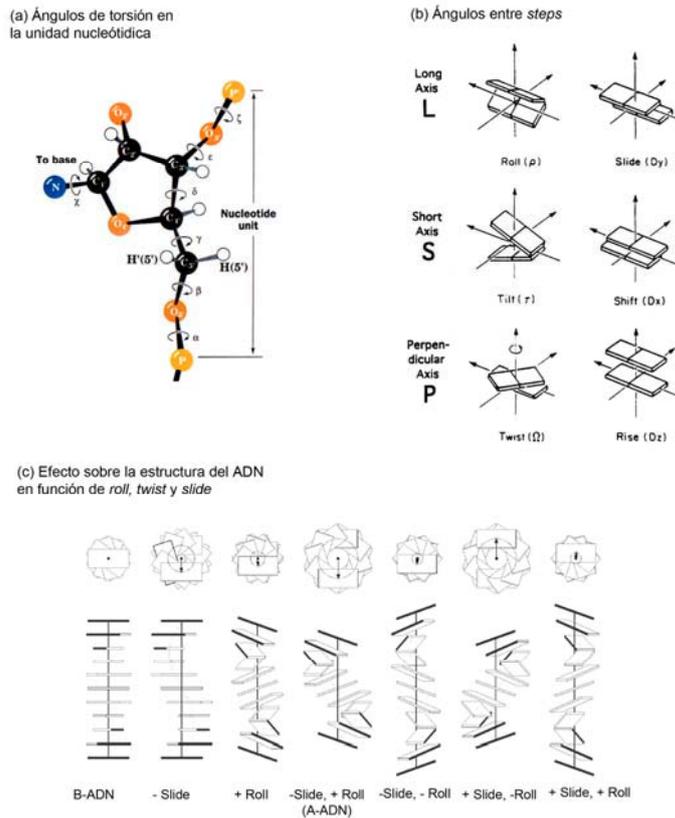


Fig_E6. Alineamiento de las regiones 1 y 2, que presentan los parámetros de *roll* y *twist* alterados. En azul se muestra la secuencia de ADN (azul fuerte, cadena C; azul cielo, cadena D). En rosa se muestran los residuos que se intercalan en el ADN, mientras que los contactos de los residuos de cada HMGbox con el ADN siguen el código de colores aplicado a las HMGbox. El alineamiento de la izquierda está hecho en base a los residuos equivalentes estructuralmente, mientras que el derecho es en base a los residuos que se intercalan y permiten diferenciar la presencia del motivo AAC.

F. RESULTADOS: ANÁLISIS DE LA FLEXIBILIDAD INTRÍNSECA DE LSP22

El mecanismo por el cual una proteína o droga puede reconocer una secuencia de ADN lleva años siendo estudiado a través de las estructuras de ADN disponibles y las estructuras de ADN en complejo con proteínas. El primero en ofrecer una explicación a este proceso fue el grupo de Rich (Seeman, Rosenberg et al. 1976), quienes postularon que la dependencia de secuencia podía venir mediado por el reconocimiento del patrón de puentes de hidrógeno entre las cadenas laterales de la proteína a los largo de los pisos que conforman el surco menor y mayor. Posteriormente, Travers, *et al.* descubrieron mediante el estudio de nucleosoma que existe la tendencia de algunas secuencias de ADN a formar estructuras cóncavas, mientras que otras frecuentemente forman estructuras convexas (Travers and Drew 1997). Más tarde Dickerson & Drew (Dickerson and Drew 1981) realizaron el primer análisis de una estructura cristalográfica de ADN y la dependencia entre posibles deformaciones de la hélice y su secuencia. En particular, los pasos que implica pirimidina-purina (C-G,T-A) tienden a presentar un *roll* positivo, es decir, a doblarse hacia el surco mayor a través del eje largo o *long axis*. Mientras que los pasos purina-pirimidina (C-G, A-T) muestran una tendencia hacia un *roll* negativo comprometiendo el surco menor. A partir de ese momento surgió el concepto de la tendencia local a una deformación en función de la secuencia de ADN, es decir, la flexibilidad intrínseca del ADN dependiente de su secuencia. Por ello en este apartado, se quiso examinar las características físico-químicas de los pares de bases que conforman la secuencia de unión de LSP para TFAM.

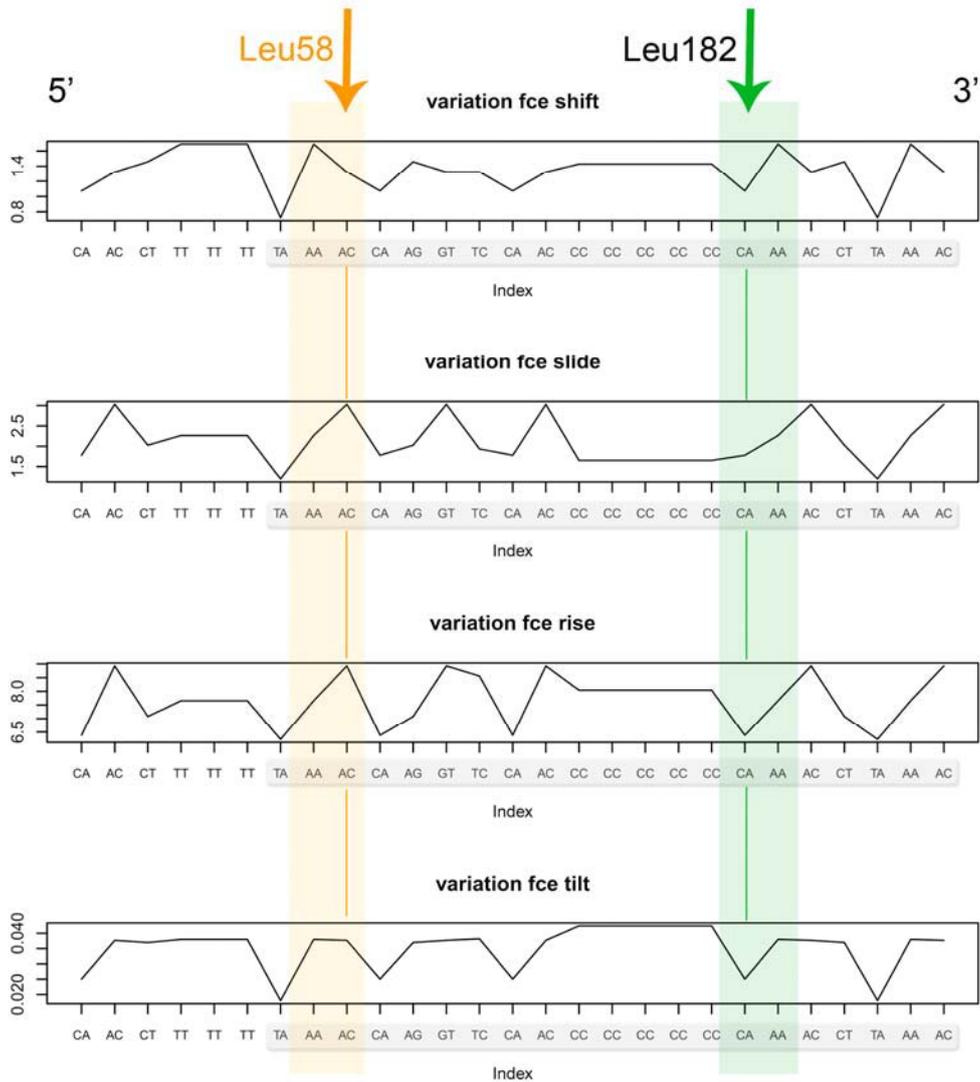
La conformación de un nucleótido depende de siete ángulos de torsión (Fig_F1(a)), estos ángulos de torsión se encuentra definidos por ángulos entre la base nitrogenada y el azúcar (χ), entre el O₅ y el grupo fosfato del nucleótido anterior (α), entre O₃ y el grupo fosfato posterior (ζ),... La dependencia entre estructura del ADN y la secuencia del ADN puede ser explicada fácilmente a partir del análisis a nivel del par de base (base-pair level), donde es posible definir 6 parámetros: *shift*, *tilt*, *slide*, *roll*, *rise* y *twist* (Fig_F1 (b)). Los parámetro de *roll*, *twist* y *slide* juegan un papel importante en la forma final de la estructura de ADN, dando lugar a hebras de ADN más largas o cortas, o con curvaturas iniciales en determinados puntos. Mediante el estudio de estructura cristalográfica ya resueltas se ha podido constatar que *roll*, *twist* y *slide* exhiben una sutil dependencia con la secuencia de ADN, por ejemplo, el par de base CA·GT es intrínsecamente menos rígido que AC·TG, permitiendo adquirir deformaciones que afectan a la supra estructura helicoidal final.



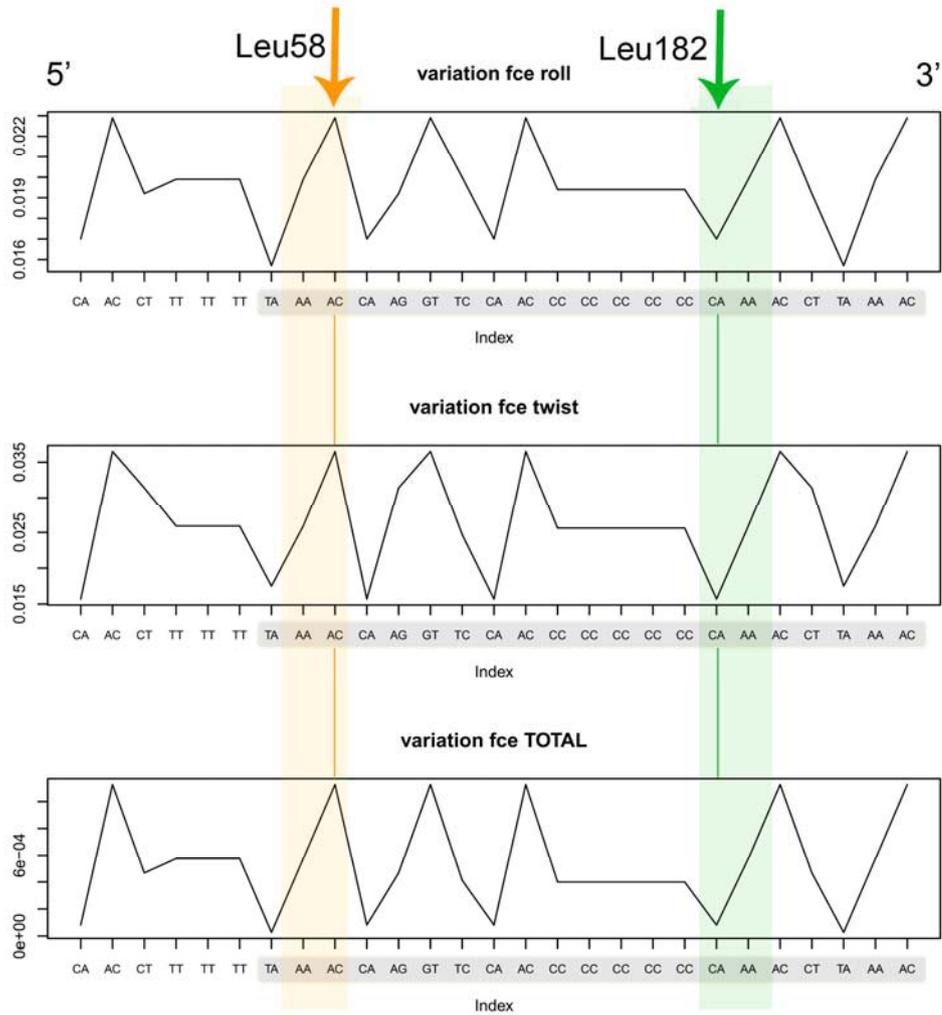
Fig_F1. Ángulos que definen la estructura local del ADN. (a) ángulos de torsión en la unidad nucleotídica. (b) Ángulos que definen los *steps* del ADN (c) Ejemplo de como los parámetros de *roll*, *twist* y *slide* afectan la estructura local del ADN.

En colaboración con Federica Battistini del grupo de Modesto Orozco en Intitut de Recerca Biomèdica (IRB), se realizó el estudio de la secuencia de 29 pb definida mediante ensayo de *footprinting in organello* por Shadel; *et al*. En concreto se realizó el análisis a nivel de par de base de los valores de: *shift*, *tilt*, *slide*, *roll*, *rise* y *twist*, mediante la base de datos presente en este grupo (Fig_F2, 3). Mediante el análisis de los valores para estos parámetros es posible diferenciar que existen diversos mínimos energéticos en la secuencia usada para la cristalización (sombreada en gris) que coinciden para todos los parámetros analizados. Cabe destacar que las regiones con mínimos energéticos son regiones menos rígidas, con tendencia a presentar deformaciones en su estructura, como apertura de las bases relacionado con una pre-curvatura intrínseca del ADN. Si intentamos relacionar estos mínimos energéticos con los motivos AAC de contacto con HMG1 (en naranja) y con HMG2 (en verde). En el caso de la intercalación de Leu182 en el motivo AAC invertido (5'-CAA-3'), el punto de intercalación coincide exactamente con el mínimo energético, por lo que la intercalación de Leu182 se vería favorecida por la forma inicial del ADN. Para el caso de interacción de HMG1 con el ADN, el punto de intercalación por parte de Leu58 no coincide exactamente con el mínimo energético. Pero es posible que exista una pre-curvatura inicial del ADN, sugerida por el mínimo energético previo, que facilite la interacción de los residuos de HMG1 con el ADN y ello promueva la intercalación final de Leu58. Por lo tanto, cabe destacar que aunque existe este motivo AA'C-10bp-C'AA con similitud en el patrón de intercalación de los residuos de TFAM, la secuencia de ADN presenta característica físico-químicas diferentes entre el primer motivo AAC y el segundo

motivo invertido CAA. Pese a ello, la curvatura final inducida por los dos dominios en el ADN es similar.



Fig_F2. Parámetros de *shift*, *slide*, *rise* y *tilt* para la secuencia LSP de 29 pb. El análisis se realiza para cada par de bases secuencial. La secuencia sombreada en gris, hace referencia a la región de ADN con la que se cristalizó TFAM, LSP22. Las regiones en naranja y verde indican el motivo AAC, en naranja para la región de contacto de HMG1 y en verde para la región de contacto de HMG2. Con flechas en la parte superior se indica el sitio de intercalación de Leu58 para HMG1, y Leu182 para HMG2.



Fig_F3. Parámetro de *roll*, *twist* y valor global energético (*variation fce*) para la secuencia LSP de 29pb. El análisis se realiza para cada par de bases secuencial. La secuencia sombreada en gris, hace referencia a la región de ADN con la que se cristalizó TFAM, LSP22. Las regiones en naranja y verde indican el motivo AAC, en naranja para la región de contacto de HMG1 y en verde para la región de contacto de HMG2. Con flechas en la parte superior se indica el sitio de intercalación de Leu58 para HMG1, y Leu182 para HMG2.

G. CARACTERÍSTICAS BIOQUÍMICAS Y BIOFÍSICAS DE LOS DOMINIOS HMGbox

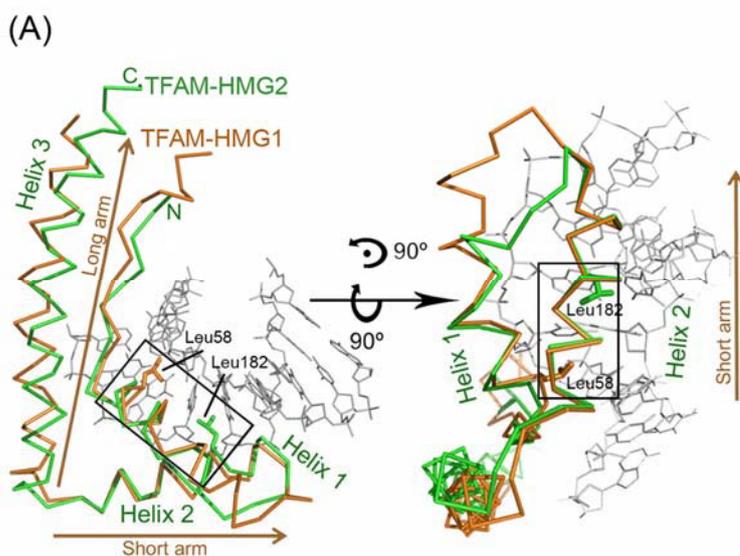
En este apartado de resultados se mostrará la caracterización bioquímica y estructural de los dominios HMGbox presentes en TFAM.

1. Análisis estructural dominios HMGbox de TFAM

Al superponer los dos dominios HMGbox de TFAM es posible observar que ambos presentan el mismo plegamiento en forma de L (r.ms.d. 0.95Å), pero no son idénticos. Las tres hélices involucradas en el plegamiento en forma de L presentan diferente ángulo entre los ejes helicoidales, dando lugar a una forma más estrecha y elongada para HMG2 comparado con HMG1.

Resulta interesante observar que la hélice1 de HMG2 es una vuelta más corta que HMG1, característica que se mantiene después de la unión al ADN, y que diferencia este dominio de otros dominios HMGbox (Fig_G1). Esta hélice 1 más corta en HMG2 se mantiene a lo largo de todas la secuencias de TFAM de *metazoos*. A pesar de esa diferencia estructural entre HMG1 y HMG2, ambas intercalan residuos no polares en el ADN.

También es importante destacar la presencia de los residuos intercalantes del ADN (Leu58 y Leu182) se encuentran situados en diferentes hélices en sus correspondientes dominios, ya que en el caso de Leu58 se encuentra localizada en Hélice1 y para Leu182 en Hélice2.



Fig_G1. Superimposición estructural de los dominios HMGbox para TFAM. (A) Representación estructural de los dominios HMGbox en complejo con ADN, donde en naranja se muestra HMG1 y en verde HMG2, con las diferentes hélices indicadas y las leucinas intercalantes representadas en forma de barras.

2. Análisis de unión a ADN de los dominios HMGbox de TFAM

2.1. Diferente patrón de reconocimiento

Estudios realizados a lo largo de 40 años han mostrado que los dominios de unión al ADN (*Vds., DNA Binding Domains*) reconocen el ADN en una gran variedad de formas. Algunos

DBDs como los represores bacterianos, se encuentra perfectamente plegados antes de unir a los sitios de reconocimiento, exhibiendo así una unión específica de secuencia con una gran afinidad y tiempos de residencia largo, mostrando un mecanismo conocido como cerradura-llave (*lock-and-key mechanism*). En eucariotas la situación es un poco más compleja, frecuentemente existen una relajación en el reconocimiento de secuencia específica para poder ofrecer una mayor grado de adaptabilidad conformacional.

Aunque los dominios HMGbox muestran una similitud funcional (actuando como factores de transcripción, proteínas arquitecturales, etc.), la selectividad por la secuencia de ADN es la principal diferencia. Mientras que los dominios HMGbox de unión específica de secuencia (*HMGbox SS*) contiene solo un sitio de intercalación y diversos contactos con las bases, los dominios HMGbox de unión inespecífica (*HMGbox NSS*) contiene dos puntos de intercalación. Además *HMGbox SS* son más inestables en solución en su forma no unida al ADN (Dragan, Read et al. 2004).

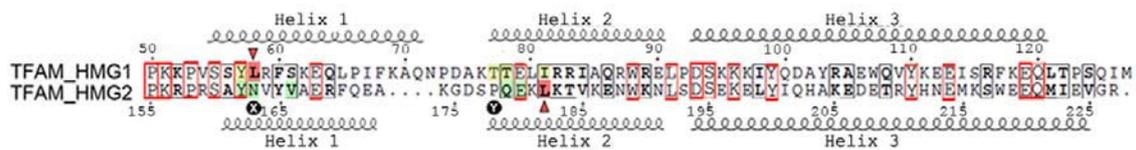
Las estructuras de dominios HMGbox libres no muestran un único estado de transición, sino que existen de tres a dos subdominios involucrados en el plegamiento, por ejemplo, para el caso de *HMGbox SS* son más inestables. Sin embargo la interacción con ADN estabiliza estos subdominios y da lugar a un único complejo cooperativo. En el caso de *HMGbox SS* que en su forma no unida son extremadamente flexibles, al formar complejo muestran una interfaz bien plegada que realiza contactos van der Waals entre residuos apolares, que determinan la unión específica de secuencia. En cambio en el caso de *HMGbox NSS*, éstos son más estables en solución en su forma no unida, es decir, más rígidos, y la mayor fuerza de unión al ADN proviene de interacciones de tipo electrostático que son independientes de la secuencia de ADN. Las interacciones de tipo no electrostático en *HMGbox SS* y las de tipo electrostático de *HMGbox NSS* son las que dirigen la curvatura del ADN mediante la intercalación de residuos (o cuñas) de sus dominios globulares entre las bases del ADN (proceso termodinámicamente poco favorable). Para ambos tipos de HMGbox, las curvaturas son inducidas más adelante mediante neutralización asimétrica de los grupos fosfatos por parte de las extensiones básicas (Dragan, Read et al. 2004).

De entrada TFAM es una proteína interesante en cuanto a su clasificación dentro de la familia de proteínas HMGB, ya que pese a presentar dos dominios HMGbox característica asociada a los dominios *HMGbox NSS*, su actividad se encuentra asociada a unión específica de secuencia (por ejemplo en LSP) e inespecífica de secuencia. Este patrón de reconocimiento de HMGbox puede establecerse según el número y posición de los residuos de los dominios HMGbox que se intercalan en el ADN (importantes para estabilizar el grado de curvatura del ADN). Como se ha comentado anteriormente, los dominios HMGbox contienen un residuo no-polar en la Hélice1 que se intercala en el ADN (llamado **X** en la Fig_G2). En el caso de los dominios *HMGbox NSS* presentan un segundo residuo intercalador adicional localizado en la Hélice2 (llamado **Y** en la Fig_G2), a diferencia del *HMGbox SS* que en esta posición muestran un residuo que forma puentes de hidrógeno con las bases del ADN. Por lo tanto, en el caso de *HMGbox NSS* las posiciones **X** e **Y** son residuos hidrofóbicos, mientras que para *HMGbox SS*, **X** es un residuo apolar que se intercala entre las bases del ADN e **Y** es un residuo polar que forma puentes de hidrógeno con las bases del ADN. Además, el sitio **X**, tanto para *HMGbox SS* como *NSS*, se encuentra normalmente precedido de un residuo aromático, como Phe o Tyr.

Mediante esta clasificación es posible determinar que HMG1 de TFAM puede ser catalogado como HMGbox de unión específica de secuencia (*HMGbox SS*), ya que en la posición **X** aparece un residuo no polar (Leu58) que se intercala entre las bases del ADN, precedido por un residuo aromático (Tyr57), mientras que en la posición **Y** presenta un residuo polar (Thr77) que forma un enlace tipo puente hidrógeno con el ADN (Fig_G2).

Sin embargo, para el caso de HMG2, el residuo en la posición **X** es un residuo polar, Asn163, que forma puentes de hidrógeno con el ADN, en lugar de intercalar el ADN. Además en la posición **Y** existe un residuo apolar, Pro178, que no se intercala entre el ADN, aunque de Pro178 a Glu180 existe contactos con las bases del ADN que favorecen la curvatura inducida por el residuo intercalante Leu182. Este residuo Leu182 no encaja con ninguna clasificación de residuos intercalantes en dominios HMGbox y parece ser único (Fig_G2), pero conservado entre los homólogos de TFAM.

Además de presentar un patrón de reconocimiento diferente para el ADN, donde HMG1 encaja como HMGbox SS y HMG2 se encuentra en una situación intermedia de clasificación, cabe destacar la diferencia en afinidad de estos dos motivos para su unión con el ADN.

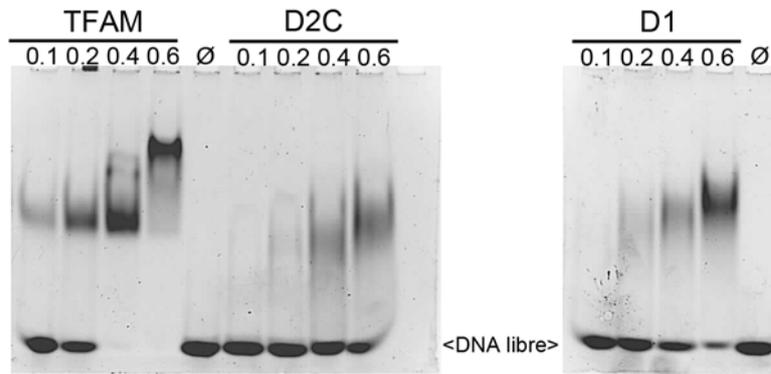


Fig_G2. Alineamiento secuencial en base a la estructura de los dos dominios HMG1 y HMG2. Las posiciones X e Y indican los sitios habituales de intercalación para HMGbox, mientras que los triángulos rojos indican los sitios reales de intercalación para HMG1 y HMG2. En naranja y verde se muestran los residuos que contactan con las bases, manteniendo el código de colores (naranja para HMG1, verde para HMG2). Los marcos rojos indican los residuos 100% conservados entre HMG1 y HMG2, mientras que los marcos negros indican residuos moderadamente conservados.

2.2. Diferente afinidad por la secuencia LSP22

Mediante ensayos EMSA se pudo comprobar la diferente afinidad de los dominios HMGbox de TFAM por la secuencia LSP22, utilizada para la cristalización (Fig_G3). Para el caso de TFAM observamos una formación del complejo total a 0.4 nmol TFAM:0.1 nmol ADN, mientras que para el caso del dominio HMG1 (constructo D1) observamos una desaparición casi total de ADN libre en ratio 0.6 nmol D1: 0.1 nmol ADN y en el caso del dominio HMG2 más la cola C-terminal (D2C), la formación del complejo implica una mayor cantidad de ADN.

Este pequeño ensayo sugirió que la afinidad del dominio HMG1 por ADN es mayor que para HMG2, pero siempre inferior al de la proteína nativa TFAM. Otros grupos han constatado esta diferencia de afinidad mediante ensayos de calorimetría (Malarkey, Bestwick et al. ; Ngo, Kaiser et al.).



Fig_G3. Ensayo EMSA para comparar la afinidad de TFAM, el constructo D1 y D2C frente a una concentración de ADN constante de 0.1 nmol. En la parte superior se indica los ratios de proteína (0.1, 0.2, 0.4 y 0.6 nmol). El carril Ø, corresponde a ADN libre.

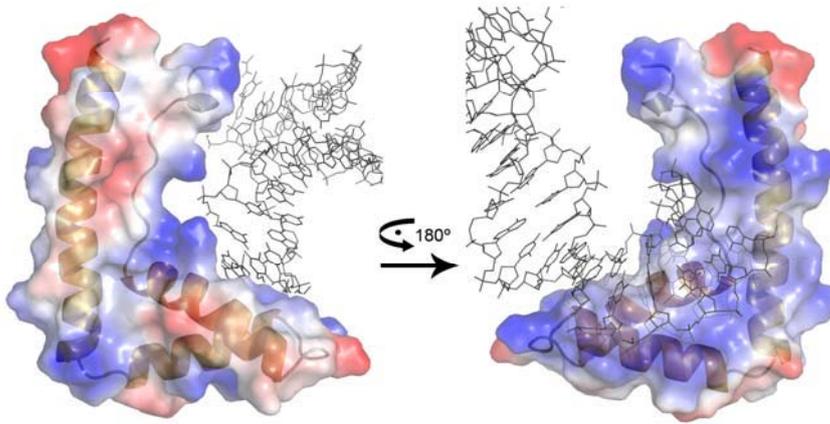
El diferente patrón de interacción de los residuos para HMGbox y la diferencia en las características de la secuencia LSP contactada por cada dominio HMG de TFAM, permiten potenciar la idea de que ambas HMGbox son diferentes en su unión a ADN.

3. Análisis electrostático de los dominios HMGbox

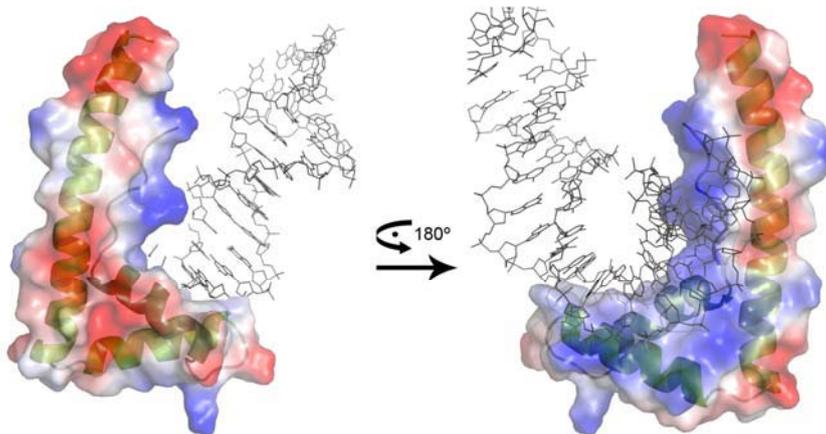
Cuando se analizan los puntos isoeléctricos teóricos de ambos dominios HMG1 (Ser56-Gln120) y HMG2 (Ala161-Val225) es posible observar la gran diferencia existente, resultando HMG1 extremadamente básico (pI 9.87) y HMG2 moderadamente ácido (pI 5.71). Esta característica es del todo sorprendente, ya que se trata de dominios de unión a ADN, por lo tanto, se esperaría que ambos (y no sólo HMG1) sean ricos en residuos básicos para favorecer la interacción con el ADN, cargado negativamente. Esta diferencia de potencial entre dominios se encuentra extendida en otros dominios HMGbos de dos o más dominios, como es el caso de UBF.

El cálculo de la superficie electrostática muestra como en la superficie de contacto con el ADN (cara interna de la bota) de ambos dominios tienen un elevado contenido de carga positiva o potencial positivo (en azul ver Fig_G4), en concreto, la región que incluye la Hélice 1 y Hélice 2. Mientras que la superficie electrostática de la parte externa de la bota, no expuesta al ADN, no muestra una concentración de cargas positivas tan alta. Sobretodo en el caso de HMG2, donde se puede diferenciar que la Hélice 3 presenta una mayor carga negativa o mayor potencial negativo (mostrado en color rojo, ver Fig_G4) en comparación con la Hélice 3 de HMG1. Quizás es esta diferencia de carga lo que provoca que HMG1 presente una mayor afinidad por el ADN que HMG2, que en lugar de presentar una constante de afinidad del orden nanomolar es micromolar (Ngo, Kaiser et al.)

A) HMG1 electrostatic surface with DNA



B) HMG2 electrostatic surface with DNA



Fig_G4. Representación del potencial electrostático en una superficie de Connolly para ambos dominios HMGbox de TFAM, en complejo con el LSP (en gris) en dos orientaciones relacionadas por 180° (azul positivo, rojo negativo). A la derecha encontramos la orientación donde es posible observar la parte del dominio que contacta con el ADN, mientras que a la izquierda se puede diferenciar la parte externa de éste. (A) Superficie electrostática para HMG1, donde se diferencia las tres hélices en naranja. (B) Superficie electrostática para HMG2, donde se diferencia las tres hélices en verde.

4. Representación en forma de rueda de hélices alpha

La representación de la distribución de los residuos para cada hélice de HMGbox en un plano único, mediante la representación conocida como *Helical Wheel projection* generada por *Don Armstrong* y *Raphael Zidovetzki* (Eisenberg, Weiss et al. 1982) (Fig_G5), permite obtener más información acerca de las diferencias estructurales entre hélices.

- Hélice 1

La comparación entre Hélice 1 de HMG1 (Ser56-Gln71) y HMG2 (Ala164-Glu190) muestra que la Hélice 1 de HMG1 es una vuelta más corta (tal y como se ha comentado anteriormente, 3 vueltas para HMG1 y 2 vueltas para HMG2). Además para el caso de HMG2 contiene menos residuos con carga positiva que para HMG1, así como la

distribución de cargas positivas es muy diferente. Cabe destacar la presencia del residuo Leu58, responsable de la intercalación en el motivo AA'C.

- Hélice 2

Para HMG1, la hélice 2 incluye del residuo Thr77 a Glu90, y para HMG2 es del residuo Pro178 a Asn191, que incluye el residuo Leu182, responsable de la intercalación en el motivo AA'C. En este caso ambas hélices presenta el mismo número de vueltas, que son 3. La carga global es similar (4 residuos positivos, 2 negativos) y su distribución más conservada.

- Hélice 3

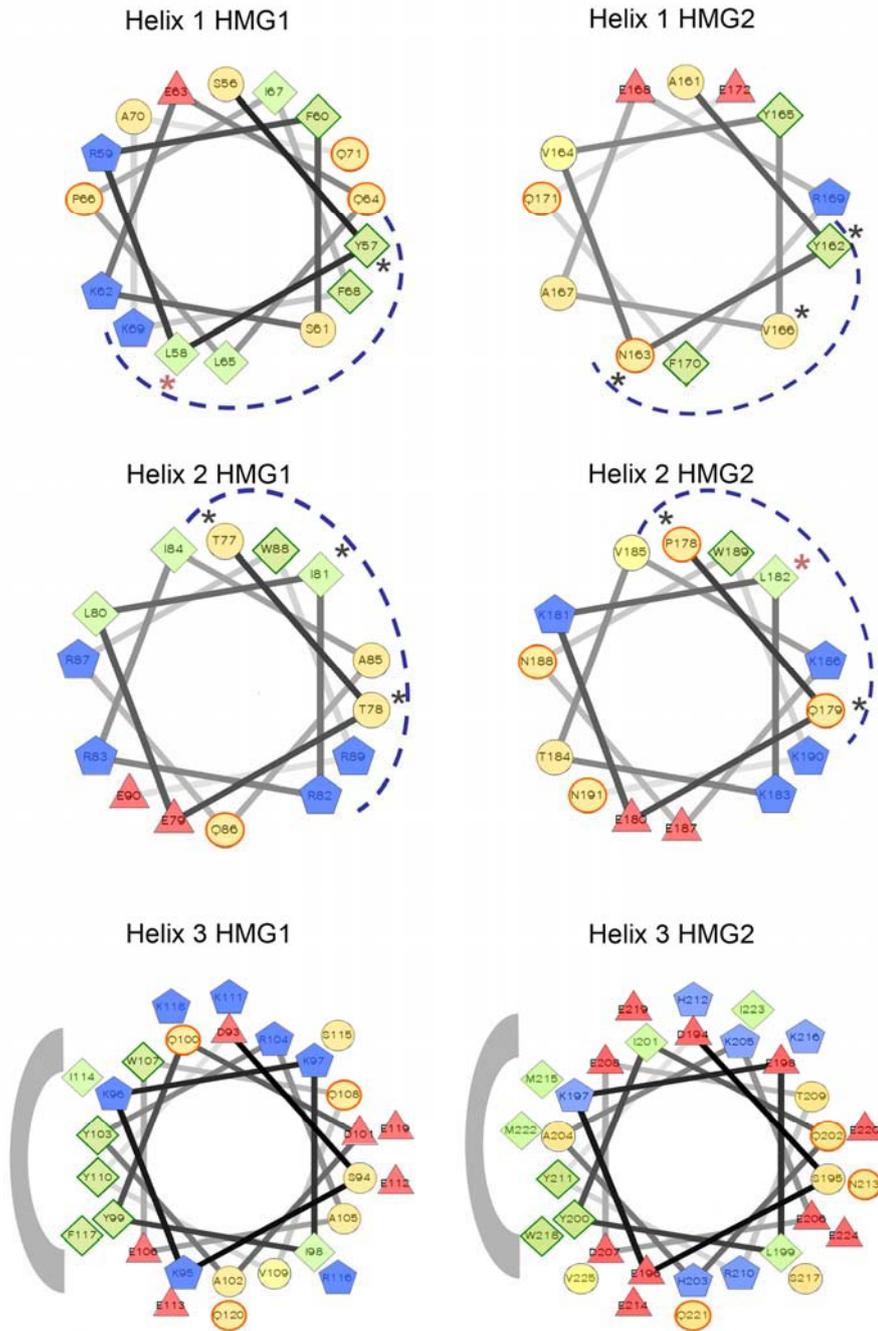
Para HMG1, la hélice 3 incluye del residuo Asp93 a Gln120, y para HMG2 va de Asp194 hasta Val225. La hélice 3 de HMG1 es un paso de vuelta más corta, frente a las 7 vueltas de la hélice 3 de HMG2. Es posible observar que estas hélices presentan una superficie hidrofóbica en la región izquierda de la hélice, conservada entre ambas y que corresponde con la parte interior del brazo largo de la estructura en forma de L. Por lo que se refiere a las cargas, la Hélice 3 de HMG2 presenta un mayor número de cargas negativas (10 residuos cargados negativamente frente a 6 en HMG1) y similar números de residuos cargados positivamente. La distribución de estas cargas es alrededor de la parte accesible al solvente.

Por lo tanto, esta representación de las hélices permite visualizar de una forma rápida las diferencias químicas entre HMG1 y HMG2, que afectaran al patrón de reconocimiento del ADN por parte de los dos dominios. La Hélice2 de HMG2 y HMG1 presenta una distribución de cargas similar, mientras que Hélice1 difiere. Además en Hélice1 de ambas HMGbox, el número de residuos cargados positivamente frente a los ácidos difiere, en HMG1 3:1 frente 1:2 en HMG2.

En el caso de la Hélice2, la superficie de contacto con el ADN cubre un mayor número de residuos en el caso de HMG1. Cabe destacar la presencia de los residuos Trp88 y Trp189, altamente conservados entre TFAM de metazoos, y que curiosamente presentan la misma posición en la proyección, sugiriendo la importancia de este residuo.

Finalmente en el caso de la Hélice 3, que constituye el brazo largo de HMGbox, presenta un superficie hidrofóbica (cara interna) conservada entre HMG1 y HMG2, pero una distribución de cargas muy diferente para ambos casos. En el caso de HMG2, esta hélice muestra un mayor número de cargas negativas, lo cual explica porque la superficie electrostática de la superficie externa (contacto con el solvente) en hélice 3 de HMG2 es más negativa que en HMG1.

Las diferencias en cuanto a la superficie electrostática y la distribución de cargas podría explicar porqué HMG1 y HMG2 presentan una diferente afinidad por el ADN y un diferente patrón de reconocimiento.



Fig_G5. Representación en forma de rueda para las hélices de HMGbox de TFAM. Los residuos hidrofílicos se muestran como circunferencias y los hidrofóbicos como rombos, los cargados negativamente como triángulos y los positivos como pentágonos. El grado de hidrofobicidad se muestra de más a menos con la intensidad del color verde. Los residuos hidrofílicos se muestran en tono naranja, cuanto más intenso más hidrofílico es el residuo. Finalmente, los residuos cargados negativamente se muestran en rojo claro y los positivamente en azul. En el caso de la Hélice 3 se muestra la presencia de una superficie hidrofóbica con un medio círculo gris claro. El interlineado azul que cubre regiones en Hélice1 y Hélice2, indica la zona de contacto con el ADN. Los asteriscos grises muestran las bases que contactan con las bases, y los asteriscos rojo claro muestran los residuos intercalantes, Leu58 y Leu182.

5. Análisis estructural entre HMGbox y su núcleo hidrofóbico

Cuando se alinean las secuencias de los dos dominios HMGbox de TFAM en función de su estructura, es posible destacar un conjunto de residuos conservados (Fig_G2). Al analizar la localización de estos residuos en la superimposición de HMG1 e HMG2, la mayoría de ellos no participan en la interacción con el ADN, sino que participan en el mantenimiento del núcleo hidrofóbico y la estructuración del dominio HMGbox en forma de L (Fig_G6).

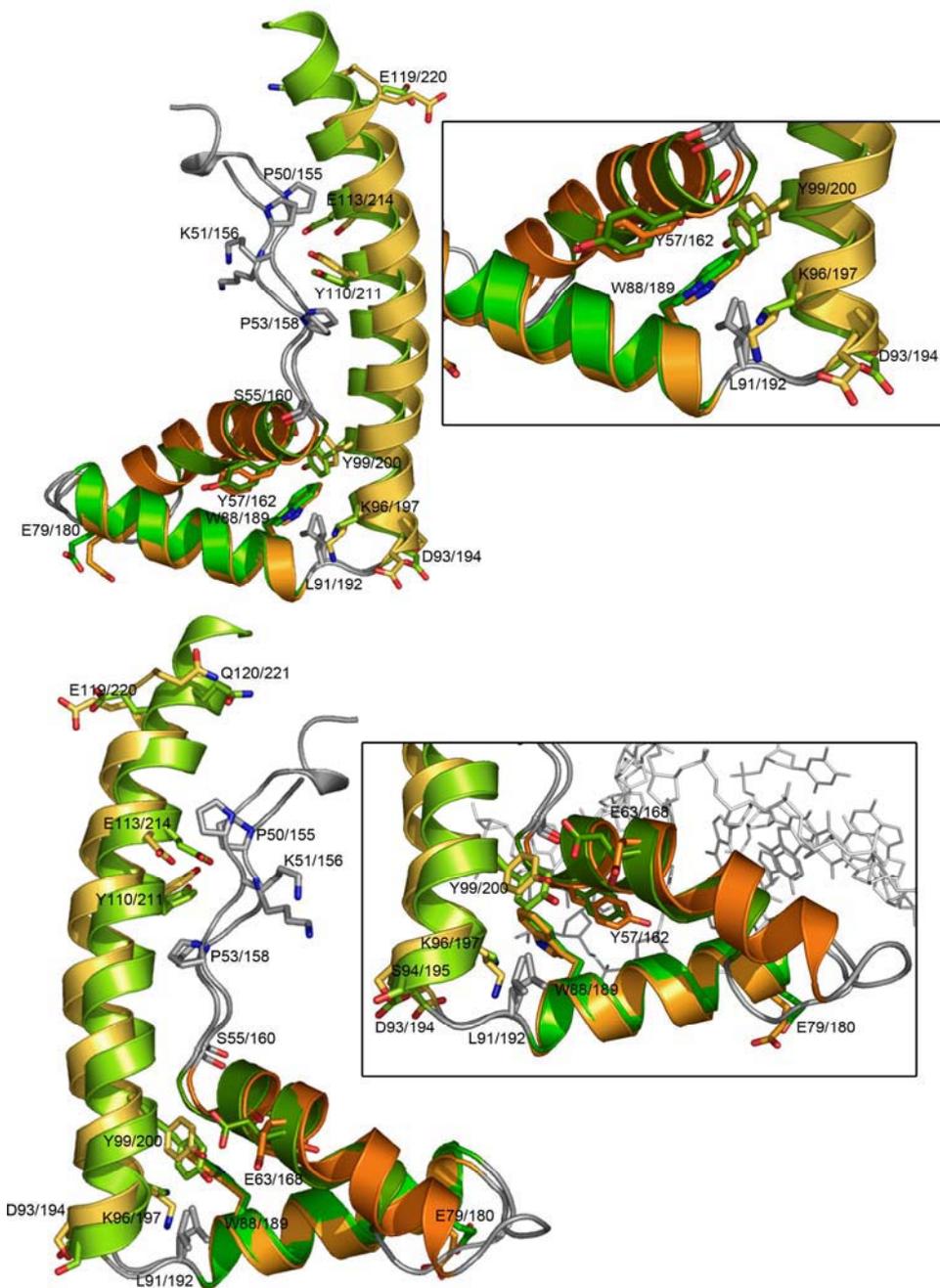
Dentro los residuos conservados cabe destacar aquellos que mantienen una orientación casi idéntica entre ambos dominios (HMG1/HMG2):

- Extensión N-terminal: Pro50/Pro155, Pro53/Pro158
- Hélice1: Tyr57/Tyr162
- Hélice2: Glu79/Glu180, Trp88/Trp189
- Loop Hélice2-Hélice3: Leu91/Leu192
- Hélice3: Asp93/Asp194, Lys96/197, Tyr99/Tyr200, Tyr110/Tyr211, Glu113/Glu220, Glu119/220

Cuando se alinea las secuencias de HMG1 y HMG2 con las estructuras resueltas de *HMGbox SS* y *HMGbox NS* es posible observar los siguientes residuos conservados. Para *HMGbox SS* son (Pro53/Pro158), (Trp88/Trp189), (Leu91/192). Mientras que para *HMGbox NSS* son (Pro50/Pro155), (Ser55/Ser160), (Trp88/Trp189), (Lys96/197) y (Tyr110/Tyr211) (ver Fig_G6). Si miramos con detenimiento la superimposición de la Fig.5, observamos que los residuos conservados entre los diferentes dominios presentan una función estructural de mantenimiento del plegamiento del dominio, donde se pueden localizar dos zonas importantes. La primera zona correspondería a los residuos localizados en la bisagra de los dos brazos, que permiten mantener el ángulo correcto entre los dos brazos. Cabe destacar las parejas de residuos aromáticos (Trp88/Trp189) (Tyr99/Tyr200) y (Leu91/Leu192) que permiten formar un núcleo ofreciendo un correcto ángulo de empaquetamiento. La segunda zona correspondería a los residuos que estabilizan la extensión N-terminal con la Hélice3: (Pro50/Pro155), (Pro53/Pro158), (Tyr110/Tyr211) (Glu113/Glu220).

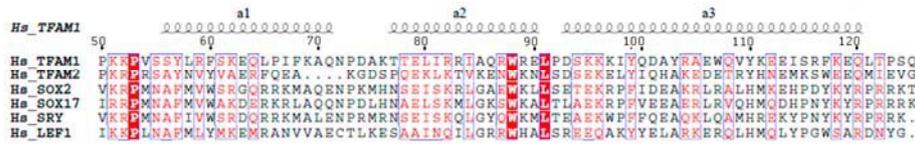
Es interesante indicar que el conjunto (Trp88/Trp189) se encuentra 100% conservado entre las proteínas TFAM de organismos multicelulares, así como los residuos Pro155, Tyr211.

Además cabe destacar la presencia de residuos aromáticos situados en la región cóncava o cerca de ella formada por los dos brazos de la L. Estos residuos presentan un papel importante en la interacción con el ADN. Por ejemplo, uno de los residuos que se encuentra conservado entre HMG1 y HMG2 es: Tyr57/Tyr162, que forma un puente hidrógeno con las bases del ADN en el surco menor del ADN. Este residuo se encuentra 100% conservado entre diferentes TFAM de organismos multicelulares, así como a través de *HMGbox SS* y *HMGbox NSS* (Fig_G7).

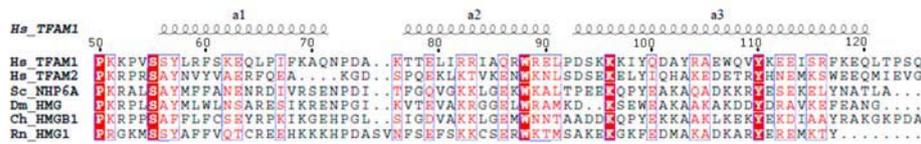


Fig_G6. Superimposición de los dominios HMG1 (naranja) y HMG2 (verde), donde se encuentran indicados mediante representación en barras los residuos conservados entre los dos dominios. La representación superior muestra la cara interna de la L, que contacta con el ADN; mientras que en la representación inferior se muestra la orientación externa de la L. Para cada representación se indican la zonas ampliadas que implican el core hidrofóbico que mantiene la correcta orientación entre las hélices de los dominios HMGbox.

(a)



(b)



Fig_G7. Alineamiento secuencial de los dominios HMG1 y HMG2 contra *HMGB1* SS con estructura resuelta (SOX2, SOX17, SRY, LEF1) en (a), y contra *HMGB1* NSS con estructura resuelta (NHP6A, HMGB1, HMGB1) en (b). Es interesante destacar que en ambos casos la pareja Trp88/189 se encuentra conservado tanto en *HMGB1* NSS como *HMGB1* SS.

6. Comparativa estructural entre dominio HMG2 en su forma unida y no unida

Previo a la resolución de la estructura de TFAM en complejo con ADN, fue resuelta la estructura cristalográfica del dominio HMG2 (Leu152-Gln221) en su forma no unida al ADN (HMG2-unbound) (Gangelhoff, Mungalachetty et al. 2009) con una resolución 1.35 Å (código de acceso al PDB 3FGH). HMG2 cristalizó en el grupo espacial C2221 con los siguientes parámetros de celdilla $a= 30.28$ Å, $b= 99.35$ Å, $c= 51.90$ Å. La estructura final incluye los aminoácidos de Gly153-Leu219, con residuos con conformaciones alternadas Ty162, Leu182, Val185, Lys197, Tyr200, Met215 y Ser217, los aminoácidos de la posición 172 a 176 fueron trazados como alanina debido a la pobre densidad electrónica.

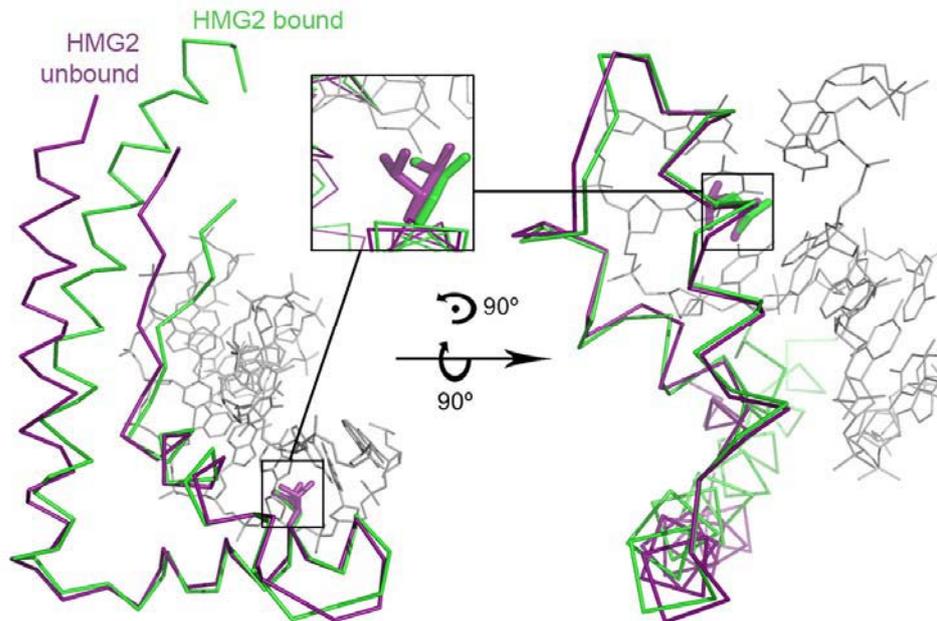
Al superponer las dos estructuras, HMG2-LSP22 y HMG-unbound, es posible observar que éstas son similares, mostrando un plegamiento similar (r.m.s.d. 2.37 Å). Al comparar las dos estructuras (Fig_G8) es posible observar que en el caso de la estructura para HMG2-LSP22 parece más compacta.

Los autores que resolvieron la estructura HMG2-unbound destacaban que dado que HMG2 en el alineamiento de secuencia con HMG1 presenta un gap en la zona N-terminal de la Hélice2, región definida como zona de contacto con el ADN, y el hecho que en la posición del residuo típico de intercalación en HMGB1, caracterizado por ser un residuo apolar que se intercala entre las bases del ADN, en HMG2 existe Asn163; todo ello inducía a pensar a los autores que este dominio HMG2 era incapaz de intercalarse en el ADN y que presentaba unas propiedades de unión al ADN muy diferentes. La resolución de la estructura en su forma unida al ADN ha permitido dilucidar algunos aspectos previamente descritos.

Por lo que se refiere a la extensión de las hélices, cuando comparamos Hélice1 y Hélice2 de las dos estructuras de HMG2, éstas coinciden en sus dimensiones y ambas son más cortas que HMG1. Pese a que Gangelhoff identificaba este gap o diferencia en longitud en la

Hélice2 de HMG1 y HMG2, la superposición de los dominios HMG1, HMG2-LSP22 y HMG2-unbound permite localizar este gap de cuatro aminoácidos (prácticamente una vuelta de hélice menos) en la Hélice1 de HMG2, tanto en su forma unida como no unida al ADN.

Además, la estructura HMG2-LSP22 permite aclarar la capacidad de intercalar un residuo apolar entre las bases del ADN por parte de este dominio. Leu182 en la Hélice2 intercala entre C₁₅-G₈ y A₁₆-T₇. Cabe destacar que este residuo Leu182 presenta en la estructura HMG2-unbound una doble conformación, una de ellas seleccionada en el complejo con el ADN para intercalar entre las bases del ADN (Fig_G8). Gangelhoff et al. sugería que los residuos Tyr162 y Tyr200, situados en el núcleo hidrofóbico y con orientaciones alternadas en HMG2-unbound, eran importantes en la interacción con el ADN. En la estructura HMG2-LSP22, Tyr162 presenta una única conformación que realiza un enlace tipo puente hidrógeno con la A16 de la cadena C, de la misma manera que el residuo topológicamente equivalente en HMG1, Tyr57. Cabe destacar el residuo altamente conservado Tyr200, que muestra doble conformación en la estructura de HMGbox2-unbound, mientras que en la estructura TFAM-LSP22, presenta una única orientación donde el anillo aromático contribuye al núcleo hidrofóbico de HMG2 y su grupo hidroxilo permanece en la superficie sin realizar interacción de ningún tipo con el ADN, a diferencia de lo que proponía Gangelhoff, et al. Además cabe destacar Tyr211 que presenta una doble conformación en HMG2-unbound, mientras que HMG2-LSP22 existe un desplazamiento de la hélice3 hacia el ADN que permite la interacción con el fosfato de la base A21.



Fig_G8. Superimposición de HMG2-LSP22 (en verde) con HMG2-unbound (en lila), donde es posible ver aumentada Leu182 con la doble conformación en HMG2-unbound y seleccionada la posición de intercalación en HMG2-unbound.

H. ANÁLISIS ESTRUCTURAL DE LAS DOS ESTRUCTURAS CRISTALOGRÁFICAS EXISTENTES PARA TFAM-LSP

Al mismo tiempo que fue publicada la estructura de TFAM en complejo con LSP22, se publicó la estructura de TFAM en complejo con un ADN de 28 pb que contenía la secuencia de LSP (LSP28) a una resolución de 2.5 Å (Ngo, Kaiser et al.). En este caso se trataba de un grupo espacial diferente, C222₁, y parámetros de celdilla a= 68.44, b= 81.91, c= 161.25; $\alpha = \beta = \gamma = 90^\circ$. Pese a ello, la visualización de la estructura muestra como TFAM coordina los dos dominios HMGbox para imponer una curvatura en el ADN con forma de U, similar a LSP22. En esta estructura, es posible constatar que las cuatro regiones de TFAM: dos dominios HMGbox, linker, y cola C-terminal, realizan extensivos contactos con el ADN. Por lo tanto, la organización de la proteína y el contacto global con el ADN es similar entre las dos estructuras. A continuación se realiza un análisis comparativo de ambas estructuras.

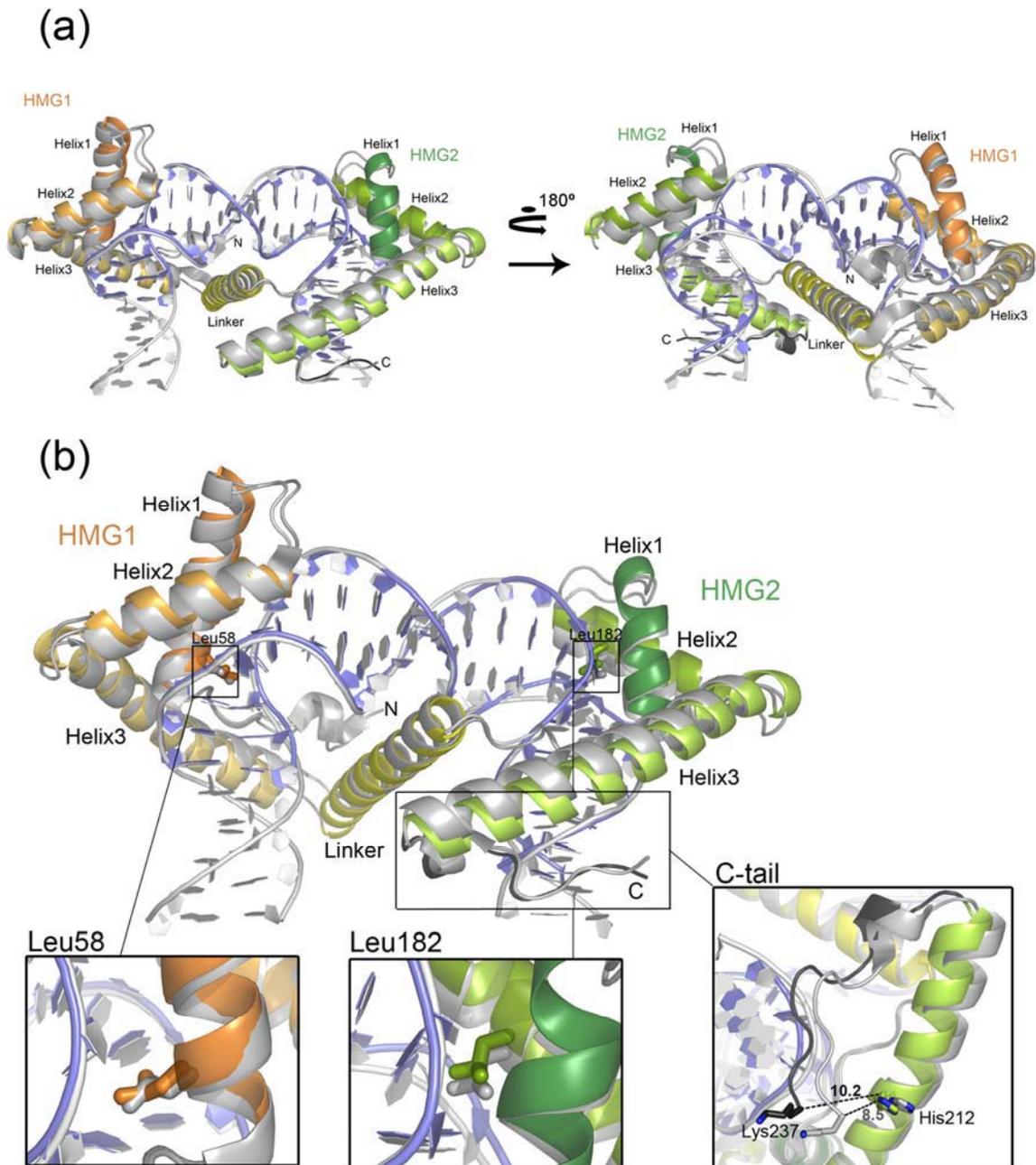
1. Superposición global

La superposición global de las dos estructuras muestra un r.m.s.d. de 1.44 Å. Ambas estructuras muestran dos dominios HMGbox con el típico plegamiento en forma de L, donde la cavidad cóncava interacciona con el surco menor del DNA provocando un curvatura de aproximadamente 90°. Ello resulta en una curvatura total del ADN de unos 180°, definida como curvatura en forma de U (Fig_H1).

Entre los dos dominios HMGbox se encuentra el linker helicoidal, que permite la reorganización de los dominios y además permite estabilizar la curvatura en forma de U, debido a que neutraliza las cargas negativas del esqueleto de fosfato del ADN. Finalmente, se encuentra la cola C-terminal que en ambos casos se encuentra en una forma parcialmente extendida y ha sido posible trazar la cadena principal hasta el residuo Lys237. Pese a que faltan los 9 residuos finales, la zona de la cola C-terminal definida no presenta ningún tipo de interacción con las bases del ADN. (Fig_H1)

De entrada es posible observar que los dos puntos intercalantes de HMGbox en el ADN, son entre los mismos residuos de ADN y por parte del mismo tipo de aminoácido: Leu 58 para HMG1 y Leu182 para HMG2. Ello confirma el motivo de reconocimiento definido anteriormente como AA'C (' indica el sitio de intercalación). Aunque es posible observar que la estructura TFAM-LSP28, muestra la proteína en una forma un tanto más compacta que para el caso de la estructura definida para TFAM-LSP22 y, por lo tanto, TFAM-LSP28 muestra distancias de interacción proteína-ADN ligeramente más cortas.

Cuando analizamos con detalle los desplazamientos estructurales existentes entre las dos estructuras resueltas, es posible observar que las mayores diferencias suceden entre la zona N-terminal del linker (residuo 119 a 128, ~2.5 Å), a lo largo de la Hélice2 del HMG2 (residuo 180-1193, ~2 Å), y entre la parte final de la Hélice3 de HMG2 y la parte inicial de la cola C-terminal.



Fig_H1. (a) Superimposición global de las dos estructura de TFAM. La versión en color corresponde con la estructura TFAM-LSP22, mientras que en gris encontramos TFAM-LSP28. El patrón de colores es naranja y verde para las dos dominios HMGbox, amarillo para linker, negro para la cola Cterminal y azul para el ADN. (b) Zoom de las zonas de intercalación al ADN para HMG1 (Leu58) y HMG2 (Leu182), y la cola C-terminal (en negro se indica la distancia entre His212 y Lys237 para TFAM-LSP22 y en gris para TFAM-LSP28).

1.1. Superposición de los dominios HMGbox

Cuando superponemos los complejos fijando los dominios HMG2 (TFAM-LSP22: 153-225, TFAM-LSP28: 111-183. Nótese que la secuencia aminoacídica de la estructura TFAM-LSP28 asigna como número 1 el primer residuo N-terminal trazado, correspondiente a Ser43) encontramos que las dos estructuras encajan con un r.m.s.d. de 0.588 Å. Mientras que cuando superpone en HMG1 (TFAM-LSP22: 43-120, TFAM-LSP28: 1-78), el r.m.s.d.

entre las dos estructuras es 0.367 Å. Por lo tanto, existe una mayor similitud entre el dominio HMG1 de las dos estructuras, que no entre HMG2 (Fig_H2).

Cuando se alinean las estructuras fijando HMG1 (Fig_H2 (a)) existe un ligero desplazamiento de HMG2, sin afectar el patrón de intercalación en el ADN. Específicamente cabe destacar el caso de la Hélice2 de HMG2, que presenta un mayor desplazamiento entre las dos estructuras (de 3.87 a 5.26 Å). Es en esta hélice donde se encuentra el residuo intercalante, Leu 182, que muestra un desplazamiento de 3.54 Å (Tabla_H1). Mientras que cuando se superponen los dominios HMG2 de ambas estructuras, es posible observar un desplazamiento de la parte N-terminal del linker de ~6 Å que prosigue en HMG1 (Fig_H2 (b)). El linker se encuentra desplazado aproximadamente a partir del residuo 138 y en dirección al dominio HMG1. Este desplazamiento llega a ser máximo a nivel del residuo Pro123, mostrando una distancia de 6.13 Å. Por lo que se refiere al desplazamiento entre los dos dominios HMG2, es posible observar otra vez que existe una conformación más compacta para HMG2-LSP28 y que este dominio presenta un desplazamiento general más agudo que en el caso anterior (Tabla_H2).

La representación de los dominios HMG1 superpuestos muestran estructuras casi idénticas, aunque es posible destacar una gran diferencia en cuanto a la conformación del residuo Leu58, donde el residuo Leu se encuentra en una orientación diferente en ambos casos. Pese a esta diferencia en la conformación, ello no afecta la intercalación en el ADN ni al patrón de reconocimiento.

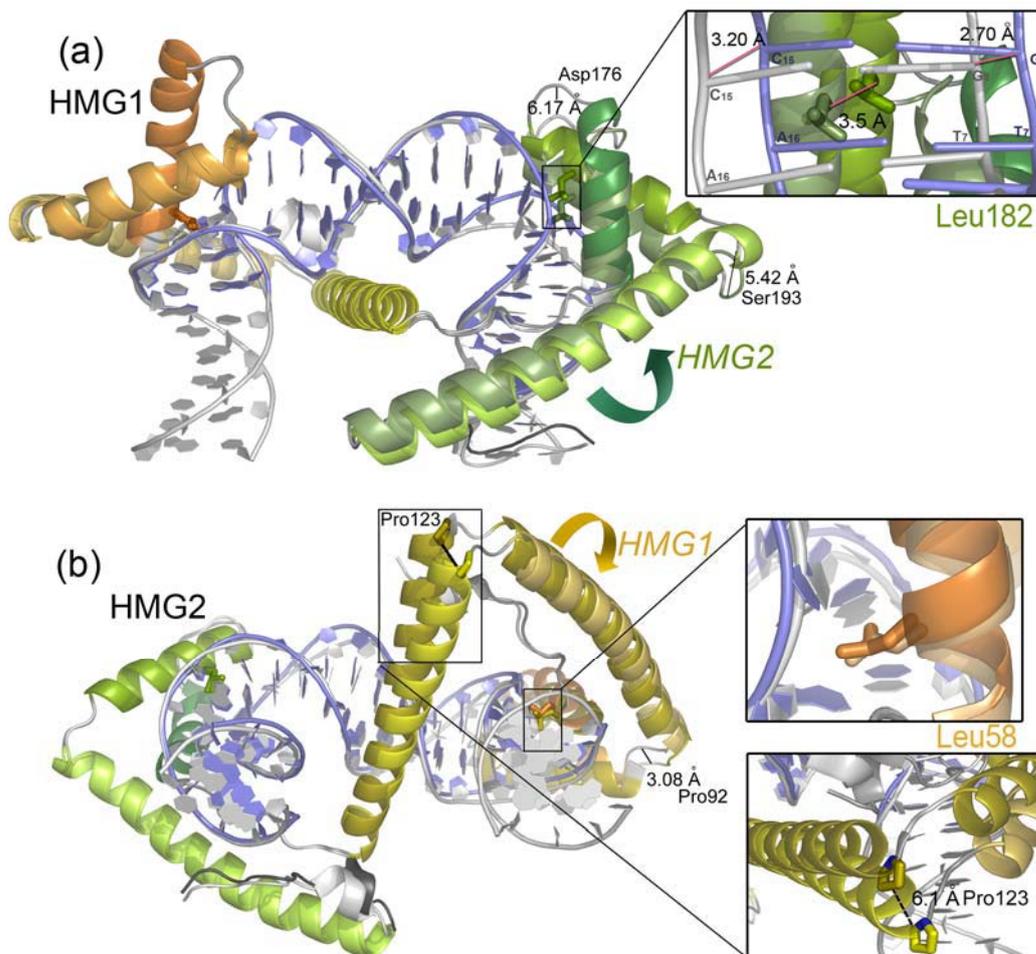
	Residuo	Desplazamiento en Å
HMG2_Hélice 1 (inicio)	Ala161	2.13
HMG2_Hélice 1 (final)	Glu172	4.41
HMG2_Hélice 2 (inicio)	Pro178	3.87
HMG2_Hélice 2	Leu182	3.54
HMG2_Hélice 2 (final)	Asn191	5.26
HMG2_Hélice 3 (inicio)	Asp194	4.60
HMG2_Hélice 3 (final)	Val225	1.88

Tabla_H1. Valores de desplazamiento para HMG2 debido a la superposición de HMG1-LSP22 vs HMG1-LSP28.

	Residuo	Desplazamiento en Å
HMG1_Hélice 1 (inicio)	Ser56	0.93
HMG1_Hélice 1	Leu58	1.15
HMG1_Hélice 1 (final)	Gln71	4.25
HMG1_Hélice 2 (inicio)	Thr77	2.55

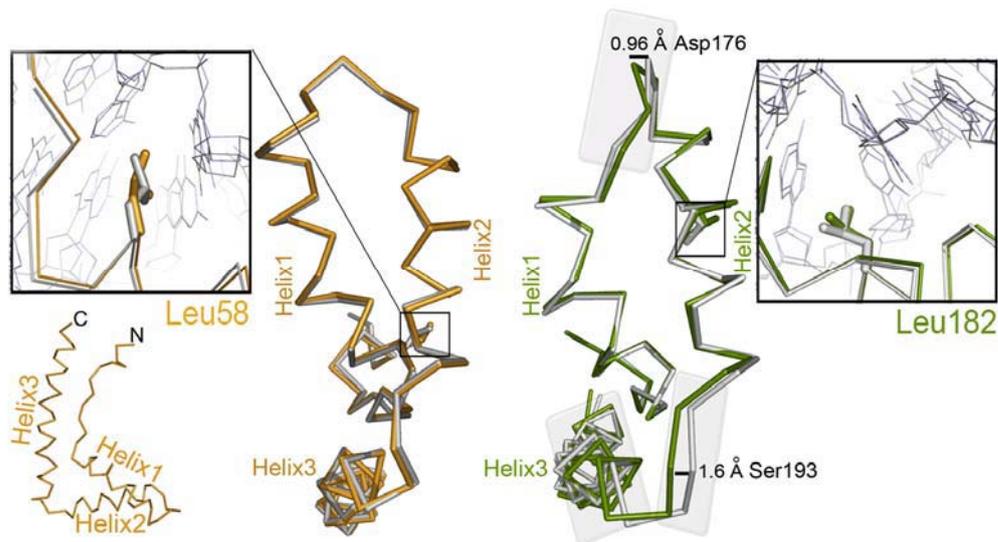
HMG1_Hélice 2 (final)	Glu90	3.04
HMG1_Hélice 3 (inicio)	Asp93	2.53
HMG1_Hélice 2 (final)	Gln120	4.99
Linker	Pro123	6.13

Tabla_H2. Valores de desplazamiento para HMG1 debido a la superposición de HMG2-LSP22 vs HMG2-LSP28.



Fig_H2. Superimposición de TFAM-LSP22 y TFAM-LSP28 para cada dominio HMGbox, (a) fijando HMG1 y (b) para HMG2. Se indican las distancias o desplazamientos máximos.

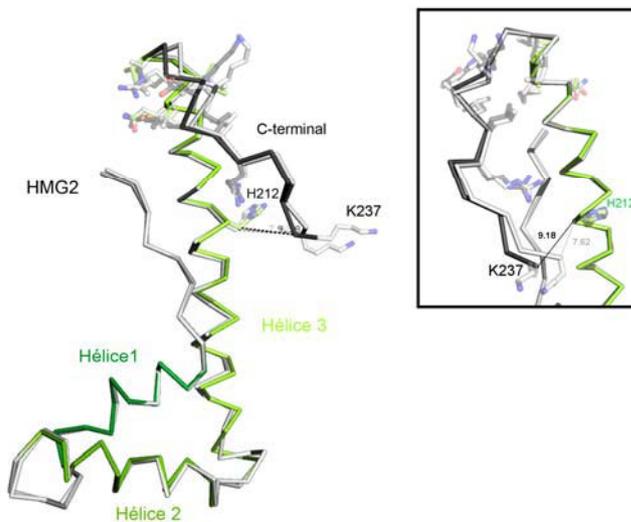
La superimposición local de HMG1 y HMG2 (Fig_H3) refleja que los dos dominios HMG1 son casi idénticos, mientras que en el caso de HMG2 existe una menor similitud entre las dos estructuras, presentando tres zonas de menor encaje. Una de ellas se encuentra en el loop situado entre Hélice1 y Hélice2 (Glu172-Asp176, se desplaza de ~ 0.6 a 0.96 Å); la otra región se trata del loop situado entre Hélice2 y Hélice 3 (Ser193-Glu198, con desplazamientos de ~ 1.6 a 1.0 Å), que afecta el talón de la L. Finalmente, es posible observar que la hélice3 en su conjunto presenta una cierta disparidad entre las dos estructuras, a diferencia de lo que se puede observar en hélice 3 de HMG1.



Fig_H3. Superposición de los dominios HMGbox por separado. A la izquierda, superposición HMG1, en naranja para TFAM-LSP22 y gris para TFAM-LSP28, con el residuo Leu58 mostrado en forma de barra. A la derecha, Superposición HMG2, en verde para TFAM-LSP22 y gris para TFAM-LSP28, con el residuo Leu182 mostrado en forma de *stick* y las zonas con mayor desplazamiento sombreadas en gris.

1.2. Organización de la cola C-terminal

Ambas estructuras carecen de los últimos nueve aminoácidos de la cola C-terminal, posiblemente debido a la flexibilidad intrínseca de esta cola C-terminal. Pese a ello se mantiene un empaquetamiento similar, con el mismo núcleo hidrofóbico entre la cola C-terminal extendida y la hélice 3 de HMG2, descrito anteriormente (Resultados E.4.3). Cabe destacar que en el caso de la estructura TFAM-LSP28 existe un empaquetamiento más estrecho entre la cola C-terminal y la hélice 3, ya que si calculamos la distancia $C\alpha$ - $C\alpha$ entre Lys237 y His212 es menor para TFAM-LSP28 (7.62 Å TFAM-LSP28 frente 9.18 Å TFAM-LSP22) (Fig_H4).



Fig_H4. Representación de la cola C-terminal para ambas estructuras, donde en negro encontramos la cola C-terminal para TFAM-LSP22 y en gris para TFAM-LSP28.

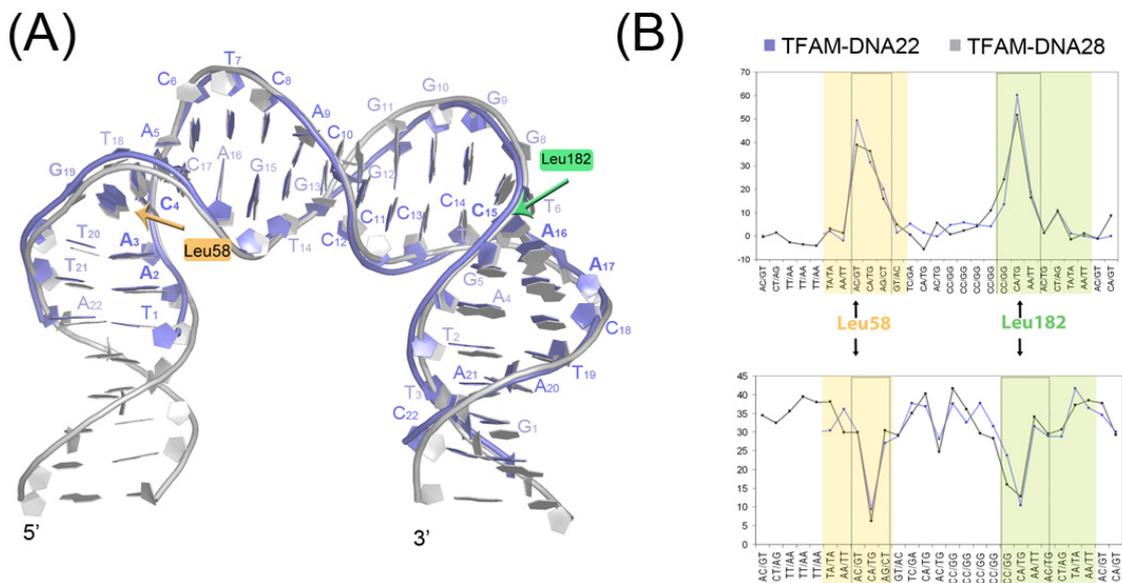
2. Análisis estructural del ADN LSP28 y LSP22

Pese a que en ambos cristales los ADN tienen diferentes longitudes (22pb y 28pb), ambos contienen la región de unión para TFAM en LSP que se había caracterizado anteriormente mediante ensayos *footprinting*. Además, tal y como se ha comentado, la intercalación de residuos por parte de los dominios HMGbox sigue el mismo patrón de reconocimiento y sucede exactamente en los mismos puntos.

El análisis estructural global del ADN para ambas estructuras es parecido (r.m.s.d. 1.15 Å), con una curvatura en forma de U, aunque es posible observar que las dos estructuras de ADN no son idénticas, sobre todo en la región donde interacciona HMG2 (C₁₄-A₂₁). Concretamente en la zona situada entre C₁₄- A₁₇, existe un desplazamiento del esqueleto fosfato del ADN entre las dos estructuras (Fig_H5A).

La conformación de las dos zonas de ADN que interactúan con los dominios HMGox presentan una desviación substancial de B-ADN canónico. En ambos sitios de intercalación existe un ensanchamiento del surco menor, por lo que existe un desenrollamiento local del ADN, aunque globalmente el ADN no está desenrollado.

El perfil para el ángulo de *roll* y *twist* son similares para ambas estructuras. En ambos casos es posible encontrar dos lugares de cambios pronunciados en los niveles de *twist* o *roll* de los pares de bases, ello es indicativo de los sitios donde existe la intercalación. Para el parámetro *roll* es posible localizar dos puntos con un incremento pronunciado, mientras que en el caso del valor *twist* existe una disminución importante en el par de bases consecutivo al lugar de intercalación. Aunque para TFAM-LSP22 los valores presentan un mayor contraste o salto en los puntos de intercalación. (Fig_H5B)



Fig_H5. Comparación estructural del ADN definido en TFAM-LSP28 y TFAM-LSP22. (A) Representación tipo *cartoon* del ADN de 22pb (en azul) y el de 28pb (en blanco), donde es posible diferenciar los puntos de intercalación que provoca la curvatura en forma de U del ADN. (B) Representación de los perfiles para los valores de *roll* y *twist* entre las diferentes base. En naranja y verde se encuentran representadas las regiones del ADN que entran en contacto con HMG1 (naranja) y HMG2 (verde). Con flechas se muestran los lugares de intercalación para las leucinas.

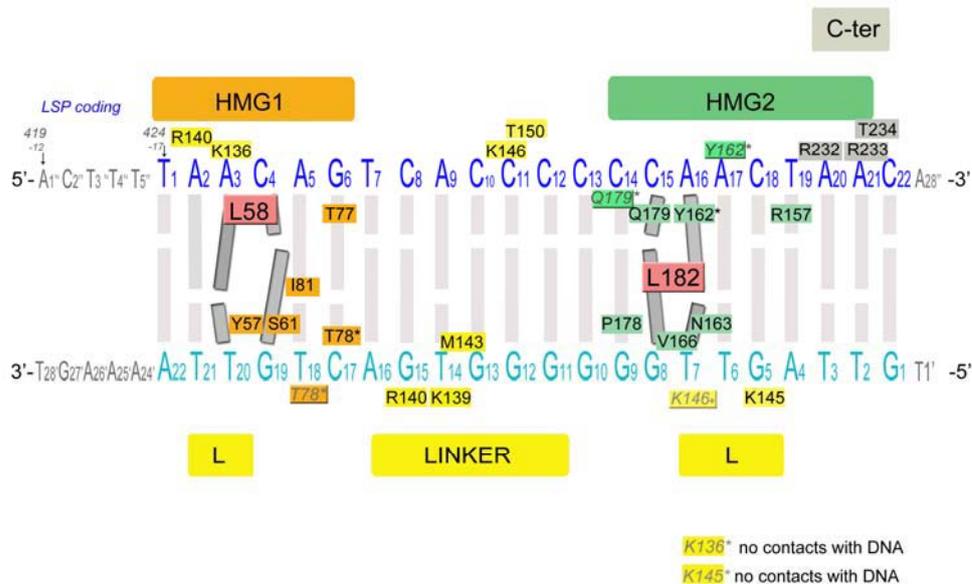
3. Contactos TFAM-ADN en TFAM-LSP22 y TFAM-LSP28

La estructura de TFAM-LSP28 pese a contener un número mayor de pb (28pb) concuerda con los análisis *footprinting* con DNAasa y los experimentos de interferencia por metilación, donde el monómero de TFAM realiza los diferentes contactos a lo largo de la secuencia definida de 23pb (al igual del caso de la estructura TFAM-LSP22). Cada uno de los dominios HMGbox une a los sitios de contacto definidos como sitios predichos de contacto para HMGbox (Ngo, Kaiser et al.), al igual que sucedía en el caso de la estructura de TFAM con la secuencia de 22pb.

En el complejo TFAM-LSP muchos de los contactos de las cadenas laterales con el ADN no son específicos de secuencia, sino que ocurren en el esqueleto fosfato-azúcar del ADN. Sin embargo, existen un pequeño número de contactos con las bases a las que la proteína accede tras abrir en el surco menor. Comparando los contactos presentes entre TFAM y ADN de ambas estructuras, es posible observar que el patrón de interacción con el ADN es prácticamente idéntico con algunos residuos que difieren en cuanto a tipo de contacto. (Fig_H6)

Mientras que Ngo, *et al.* definió Leu 58 y su papel intercalante con la base A3 (de la cadena rica en Citosinas), no identificaron la intercalación de ningún residuo para HMGbox2. Los autores concluyen que al encontrarse Asn163 en la posición del residuo Leu58 en HMG1, residuo polar en lugar de no-polar, HMG2 no realiza el mismo tipo de intercalación que HMG1. En el caso de Asn163 sólo existe un ligero contacto con T₆ (cadena rica en G), a diferencia de lo que sucedía para Leu58. Además, señalaban la presencia de Pro178, en la segunda posición típica de intercalación en HMGbox de unión inespecífica. Esta Pro178 se encuentra insertada en el surco menor y contacta con la G9 (de la cadena rica en G) y está desplazada más de 2 Å para realizar el contacto con la base en comparación con la estructura de HMG2 en su forma no unida al ADN.

Además analizando las estructuras parece que HMG1 de TFAM-LSP28 realiza contactos con distancias inferiores, mientras que para HMG2 la distancia de las interacciones es más similar. Por lo que se refiere a linker y la cola C-terminal, éstos realizan contactos con el esqueleto fosfato de ADN con el mismo tipo de residuos, que pasamos a describir en el siguiente apartado.

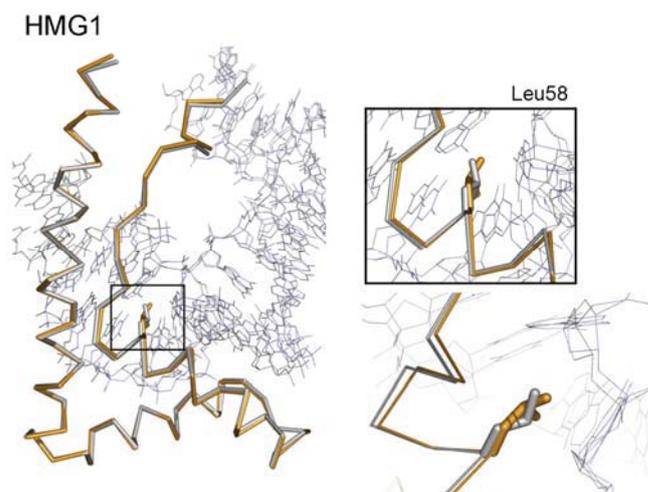


Fig_H6: Esquema con los contactos que realiza TFAM con el ADN de LSP, tanto para TFAM-LSP28 como TFAM-LSP22. En azul se muestra la secuencia de ADN de 22pb usada en ambos casos, mientras que en gris se muestran las bases extra usadas por Ngo, *et al.* Para HMG1 y HMG2 se muestran los contactos específicos con las bases, mientras que para Linker y la cola C-terminal se muestran los contactos existentes, que se refieren a contacto con el esqueleto fosfato en su mayoría. Con asteriscos se muestran los residuos que presentan diferente tipo de interacción entre TFAM-LSP22 y TFAM-LSP28, en gris se muestran los residuos para TFAM-LSP28 que presentan diferente tipo de interacción. El símbolo + y en gris, indica aquellos residuos que para TFAM-LSP28 muestran, además de la interacción existente en TFAM-LSP22, otro tipo de contacto con el ADN.

3.1. Contactos HMG1 con ADN

Cuando se comparan los contactos en HMG1 es posible observar que en TFAM-LSP28 existe una menor distancia, aunque los contactos en cuanto a reconocimiento de bases se mantienen y presentan el mismo patrón de reconocimiento. Los contactos permitirían definir la región de ADN entre las bases (T1-G6)(T22-C17) como zona de contacto con HMG1, región que concuerda con la definida anteriormente para la estructura de TFAM-LSP22. En esta región de ADN es posible destacar los residuos Leu58, Tyr57, Ser61, Thr77 y Thr78 como los principales residuos que interaccionan con las bases del ADN.

Es curioso observar que Leu58 en TFAM-28pb se encuentra intercalada entre las bases A₃-C₄ (de la cadena rica en citosinas), y presenta una conformación diferente entre las dos estructuras. En el caso de Leu58 en TFAM-LSP22, la distancia entre CD1Leu58-N1/C₄ es de 3.4 Å, mientras que para Leu58 en TFAM-LSP28 la distancia es de 3 Å para CD2Leu58_N3/A₃. Además Thr78, que en la estructura TFAM-LSP22 se encuentra formando interacción de tipo van der Waals con C₁₇, para TFAM-LSP28 se encuentra interaccionando con el esqueleto fosfato de T₁₈ debido a que Thr78-TFAM-LSP28 queda alejada de ADN (Fig_H7 y Fig_H1).



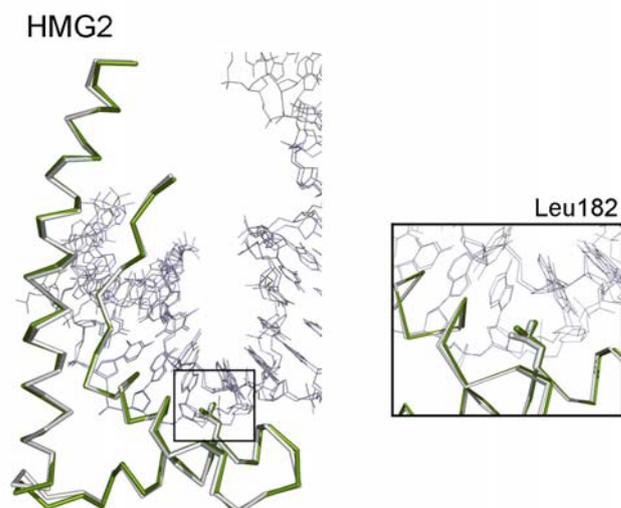
Fig_H7: Superposición del dominio HMG1 en complejo con el ADN, donde se observa la intercalación parcial del residuo Leu58.

Por lo que se refiere a los nucleótidos extra usados por TFAM-LSP28, situados más adyacentes al sitio de inicio de transcripción, tan sólo existe interacción con el esqueleto azúcar-fosfato y es por parte de dos residuos de HMG1: Tyr103 y Trp107, que se encuentran interaccionando con A₂₄.

3.2. Contactos HMG2 con ADN

Los contactos existentes en ambas estructuras permiten definir la región de ADN entre las bases (C₁₄-A₂₁)(G₉-T₂) como zona de contacto de ADN con HMG2, región que concuerda con la definida anteriormente para la estructura TFAM-LSP22. Por lo que se refiere a los contactos proteína-ADN, en este caso parece ser la estructura TFAM-LSP22 la que se encuentra más cercana al ADN. Rubio, *et al.* definieron el residuo Leu182 como intercalante y los siguientes residuos de contacto con las base del ADN: Tyr162, Val166, Pro178 y Gln179. En cambio Ngo, *et al.* sólo definió Pro178, Arg157 y Gln179 (Fig_H8).

Aunque HMG2 presenta un patrón de interacción similar entre las estructuras de TFAM-LSP22 y TFAM-LS28, existen algunas diferencias en cuanto a residuos. Tyr162 no realiza puentes de hidrógeno en TFAM-LSP28 y además interacciona con A₁₇ (en lugar de A₁₆). Leu182 se encuentra más próximo a ADN en TFAM-LSP22 (3.54 vs 3.57). Además Gln179 en la estructura TFAM-LSP22 está formando un puente hidrógeno con C₁₄, mientras que en TFAM-LSP22 se encuentra formado interacción tipo van der Waals con C₁₅. Pese a estas diferencias, cabe destacar la orientación del residuo Leu 182, que en ambos casos se encuentra intercalado totalmente entre las bases (C₁₅-A₁₆, de la cadena rica en citosinas) y (G₈-T₇, de la cadena rica en guaninas).



Fig_H8: Superposición del dominio HMG2 en complejo con el ADN, donde se observa la intercalación total del residuo Leu182.

3.3. Contactos linker con ADN

Los contactos entre linker y ADN son similares para ambas estructuras, destacándose el papel de un grupo de residuos cargados positivamente situados en el extremo C-terminal de linker. Estos residuos pueden ser clasificados en dos tipos. Mientras que Arg140 presenta el mismo patrón de interacciones aunque con distancias más cortas (Arg140-G₁₅: 3.62 Å para TFAM-LSP22, 3.12 Å para TFAM-LSP28; Arg140-A₂: 3.66 Å para TFAM-LSP22, 2.79 Å para TFAM-LSP28). Para el caso del residuo Lys146 las distancias entre Lys146-C₁₁ son similares, pero en el caso de TFAM-LSP28 existe un contacto extra con T₇ (de la cadena rica en guaninas).

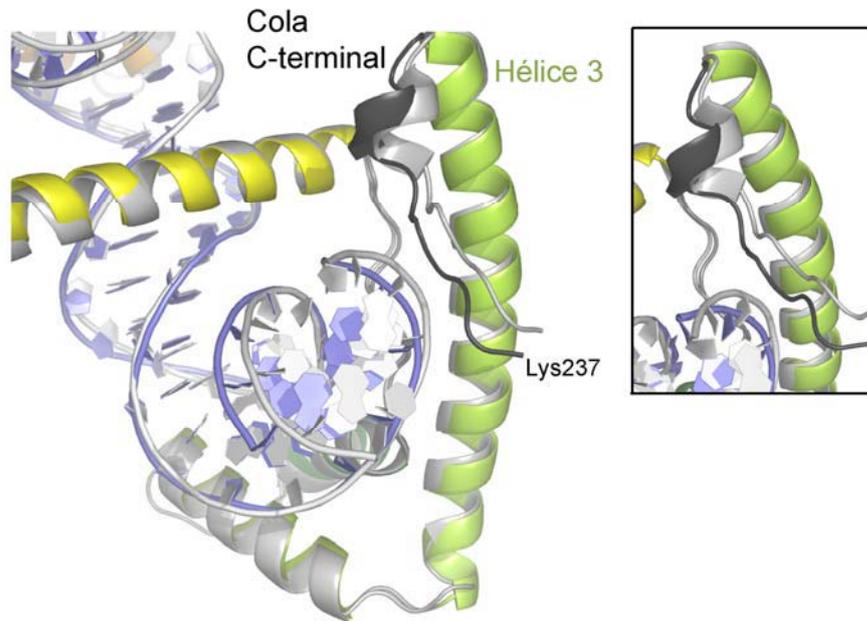
Además, es posible destacar los residuos Lys136 y Lys145 que se encuentran contactando con el surco menor en las regiones de mayor curvatura, aunque estos contactos solo están presentes para la estructura TFAM-LSP22.

Finalmente existe otro pequeño grupo de residuos Lys139, Met143, Th150 que interaccionan con el surco menor del ADN de la región rica en guaninas, con conformación similar a B-ADN canónico.

3.4. Contactos cola C-terminal con ADN

Pese a que durante muchos años se ha definido la cola C-terminal como elemento imprescindible para el reconocimiento específico de secuencia, mediante las estructuras definidas para ambos ADNs (LSP22 y LSP28) no es posible identificar contactos con las bases. Aunque sí es posible destacar una serie de contactos el esqueleto fosfato del ADN a nivel de las bases A₂₁ y C₂₂, de la cadena rica en citosinas. Los residuos que realizan estos contactos son Arg232, Arg233 y Thr234 para ambos casos. Cabe destacar que en ambos casos se ha podido trazar solo hasta el residuo Lys237, por lo que es posible que se pierda información en cuanto a la existencia de otro tipo de contacto. Sin embargo, parece difícil pensar que existe algún tipo de interacción específica con el ADN por parte de los nueve

resíduos no trazados, ya que la dirección de la cola C-terminal emerge en dirección opuesta al ADN (Fig_H9).



Fig_H9. Representación de la cola C-terminal, en gris para TFAM-LSP28 y negro para TFAM-LSP22. Es posible observar como el residuo Lys237 se encuentra orientado en sentido contrario al esqueleto del ADN.

4. Discusión final

Por lo tanto, la comparación de las dos estructuras (TFAM-LSP22 y TFAM-LSP28) permite ratificar la curvatura en forma de U con 180° que promueve la unión de TFAM al ADN. Además se confirma el patrón de reconocimiento para HMGbox, mostrando los mismos residuos de interacción con las bases de ADN y el motivo secuencial AA'C; así como el mismo tipo de contacto para linker y cola C-terminal. También se confirma que la cola C-terminal no realiza ningún tipo de interacción específica con las bases del ADN y se emerge en dirección opuesta al ADN.

Cabe destacar que el hecho más sorprendente es que una estructura respecto a la otra muestra un ligero desplazamiento de la zona N-terminal del linker que provoca una reorientación diferente para HMG2 respecto HMG1. De forma que, la estructura de TFAM-LSP28 muestra una conformación más compacta que TFAM-LSP22. Ello sugiere un posible papel del linker en la reorientación respectiva de los dominios HMGbox.

I. ANÁLISIS ESTRUCTURAL DE LA ESTRUCTURA NOVEL

BIDOMINIO HMGbox

La presencia de múltiples dominios HMGbox en una proteína es en general una característica de proteínas con unión de tipo inespecífica como HMGB1, HMGB2, ABF2...o de proteínas con una menor especificidad de secuencia como es el caso de UBF (*Upstream Binding Factor*), involucrada en transcripción y que contiene entre 4 y 6 dominios HMGbox dependiendo de la especie. Por ello cabe destacar la existencia de TFAM, como caso peculiar, ya que presenta tanto una unión de tipo específico como unión inespecífica y presenta dos dominios HMGbox.

La estructura de TFAM en complejo con ADN es la primera estructura cristalográfica de una proteína nativa con dos dominios tándem HMGbox. Aunque previamente se había resuelto por RMN (*Resonancia Magnética Nuclear*) la estructura de una proteína híbrida construida mediante ingeniería genética, llamada SRY.B (Stott, Tang et al. 2006). Stott, et al. diseñaron una proteína híbrida donde substituyeron el dominio A de HMGB1, de unión inespecífica, por el dominio HMGbox SRY de unión específica. Dejando entre ambos dominios HMGbox la región conectora o linker básica existente entre dominio A y B de HMGB1. La estructura de SRY.B fue resuelta mediante RMN en complejo con una oligo de 16pb con la secuencia consenso de unión de SRY.

En este apartado se pretende realizar un análisis comparativo de ambas estructuras (SRY.B, TFAM), para dilucidar los aspectos diferenciales y en común.

1. Interacción con ADN y curvatura final

La estructura SRY.B en complejo con el ADN ya mostraba que se trata de la longitud de ADN límite para forma complejo. El alojamiento de los dos dominios HMGbox en esta longitud de ADN es consistente con las longitudes de ADN ocupados por dominios únicos HMGbox en complejo con ADN (9pb para SRY (Murphy, Zhurkin et al. 2001), 9pb para LEF-1 (Love, Li et al. 1995), 7pb para HMG-D (Murphy, Sweet et al. 1999)).

En la estructura del complejo SRY.B con ADN, el ADN presenta una conformación intermedia entre forma A y B, debido a la introducción de una curvatura global de $\sim 101^\circ$ en forma de U incompleta (Fig_11,A), suma de la curvatura inducida por SRY y el dominio B en los dos extremos del ADN con magnitud similar. Mientras que en el caso de TFAM-LSP22, la curvatura global del ADN debida al efecto de ambos dominios es de 180° completando un U-turn.

Cuando analizamos la interacción con el ADN de los dominios en el caso de SRY.B, encontramos que el dominio SRY muestra el mismo patrón de reconocimiento e intercalación que en el caso de la estructura resuelta para SRY-DNA (Murphy, Zhurkin et al. 2001) mediante la cadena lateral de un residuo Isoleucina. Así mismo, el dominio B de SRY.B se une al ADN de la misma forma descrita para HMGB1, donde existen dos residuos intercalante (Phe, Ile) (característica de los dominios de unión inespecífica). Cabe destacar que HMG1 de TFAM muestra un patrón de reconocimiento similar a SRY, mientras que el tipo de interacción de HMG2 no es similar al dominio B. Aunque en ambos casos el dominio de unión específica (SRY o HMG1) actúa como ancla para la unión de la proteína

al ADN; dirigiendo así la posterior unión en una región adyacente por parte del dominio de unión inespecífica (B o HMG2).

2. Orientación dominios HMGbox

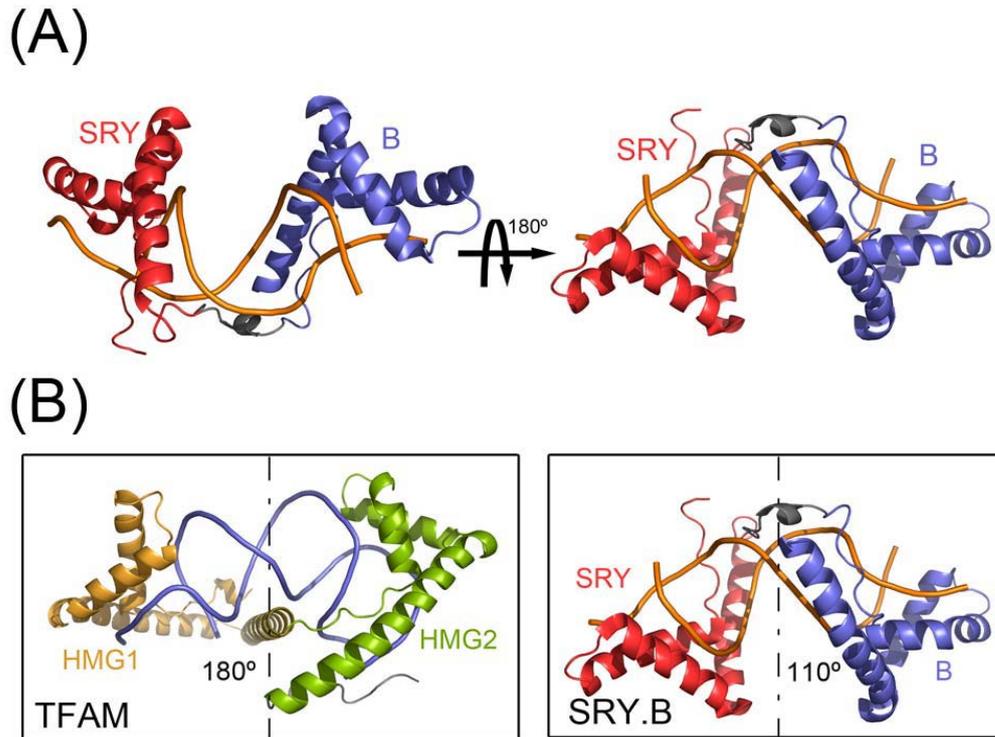
Otra de las diferencias remarcables entre las dos estructuras es la diferente orientación de los dominios HMGbox uno respecto al otro (Fig. I1,B). En el caso de TFAM, tal y como se ha comentado, los dominios se encuentran relacionados mediante un eje de simetría binario perpendicular al ADN y orientados cara a cara (*head-to-head*), con la cara interna o cóncava mirando una hacia la otra. En el caso de SRY.B, la orientación de los dominios es completamente diferente, pese a que también existe un eje de simetría binario que los relaciona. SRY.B presenta los dominios orientados mediante el brazo largo de la L, es decir, mediante la Hélice3 de ambos dominios (*tail-to-tail*).

3. Región conectora o linker

Una región de la estructura en la que es necesario incidir es la región conectora o linker. Para SRY.B, la región linker entre los dos dominios está constituida por la siguiente secuencia: KGETKKKFK, y corresponde a la región desestructurada en HMGB1. Al contrario de lo que sucede en solución para HMGB1, la región del linker en complejo ofrece una conformación con cierto grado de rigidez y se encuentra localizado en la zona del surco menor situada entre las regiones de contacto de los dominios HMGbox. Debido a que este linker presenta cierta movilidad, no se encuentra completamente definida. Este linker se encuentra desprovisto de una clara estructura secundaria, pero existen dos puntos de contacto con el ADN donde se halla definido más claramente. Estos dos puntos corresponden con la zona N-terminal del linker que contacta con el sitio de unión de SRY, y la zona C-terminal del linker que contacta con el sitio de unión del dominio B. Para TFAM la situación es una tanto diferente, ya que se trata de un linker de mayor longitud (25aa), flexible en solución (ver apartado SAXS) pero que en complejo con el ADN de 22pb yace en el surco menor del ADN, donde el extremo C-terminal realiza contactos con diferentes puntos del ADN estabilizando el *U-turn*.

Por lo tanto, parece ser que en ambos casos la región flanqueante básica entre los dominios (linker) parece potenciar la capacidad de unión al ADN proporcionando contactos con el ADN adicionales, y finalmente estabilizando la curvatura.

Además por último, es posible constatar que la diferencia en longitud del linker (9aa para SRY.B, 25aa para TFAM) y ADN (16pb para SRY.B frente 22pb para TFAM), implica una diferente posición relativa de los dominios en las dos estructuras comentadas. Es decir, el linker permite una reorganización de los dominios.



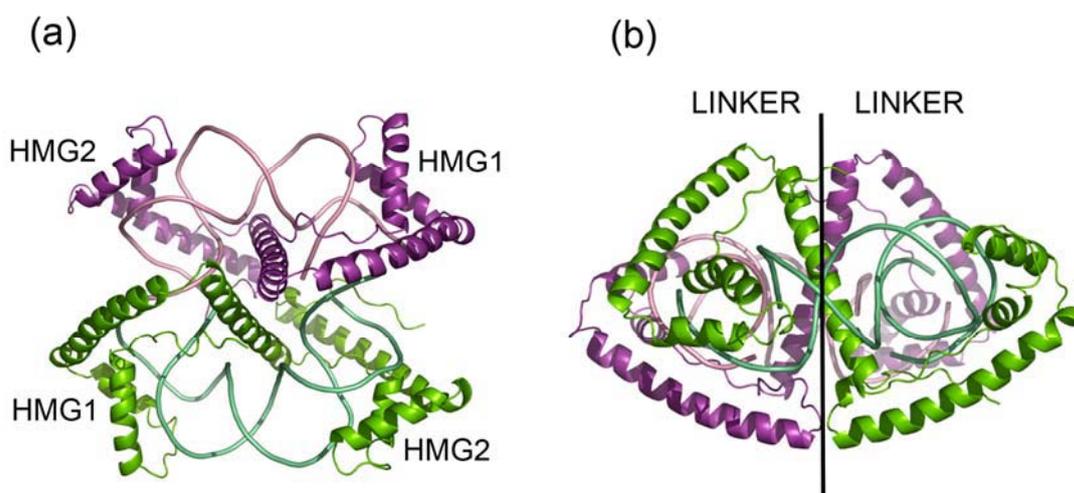
Fig_I1. (A) Representación de la estructura SRY.B en complejo con una ADN de 16pb, con la secuencia consenso de unión de SRY. En rojo se muestra el dominio SRY y, en azul el dominio B de HMGB1. Entre los dos dominios se encuentra la región conectora coloreada en gris. La estructura se representa con dos orientaciones diferentes relacionadas por un ángulo de 180° . (B) A la izquierda se encuentra representada la estructura de TFAM en complejo con LSP22, con los dos dominios HMG1 y HMG2; así como el eje de simetría y el ángulo de curvatura. A la derecha se encuentra representada la estructura de SRY.B, con el ángulo final de curvatura y el eje de simetría.

J. INTERACCIONES PROTEÍNA-PROTEÍNA

1. Empaquetamiento cristalino

En la literatura, ha sido ampliamente propuesta la supuesta dimerización de TFAM en su forma unida al ADN (Gangelhoff, Mungalachetty et al. 2009) (Kaufman, Durisic et al. 2007). Además existen diversos experimentos que demuestran la unión cooperativa de TFAM en presencia de ADN (Kaufman, Durisic et al. 2007; Wong, Rajagopalan et al. 2009). Por lo tanto, parece bastante evidente que la unión de TFAM al ADN favorece una reorganización de la proteína que podría ayudar a las interacciones proteína-proteína responsables de la cooperatividad. Es por ello que en este apartado se analiza el empaquetamiento cristalino para poder establecer cuál es la región causante de dimerización, o mejor dicho poder acotar la zona donde se dan las interacciones proteína-proteína.

En el caso de TFAM en complejo con LSP22, la celda cristalina presenta dos moléculas de proteína con sus respectivos oligos (2 cadenas de ADN por proteína) por unidad asimétrica con las siguientes dimensiones del celdilla: $a= 113.90$, $b= 117.20$, $c= 56.53$; $\alpha=\beta=\gamma= 90^\circ$ y grupo espacial $P2_12_12$. En este caso las dos moléculas de proteína se encuentran relacionada por el linker, donde la hélice α del linker está orientada con la otra en una forma antiparalela con un ligero ángulo de desplazamiento (Fig_J1). Parece que en esta interacción existen residuos de capital importancia: Met127, Glu130, Lys131 que se encuentran orientados frente a la otra molécula ofreciendo una posible interacción por enlaces salinos con los residuos contrapuestos o mediante interacciones hidrofóbicas en el caso de Met127. Por otra parte, es interesante fijarse en el hecho que las hebras de ADN de las diferentes moléculas de proteína se encuentran casi formando un perfecto *stacking* entre ellas, lo cual sugiere que quizás el ADN de alguna forma dirige la interacción entre proteínas a través del linker. No obstante, esta interacción no se observa en el caso de la estructura TFAM-LSP28.

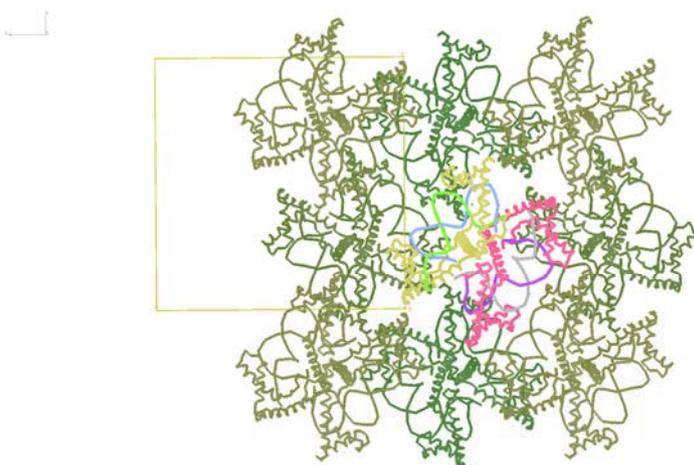


Fig_J1. Composición de la celdilla unidad para TFAM-LSP22, con dos moléculas proteicas con sus respectivos ADN (en color verde y morado) relacionadas mediante una simetría no cristalográfica. (a) Orientación de las dos complejos con el linker paralelo al eje de visión, donde es posible observar como los ADN's de los diferentes complejos parece establecer un perfecto *stacking* formando un círculo perfecto. (b) Orientación de los dos complejos donde es posible ver el papel del linker, como se recolocan de forma antiparalela.

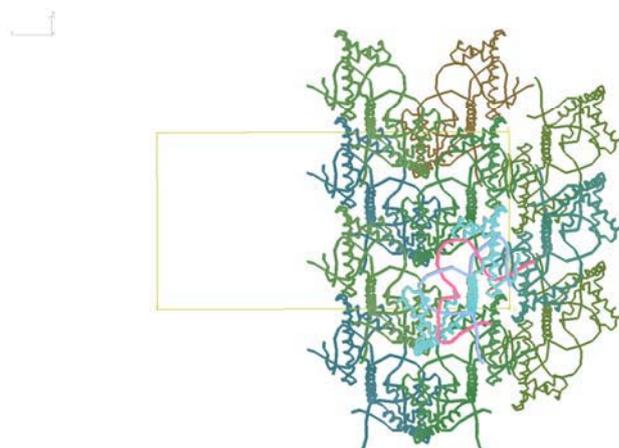
Además al visualizar el empaquetamiento cristalino (Fig_J2, parte superior) es posible observar que existe una segunda interacción proteína-proteína mediante la Hélice 3 del dominio HMG1, donde ambas Hélice 3 presentan una orientación casi antiparalela con un ligero desplazamiento, formando una X entre ellas.

En el caso de TFAM en complejo con LSP28, la celdilla unidad presenta tan solo una molécula de proteína y sus dos cadenas de ADN por unidad asimétrica con las siguientes dimensiones de celdilla: $a= 68.44$, $b= 81.91$, $c= 161.25$; $\alpha=\beta=\gamma= 90^\circ$ y grupo espacial $C222_1$. Al visualizar el empaquetamiento, es posible darse cuenta que al igual que para TFAM-LSP22 existe una orientación antiparalela entre Hélice 3 del dominio HMG1 entre las moléculas proteicas (Fig_J2).

TFAM_LSP22



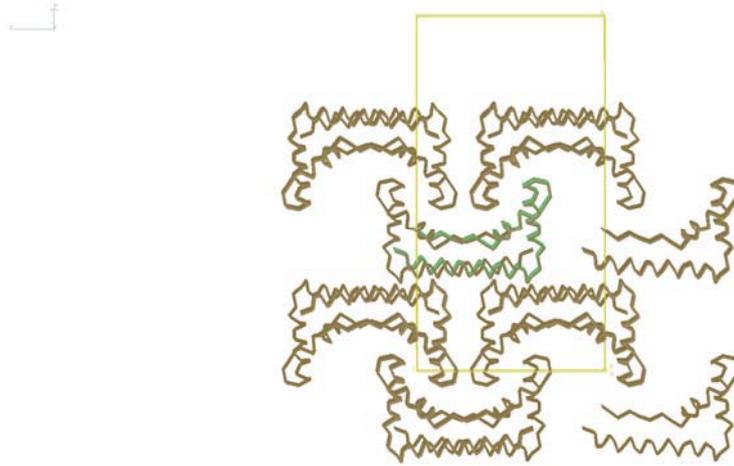
TFAM_LSP27



Fig_J2. Empaquetamiento cristalino para TFAM-LSP22 (parte superior) y TFAM-LSP28 (parte inferior). En la representación se encuentran indicadas las moléculas simétricas para las estructuras resueltas y las dimensiones de la celdilla.

También resulta interesante analizar el empaquetamiento cristalino de la estructura resuelta por *Gangelhoff, et al.* del dominio HMG2 en su forma no unida al ADN. En este caso

se encuentran dos moléculas proteicas por unidad asimétrica, perfectamente orientadas de forma antiparalela mediante la Hélice 3 del HMG2. El cristal pertenece al grupo espacial $C222_1$, con las siguientes dimensiones de celdilla: $a= 30.28$, $b= 99.35$, $c= 51.91$; $\alpha=\beta=\gamma= 90^\circ$. Lo cual da lugar a un empaquetamiento cristalino divertido donde las L o botas definidas por los dominios HMG2 dan lugar a una especie de consonante C (ver Fig_J3)

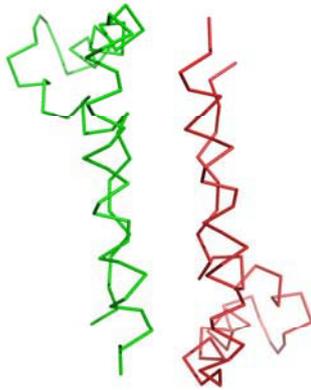


Fig_J3. Empaquetamiento del dominio HMG2 resuelto por Gangelhoff et al, con el código de acceso a pdb:3FGH

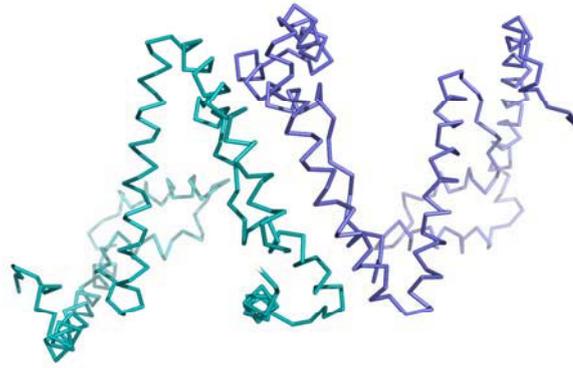
Por lo tanto, existen tres posibles zonas de interacción proteína-proteína (Fig_J4) mediante el empaquetamiento cristalino observado para las diferentes estructuras resueltas:

- Hélice 3 de HMG2-Hélice 3 de HMG2, presente en la estructura resuelta para el dominio HMG2 (Fig_J4(A))
- Hélice 3 de HMG1-Hélice 3 de HMG1, presente en TFAM-DNA28 y TFAM-DNA22 (Fig_J4 (B), (C))
- Zona central del linker, interacción presente entre las dos moléculas de la unidad asimétrica de TFAM-DNA22 (Fig_J4(C))

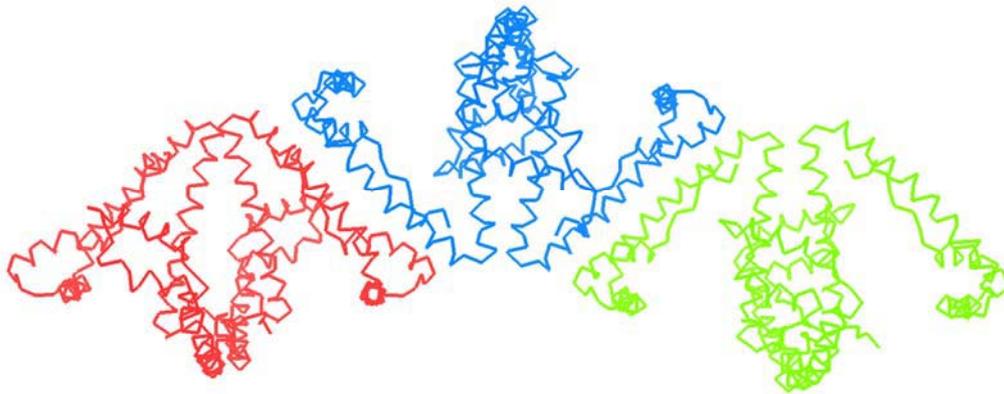
(a) HMG2, PDB= 3FGH



(b) TFAM-DNA28, PDB= 3TMM



(c) TFAM-LSP22, PDB= 3TQ6



Fig_J4. Empaquetamiento cristalino de TFAM en diferentes estructuras, con rectángulos de color gris donde se indican las interacciones proteína-proteína existentes en la unidad asimétrica o con moléculas simétricas dentro del empaquetamiento cristalino. (A) Empaquetamiento cristalino en la estructura de HMG2-unbound, con el código de acceso: 3FGH. (B) Interacción proteína-proteína en la estructura TFAM-LSP28, que implica la interacción entre HMG1 de la molécula con su simétrica. (C) Interacción proteína-proteína en la estructura TFAM-LSP22, que implica la interacción entre linker de las dos moléculas de la unidad asimétrica, más la interacción a través de HMG1 con la molécula simétrica.

2. Análisis de la superficie de interacción y predicción de ensamblajes

El servidor PISA permite analizar estructuras y realizar predicciones de posibles complejos estables mediante el análisis de propiedades fisico-químicas como son los enlaces de hidrógeno, energía de ganancia por solvatación, cálculo del área de interfaz, enlaces salinos entre la interfaz, especificidad hidrofóbica, etc. (Krissinel and Henrick 2007).

La superficie accesible a solvente para el ensamblaje de la unidad asimétrica ([A:C:D]+[B:E:F]) (dónde A y B son las cadenas de proteína, y C/D/E/F son las cadenas de ADN) es de 33294 Å² con una energía libre de solvatación (ΔG^{int}) después del ensamblaje de -85.2 kcal/M.

El cálculo de la interfaz permite obtener la superficie de la región de contacto entre las dos moléculas proteicas de la unidad asimétrica, que contactan a través del linker y extremo C-terminal de la Hélice 3 de HMG2, la interfaz se calcula como la diferencia entre el total de superficie expuesta al solvente de las estructuras por separado y juntas dividido entre las dos: 593.1 Å². Además, el valor estadístico P para la ganancia por energía libre de solvatación, mediada de la especificidad de la interfaz (relacionado con la hidrofobicidad) es Δ^iG P-value=0.084, indicando una elevada probabilidad que suceda. Permite predecir dos potenciales enlaces salinos (Tabla_J2) y 6 puentes de hidrógeno (Tabla_J1) y el número de residuos implicados son 16 para molécula B y 17 para molécula A. Además existe el parámetro CSS (*Complex Formation Significance Score*) que presenta valores de 0 a 1 como indicador de la relevancia de la interfaz para la formación del complejo. En el caso de la interacción por el linker, el valor CSS es de 0.862. Ello indica que esta interfaz juega un papel esencial en la formación del complejo.

Puentes hidrógeno	Molécula B	Distancia (Å)	Molécula A
1	Ser 124 (OG)	2.80	Ser124 (OG)
2	Arg227 (N)	3.79*	Gly226 (O)
3	Lys228 (N)	3.05	Gly226 (O)
4	Lys228 (NZ)	2.80	Val225 (O)
5	Gly226 (O)	3.05	Lys228 (N)
6	Val225 (O)	3.21	Lys228 (NZ)

Tabla_J1. Potenciales puentes de hidrógeno entre las moléculas A y B de la unidad asimétrica a través de linker y el extremo C-terminal de la Hélice 3 de HMG2. El asterisco indica distancias al límite de una distancia estandar para un puente de hidrógeno o una interacción electrostática (2.5 Å a 3.5 Å).

Enlaces salinos	Molécula B	Distancia (Å)	Molécula A
1	Lys131 (NZ)	3.37	Glu130 (OE2)
2	Glu130 (OE2)	3.58	Lys131 (NZ)

Tabla_J2. Potenciales enlaces salinos entre las moléculas A y B de la unidad asimétrica a través de linker y el extremo C-terminal de la Hélice 3 de HMG2.

En el mismo programa PisaServer la herramienta “*assemblies*” nos permite calcular información de la estructura cuaternaria, indicando posibles interfaces favorecidas. Mediante esta herramienta es posible encontrar la posible interfaz definida entre HMG1 de moléculas simétricas en el empaquetamiento cristalino que presenta una superficie de 1025.9 Å² y un P-value de energía libre por solvatación de Δ^iG P-value=0.367. Con 29 residuos para cada molécula implicados en la interfaz, 7 potenciales puentes de hidrógeno (Tabla_J3) y 8 enlaces salinos (Tabla_J4). En este caso el valor CSS es de 0.274, lo cual implica que esta interfaz juega un papel auxiliar en la formación del complejo.

Puentes hidrógeno	Molécula B	Distancia (Å)	Molécula A
1	Arg59 (NH2)	2.79	Glu113 (OE2)
2	Tyr99 (OH)	3.15	Glu113 (OE2)
3	Gln125 (NE2)	3.86	Gln71 (OE1)
4	Glu63 (OE1)	2.77	Arg116 (NH2)
5	Gln71 (OE1)	2.94	Gln125 (NE2)
6	Tyr99 (OH)	3.05	Arg116 (NH2)
7	Glu113 (Oe2)	3.16	Tyr99 (OH)

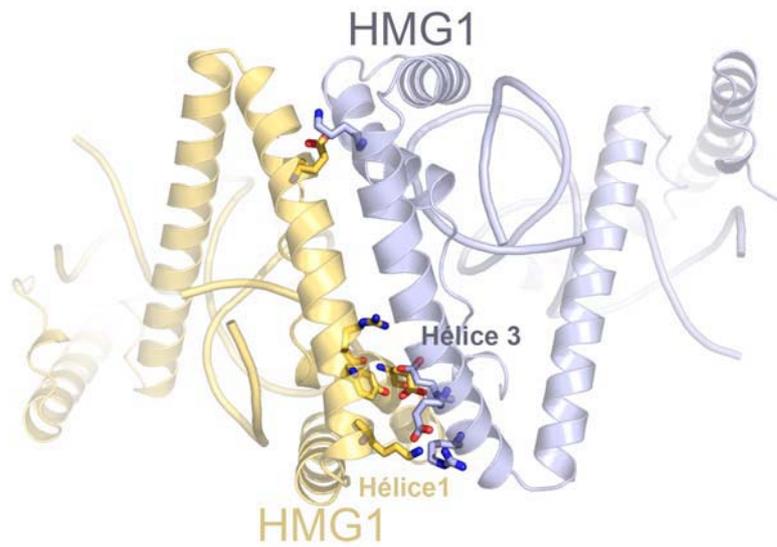
Tabla_J3. Potenciales puentes de hidrógeno entre las moléculas A y B a través del dominio HMG1.

Enlaces salinos	Molécula B	Distancia (Å)	Molécula A
1	Arg59 (NH2)	2.79	Glu113 (OE2)
2	Arg59 (NH2)	3.41	Glu113 (OE1)
3	Lys95 (NZ)	3.53	Glu112 (OE2)
4	Lys95 (NZ)	3.53	Glu112 (OE1)
5	Glu63 (OE1)	2.77	Arg116 (NH2)
6	Glu63 (OE2)	3.86*	Arg116 (NH2)
7	Glu112 (OE1)	3.25	Lys95 (NZ)
8	Glu112 (OE2)	4	Lys95 (NZ)

Tabla_J4. Potenciales enlaces salinos entre las moléculas A y B de la unidad asimétrica a través del dominio HMG1. El asterisco indica distancias al límite de una distancia estandar para un puente de hidrógeno o una interacción electrostática (2.5 Å a 3.5 Å).

Si comparamos las superficies de la interfaz y la ganancia por energía libre de solvatación de la interacción a través de HMG1 de dos moléculas simétricas se ve más favorecida respecto a sus propiedades fisicoquímicas y área de la superficie que la interacción a través del linker. Aunque el valor CCS para la interacción a través del linker presenta un valor más relevante.

Para el contacto por HMG1, los residuos que participan en el mismo se encuentra en la superficie exterior (es decir, la región que no contacta con el ADN) del brazo corto de la HMGbox formado por la Hélice 1 y Hélice2 para una molécula, y en la Hélice 3 para la estructura simétrica (Fig_J5). Por lo tanto, existe una orientación casi antiparalela entre la Hélice 3 de HMG1 de diferentes estructuras que permite la interacción de Hélice 3 con el brazo corto de la HMG1 contigua. Este tipo de interacción parece ser dirigida por la fuerza iónica del entorno.



Fig_J5. Representación de dos estructuras simétricas de TFAM, donde existe el contacto entre HMG1 de ambas. Mediante el servidor PisaServer o por inspección visual midiendo las distancias entre átomos, fue posible determinar los posibles residuos de contacto con interacciones de tipo enlace salino o puente de hidrógeno. Los residuos con este tipo de interacciones están representados en forma de barra.

Para verificar estos posibles contactos es necesario realizar experimentos bioquímicos, por ejemplo: *crosslinking* con el ADN y la proteína nativa y mutantes, que rompan estas posibles interacciones, o análisis mediante EMSA, para entender mejor las interacciones proteína-proteína que juegan un papel importante en el proceso de empaquetamiento del mtADN.

K. RESULTADOS: PERMUTACIÓN CIRCULAR

1. Técnicas de medida de la curvatura del ADN

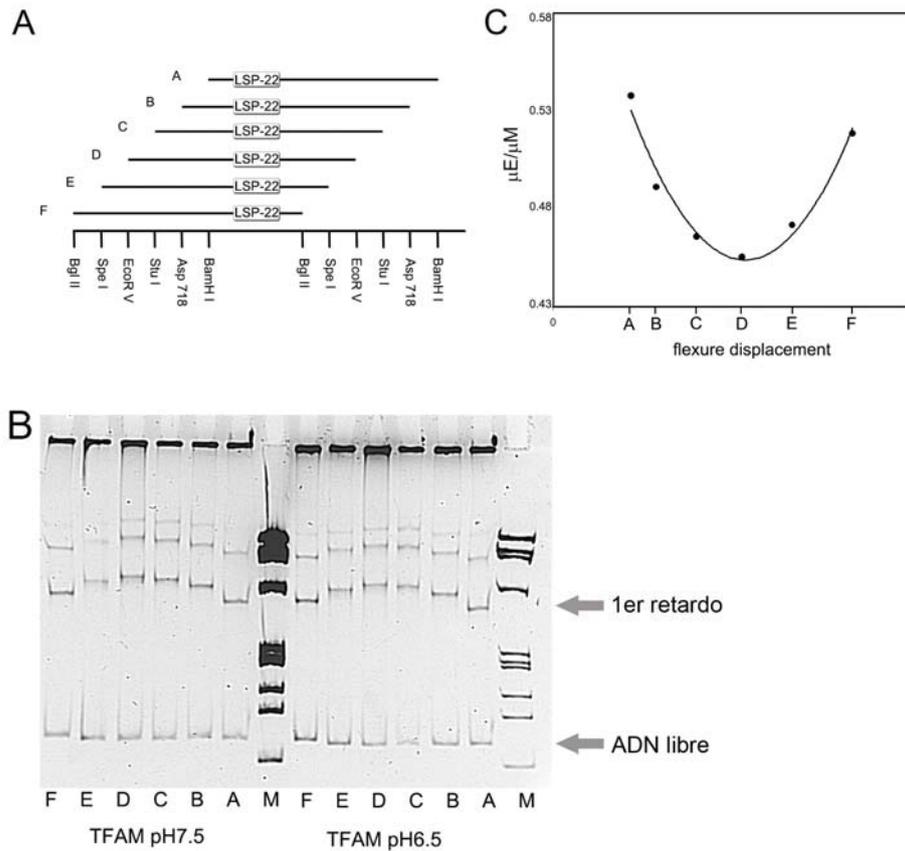
Para el ensayo de la curvatura inducida por una proteína sobre el ADN existen diversas metodologías: métodos cristalográficos permiten calcular un ángulo preciso pero dicho ángulo puede verse distorsionado por las fuerzas del empaquetamiento cristalino; RMN (Resonancia Magnética Nuclear) permite el cálculo del ángulo en solución pero existe una limitación en el tamaño del complejo; FRET (*Fluorescence Resonance Energy Transfer*) también permite el cálculo del ángulo en solución pero implica un diseño más elaborado del experimento. Finalmente, existe una técnica llamada permutación circular que permite el cálculo del ángulo de curvatura mediante ensayos de movilidad electroforética. Esta técnica fue muy usada a finales de la década de los ochenta para el estudio de curvatura de proteínas HMGB como SRY (Ferrari, Harley et al. 1992) o IHF (Thompson and Landy 1988) o TCF1 α (Read 1994), debido a que los requerimientos y material necesarios eran fáciles de alcanzar.

2. Ensayo de permutación circular para el complejo TFAM-fragmento ADN wt y mutado

Esta parte del proyecto se realizó bajo la supervisión del Dr. Jordi Bernués en el IBMB-CSIC.

Para analizar la curvatura que TFAM induce sobre LSP22, se aplicó la técnica conocida como permutación circular, para la que se generaron fragmentos de ADN de 143 pb con la secuencia de LSP22 situada en diferentes posiciones, en concreto 6 variantes diferentes (Fig_K1, A), y se analizó la movilidad electroforética de los complejos. Curiosamente, era el fragmento que presenta la diana LSP22 en la región central, el que presenta un retardo mayor en los geles EMSA, ya que implicaba una conformación del complejo con un volumen mayor y, por tanto más retardado. Posteriormente, se representaron los valores de movilidad electroforética relativa ($\mu\text{M}/\mu\text{E}$) en función del lugar de flexibilidad (*flexure displacement*) o posición relativa de LSP22 dentro del fragmento (D/L) (Fig_K1, C)

La curvatura calculada a través de este método es de $\sim 67^\circ$, lo cuál no corresponde con el ángulo observado por la estructura cristalográfica. El experimento se realizó a diferentes concentraciones de NaCl (20mM y 100mM) y diferentes pH (6.5 y 7.5), sin que existiera ningún cambio significativo (Fig_K1, B). Cabe destacar la presencia de bandas superiores de retardo, posiblemente debido a la presencia de diferentes estados de oligomerización. Este detalle también fue descrito anteriormente mediante la aplicación de esta técnica para TFAM, pese a que en este caso no pudieron identificar el grado de curvatura debido a que identificaron dos mínimos aparentes (Fisher, Lisowsky et al. 1992).



Fig_K1. Ensayo de permutación circular con TFAM. (A) Representación esquemática de los seis fragmentos generados a partir de la digestión con diferentes enzimas de restricción: Bgl II, Spe I, Eco RV, Stu I, Asp 718, Bam HI, que serán llamados A, B, C... en función de la enzima usada. Es posible identificar la secuencia para el oligo LSP22 situada en diferentes posiciones a lo largo de los 143 pb totales. (B) Gel nativo PAGE al 8%, para los complejos de TFAM con diferentes fragmentos, a dos pH diferentes (7.5 y 6.5). El carril M indica marcador de ADN. En los pocillos se identifican estados de agregación, relacionados con la unión de TFAM al ADN restante producto de la digestión del plásmido (~ 1.9Kb), que permiten arrastrar el tipo de unión inespecífica. (C) Representación de $\mu\text{M}/\mu\text{E}$ vs la posición relativa de LSP22 en el fragmento para el caso de TFAM en la condición de pH 7.5, 100mM NaCl. Es posible ajustar una función polinómica $y=0.7608x^2-0.8128x+0.6707$ con un coeficiente de regresión $r^2= 0.97$.

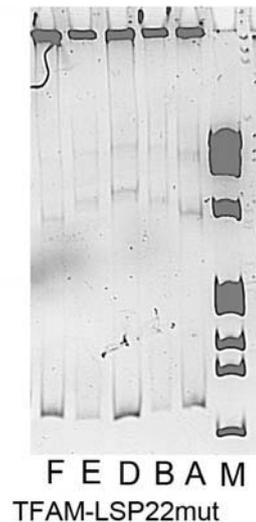
Además, se realizó el mismo experimento con una secuencia de ADN modificada para la zona central, ya que existía la sospecha de que la región central de seis guaninas podría generar la formación de supra estructuras llamada G-cuadruplex que ayudasen a una pre-curvatura inicial del ADN. La secuencia mutada era: 5'-TAACAGTCACCAACCAACTAAC-3', con las dos citosinas centrales mutadas por dos adeninas. Ello no dió lugar cambios visibles en la curvatura del ADN, observándose un ángulo similar (Fig_K2).

LSP22

5'-TAACAGTCACCCCCCAACTAAC-3'

LSP22-mut

5'-TAACAGTCACCAACCAACTAAC-3'

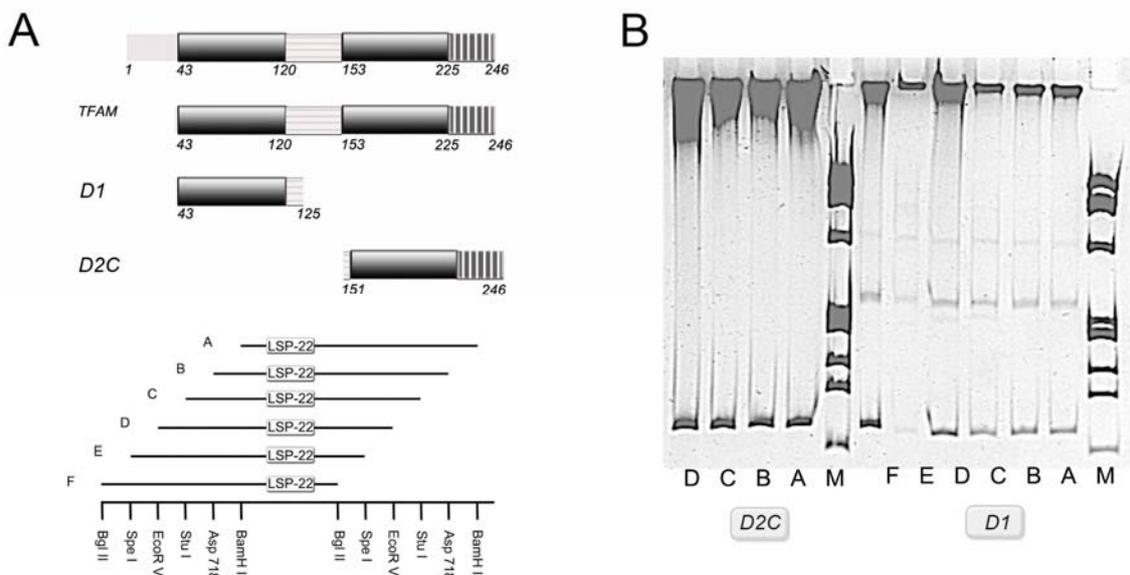


Fig_K2. Secuencia LSP22 nativa y LSP22 mutada para las guaninas centrales. El gel de retardo muestra una apariencia similar, la representación de los valores experimentales permite fijar la siguiente función polinómica de segundo grado $y = 1.2074x^2 - 1.0723x + 0.96893$ con un coeficiente de regresión $r^2 = 0.94$, dando lugar a un ángulo similar que en el caso TFAM-LSP22.

3. Ensayo con dominios

También se estudió la capacidad de unión y curvatura de HMG1 y HMG2, en concreto con los dominios definidos como D1 (Ser44-Gln125) y D2C (Leu151-Cys246). En el caso de D1 presentaba unión específica por el ADN, pero una curvatura no significativa; y D2C no mostraba unión al ADN de 143 pb (Fig_K3).

Este estudio permitió aclarar que la capacidad de torsión del ADN por parte de TFAM no está solo gobernada por los dominios HMGbox, sino también por una acción cooperativa de los demás segmentos.



Fig_K3. Permutación circular para los dominios HMGbox de TFAM. (A) Representación esquemática de los constructos usados para el ensayo y los fragmentos de 143pb usados (idénticos a los casos anteriores). (B) Gel nativo PAGE al 8%, para los complejos de los dominios (D1 y D2C) con diferentes fragmentos.

4. Limitaciones de la técnica

Como se ha comentado más arriba el ángulo de curvatura observado mediante permutación circular fue de $\sim 67^\circ$ para TFAM. Ello difiere significativamente del ángulo determinado mediante la estructura cristalográfica. Además, para el dominio HMG1 (D1) no se observó curvatura sobre el ADN, pero si unión específica, lo que resulta impactante, ya que está ampliamente descrito que los dominios HMGbox provocan una distorsión del ADN después de su unión. Es interesante remarcar que esta técnica está sujeta a varias limitaciones. Antes de todo cabe destacar que esta técnica permite calcular el ángulo de curvatura, pero no con la misma precisión que por ejemplo la estructura cristalográfica.

Existen dos factores que pueden haber afectado en esta variación en el ángulo de curvatura entre las dos técnicas. Primero de todo, ha sido descrito que la longitud del fragmento de partida (L) es importante (Thompson and Landy 1988). Fragmentos demasiados largos pueden dar lugar a una disminución del ángulo, debido a la existencia de una mayor flexibilidad del ADN; mientras que fragmentos demasiado pequeños se encuentran en el límite donde el coste elástico para que se produzca la curvatura no es asumible. Esta distancia es conocida como "longitud de persistencia" (*persistence length*), que corresponde con una longitud de 50 nm. Si calculamos la longitud del fragmento de 143pb, encontramos que esta distancia es de 48.6nm. Por lo tanto, es posible que los fragmentos generados para usar la técnica fueran demasiado cortos para poderse producir una curvatura extrema debido a la rigidez intrínseca del ADN. Otro de los factores que pueden haber afectado al cálculo del ángulo es un problema más teórico, y es que la base de esta técnica tiene en cuenta un modelo de curvatura que recuerda a un triángulo, con tres vértices y el ángulo θ (ver materiales y métodos, apartado permutación circular). Si observamos la estructura cristalográfica de TFAM en complejo con LSP22, no existe un modelo tipo triángulo con un solo ángulo (un lugar de intercalación), sino una forma de U con dos ángulos (con dos lugares de intercalación).

L. RESULTADOS: DISPERSIÓN DE RAYOS X A BAJO ÁNGULO (SAXS)

Estudios previos basados en dicroísmo circular mostraron un incremento en el contenido de hélices α de TFAM después de su unión con el ADN (Wong, Rajagopalan et al. 2009). Además la estructura cristalográfica de TFAM-LSP22 muestra como las dos moléculas, proteína y ADN, se encuentran entrelazadas sugiriendo que existe una reorganización estructural previa que facilita el encaje entre las dos moléculas.

Es por ello que se decidió analizar TFAM y el complejo TFAM-LSP22 en solución mediante el uso de la dispersión de rayos X a bajo ángulo (SAXS: *Small angle X-ray scattering*), para hallar nueva información que permitiera entender mejor el mecanismo de unión.

Esta parte del proyecto se realizó en colaboración con Pau Bernadó, *Centre de Biochimie Structurale*, Montpellier (Francia), quién ofreció el soporte técnico, metodológico y de análisis fundamental para llevar a cabo estos experimentos.

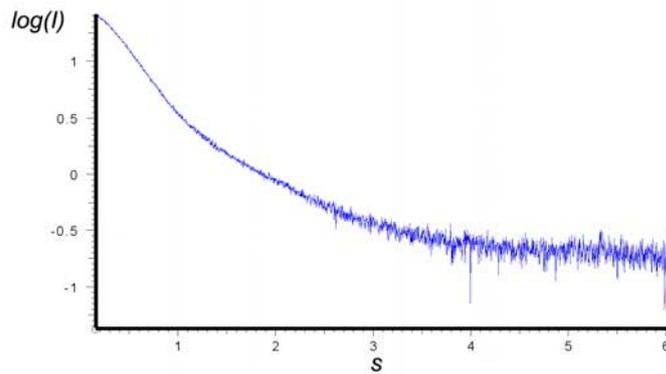
1. Descripción de la técnica

La dispersión de rayos X a bajo ángulo (SAXS: *Small angle X-ray scattering*) ofrece información complementaria sobre el plegamiento o desplegamiento, la agregación, la posibilidad de conformación extendida o existencia de linker flexible entre dominios; aunque en un rango de resolución bajo entre 50 Å y 10 Å, pero sin limitaciones de tamaño como ocurre en RMN y microcopia electrónica.

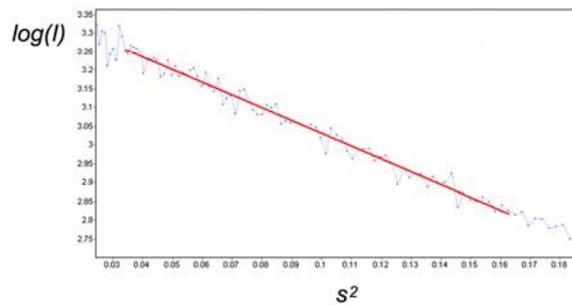
2. Caracterización de TFAM libre en solución mediante SAXS

El análisis de la curva de dispersión mediante la representación de Guinier a partir de la curva experimental (Fig_L1) permitió concluir que TFAM en solución era una partícula con un peso molecular aparente de 24 kDa, coherente con el peso molecular de la proteína TFAM expresada de 25.6 kDa. Por lo tanto, TFAM en solución mediante experimentos de SAXS se presentaba en forma de monómero. Además es posible calcular el radio de giro (R_g) de la partícula sujeta al estudio, TFAM presentaba un R_g de 32.0 ± 0.3 Å, superior al R_g estimado para una proteína de ~24 kDa, alrededor de 18 Å (según la ecuación de Flory $R_g = 3 \times N \times 10^{0.33}$). Cabe destacar que durante las primeras recogidas de datos también se realizaron mediadas de los constructos con la cola C-terminal delecionada, llamados CTT's, ya que existía la posibilidad de que TFAM formase dímeros a través de la cola C-terminal (hipótesis planteada por Kaufman et al). El análisis de estas curvas reflejó que no existía diferencia alguna en parámetros como $I(0)$, R_g , D_{max} (independientemente de la concentración testada) entre los constructos CTT'S y TFAM (Tabla_L1).

(a) Curva experimental



(b) Representación Guinier

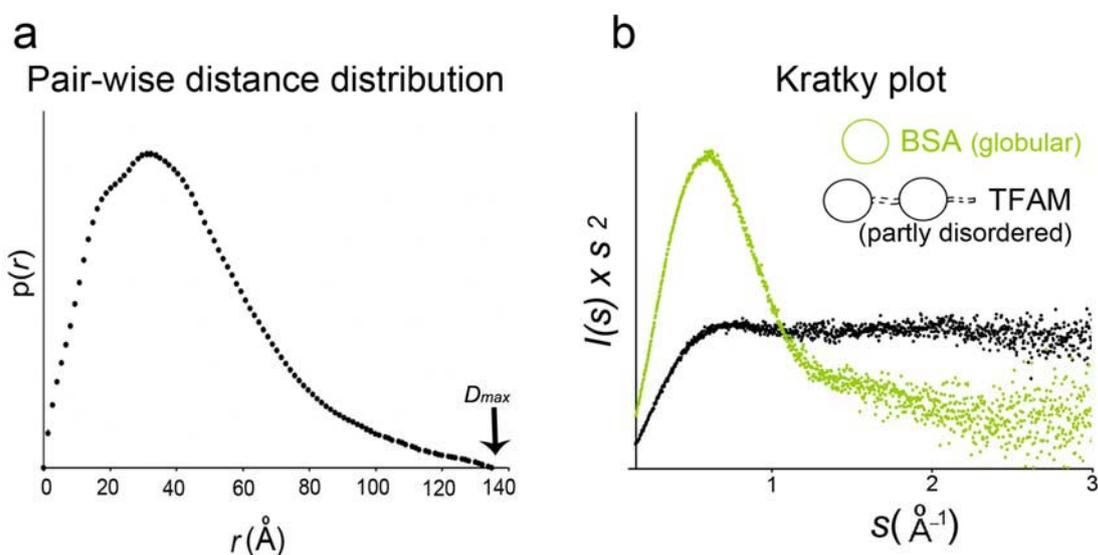


Fig_L1. (a) Curva de dispersión fusionada para TFAM, $\log(I)$ vs s . Para resoluciones inferiores se usó la curva correspondiente a una menor concentración (1.5mg/ml) de proteína, para evitar efectos entre partículas, mientras que para resoluciones superiores (1.5 a 3 nm^{-1}) se usaron concentraciones mayores de proteína (3mg/ml) y para la parte final (resoluciones de 3 a 5 nm^{-1}) de 5.3mg/ml. Es posible observar que la parte final de la curva, correspondiente a una mayor resolución, presenta un mayor ruido. (b) Representación Guinier, $\log(I)$ vs s^2 cogiendo los puntos experimentales de baja resolución. A partir de la pendiente de la ecuación lineal aplicada es posible extraer el valor de R_g y extrapolando en $y=0$, el valor de $I(0)$.

	mg/ml	R_g	$I(0)$	MW (kDa)
TFAM	1.3	3.10	23.261	24
	5.5	3.36	28.082	29
	7.4	3.39	27.112	28
CTT2	1.1	3.24	23.090	24.6
	3.9	3.9	21.484	22.15
	4.6	3.26	29.674	30.6
CTT3	3	3.24	23.090	24.6
	5.4	3.13	23.7	24.6
	7.1	3.20	25.573	26.365

Tabla_L1. Resumen de los valores obtenidos entre diferentes concentraciones y constructos para TFAM. Los valores no difieren entre los diferentes constructos ensayados, quedando descartada la posible dimerización de TFAM a través de la cola C-terminal, tal y como había propuesto Kaufman et al. El valor $I(0)$ y en consecuencia MW, presenta una pequeña variabilidad con la concentración debido a posibles efectos de interacción entre partículas a concentraciones mayores.

La representación de la *función de distribución* ($p(r)$) (Fig_L2(a)), que refleja la distribución de las distancias interatómicas dentro de la partícula, presentaba un largo hombro final que se extiende dando como resultado un valor de longitud máxima en la partícula (D_{max}) muy grande, 135 ± 5 Å. Esto reflejaba que TFAM en su forma no unida en solución era una partícula elongada y con un gran grado de flexibilidad, presentando diferentes conformaciones en solución. Además la representación de *Kratky* (Fig_L2(b)), da lugar a un perfil correspondiente de proteína globular, es decir, con dominios bien plegados; aunque también se mostraba una zona del perfil final totalmente plana correspondiente a zonas de la proteína desplegadas. Por lo tanto, TFAM se encontraba parcialmente desestructurada.

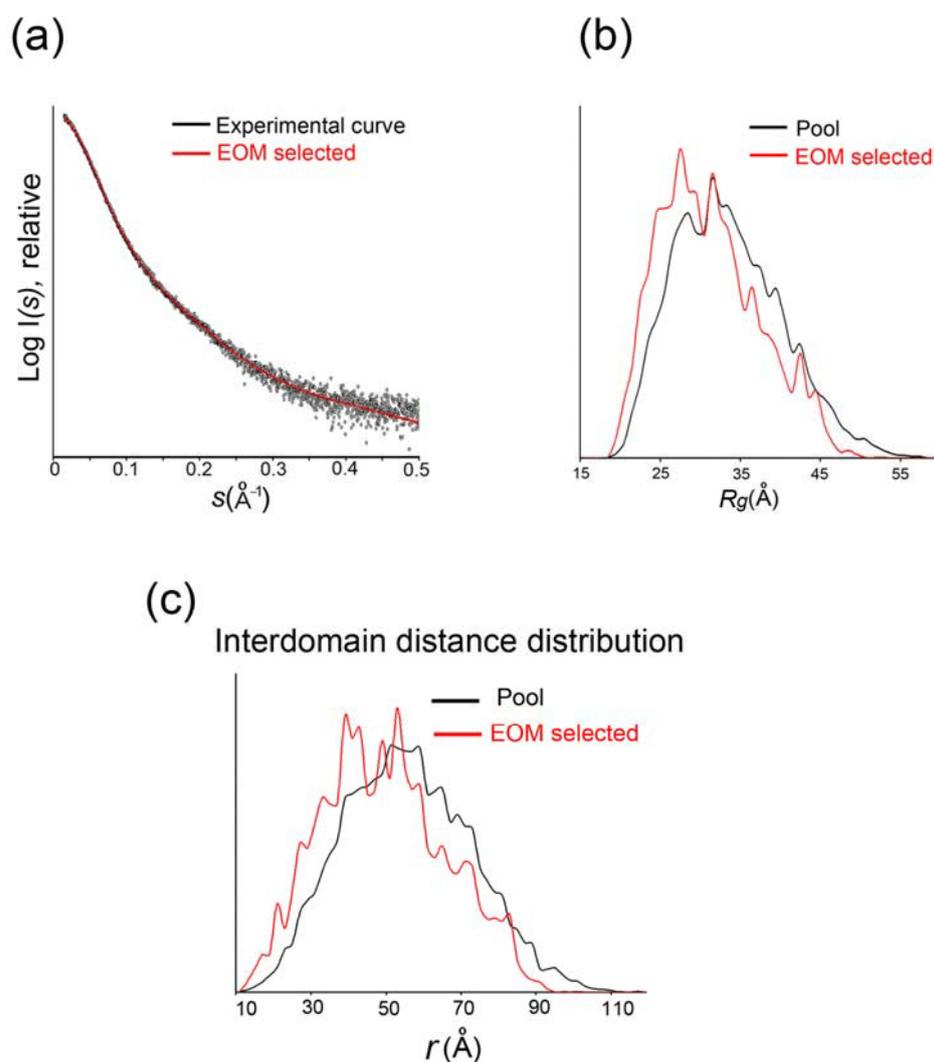


Fig_L2. Análisis de los datos de SAXS para TFAM en su forma no unida. (A) Representación de la función de distribución ($p(r)$), que muestra un perfil con una cola alargada que da lugar a un parámetro D_{max} mayor. (B) Representación de *Kratky* para una proteína típicamente globular, como BSA (en verde), con su perfil gaussiano; y TFAM (en negro) que parece mostrar la presencia de regiones parcialmente desestructuradas.

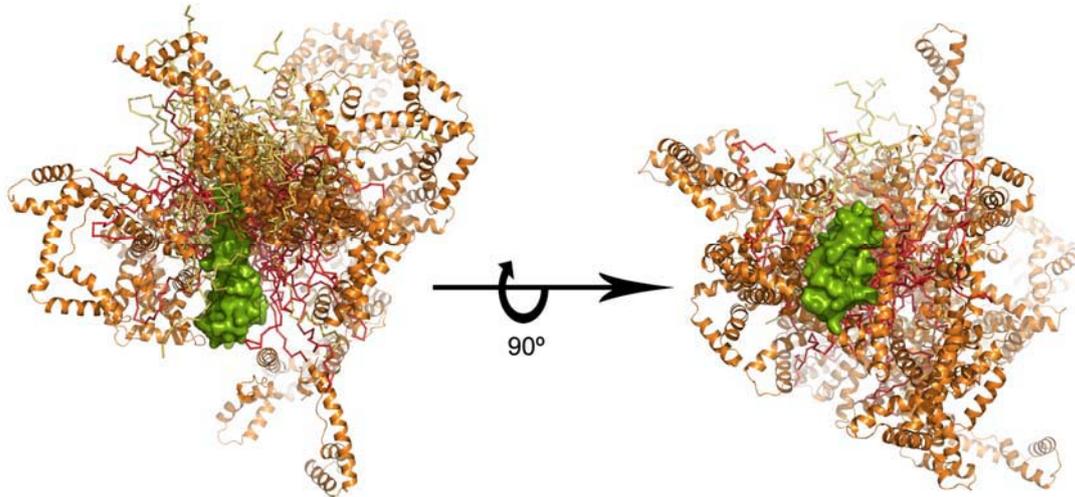
2.1. Determinación de la flexibilidad en TFAM

Para estudiar la flexibilidad y la coexistencia de diferentes conformaciones existentes para TFAM se aplicó el método conocido como EOM (*Ensemble Optimization Method*). Mediante el programa RANCH se generó un conjunto de 10.000 estructuras basadas en las coordenadas cristalográficas de la proteína, donde se permitió total libertad de conformación para la región del linker y la cola C-terminal, es decir, se definieron como cuerpos rígidos los dos dominios HMGbox. Seguidamente mediante el uso del EOM (ver materiales y métodos) y sus algoritmos genéticos fue posible seleccionar una subpoblación de 50 conformaciones, que de forma conjunta, permitían un buen ajuste con la curva experimental de SAXS para TFAM mediante el programa CRY SOL ($\chi^2 = 0.59$) (ver Fig_L3.a). Ello indicaba que TFAM en solución presentaba un elevado grado de plasticidad para la región del linker y su cola C-terminal. La distribución de R_g de la subpoblación era

amplia y similar a la distribución obtenida para el conjunto inicial de 10.000 conformaciones, y lo mismo sucedía para el perfil de la distancia entre dominios (ver Fig_L3, b y c). Ello indicaba que dentro de la subpoblación, las 50 conformaciones eran muy diferentes entre ellas y, por lo tanto, parece imposible definir TFAM como un sistema de conformación única en solución. De hecho al representar gráficamente las 50 conformaciones seleccionadas por EOM, se constataba la gran variedad de conformaciones en el espacio y las diferentes posiciones relativas que adoptan HMG1 y HMG2 debido a la flexibilidad del linker (Fig_L4)



Fig_L3. Determinación de la flexibilidad en TFAM. (a) Ajuste de la curva teórica (en rojo) mezcla de las 50 conformaciones seleccionadas por EOM con la curva experimental (en negro) mediante el uso del programa CRY SOL, donde es posible observar que el ajuste es perfecto a lo largo de toda la curva, excepto en la región final debido al ruido presente en la curva experimental. (b) Distribución del radio de giro (R_g) para la población inicial de 10.000 conformaciones (en negro) y la subpoblación de 50 conformaciones seleccionada a través de EOM (en rojo). (c) Distribución de la distancia entre dominios para la población inicial de 10.000 conformaciones (en negro) y la subpoblación de 50 conformaciones seleccionada a través de EOM (en rojo).



Fig_L4. Representación de las 50 conformaciones seleccionadas por EOM superpuestas, alineadas en HMG2 (en verde) donde es posible observar como la posición relativa de HMG1 (en naranja). La representación muestra una amplia gama de orientaciones para HMG1, linker (en rojo) y cola C-terminal (en amarillo) dejando constancia de la gran flexibilidad de TFAM en solución.

Una vez definidas las dos regiones que confieren la flexibilidad intrínseca a TFAM, linker y cola C-terminal, se decidió acotar la flexibilidad en la región del linker. La posibilidad de que el linker de TFAM pudiese formar una hélice α estable fue explorada mediante la generación de una población de 2.000 conformaciones mediante pre-BUNCH, donde se definieron como cuerpos rígido HMG1, HMG2 y el linker helicoidal, donde las regiones flanqueantes del linker se definieron como bisagras flexibles. Seguidamente mediante EOM se seleccionó la subpoblación de 50 conformaciones que mejor se ajustaban a la curva experimental, pero en este caso el ajuste empeoró ($\chi^2 = 2.93$). También se realizó el ensayo inverso, ofreciendo flexibilidad a la región central del linker y definiendo dos cuerpos rígidos que involucraban ambos HMGbox y las regiones flanqueantes, en este caso el ajuste tampoco fue mejor

Por lo tanto, según los resultados de SAXS, TFAM en su forma no unida en solución presenta dos regiones no estructuradas, el linker y la cola C-termina. Si combinamos esta información con la estructura cristalográfica en complejo con el ADN, es posible comprender que existe una reorganización del linker después de la unión al ADN que implica la adquisición de una conformación helicoidal junto con la interacción con el ADN.

3. Caracterización de TFAM-LSP22 en solución mediante SAXS

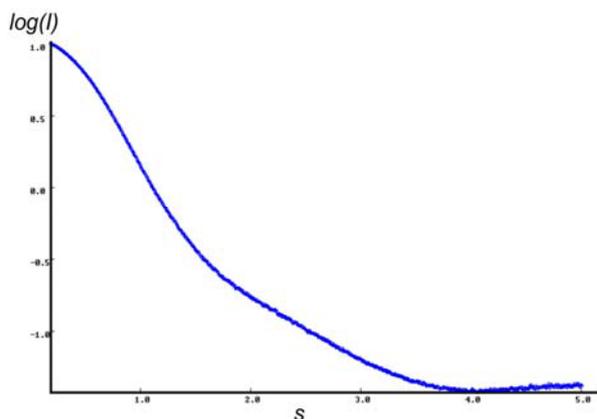
Una vez resuelta la estructura cristalográfica del complejo TFAM-LSP22, que sugería un sistema rígido donde la proteína rodeaba el ADN induciendo curvatura en éste de 180° , y con la observación del extremado grado de plasticidad en su forma no unida en solución, se decidió aplicar la metodología en solución SAXS para el estudio del complejo TFAM-LSP22.

3.1. Características del sistema TFAM-LSP22

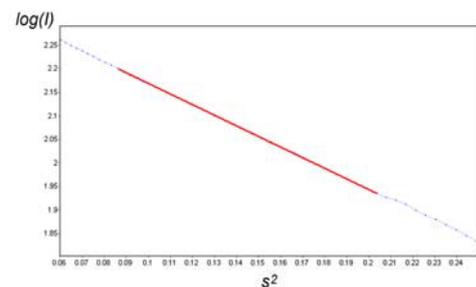
El cálculo del R_g de la muestra (Fig_L5) era aproximadamente de 26 \AA , en concordancia con el R_g teórico según la ecuación de Flory, 28 \AA . El R_g del complejo TFAM-LSP22 es inferior al de la proteína en su forma no unida, indicando que en el caso de TFAM-LSP22

en solución es una partícula mucho más compacta que la proteína libre, al igual de lo que se observaba en la estructura cristalográfica. Además, cuando se analiza la función de distribución, $p(r)$, ésta presentaba una forma más similar a una campana de gauss típica, sin presencia del hombro final existente para TFAM libre (Fig_L5 (c)). Ello implica que el valor final D_{\max} era aproximadamente de 82 Å, a diferencia del valor para TFAM sola de 135 Å. Reflejando así en solución, que el complejo TFAM-DNA22 es una molécula mucho más compacta que TFAM en su forma no unida. Finalmente, los ensayos de SAXS permitieron dilucidar que TFAM en solución se une de forma monómerica al ADN, con rangos de concentración proteína en el complejo de 1 mg/ml a 11 mg/ml.

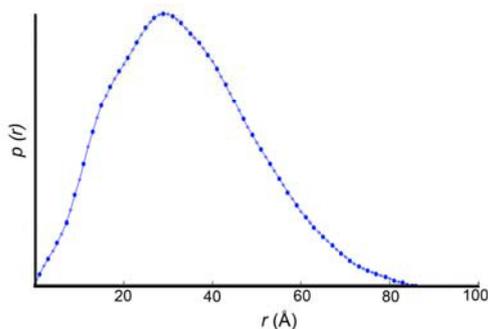
(a) Curva experimental TFAM-LSP22



(b) Representación Guinier TFAM-LSP22

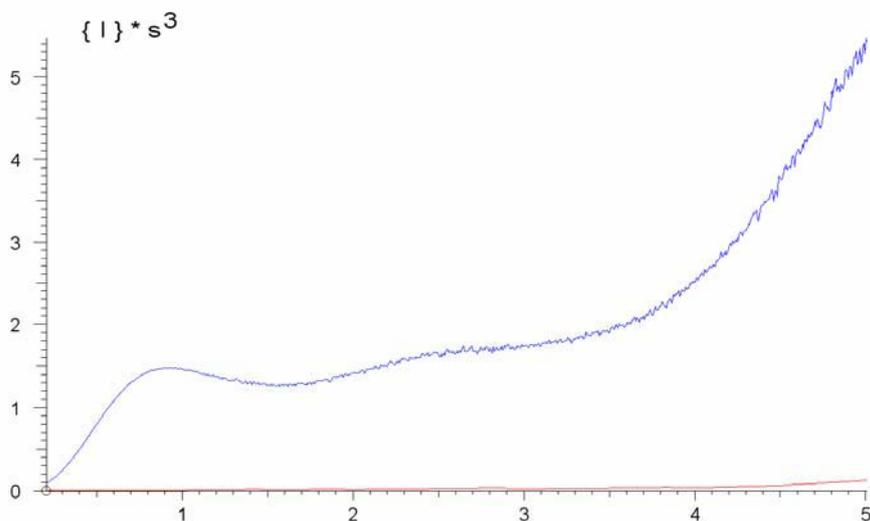


(c) Función de distribución, $p(r)$



Fig_L5. (a) Curva de dispersión experimental para TFAM-LSP22 a 5 mg/ml, $\log(I)$ vs s . (b) Representación Guinier, $\log(I)$ vs s^2 cogiendo los puntos experimentales de baja resolución. A partir de la pendiente de la ecuación lineal aplicada es posible extraer el valor de R_g y extrapolar en $y=0$, el valor de $I(0)$. (c) Función de distribución para TFAM-LSP22, con un valor D_{\max} inferior que para TFAM.

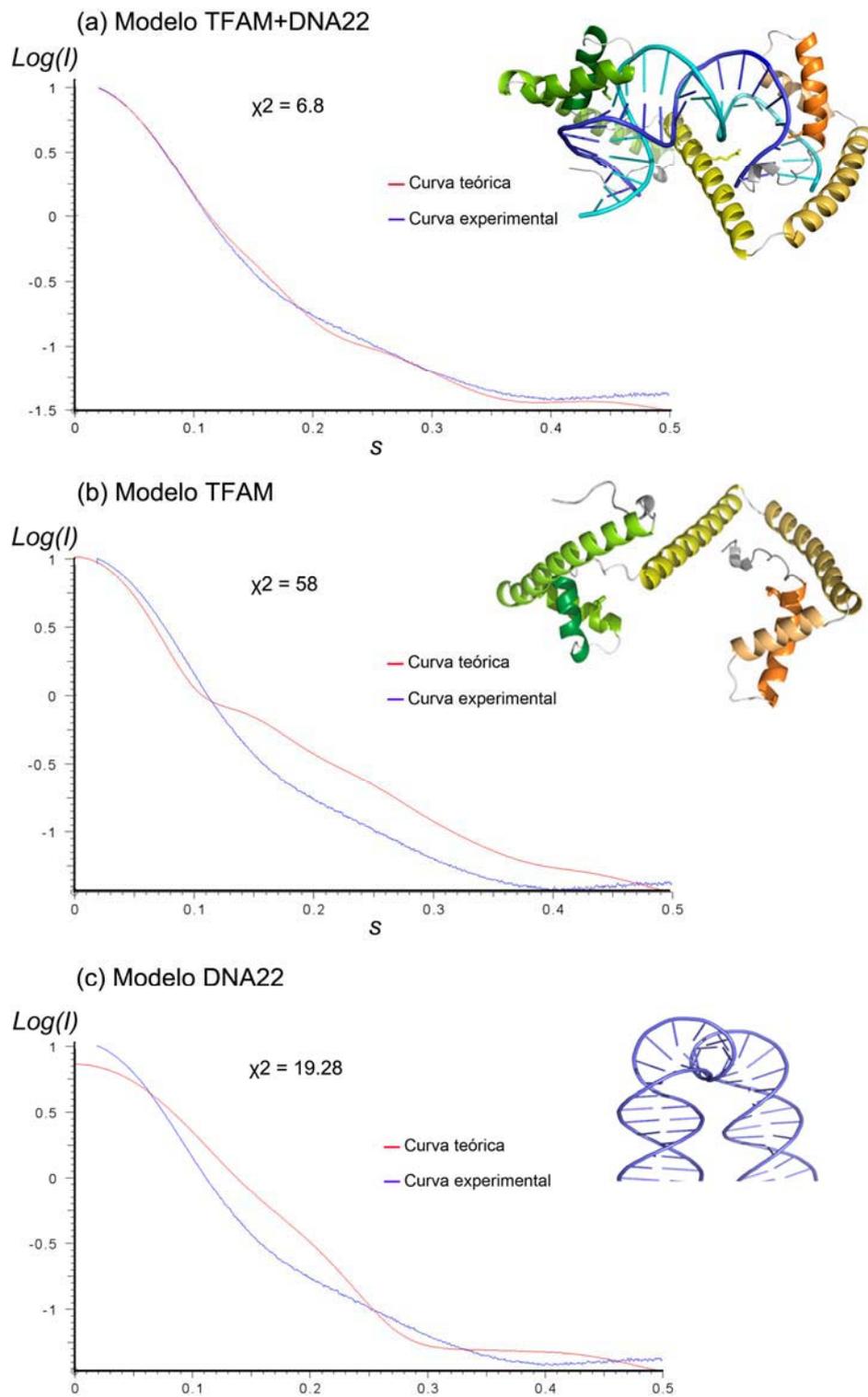
Por otra parte la representación de *Kratky* (Fig_L6) dió lugar a un perfil correspondiente al de una proteína globular con ciertas regiones flexibles. Si comparamos el perfil *Kratky* de TFAM y TFAM-LSP22 observamos que en ambos casos se trata de sistemas flexibles, pese a que en el caso de TFAM-LSP22 el sistema es mucho más compacto.



Fig_L6. Representación *Kratky* para el complejo TFAM-LSP22.

3.2. Comparación de SAXS con modelo cristalográfico

Dado que existía la estructura de TFAM-LSP22 resuelta, mediante el programa CRY SOL y el pdb disponible, se ajustó la curva experimental de SAXS con la curva teórica originada a partir del pdb TFAM-LSP22, pero el valor de discrepancia obtenido era muy alto, $\chi^2=6.8$ (Fig_L7). También se ajustó a la curva experimental a la curva teórica correspondiente a las coordenadas de la proteína sola y el ADN solo, presentado valores de discrepancia mayores. El hecho que la estructura cristalográfica no pudiese explicar los datos de SAXS en solución sugería la presencia de conformaciones ligeramente diferentes o movimientos moleculares sutiles.

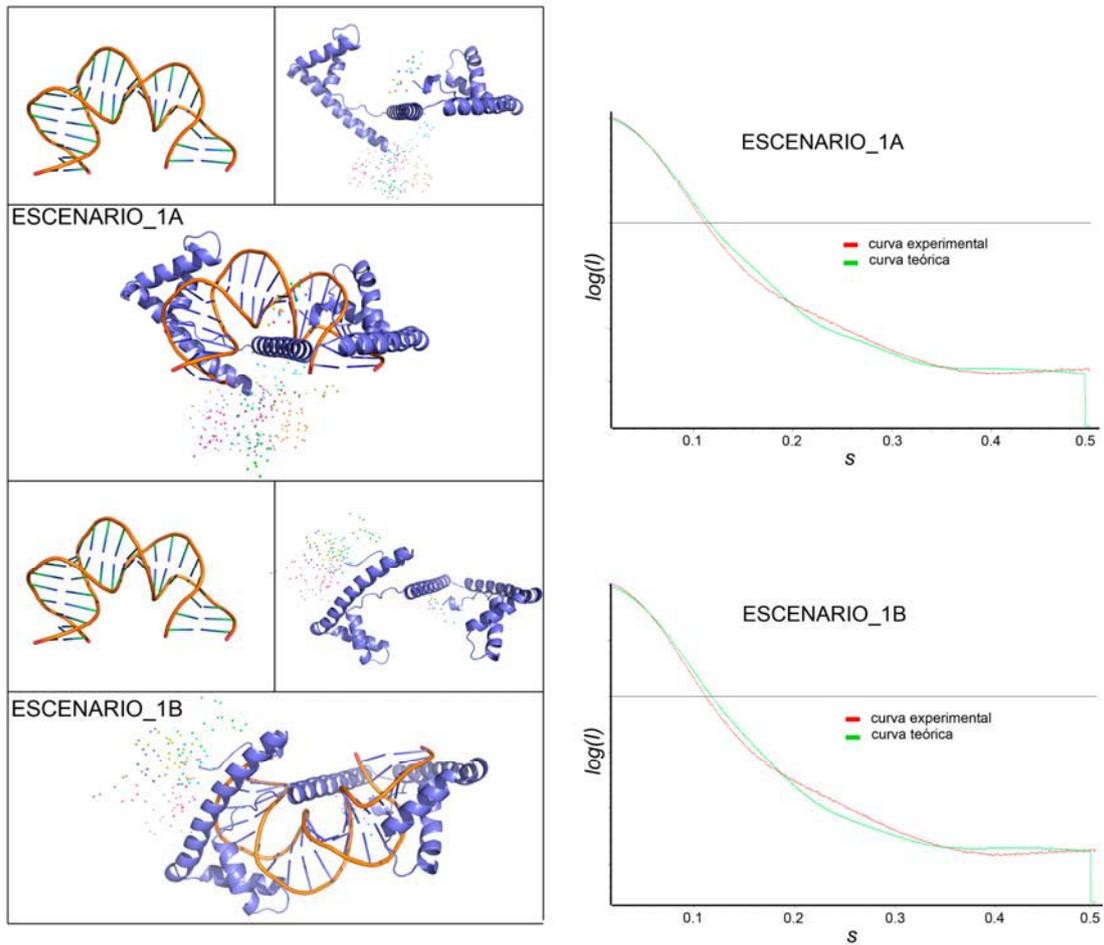


Fig_L7. Comparación de la curva experimental de SAXS con las estructuras cristalográficas disponibles y sus valores de discrepancia (χ^2). (a) Comparación con la estructura TFAM-LSP22. (b) Comparación con las coordenadas cristalográficas para la proteína. (c) Comparación con las coordenadas cristalográfica para el ADN.

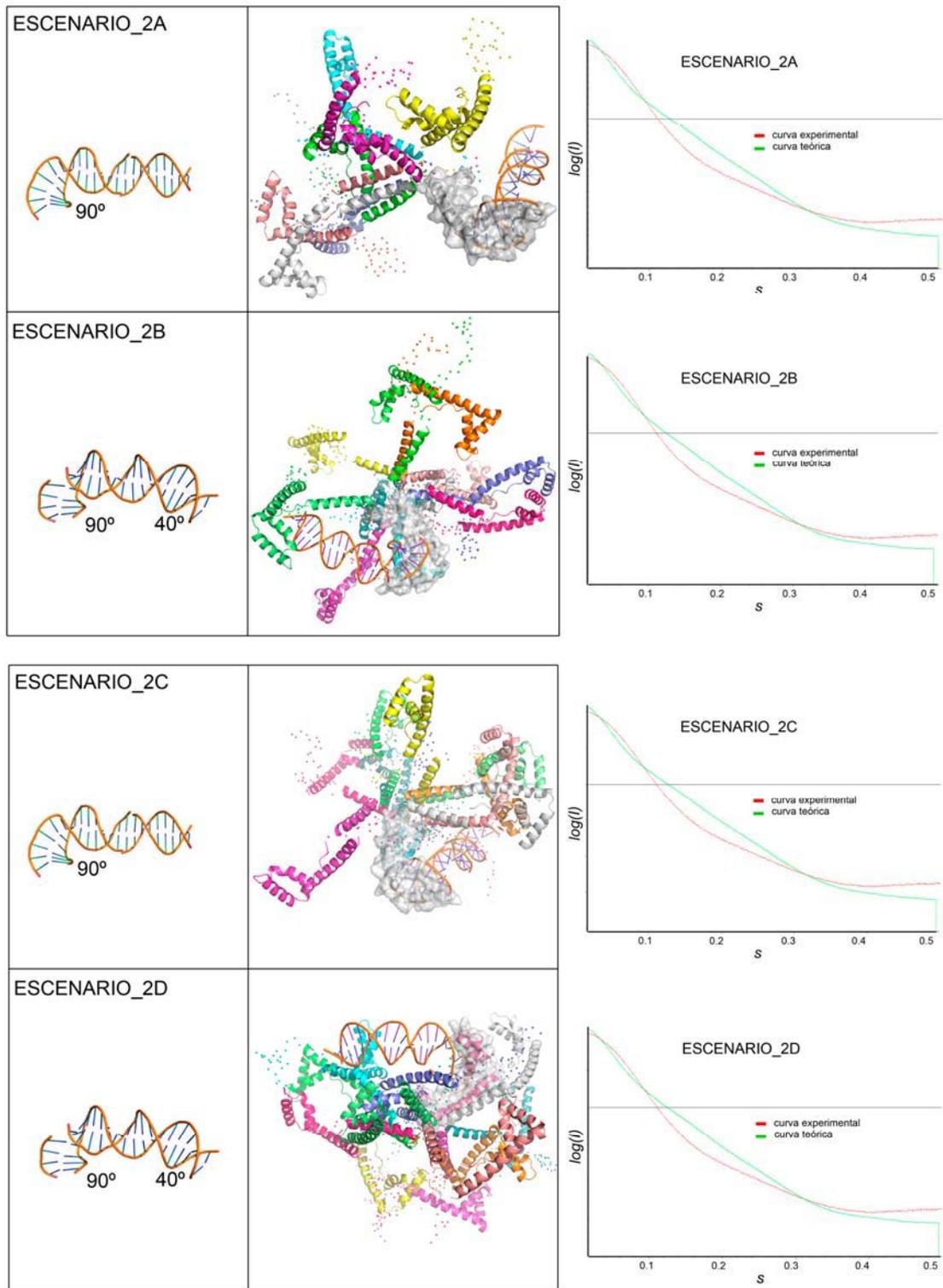
3.3. Estudio de la dinámica del complejo TFAM-LSP22 en solución

Para confirmar un modelo dinámico de unión y curvatura del ADN se calculó el % de diferentes poblaciones con conformaciones diferentes para el complejo TFAM-LSP22 en solución. Para ello primero fue necesario generar diferentes “escenarios”, basados en el modelo de definido para TFAM y su participación en el inicio de transcripción debido a los resultados obtenidos. Los diferentes escenarios implicaban flexibilidad tanto en la conformación de la proteína como en la curvatura del ADN (ver definición de los diferentes escenarios en materiales y métodos: SAXS). Una vez obtenidos los modelos para los diferentes escenarios se comparaba la curva teórica promedio para los diferentes modelos con la curva experimental de SAXS.

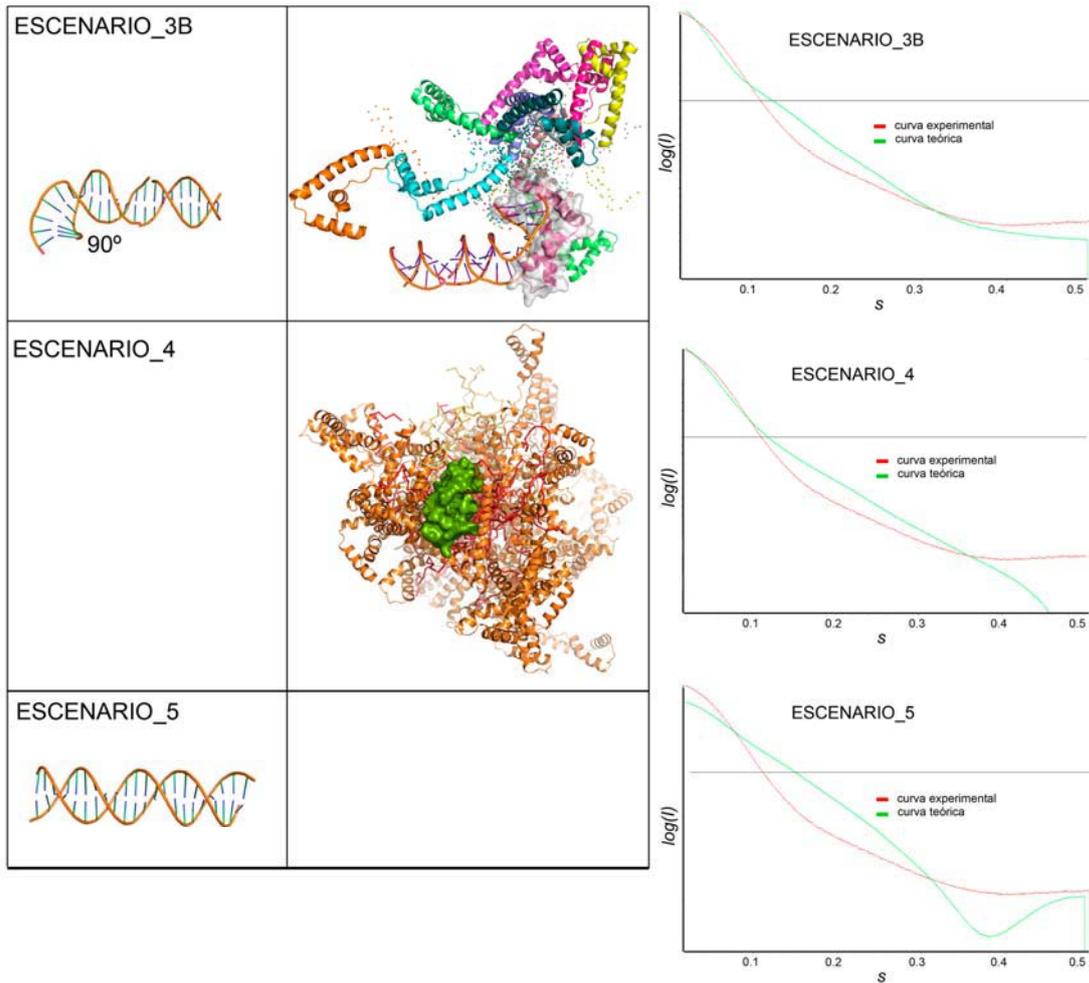
Fue en el caso del escenario 1 (Fig_L8), concretamente 1A, donde se obtuvo un mejor ajuste con la curva experimental. En este escenario el ADN era idéntico al cristalográfico, con la curvatura de 180° , y se definió como cuerpo rígido toda la proteína excepto para la cola C-terminal con total libertad (resíduos 225-246). Otro escenario que también se ajustaba aunque con una mayor discrepancia a la curva experimental era el escenario 6 (Fig_L11). En este escenario la proteína fue definida como un cuerpo rígido para HMG1 y la región del linker, ofreciendo flexibilidad para la zona de HMG2. El ADN utilizado presentaba una curvatura de 90° en la región de HMG1, para el escenario 6A; en el caso del escenario 6B, el ADN presentaba la misma curvatura de 90° en la zona de HMG1 más una curvatura más débil de $\sim 40^\circ$ en la región correspondiente a la unión de HMG2. Para el caso del escenario 2 (Fig_L9) el planteamiento era fijar la unión de HMG1 al ADN y definir como dominios HMG1, HMG2 y la región del linker. El escenario_2A/B presentan posibilidad de contacto con el ADN hasta el residuo final de HMG1, mientras que escenario_2C/D presentan posibilidad de contacto hasta el final del linker. Pero en este caso la discrepancia era mayor, hecho que era presumible ya que el análisis de la curva experimental nos indicaba que se trata de un sistema ligeramente compacto y el volumen del escenario 2 es mucho mayor al experimental. Finalmente también se realizó el ajuste con el escenario 3, 4 y 5; aunque el ajuste no era óptimo (Fig_L10).



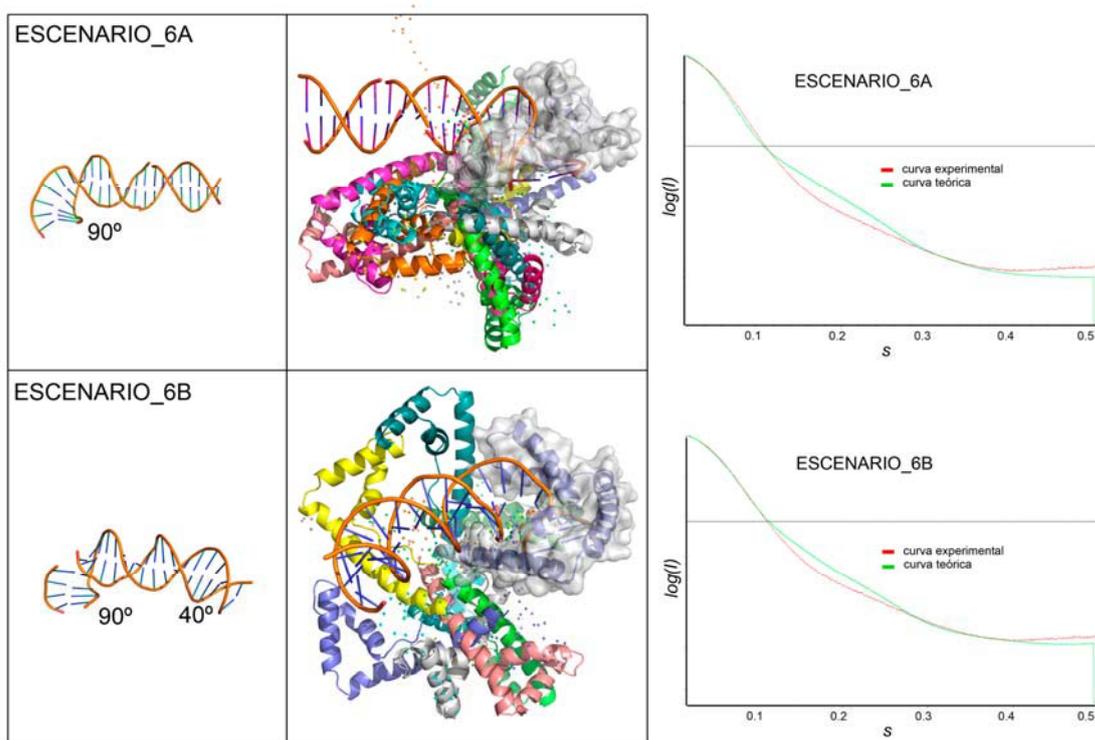
Fig_L8. **Escenario1 y su ajuste a la curva experimental.** En el panel izquierdo se presentan el ADN utilizado, idéntico al ADN con 180° de curvatura de la estructura cristalográfica, mientras que la proteína presentaba flexibilidad para la totalidad de su cola C-terminal (**escenario_1A**) o parcial (**escenario_1B**) (sólo se indican 10 modelos, para simplificar la figura). En el panel derecho se muestra el ajuste para los subescenarios, en rojo la curva experimental y verde la curva teórica.



Fig_L9. **Escenario2 y su ajuste a la curva experimental.** En el panel izquierdo se representa el ADN utilizado, seguido de la forma final que presentan los complejos con diferente flexibilidad intrínseca en la proteína y contacto para el ADN. El **escenario_2A/B** presentan posibilidad de contacto con el ADN hasta el residuo 120, mientras que **escenario_2C/D** presentan posibilidad de contacto hasta el residuo 220. En el panel derecho se muestra el ajuste para los subescenarios, en rojo la curva experimental y verde la curva teórica.



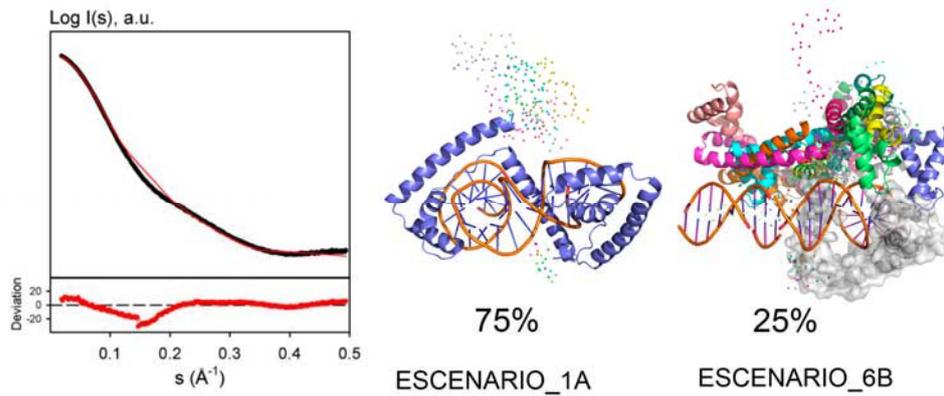
Fig_L10. **Escenario3, 4 y 5, con el ajuste a la curva experimental.** El **escenario3** presenta un linker no helicoidal con flexibilidad y posibilidad de unión al ADN hasta HMG1. El **escenario4** está formado por las 50 conformaciones que mejor definen TFAM en solución y, finalmente, el **escenario5** está formado por un ADN lineal de forma B. En el panel derecho se muestran los ajustes para los diferentes escenarios.



Fig_L11. Escenario 6 y su ajuste con la curva experimental. En este caso se define el conjunto HMG1 y linker como un cuerpo rígido unido al ADN, mientras que se permite libertad para HMG2. El escenario 6A y 6B, difieren en el tipo de ADN usado. En el panel derecho se muestran los ajustes para los diferentes escenarios.

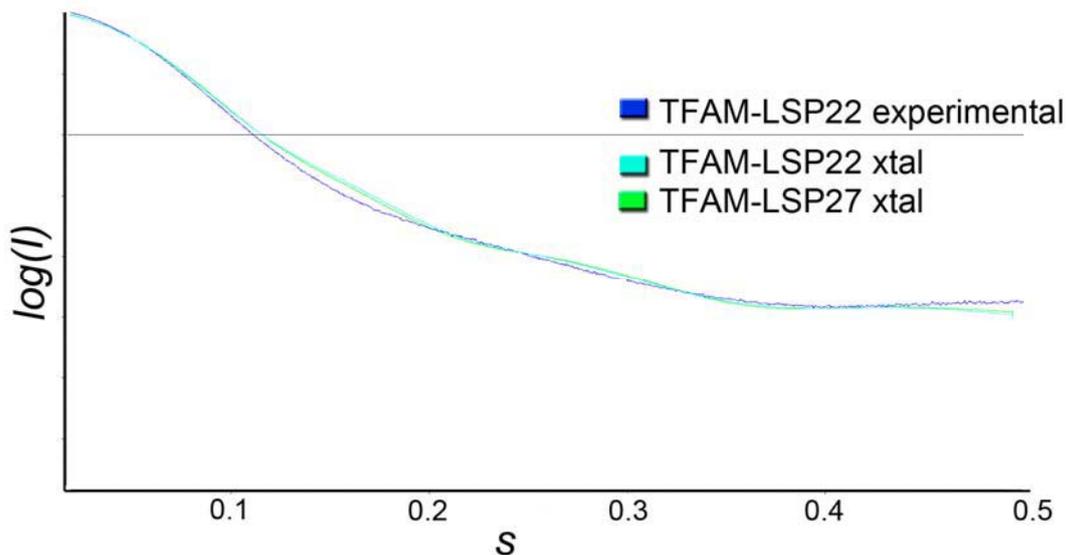
A partir de los diferentes escenarios creados y su ajuste se intentó definir poblaciones mixtas dentro de la curva experimental mediante el programa EOM, es decir, a partir de las curvas promedio para los diferentes escenarios se optimizaron las poblaciones para que el ajuste con los datos experimentales fuera óptimo (Fig_L12).

Pese a ello nuestros modelos simplistas para el mecanismo dinámico no fueron suficientes para mostrar una concordancia con los datos de SAXS, ya que existe un elevado porcentaje de especies compactas. Posiblemente los fenómenos dinámicos dentro del complejo TFAM-LSP22 son diferentes a los testados y mucho más sutiles (por lo tanto, más difíciles de predecir) y aunque la estructura global del complejo en solución es más bien compacta, SAXS es sensible a los pequeños cambios estructurales (por ejemplo, ángulos de curvatura que oscilan entre 180-160° o desplazamientos leves de HMG2 y linker).



Fig_L12. Ajuste de poblaciones para el escenario 1A y escenario 6B, con diferente porcentaje. En la gráfica es posible diferenciar la curva experimental, en negro, y la curva teórica producto de los diferentes escenarios en rojo. Existen dos puntos de máxima desviación, en la parte inicial y la mitad de la curva.

Por otra parte también se probó con la estructura de TFAM en complejo con un ADN de 28 pb (Ngo, Kaiser et al.), descrita en el apartado de análisis estructural. Mediante el análisis comparativo de las dos estructuras fue posible distinguir la presencia de una molécula más compacta en el caso de TFAM-LSP28, con un desplazamiento del linker y el dominio HMG2. Por ello, se usó esta estructura con el ADN sobrante recortado para ajustar a la curva experimental de SAXS. Como se puede observar en la fig_L13, el ajuste de la curva experimental de TFAM-LSP27 es prácticamente idéntico al de TFAM-LSP22. Por lo tanto, queda por dilucidar el comportamiento del sistema TFAM-LSP en solución mediante otro tipo de herramientas que nos permitan entender los datos experimentales de SAXS.



Fig_L13. Ajuste de las estructuras del complejo TFAM-LSP22 (en turquesa) y TFAM-LSP27 (en verde), con los datos experimentales de SAXS (en azul).

M. spFRET PARA EL ESTUDIO DE LA CURVATURA DEL ADN EN SOLUCIÓN

Como se ha comentado anteriormente el fenómeno físico llamado FRET (*Fluorescence Resonance Energy Transfer*) puede ser usado para determinar cuantitativamente el ángulo de torsión que produce una proteína al formar complejo con un ADN marcado en sus extremos con fluoróforos. Es por ello que una vez determinada la estructura de TFAM en complejo con DNA-LSP22, se optó por realizar una colaboración para estudiar esta torsión en solución mediante spFRET. Los resultados descritos en este apartado fueron fruto de una estancia de 3 meses dentro del programa de movilidad JAE-preDoc en el laboratorio del Prof. Jörg Lagowski en la unidad de Biofísica de Macromoléculas de *DKFZ (German Cancer Research Center)*, Heidelberg, Alemania. La supervisión del trabajo fue llevada a cabo por la Dra. Katalin Tóth.

La primera parte del proyecto consistió en el estudio de TFAM y la torsión que ésta crea en la secuencia LSP. Además se quiso constatar el modelo de dos puntos de curvatura por parte de ambos dominios HMGbox, y fue por ello que el estudio implicó el uso de dos secuencias de ADN de diferente longitud: 30pb y 50pb. Metodológicamente implicó el marcaje con fluoróforos de los diferentes ADN y la titración del sistema para saber cuánta TFAM era necesaria para llegar a la fase plateau de E_{FRET} . Para el marcaje de los fluoróforos fue necesario evaluar el tamaño del ADN a estudiar (con la región central de 22 pb usada para la cristalización) y si existía libre rotación de al menos una de las sondas. Para la titración fue necesario estudiar nuestro sistema y comprender si la formación de una doble banda de retardo por EMSA, es decir, la presencia de diferentes estados oligoméricos, afectaba el estudio mediante FRET. También fue necesario establecer las condiciones ideales para estabilizar los complejos TFAM-ADN bajo las condiciones experimentales de spFRET.

La segunda parte del proyecto consistió en caracterizar la curvatura ocasionada por el dominio 1 (HMG1) con o sin presencia de linker, para entender el papel del linker en la unión de TFAM a la secuencia LSP22 y su perturbación estructural. También se quiso realizar el estudio del segundo dominio HMGbox, HMG2, y su unión y posible curvatura del ADN. Este estudio implicó el uso de diferentes mutantes puntuales de los residuos que juegan un papel más importante en la unión de HMG2 con LSP22.

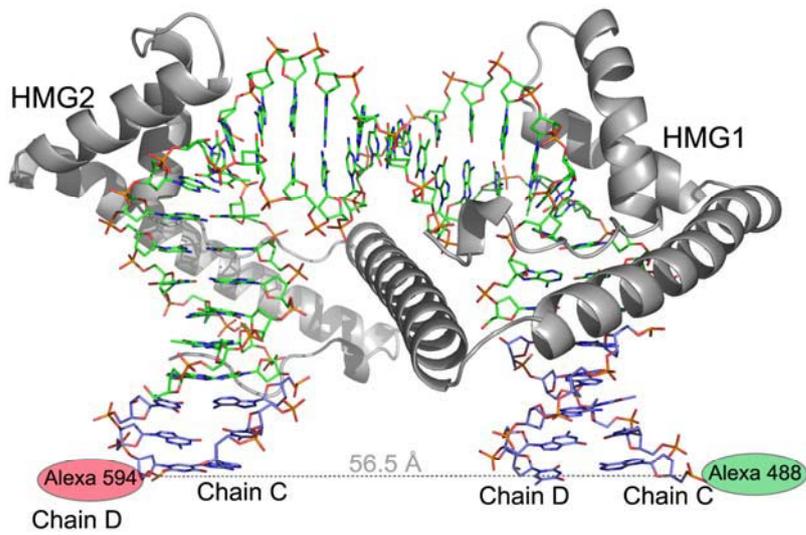
Finalmente, se intentó realizar un estudio cinético del mecanismo “*step-wise*” propuesto para la unión de TFAM a LSP mediante el uso de time-resolved FRET dentro del sistema spFRET. En la literatura científica encontramos diferentes ejemplos de uso de esta técnica para este tipo de estudios, uno de los ejemplos es para el complejo TBP-TATA donde se encontraron especies intermedias dentro de la curvatura del ADN hasta inducir la curvatura final del complejo (Parkhurst, Richards et al. 1999).

1. Modelos de ADN para aplicar en estudios spFRET

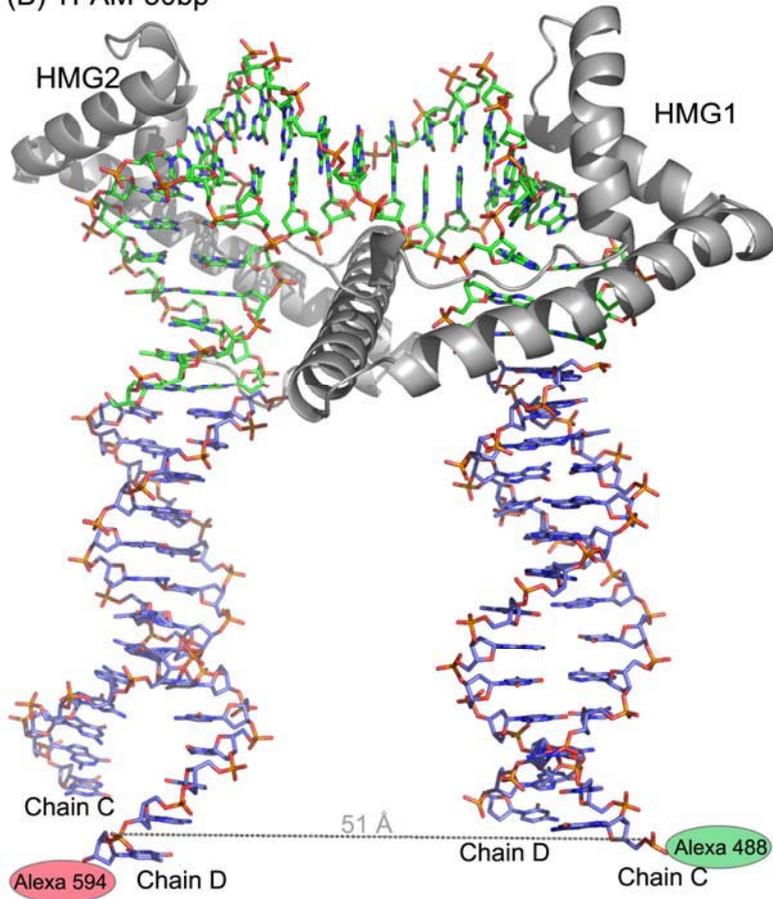
El primer paso para llevar a cabo los experimentos de FRET fue el diseño de los oligos, en base a los principios técnicos del fenómeno y el conocimiento de la estructura cristalográfica.

El marcaje del ADN se realizó en la cara externa del ADN, es decir, en la cadena C (que corresponde al oligo rico en citosinas) se marcó con el fluoróforo Alexa 488 en su extremo 5'; mientras que para la cadena D (oligo rico en guaninas) se marcó en extremo 5' con Alexa 594. En este sistema Alexa 488 es el donador y Alexa 594 es el aceptor. El marcaje de los ADNs implicaba la adición de un poli-linker para favorecer la libertad de rotación de los fluoróforos y, por lo tanto, intentar reducir la anisotropía (κ^2). Como se ha comentado anteriormente (materiales y métodos: FRET), se diseñaron dos construcciones de ADN con diferentes longitudes (30 pb o 50pb) para los experimentos de FRET, con la secuencia de ADN conocida como LSP22 en la zona central y las bases extra a los lados (Fig_M1). De esta forma tenemos un ADN que, en su forma lineal, presentaría una distancia entre fluoróforos de $\sim 97 \text{ \AA}$ para un ADN de 30 pb y $\sim 162 \text{ \AA}$ para 50pb y, por lo tanto, no se daría el fenómeno de transferencia energética conocido como FRET. Si existe la curvatura de 180° descrita en el cristal, los modelos construidos indican que la distancia entre los dos fluoróforos se acortaría resultando en una distancia de $\sim 56 \text{ \AA}$, dando lugar al fenómeno FRET. Por lo tanto, la medida de E_{FRET} para los complejos TFAM- ADN permite obtener un cálculo aproximado de la distancia entre los fluoróforos y, en consecuencia, del ángulo de torsión que ocurre en el ADN. Si el mecanismo de unión de TFAM en solución es similar a la estructura cristalográfica, la distancia entre ambos fluoróforos (E_{FRET}) será similar entre los complejos con diferentes ADN. Por lo tanto, el ADN de 30pb permitirá llevar a cabo los experimentos para dilucidar la E_{FRET} , permitiendo así obtener la distancia entre los fluoróforos y el cálculo final de curvatura. Mientras que el ADN de 50pb nos permitirá dilucidar si se trata de un sistema con dos puntos de curvatura como describe la estructura cristalográfica (forma de U) manteniéndose una E_{FRET} similar a ADN 30pb. O un sistema con un solo punto de curvatura (forma de V), difiriendo la E_{FRET} entre los dos ADNs.

(A) TFAM-30bp



(B) TFAM-50bp



Fig_M1. Modelo de FRET en el complejo TFAM-ADN en base a la estructura cristalográfica. La secuencia de ADN tiene dos códigos de colores: verde para las bases dentro de LSP22 y azul para las bases extra añadidas para complementar la longitud final. En verde se indica la posición para el fluoróforo Alexa 488 y rojo para el fluoróforo Alexa 594. La distancia entre las sondas es de $\sim 56 \text{ \AA}$ ($\pm 10 \text{ \AA}$), si el sistema en solución coincide con el complejo cristalográfico. (A) complejo entre TFAM y el ADN de 30pb. (B) Complejo entre TFAM y el ADN de 50pb.

2. Primeros experimentos TFAM-30bp

Antes de la adquisición de datos definitiva fue necesario poner a punto el sistema para el comportamiento de nuestra muestra. Los ensayos spFRET se realizaron a concentraciones del orden pM (20-50pM) con lo cual fue necesario estabilizar la formación del complejo y estabilizar la fluorescencia de las sondas en la solución durante el tiempo del experimento (ver material y métodos: FRET).

2.1. Comportamiento del ADN

El experimento de FRET requiere el uso de controles negativos, en este caso existían dos tipos de controles denominados **Donor only** o **C'D dúplex**. El control negativo **Donor only** hace referencia a la hebra de ADN marcada con el donador, mientras que **C'D dúplex** es la doble hebra de ADN marcada sólo con el donador (donde ' indica el marcaje con fluoróforo en el oligo, **C** el oligo rico en citosinas y **D** el oligo rico en guaninas).

Al analizar el histograma de la medida del control negativo **C'D dúplex** (doble hebra marcada con donador) con **C'D' dúplex** de 30pb (doble hebra marcada con los dos fluoróforos, con una distancia de ~ 100 Å entre ellos) (Fig_M2A) es posible observar que C'D dúplex presenta un solo pico, mientras que C'D' dúplex está formado por dos picos. Esto sugiere que el ADN presenta un comportamiento intrínseco característico, donde sería posible la existencia de una leve pre curvatura intrínseca.

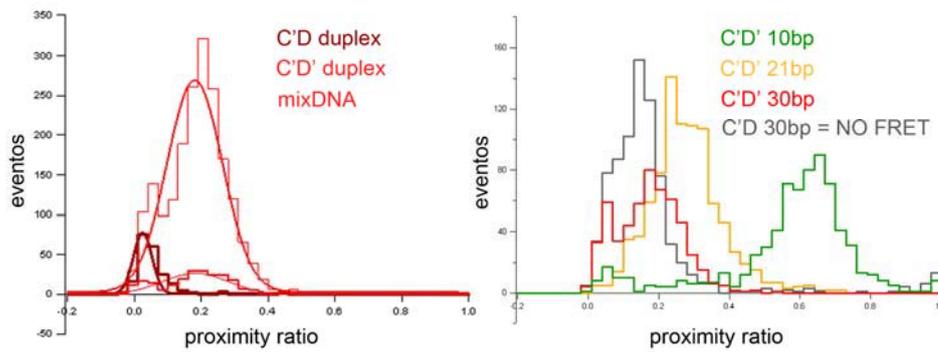
También se realizaron controles positivos, se disponía de ADN de 30pb con los fluoróforos colocados a 21 pb o ~ 68 Å (llamado C'D' 21pb) y 10pb o ~ 32 Å (llamado C'D'10pb) (Fig_M2B). Estos ADN presentaban diferente E_{FRET} debido a la diferencia entre la distancia entre sondas y, por lo tanto, un valor diferente para *proximity ratio*. Estas medidas permiten calibrar el sistema para poder extraer una distancia entre sondas a partir del valor *proximity ratio* experimental. En la figura también se muestra el valor de *proximity ratio* para C'D'30pb, una vez adquiridos los datos y visualizados, será necesario reprocesarlos fijando un valor de *proximity ratio* zero para C'D' duplex. Este proceso es conocido como determinación *crossstalk correction*.

2.2. TFAM en complejo con C'D' 30bp

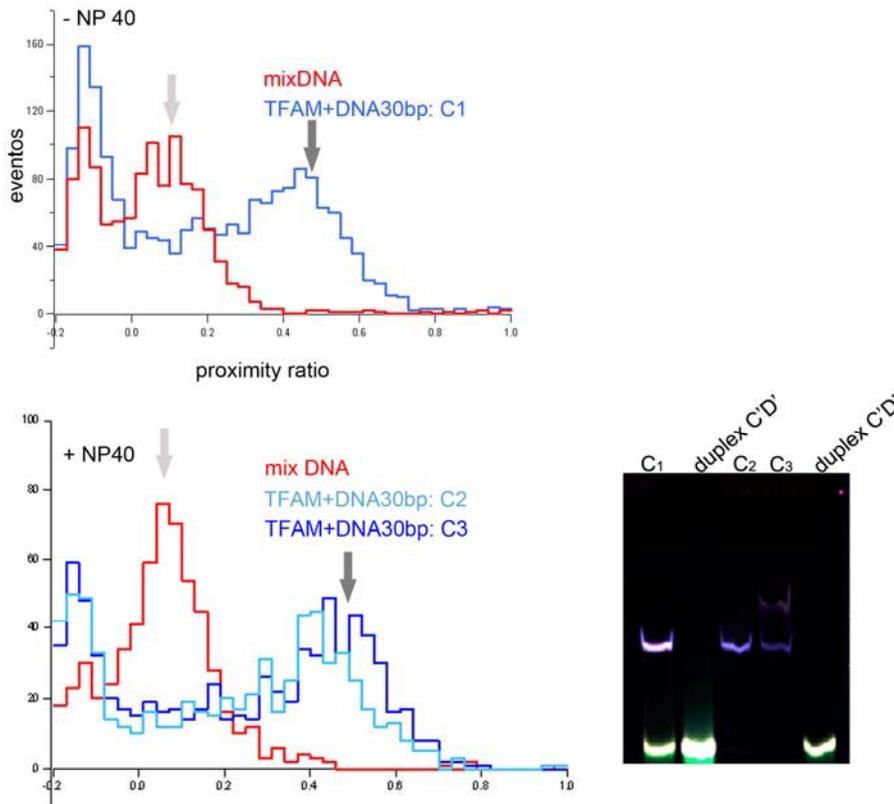
Para las medidas de spFRET fue necesario llevar a cabo la formación del complejo mediante diálisis O/N. También fue necesario establecer un ratio proteína:ADN óptimo, que era previamente analizado mediante geles EMSA. En la figura M_2B, se muestran el número de eventos de FRET de los complejo C1, C2 y C3 con ratios proteína:ADN crecientes, es decir, la cantidad de proteína frente ADN es mayor de C1 a C3. En rojo observamos el valor de *proximity ratio* para el ADN, con dos picos muy próximos. Cuando se añade TFAM observamos un primer pico en concordancia con el ADN, debido a la presencia de ADN libre, y un segundo pico debido al fenómeno FRET en presencia de TFAM donde es posible fijar el valor medio de la campana Gauss sobre $x = 0.45$ para *proximity ratio*. Para los tres complejos el valor medio era similar, aunque en el caso de menor cantidad de TFAM existía una mayor cantidad de ADN libre y, por lo tanto, el histograma mostraba un mayor número de eventos para la señal del ADN libre. Además también se ensayó el efecto del detergente NP40 para estabilizar los complejos proteína-ADN, viendo que no provocaba efectos negativos sobre el fenómeno FRET.

También se estudió el efecto de concentraciones crecientes de NaCl en el fenómeno FRET (Fig_M3C). A partir de concentraciones de NaCl alrededor de 250mM el valor de *proximity ratio* para el complejo prácticamente desaparece y solo es posible distinguir el pico referente al ADN (levemente desplazado debido a que la sal añade ruido a la medida, Fig_M3A). Esta desaparición del pico referente al fenómeno FRET sugiere que bajo concentraciones elevadas de NaCl existe una disociación del complejo. Por otra parte cabe destacar que durante los primeros experimentos de spFRET se determinó que el agente reductor usado típicamente para TFAM (5 mM DTT) podría provocar daños sobre los fluoróforos: el análisis del FRET bajo varias concentraciones de DTT, indicó que el DTT enmascaraba la señal del fluoróforo donador en el aceptor (Fig_M3B). Por ello se decidió usar concentraciones de DTT de 1 mM.

(A) Comportamiento diferentes ADN

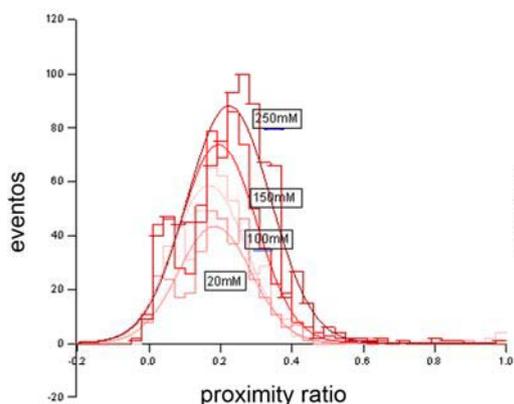


(B) Comportamiento del complejo TFAM-30bp ± NP40

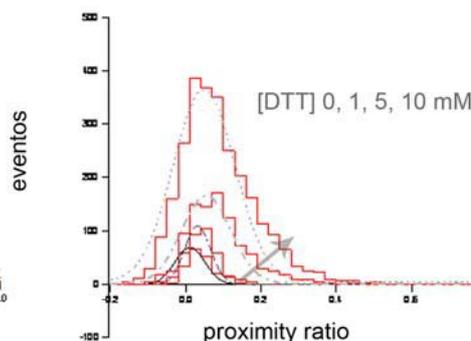


Fig_M2. Histograma spFRET donde se representa *proximity ratio* vs número de eventos. (A) Comportamiento de los diferentes ADN. En el panel izquierdo encontramos la fluorescencia detectada para diferentes tipos de dúplex de ADN: **C'D dúplex**, solo el oligo C está marcado con el donador (muestra conocida como Donor only); **C'D' dúplex**, ambos oligo están marcados (C con donador, D con aceptor); **mixDNA**, combinación de dúplex marcado con los dos fluoróforos y dúplex no marcado. En el panel derecho, histograma de fluorescencia para ADN de diferente longitud marcado (**C'D' 10pb**, **C'D' 21pb**, **C'D' 30pb**) y el dúplex control negativo llamado Donor only (**C'D**). (B) Comportamiento del complejo TFAM-30pb con presencia o ausencia de detergente NP40 en el tampón de adquisición de datos. Los diferentes complejos presentan un ratio TFAM-ADN diferente, **C1**, 0.1 nmol TFAM:0.1 nmol ADN; **C2**, 0.2 nmol TFAM: 0.1 nmol ADN; **C3**, 0.4 nmol TFAM: 0.1 nmol ADN. Las flechas indican los picos del histograma, en gris claro correspondiente al valor *proximity ratio* para mixDNA, y en gris oscuro, valor de *proximity ratio* para el complejo TFAM+DNA30pb.

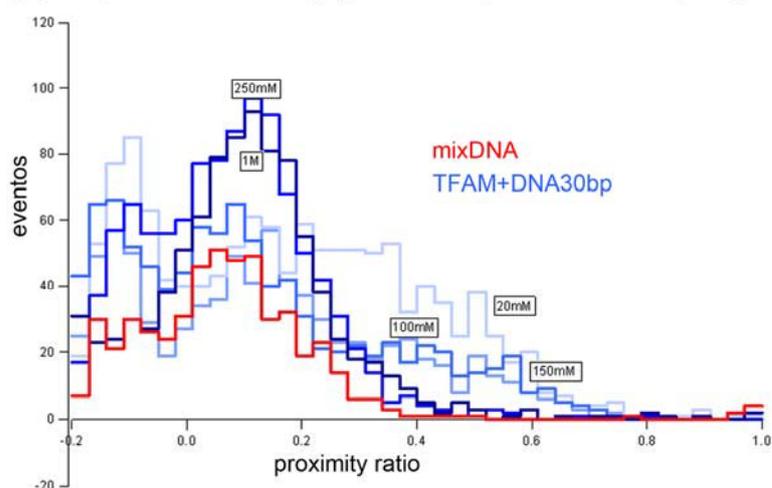
(A) Efecto NaCl sobre ADN marcado



(B) Efecto DTT sobre ADN marcado



(C) Comportamiento del complejo TFAM-30bp frente diferentes [NaCl]



Fig_M3. (A) Histograma spFRET para el ADN con concentraciones crecientes de NaCl (20, 100, 150 y 250 mM NaCl). (B) Efecto de DTT sobre el ADN marcado con los fluoróforos y su fluorescencia. La flecha gris indica el aumento en la concentración de DTT de 0 a 10 mM. (C) Comportamiento del complejo TFAM-30pb (en diferentes tonalidades de azul) frente diferentes concentraciones de NaCl y su disminución en el fenómeno FRET.

3. FRET para TFAM, HMG1 y mutantes puntuales en HMG2 para el complejo con ADN 30pb

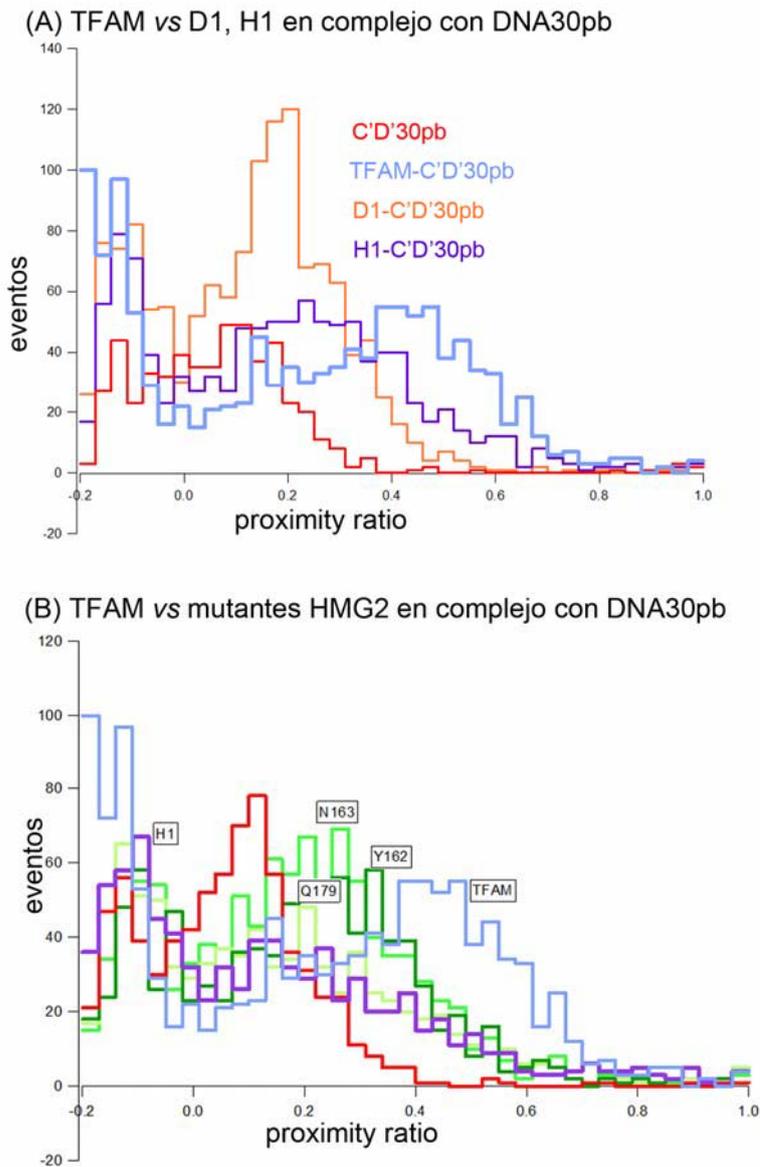
Una vez obtenidos datos para lo complejo TFAM-30pb se llevaron a cabo experimentos spFRET para dominios y mutantes puntuales de TFAM.

3.1. Dominios para HMG1

Se comparó el fenómeno FRET debido al dominio D1 (HMG1: Ser44-Gln125) y dominio H1 (HMG1+linker: Ser44-Leu152) con el de la proteína entera TFAM (Fig_M4A). Si fijamos los picos asociados a cada muestra, es posible observar que en el caso de TFAM existe un mayor valor de *proximity ratio* que para D1 y H1, ya que, la curvatura inducida por la proteína entera es mucho mayor. Además en el caso de D1, sin el linker, el valor de *proximity ratio* es menor que para H1 (por lo tanto, en D1 hay una mayor distancia entre fluoróforos), indicando que el linker incrementa la curvatura del ADN.

3.2. Mutantes puntuales para la unión de HMG2

Durante la descripción de la estructura fue posible definir contactos proteína-ADN en el dominio HMG2 que parecían importantes para la unión y curvatura del ADN, que incluían Tyr162, Asn163, Val166, Pro178, Gln179 y Leu182. En este apartado concretamente se analizó el efecto de mutantes puntuales de los residuos: Gln179, Tyr162 y Asn163, substituidos uno a uno por Ala (Fig_M4B). Ninguno de los mutantes mostraba el mismo valor de *proximity ratio* que para la proteína nativa TFAM, indicando la participación de los residuos en la promoción de la curvatura.



Fig_M4. Histograma spFRET para D1, H1, mutantes puntuales de HMG2 con ADN 30pb. (A) TFAM vs D1, H1 en complejo con ADN 30pb. En rojo se muestra el perfil para el ADN, en naranja para D1 y en púrpura para H1. (B) TFAM vs mutantes puntuales para los residuos de interacción específica con ADN de 30pb. En diferentes tonalidades de verde se encuentran los mutantes de HMG2: N163, Y162, Q179.

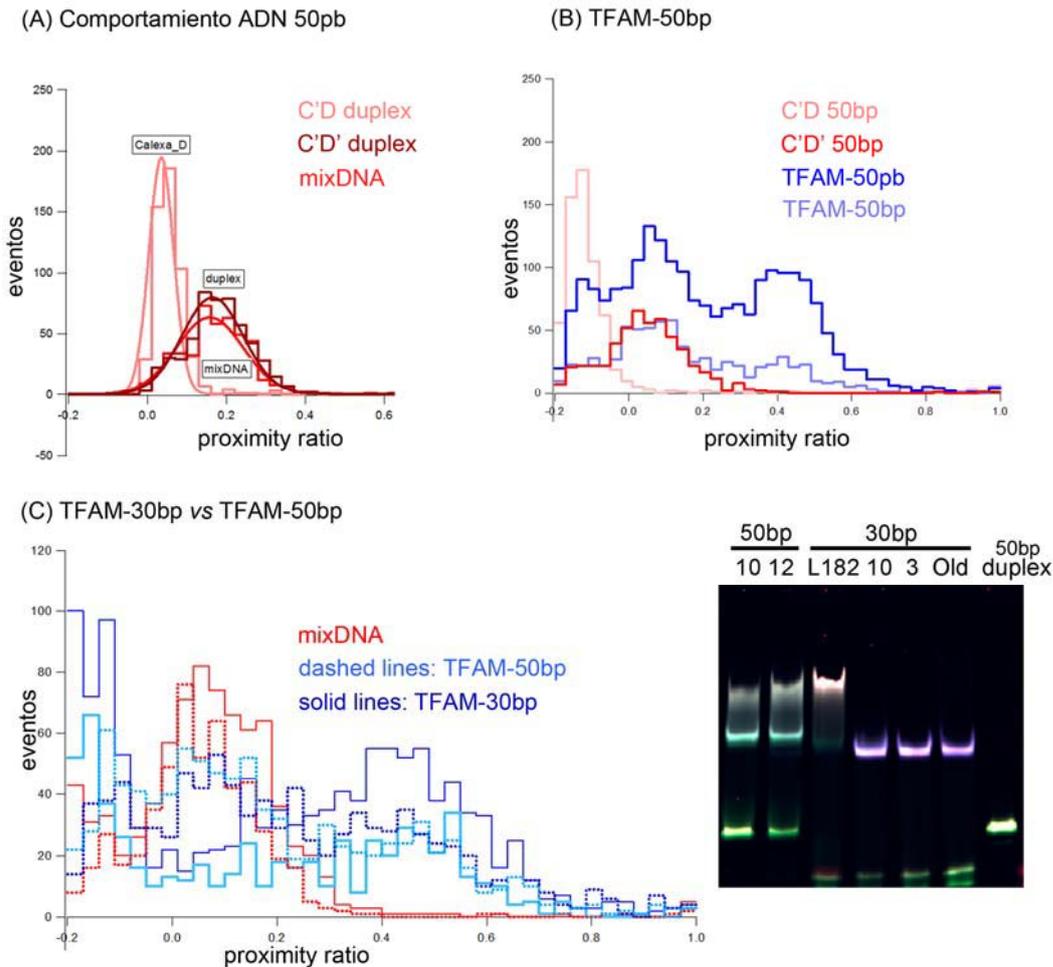
4. FRET para TFAM, HMG1 y mutantes puntuales de HMG2 en complejo con ADN 50pb

4.1. TFAM

Al igual que en el caso de ADN de 30pb, para las medidas de los complejo con ADN de 50pb fue primero necesario analizar el comportamiento de los diferentes dúplex. En la figura M5(A) se puede observar otra vez la diferencia en el perfil de FRET para el dúplex C'D-50pb (en rosa) y el dúplex C'D'-50pb (en morado), al igual que sucedía para ADN de 30pb, donde existía un doble pico. En este caso los fluoróforos se encuentran teóricamente a una distancia de $\sim 162 \text{ \AA}$, con lo cual resulta poco probable que este ADN presente fenómeno FRET debido a, por ejemplo una leve curvatura intrínseca. Diversas hipótesis podrían explicar este comportamiento del ADN: una curvatura intrínseca, efecto del clúster de guaninas en la distribución en solución de los fluoróforos y sus dipolos, o la presencia de supraestructuras de ADN que permitiesen la excitación del aceptor, entre otras.

Los experimentos de TFAM en complejo con ADN de 50pb daba lugar a histogramas similares al ADN 30pb, con presencia de 3 campanas de Gauss en un mismo histograma (Fig_M5B). Además parecía que el valor de *proximity ratio* para el fenómeno FRET debido al complejo TFAM-50pb era similar al de 30pb (Fig_M5C). En el gel EMSA, se distinguían los complejos con ADN de 30pb con los complejos para ADN de 50pb. Aunque la banda de ADN libre migra de forma diferente para las diferentes longitudes de ADN, la migración del complejo TFAM-ADN da lugar a una banda prácticamente a la misma altura. En el caso del complejo TFAM-ADN 50pb fue necesario afinar el ratio proteína: ADN debido a que la presencia de bandas superiores y muchos más difusas era más corriente, posiblemente por la mayor longitud de ADN se favorecía la unión de forma dimérica de TFAM al ADN.

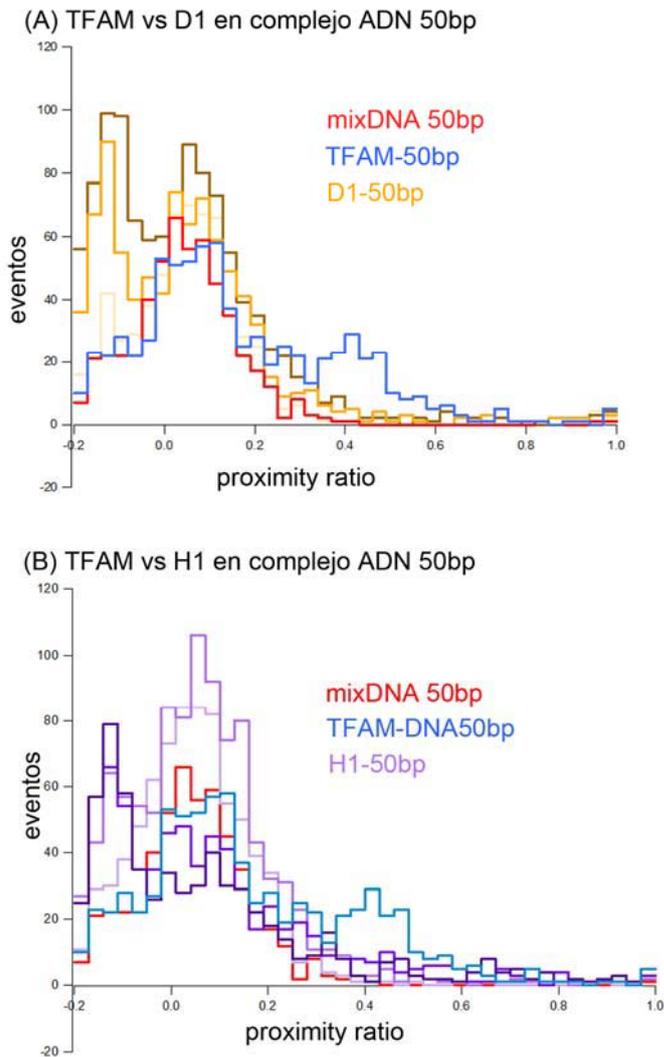
El hecho que el valor de *proximity ratio* sea similar para TFAM-30pb y TFAM-50pb indica que la distancia entre fluoróforos es independiente de la longitud del ADN, lo que permite sugerir un sistema de unión parecido al de la estructura cristalográfica, donde la proteína promueve una forma de U en el ADN debido a la presencia de dos puntos de curvatura.



Fig_M5. Experimentos spFRET para TFAM en complejo con ADN de 50pb. (A) Comportamiento ADN marcado de 50pb. Histograma para dúplex marcado solo con donador (C'D dúplex 50pb, en rosa) frente al histograma con dúplex con ambos fluoróforos (C'D' dúplex 50pb, en morado) y una mezcla de ADN marcado y no marcado (mixDNA, en rojo). (B) Histograma de TFAM-50pb (en azul) frente a ADN de 50pb marcado con ambos fluoróforos (en rojo) o solo donador (en rosa). (C) Comparación entre perfiles de spFRET para TFAM-30pb (interlineado azul) vs TFAM-50pb (línea sólida azul). En el panel derecho, EMSA gel al 12% para las muestras representadas en el histograma. Las muestras 10 y 12 son para el complejo TFAM-50pb, y 10, 3, Old, para el complejo TFAM-30pb.

4.2. Dominios para HMG1

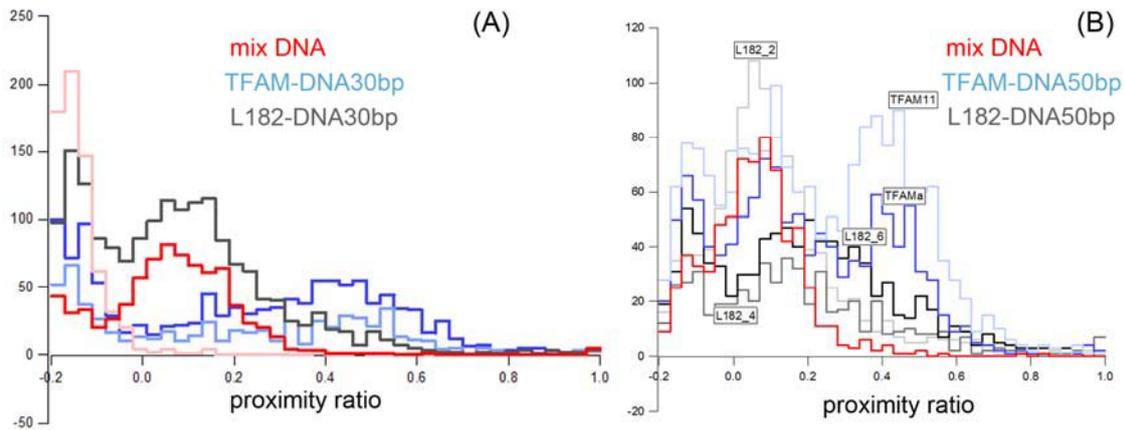
En el caso del ADN 50pb también se realizaron experimentos con el dominio D1 y H1. Los perfiles de distribución del valor *proximity ratio* para D1 y H1 (Fig_M6A y B) son prácticamente similares al perfil con el ADN 50pb. En este caso los perfiles son muy diferentes al caso del ADN de 30pb.



Fig_M6. Histograma spFRET para D1, H1 y TFAM. (A) Histograma de la distribución de proximity ratio para diversas muestras de D1 (con diferentes tonalidades de naranja), para ADN en rojo y para TFAM en azul. (B) Histograma para diversas muestras de H1 (con tonalidades de púrpura).

4.3. Mutante puntual para L182

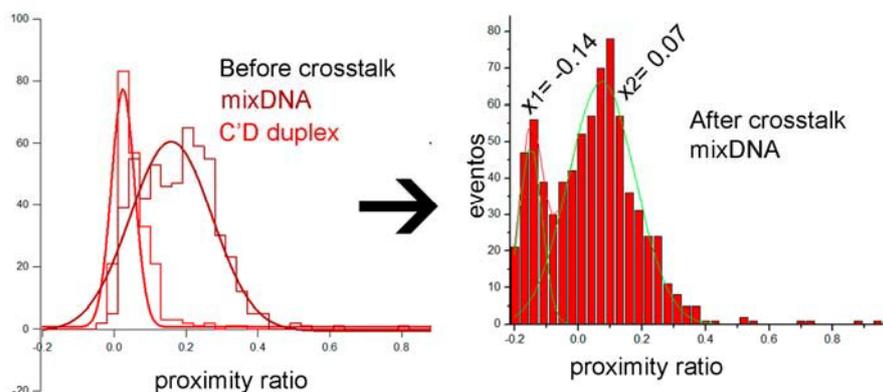
Otra de las medidas realizadas fue para el mutante puntual del residuo intercalante para el dominio HMG2, Leu182. El perfil de distribución para Leu182 en ambos casos (ADN 30pb, ADN 50pb) mostraba la presencia de dos picos con una cola extendida hasta valores de *proximity ratio* mayores que para el caso de la medida del ADN sin proteína (Fig_M7 A y B). Por lo tanto, la mutación del residuo Leu182 parece mostrar fenómeno FRET, pero menor que para TFAM nativa. Por lo tanto, podemos sugerir que la promoción de la curvatura por parte de HMG2 es un proceso cooperativo donde sería el efecto conjunto de todos los residuos el que permite obtener la curvatura final.



Fig_M7. Histogramas de spFRET para mutantes puntuales de Leu182 en complejo con ADN 30pb (A) y ADN de 50pb. Los complejos TFAM-ADN se muestran con diferentes tonalidades de azul. En rojo aparece la media para el ADN, sin presencia de proteína. (B) Con diferentes tonalidades de gris se muestran las medidas para el mutante Leu182.

5. Modificación del cálculo *crosstalk correction*

Después de la obtención de los perfiles de distribución de *proximity ratio* para los diferentes complejos y ADN, es necesario normalizar los valores de *proximity ratio* en función del control negativo usado en cada medida. Es decir, con el programa IGOR se fija a cero el valor de *proximity ratio* asociados para el ADN sin presencia de proteína. Normalmente se usa la muestra descrita como *Donor only* (C'D dúplex), ADN con presencia solo del donador, y/o el ADN con los dos fluoróforos pero sin presencia de proteína (C'D dúplex o mixDNA). Tal y como se ha comentado anteriormente en la muestra mixDNA o C'D dúplex existen dos picos en el histograma. Para el procesamiento de las siguientes muestras se tomó la posición central de los dos picos como valor cero (Fig_M8). Es posible observar que después de la corrección ninguno de los dos picos se centra en la posición cero, por ello fue necesario reprocesar los datos tomando como *crosstalk correction* la posición central del pico para *Donor only*, es decir, el primer pico del histograma para mixDNA. Una vez obtenido el valor para *crosstalk correction* se deben reprocesar los datos mediante el programa FRETtchen y posteriormente, visualizarlo y analizarlo mediante el programa IGOR y ORIGIN.



Fig_M8. Histograma del perfil de distribución de los valores de *proximity ratio* para dos tipos de ADN: C'D dúplex (ADN marcado solo con el donador) y mixDNA (mezcla de hebras de ADN marcadas con los dos

fluoróforos y sin marcaje). A la izquierda se observa el histograma antes de *crossstalk correction* y el histograma después de *crossstalk correction* con los valores de *proximity ratio* para cada posición del pico.

6. Resultados finales

Una vez reprocesados los datos con los valores de *crossstalk correction* se visualizaron los histogramas con el programa IGOR que permite obtener los valores numéricos para cada muestra. Posteriormente, mediante el programa ORIGIN se determinó para cada curva de distribución la **posición central (x)**, que corresponde al pico de la campana de Gauss y determina el valor de *proximity ratio* para la población de la muestra. El **área (A)** de la campana es la medida del porcentaje representativo de esta conformación en la muestra. Finalmente, la **anchura (w)** a la mitad del pico que ofrece información sobre la heterogeneidad de la población, es decir, si se trata de una conformación dinámica o más bien estática.

6.1. Valores de *proximity ratio* obtenidos para los diferentes complejos

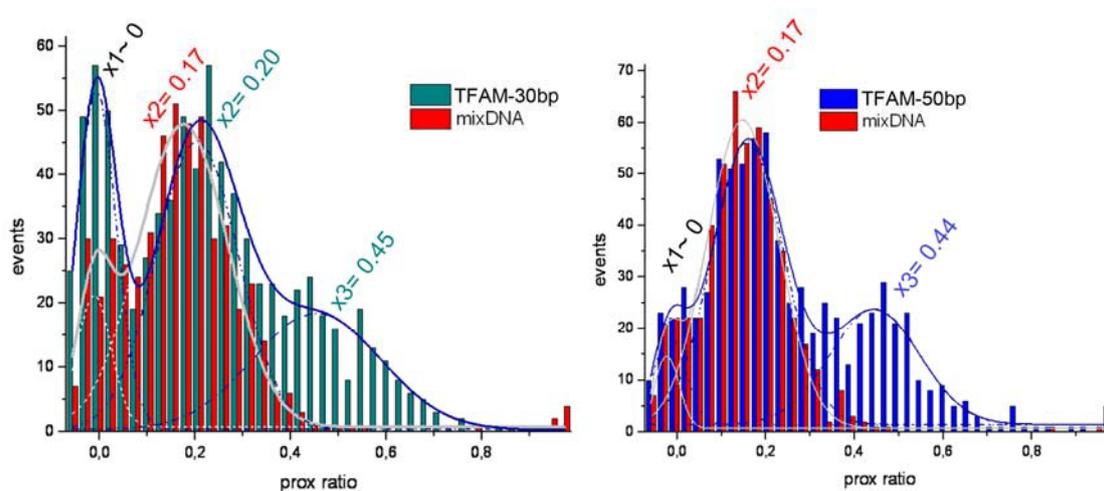
Las diferentes curvas fueron analizadas por separado en función del día de obtención de los resultados y su correspondiente *crossstalk*. Seguidamente se obtuvieron los valores de *proximity ratio*, que luego fueron promediados. A continuación (Tabla_M1) aparecen los valores de *proximity ratio*, o la posición central definida para cada pico, para cada tipo de complejo con los parámetros *crossstalk correction* definidos previamente.

	Proximity ratio
TFAM_30pb	0,21
intermediate	
TFAM_30bp	0,48
TFAM_50bp	0,45
D1_30bp	0,24
D1_50bp	0,16
H1_30bp	0,31
H1_50bp	0,16
L182_30bp	0,28
L182_50bp	0,16
N163	0,28
Y162	0,24
Q179	0,39

Tabla_M1. Valores de *proximity ratio* para cada tipo de complejo. Los valores más bajos de *proximity ratio* indican mayor distancia entre fluoróforos, mientras que valores más altos indican menor distancia entre fluoróforos.

6.1.1. Valores de proximity ratio para TFAM y ADN 30pb/ 50pb

En el caso de TFAM con ADN de 30pb era posible diferenciar 3 perfiles o campanas de Gauss, indicando la presencia de heterogeneidad en la muestra (Fig_M9). El primer pico correspondía con el del ADN-*Donor only* con un valor $x \sim 0$ (no FRET), debido a la presencia de ADN libre en la muestra. El segundo pico se encuentra desplazado referente respecto al pico del ADN marcado con los dos fluoróforos con un valor $x = 0.21$ (valor promedio, 0.20 para el caso de la Fig_M9). Y, finalmente, existe un tercer pico con un valor de *proximty ratio* de 0.45 (valor promedio, 0.48 para el caso de la Fig_M9).. Por lo tanto, parece existir dentro de TFAM en complejo con ADN 30pb, dos tipos de poblaciones con diferente eficiencia para el fenómeno FRET, sugiriendo diferente grado de curvatura en el ADN. El complejo TFAM-ADN50pb no muestra esta población intermedia con valor *proximity ratio* superior al pico del ADN, aunque sí muestra un valor similar a una de las poblaciones descritas para TFAM-ADN30pb ($x = 0.45$ para TFAM-ADN50pb, $x=0.48$ para TFAM-ADN30pb). La ausencia de la población intermedia en el caso de TFAM-ADN50pb es debido a que la distancia entre los fluoróforos para cada ADN es diferente, por lo que, una curvatura de 90° en el ADN de 30pb implicaría una distancia entre sondas (R) que da lugar a un valor de *proximity ratio* x . Para el ADN de 50pb, esta curvatura se traduciría en una mayor distancia entre sondas y, por lo tanto, un valor menor de *proximity ratio*.



Fig_M9. Distribución de proximity ratio para el complejo TFAM-DNA30pb (a la izquierda) y complejo TFAM-DNA50pb (a la derecha), donde se indica con línea sólida el perfil de la distribución global y con interlineado los ajustes para los diferentes picos que conforman la distribución global. En el caso del complejo TFAM-30pb es posible observar la presencia de una nueva población con valor x_3 y una población intermedia con valor x_2 , entre el perfil del ADN y el perfil para el complejo.

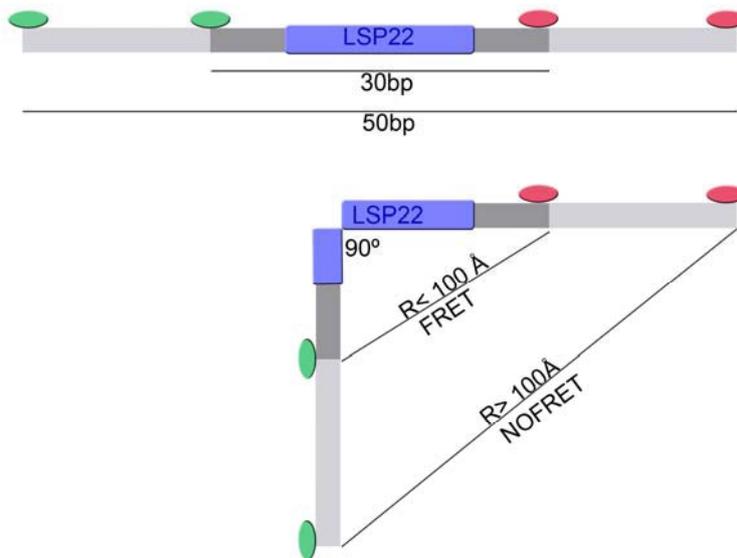
6.1.2. Valores de proximity ratio para los dominios y sus mutantes puntuales

Los complejos de los diferentes dominios y mutantes con ADN 30pb muestran una reducción del fenómeno FRET respecto de TFAM, ya que su valor de *proximity ratio* es menor. Cabe destacar que el dominio D1 (HMG1) presenta un valor de *proximity ratio*

menor al dominio H1 (HMG1+linker), confirmándose el hecho que la adición del linker a HMG1 incrementa la capacidad de este dominio de curvar el ADN.

Si comparamos los valores *proximity ratio* para los diferentes mutante puntuales N163, Y162, Q179 y L182 (Tabla_M1), es posible observar que todos ellos presentan una reducción del valor *proximity ratio* comparado con TFAM nativa, es decir, una disminución de la curvatura en el ADN. Aunque es en el caso del mutante Y162 donde la reducción es mayor, pese a que se esperaría que el mutante para el residuo intercalante (L182) causará una mayor perjuicio en la curvatura final.

Para el ADN de 50pb el comportamiento de los mutantes es muy distinto, mostrando un valor de *proximity ratio* prácticamente idéntico al ADN control y sugiriendo que no existe FRET bajo estas condiciones (al igual que sucedía para la población intermedia de TFAM-30pb que desaparecía en TFAM-50pb). Ello es debido a que en presencia de un mismo ángulo de curvatura, la distancia entre los fluoróforos (R) para cada ADN puede resultar diferente. Por ejemplo, una posible curvatura de 90° (similar a la curvatura inducida para cada HMGbox) en el ADN de 30pb implica una distancia entre sondas que da lugar a un valor de *proximity ratio* x. Para el ADN de 50pb, esta curvatura se traducirá en una mayor distancia entre sondas provocando la disminución del fenómeno FRET o su ausencia (Fig_M10).

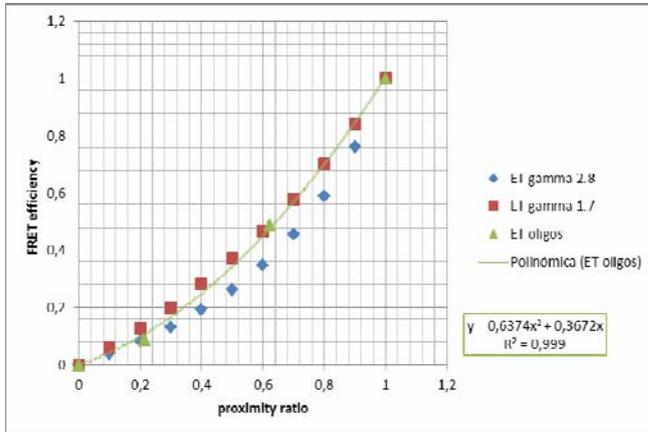


Fig_M10. Esquema explicativo del sistema FRET mediante el uso de ADN con diferente longitud, donde R indica la distancia entre fluoróforos (señalados con círculos verde para Alexa 488 y rojo para Alexa 594). En la parte superior aparece el ADN sin presencia de curvatura, donde $R > 100 \text{ \AA}$ y no existe fenómeno FRET. En la parte superior existe una curvatura en la región LSP22 de 90°, que provoca un acortamiento de la distancia entre fluoróforos para ambos ADN's, que en el caso de ADN de 30pb es suficiente para que exista fenómeno FRET.

6.2. Proximity ratio vs E_{FRET}

Como se ha comentado en la descripción del fundamento (Materiales y métodos: FRET), la relación entre P y E_{FRET} , depende del factor de detección del sistema (γ), definido por la eficiencia de detección y el rendimiento cuántico de donador y aceptor. Cuando $\gamma=1$, sucede que $P = E_{\text{FRET}}$. En nuestro caso se realizó la calibración para nuestros oligos y el

sistema (en verde en la gráfica de la Fig_M11), y a partir de la función polinómica obtenida se extrapolaron los valores de proximity ratio para cada complejo. Así se obtuvieron los valores de Eficiencia de FRET (E_{FRET}).



Fig_M11. Gráfica de la relación entre E_{FRET} (definida como ET en la leyenda) y proximity ratio para diferentes valores del factor de detección del sistema (γ). En rojo para $\gamma = 1.7$, en azul para $\gamma = 2.8$ y en verde para el sistema FRET con nuestros oligos.

Los valores de E_{FRET} permiten llegar a las mismas conclusiones que los valores de *proximity ratio* (Tabla_M2). La mayor eficiencia del fenómeno sucede para el caso de TFAM nativa y su valor es similar para ambos ADN, mientras que los dominios y mutantes presentan diferentes valores.

	Proximity ratio	E_{FRET}
TFAM_30pb		
intermediate	0,21	0,11
TFAM_30pb	0,48	0,32
TFAM_50pb	0,45	0,29
D1_30pb	0,24	0,12
D1_50pb	0,16	0,08
H1_30pb	0,31	0,17
H1_50pb	0,16	0,08
L182_30pb	0,28	0,15
L182_50pb	0,16	0,08
N163	0,28	0,15

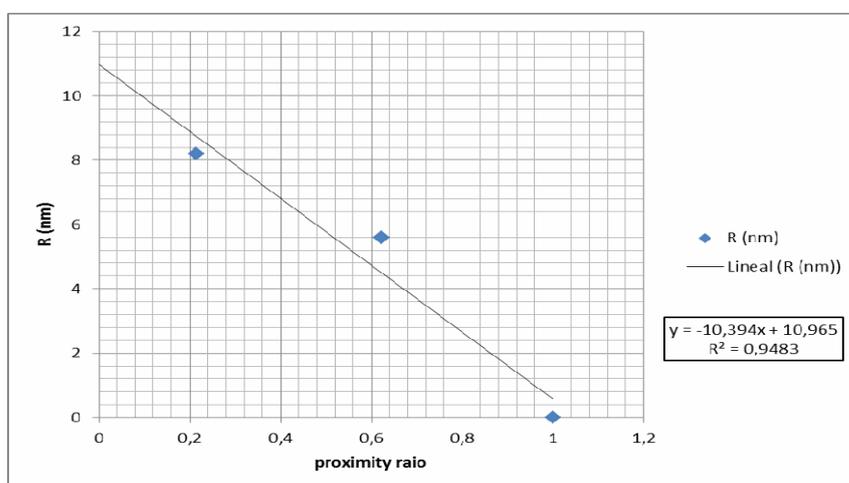
Y162	0,24	0,13
Q179	0,39	0,24

Tabla_M2. Valores de E_{FRET} , para los diferentes complejos.

6.3. Cálculo de la distancia (R) entre fluoróforos

Para el cálculo de la distancia entre los fluoróforos se optó por la calibración del sistema, es decir, se realizaron la medida de ADN's con los fluoróforos colocados a distancias conocidas. Por lo tanto, es posible representar los valores de proximity ratio obtenidos por nuestro sistema *versus* la distancia en nanómetros entre los fluoróforos (R) (Fig_M12). A esta representación se pueden extrapolar los valores de proximity ratio obtenidos para la muestras, obteniendo así la distancia entre fluoróforos. Pese a ello cabe destacar que existe un parámetro que influye en el error asociado al cálculo de distancia conocido como anisotropía, que hace referencia a la orientación entre los momentos dipolos de los fluoróforos. Este parámetro se encuentra entre valores de 0 (dipolos paralelos, esta situación aunque es la ideal no ocurre experimentalmente) a 4 (para dipolos orientados de forma perpendicular). El cálculo de la anisotropía ($\kappa^2= 2.86$ para TFAM-30pb y $\kappa^2= 2.32$ para TFAM-50pb), permitió realizar un cálculo del error o incertidumbre de la medida de la distancia de -22% a +27% para TFAM-30pb y -15% a +23% para TFAM-50pb.

A partir de estos valores R, es posible hacer un cálculo aproximado del ángulo de curvatura asociado a determinados valores R. Ello consiste en el uso de la información cristalográfica y los modelos construidos, ya que, si la curvatura del ADN es similar a la cristalográfica con 180° sabemos que $R \sim 56 \text{ \AA}$. Por lo tanto, podemos hacer un cálculo aproximado del ángulo de curvatura.



Fig_M12. Recta de calibración entre proximity ratio vs R (nm), con la siguiente ecuación lineal asociada .

La menor distancia entre fluoróforos se da en el complejo TFAM-30pb y TFAM-50pb, que permite calcular un ángulo final de $\sim 170^\circ$ (Tabla_M3).

Al comparar la distancia entre fluoróforos que ofrece el complejo TFAM-ADN con los valores para los complejos con el dominio D1 y H1, es posible observar que son inferiores. Además la distancia para el dominio H1 es menor al dominio D1, por lo que la curvatura inducida por el dominio H1, es decir, HMG1 con linker; es mayor. Es interesante destacar

que el ángulo aproximado calculado para el dominio D1 es similar a la conformación intermedia de curvatura para TFAM (ver más adelante).

Las mutaciones puntuales de los residuos de interacción de HMG2 con el ADN permiten contrastar que la mutación que ofrece un mayor efecto reductor de la curvatura es Y162 seguido del mutante L182. Mientras que el mutante para el residuo Q179 no implica una distorsión de la curvatura tan grande. En el caso de los complejos con los mutantes para L182, N163 y Y162, el valor de la distancia entre fluoróforos es similar al dominio H1 y D1. Ello refleja que en el caso de estos mutantes puntuales, el dominio HMG2 no puede promover una curvatura estable en el ADN. Estos valores de distancia muestran que no existe un residuo capital en la intercalación al ADN en HMG2, sino que se trata de un proceso cooperativo suma del efecto de los diferentes residuos de interacción. Por lo tanto, los experimentos de FRET permitieron profundizar en el mecanismo de unión y curvatura de HMG2 al ADN.

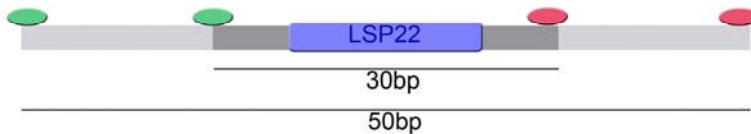
	Proximity ratio	R (nm)	R(Å)	curvatura
TFAM_30pb intermediate	0,21	8,76	87,61	81,37
TFAM_30pb	0,48	5,96	59,60	170,44
TFAM_50pb	0,45	6,28	62,79	160,31
D1_30pb	0,24	8,47	84,75	90,47
H1_30pb	0,31	7,76	77,62	113,15
L182_30pb	0,28	8,09	80,86	102,85
N163_30pb	0,28	8,02	80,24	104,83
Y162_30pb	0,24	8,44	84,38	91,65
Q179_30pb	0,39	6,88	68,84	141,08

Tabla_M3. Valores de proximity ratio, distancia entre fluoróforos y curvatura del ADN.

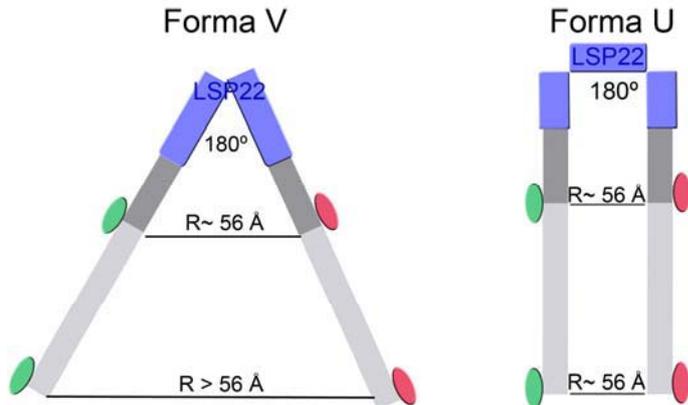
6.4. Confirmación del sistema en forma de U

La correlación entre los valores de *proximity ratio*, E_{FRET} y R entre los complejos TFAM-30pb y TFAM-50pb, permite ratificar el sistema de curvatura descrito en la estructura cristalográfica donde existen dos punto de curvatura en el ADN (Fig13). Por lo tanto, el mecanismo de curvatura de ADN promovido por TFAM es en forma de U y no de V como se había descrito anteriormente por otros grupos (Malarkey, Bestwick et al.).

(A) DNA, NO FRET



(B) DNA, FRET



Fig_M13. Esquema explicativo del sistema FRET mediante el uso de ADN con diferente longitud, donde R indica la distancia entre fluoróforos (señalados con círculos verde para Alexa 488 y rojo para Alexa 594). (A) Aparece el ADN sin presencia de curvatura, donde $R > 100 \text{ \AA}$ y no existe fenómeno FRET. En azul aparece la secuencia LSP22 en la región central, en diferentes tonalidades de gris se muestra el ADN de 30pb y ADN de 50pb. (B) Dos casos donde existe un acortamiento de la distancia entre fluoróforos debido a dos tipos de curvatura muy distintos: en forma de U o forma de V. En el caso de forma V, la curvatura en la zona central de LSP22 provoca un acortamiento de distancias mayor entre ADN de 30pb que en ADN 50pb. Mientras que en el caso en forma de U, la curvatura final de 180° (suma de dos puntos de curvatura) provoca una distancia entre fluoróforos similar para ambos ADN.

6.5. Presencia de conformaciones múltiples

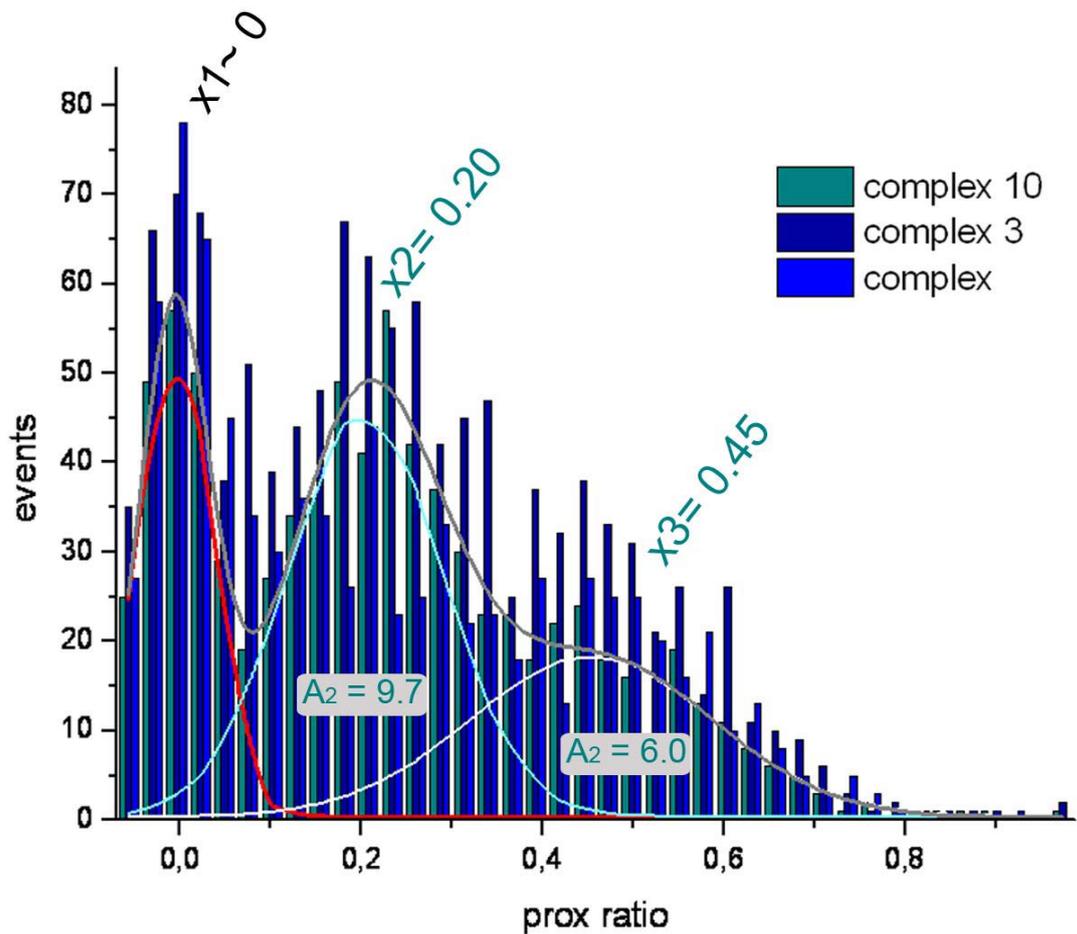
Los ensayos spFRET permiten distinguir la existencia de diferentes especies o conformaciones en la muestra. La presencia de conformaciones diferentes se muestra por la presencia de múltiples picos dentro del perfil global. En el caso del complejo TFAM-30pb (Fig_M14) se obtiene un perfil donde es posible determinar tres campanas de gauss con diferentes posiciones centrales o valores de proximity ratio: $x_1 \sim 0$, hace referencia a la presencia de ADN libre; $x_2 \sim 0.20$ y $x_3 \sim 0.45$. Por lo tanto, existen dos conformaciones dentro de la dinámica de curvatura del ADN. Una primera conformación con una distancia entre sondas de $\sim 87 \text{ \AA}$, lo cual indica una curvatura de $\sim 81^\circ$. La segunda conformación presenta una distancia entre sondas de $\sim 59 \text{ \AA}$ y, por lo tanto, un ángulo de $\sim 170^\circ$. Por lo tanto, esta tercera conformación presentaría una similitud con la estructura cristalográfica, ya que los modelos creados a partir de esta pronosticaban una distancia de $\sim 56 \text{ \AA}$ entre sondas.

A partir de los perfiles también es posible obtener información de la representación de las diferentes conformaciones en la muestra total, en forma de área (A) de la campana de Gauss, así como obtener información sobre la heterogeneidad de la conformación en base a la anchura de la curva a la mitad de la altura del pico (w). La población similar a la estructura cristalográfica muestra un área inferior a la población de curvatura intermedia

y además muestra una mayor heterogeneidad dentro de la población, ya que presenta un valor w mucho mayor a la conformación intermedia.

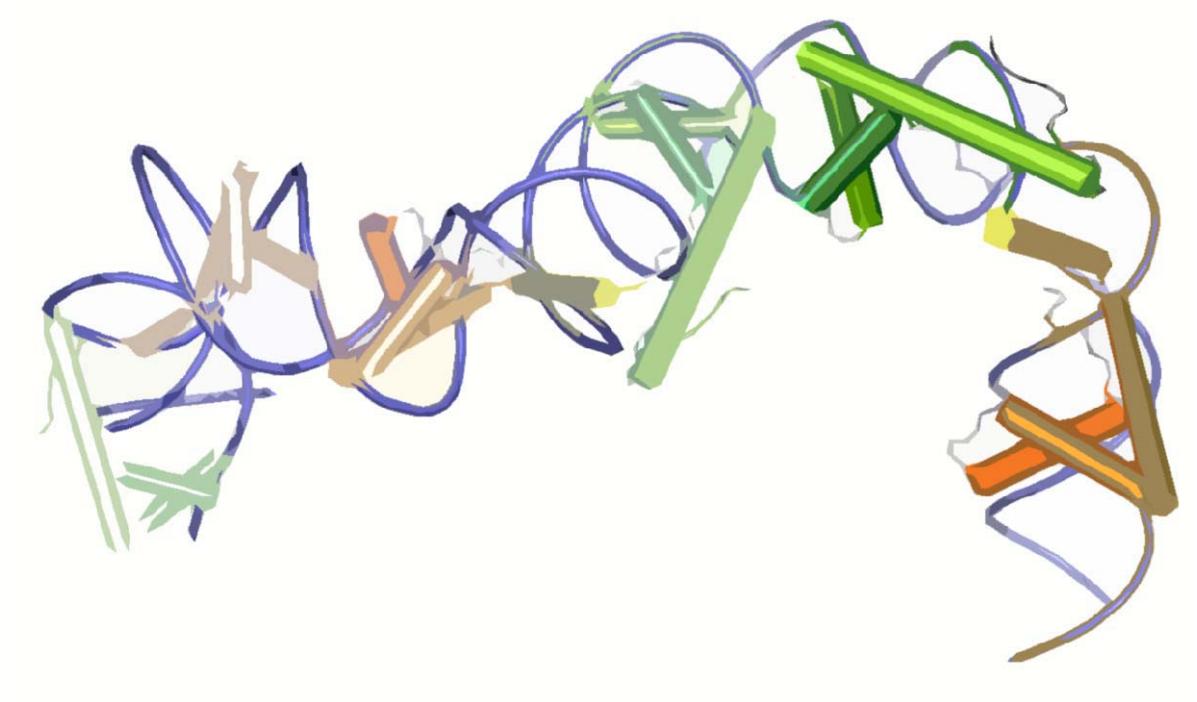
Cabe destacar que la existencia de este población intermedia no se observaba en todo los ensayos.

Conformación	x	A	w
Curvatura ~ 80°	0.20 ± 0.01	9.66 ± 1.74	0.17 ± 0.02
Curvatura ~ 170°	0.45 ± 0.043	6.03 ± 2.14	0.27 ± 0.07



Fig_M14. Histograma TFAM-DNA 30pb para 3 complejos diferentes (en diferentes tonalidades de azul). Dentro del histograma con un perfil global (indicado con una línea gris) se ajustaron diferentes picos asociadas a cada campana de Gauss (en rojo, celeste y blanco) con el programa ORIGIN. Se indican los valores para cada posición central y el área relativa para los picos debidos a la presencia de diferentes conformaciones.

DISCUSIÓN



A. TFAM IMPONE UNA CURVATURA DE 180° SOBRE LSP

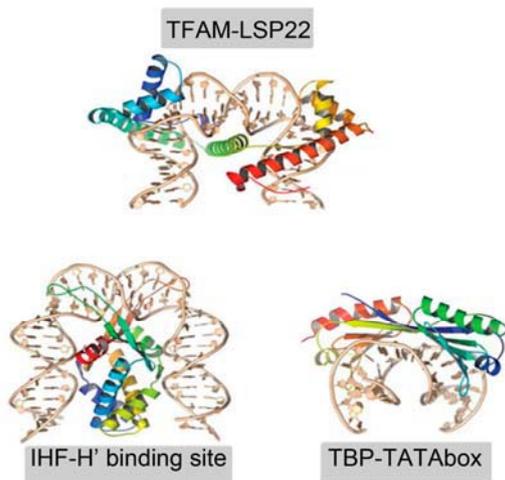
1. TFAM provoca una torsión de 180° en LSP

La resolución de la estructura cristalográfica permite visualizar la torsión que TFAM provoca sobre la secuencia LSP. Esta perturbación topológica queda también reflejada mediante estudios de FRET, ratificándose el modelo de unión de TFAM con dos puntos de curvatura realizados por cada dominio HMGbox por separado.

La forma en la que TFAM curva el ADN muestra una similitud con las proteínas procariotas HU (*Heat-Unstable nucleoid*) y IHF (*Integration Host Factor*) que compactan y empaquetan el ADN procariota mediante la introducción de curvaturas en forma de U (Fig_A1) (Rice, Yang et al. 1996; Mouw and Rice 2007). IHF es una proteína pequeña heterodimérica que provoca una curvatura de ~ 180° mediante la intercalación de dos residuos de prolina muy conservados en dos puntos diferentes del surco menor (uno por cada monómero). Debido a que TFAM es una de las proteínas principales en el mantenimiento y empaquetamiento del mtDNA fácilmente podemos llegar a la conclusión que la curvatura que TFAM promueve después de la unión con el ADN es un mecanismo para su empaquetamiento.

Además, también se trata de un factor de transcripción mitocondrial. El papel topológico que TFAM provoca sobre LSP recuerda al factor de transcripción nuclear de unión a la caja TATA (TBP) (Fig_A1). TBP se une de forma específica a una secuencia de ADN llamada caja TATA situada 30 pb río arriba del sitio de inicio de la transcripción, similar a TFAM que une la región -35 a -15 pb. La unión de TBP provoca una distorsión en el ADN mediante la intercalación de residuos de fenilalanina de sus hojas β en el surco menor, provocando un desenrollamiento local y una curvatura final de 80° (Nikolov, Chen et al. 1996). Esta curvatura del ADN permite el reclutamiento de la maquinaria para el inicio de la transcripción y promueve una apertura en la zona rica en AT cercana al sitio de inicio de la transcripción.

Tanto en IHF y TBP la curvatura viene dada por la intercalación de residuos hidrofóbicos en el surco menor del ADN, con dos puntos de hendidura, dando lugar a una curvatura similar. Al igual que sucede en TFAM.



Fig_A1. Estructura de la proteína TFAM, IHF y TBP en complejo con sus respectivos DNA's. Es posible observar como en todos los casos existe una curvatura del ADN provocada por la unión de la proteína. TFAM-LSP22 (Rubio-Cosials, Sidow et al.), IHF-H' sitio de unión (Mouw and Rice 2007), TBP humana en complejo con la caja TATA (Nikolov, Chen et al. 1996).

2. Dinámica del complejo TFAM-LSP

El hecho que la estructura cristalográfica no pudiese explicar los datos experimentales de SAXS en solución para el complejo TFAM-LSP22 indica cierta dinámica en el complejo TFAM-LSP22. Pese a ello, los dos escenarios que mejor ajuste presentaban eran, por una parte, para la situación en que el ADN era idéntico al cristalográfico, con la curvatura de 180° , y se definía como cuerpo rígido toda la proteína excepto la cola C-terminal con total libertad (resíduos 225-246) (escenario 1). Para el segundo escenario, la proteína era definida como un cuerpo rígido para HMG1 y la región del linker, dando flexibilidad a HMG2, y con el ADN curvado 90° para la región de unión de HMG1 (escenario 6). Por lo tanto, estos experimentos *in silico* sugerían un sistema con conformaciones más o menos compactas con la presencia de posibles movimientos moleculares sutiles entre ellas, ya fuese un ligera respiración del ADN o un desplazamiento de HMG2.

En relación con esta sospecha, los ensayos spFRET permitieron distinguir la existencia de conformaciones diferentes en la muestra en relación a la curvatura del ADN. FRET indica que existen dos conformaciones de ADN en complejo: una conformación con un elevado grado de heterogeneidad (su histograma muestra un pico muy ancho) con una distancia entre sondas de $\sim 59 \text{ \AA}$ y, por lo tanto, un ángulo de $\sim 170^\circ$. Esta conformación está en concordancia con las estructuras cristalográficas descritas. Por otra parte existe una conformación presente en alguna de las medidas, con una distancia entre sondas de $\sim 87 \text{ \AA}$, lo cual indica una curvatura de $\sim 81^\circ$. Debido a que HMG1 presenta mayor afinidad que HMG2 por ADN y que en general la unión de los dominios HMGbox típicamente provoca un ángulo de torsión entre 70 a 110° , podríamos pensar que esta curvatura en el ADN de $\sim 80^\circ$ es debida a la unión de HMG1 al ADN. Además coincide con los valores de curvatura obtenidos para el dominio HMG1 en spFRET.

Ello sugiere la existencia de una conformación intermedia de unión que provoca una curvatura al ADN más inestable, donde existe la unión de HMG1 al ADN dando lugar a una primera hendidura entre pares de bases. Posteriormente, se daría la unión de HMG2 que conjuntamente con HMG1 daría lugar a la curvatura de 180° , más estable energéticamente.

Este tipo de mecanismo de unión con especies intermedias se ha descrito también para el caso de IHF y TBP mediante experimentos de spFRET. Ensayos de spFRET en complejo de TBP con TATABox muestran un elevado dinamismo con transiciones abruptas entre múltiples estados de FRET (Schluesche, Heiss_2008). El histograma de distribución en tiempos de 750ms revela la presencia de hasta seis discretos estados, mostrando que el mecanismo de unión de TBP al ADN es más complejo de lo que se pensaba (Fig2). Para IHF el uso de spFRET *time-resolved* permitió entender la cinética de curvatura/no curvatura en el ADN, donde existe un estado intermedio con un ADN parcialmente curvado (Kuznetsov, Sugimura et al. 2006)(Fig2).

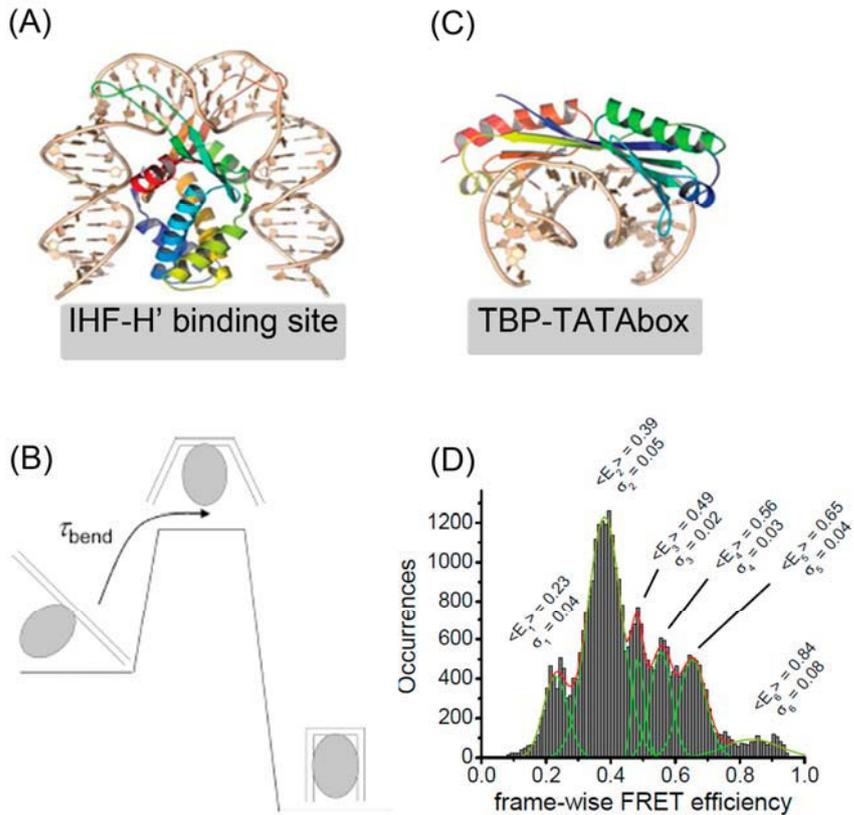


Fig2. Dinámica de los complejos para IHF y TBP con sus respectivos ADN mediante estudios de spFRET. (A) (B) Para IHF fue posible definir un estado intermedio mediante el uso de técnicas de *time-resolved* con tiempos del orden de ms. (C) (D) Para TBP, el diagrama de distribución de E_{FRET} en función del tiempo permite distinguir diferentes especies de curvatura diferente.

B. TFAM Y SU RECONOCIMIENTO ESPECÍFICO DE SECUENCIA

En este apartado de discusión se pretende profundizar en el reconocimiento específico de secuencia por parte de TFAM en LSP y las implicaciones en la transcripción. Tal y como se ha comentado anteriormente, TFAM interacciona directamente con los nucleótidos situados entre -35 y -17 (Fisher, Topper et al. 1987). Además se ha descrito que la distancia entre el sitio de unión a TFAM y el sitio de inicio de transcripción es esencial para la actividad de la secuencia promotora (Dairaghi, Shadel et al. 1995).

1. Secuencias putativas de unión por dominios HMGB

Tal y como se ha comentado existen dos tipos de dominios HMGB dentro de la familia de proteínas HMGB: unión específica de secuencia (HMGB *ss*) y unión inespecífica de secuencia (HMGB *nss*) (ver Introducción, apartado C.6). Mediante la comparación de los sitios de unión para HMGB *ss* (por ejemplo: SOX, SRY, LEF) fue posible definir una secuencia específica de unión en la región del surco menor para los dominios HMGB de unión específica (Laudet, Stehelin et al. 1993):

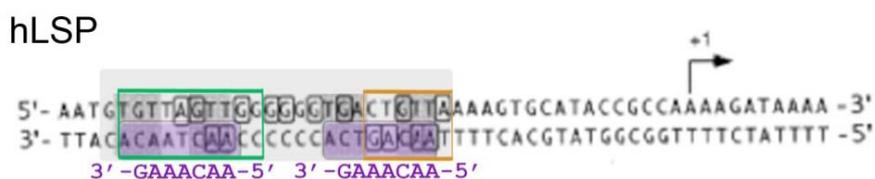
5'-A/T A/T C A A A G-3'

Tanto la región promotora LSP y HSP, para humanos (hLSP, hHSP) y ratón (mLSP, mHSP), contienen las secuencias putativas de interacción para HMGbox. Estos sitios se encuentran colocados en tándem en los promotores y se encuentran distanciados entre 5 y 6 bases. La orientación de los sitios de unión de HMGbox es opuesta para HSP y LSP, lo cual se relacionó con el carácter bidireccional en los sitios de unión de TFAM en los promotores (Fisher, Topper et al. 1987). Debido a que los pares de base A·T y T·A son indistinguibles en el surco menor, el extremo 3' de los sitios de unión para HMGbox se encuentra particularmente conservados en los promotores mitocondriales (Fig_B1). Además los ensayos de interferencia por metilación mostraban regiones de contacto para TFAM en las zonas donde se encuentran estas secuencias putativas de interacción para HMGbox ss (Fig_B1)



Fig_B1. Regiones promotoras para la cadena pesada (HSP) y ligera (LSP), con las regiones de unión para TFAM. En lila se muestran las cajas predichas para la unión de los dominios HMGbox de unión específica (5'-A/T A/T CAAAG-3'). Los nucleótidos rodeados con un círculo son aquellos que muestran interacción con la proteína mediante ensayos de protección por metilación. (Fisher, et al.1987)

Con la estructura resuelta de TFAM en complejo con la secuencia LSP22, es posible atribuir estas secuencias predichas de unión específica a los dominios HMGbox con la regiones de LSP22, definidas como región 1 y región2 (Fig_B2).

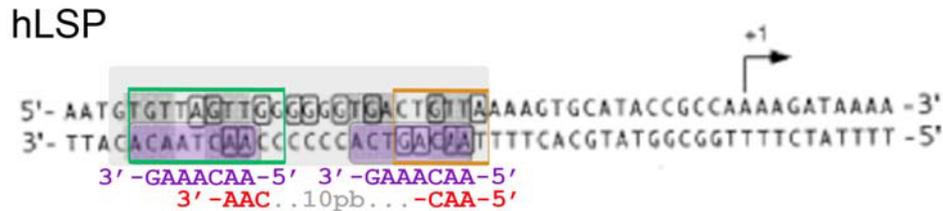


Fig_B2. Secuencia promotora de LSP, con la región de unión de TFAM identificada en los cristales (Rubio-Cosials, Sidow et al.) sombreada en gris. Las cajas purpura indican las secuencia predicha de unión específica para los dominio HMGbox (Laudet, et al.1993), mientras que las cajas naranja y verde indican las zonas de contacto de los dominios HMG1 y HMG2 de TFAM, respectivamente. Fisher, et al.

2. Motivo invertido AAC

Durante la descripción de la interacción de TFAM con LSP22, es posible definir un motivo invertido AA'C-10pb-C'AA (donde ' indica los lugares de intercalación), que sigue la

simetría binaria similar de los HMGbox. A continuación se discute la existencia de este motivo AAC, que se encuentra estrechamente relacionado con la presencia en LSP22 de las secuencias definidas como secuencias consenso de unión para dominios HMGbox ss (Fig_B3), aunque la superposición de las cajas definidas para HMGbox en hLSP realizada por Fisher, *et al.* no encaja perfectamente en el segundo motivo AAC.

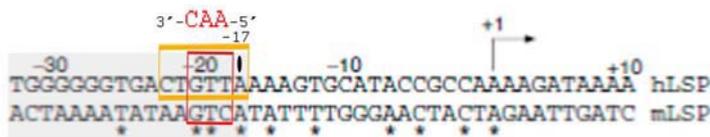


Fig_B3. Secuencia promotora de LSP, con la región de unión de TFAM sombreada en gris. Las cajas púrpura indican las secuencia predicha de unión específica para los dominio HMGbox (Laudet, *et al.*), mientras que las cajas naranja y verde indican las zonas de contacto de los dominios HMG1 y HMG2 de TFAM, respectivamente. En rojo se muestra el motivo invertido (5'AA'C-10pb-C'AA). Fisher, *et al.*

2.1. Conservación entre mLSP y hLSP, e intercambio funcional

Al comparar las secuencias que gobiernan la especificidad de la transcripción de ratón (mLSP) y humano (hLSP) es posible diferenciar dos regiones que al ser mutadas alteran el nivel de transcripción (Gaspari, Falkenberg et al. 2004): *PPE* o *Proximal Promoter Element*, la región promotora que alberga las posiciones -1 a -4, que es especialmente sensible a mutaciones; y *DPE: Distal Promoter Element*, que comprende la región de -17 a -20. Esta región distal se encuentra incluida dentro de la región de unión para TFAM.

El alineamiento de hLSP y mLSP revelaba que los nucleótidos situados en -17, -19 y -20 se encuentran conservados entre las dos regiones promotoras, mientras que en el caso del nucleótido situado en la posición -18 es una T para hLSP y C para mLSP. Esta transición de T a C provoca la inhabilidad de hTFAM para activar la transcripción de mLSP (Gaspari, Falkenberg et al. 2004) (Fig_B4).



Fig_B4. Comparación entre la secuencia hLSP y mLSP. Los asteriscos representan los nucleótidos conservados entre las dos secuencias. Con la región de unión de TFAM a LSP (-17 a -35pb) sombreada en gris y la región de contacto de HMG1-TFAM en naranja. La caja roja indica el motivo AAC invertido. Gaspari, *et al.*

Gaspari, *et al.* observaron que el sistema humano no puede llevar a cabo la transcripción a partir de la secuencia salvaje de mLSP. Sin embargo, cuando se introduce la región de hLSP de -40 a -11 pb, existe una fuerte transcripción. Además, mediante la construcción de secuencias promotoras híbridas entre ratón y humano, fue posible observar que la mutación de la región de -17 a -20 anulaba la transcripción.

Esta región de gran importancia para la activación de la transcripción (nucleótido de -17 a -20pb) contiene la secuencia 5'-AAC-3', y coincide con la región de intercalación de HMG1

de TFAM en la secuencia LSP22, concretamente con el motivo 5-AA'C-3 propuesto a partir de la resolución de la estructura.

2.2. Conservación del motivo AAC entre LSP de hominoides

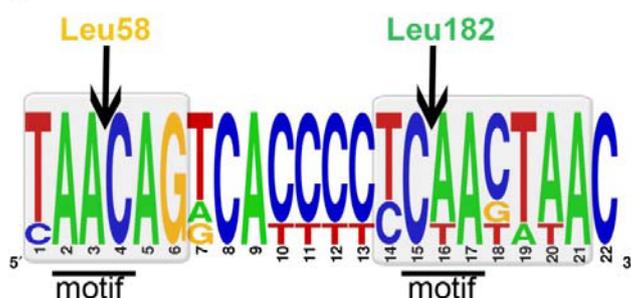
Cuando se representan las secuencias de LSP de 5 hominoides (humana, gorila, chimpancé común, chimpancé Pygmy, orangután) es posible observar que el primer motivo AAC de -18 a -20pb se encuentra conservado entre las cinco secuencias (Fig_B5 (A)). Mientras que para el segundo motivo AAC invertido de -31 a -33 pb, existe una mayor divergencia en el caso de la base A en la posición -32 (posición 16 en el Logo de la Fig_B5), aunque siempre se mantiene la distancia de 10pb entre los motivos AAC.

2.3. Conservación del motivo AAC entre las secuencias de unión a TFAM

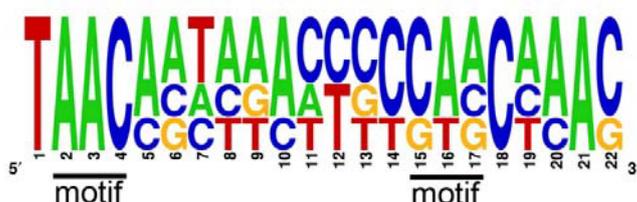
Cuando se alinea las diversas secuencias de unión específica de TFAM en el mtADN humano (descritas anteriormente, ver Introducción, apartado C.5.1.2.): LSP, HSP1, X e Y; es posible observar que el número de nucleótidos entre las repeticiones de trinucleótidos invertidas es sistemáticamente 10pb (Fig_B5 (B)).

Ello sugiere que el modo de unión de TFAM a estas secuencias podría ser topológicamente similar. Sin embargo, mientras que el primer motivo AAC es idéntico entre las cuatro secuencias de unión, el segundo motivo se encuentra menos conservado (tan sólo se mantiene conservado entre HSP1 y LSP). Por lo tanto, este primer motivo AAC parece ser imprescindible para la unión específica de secuencia por parte del dominio HMG1 de TFAM.

(A) Secuencias LSP



(B) Secuencias de unión específica hTFAM

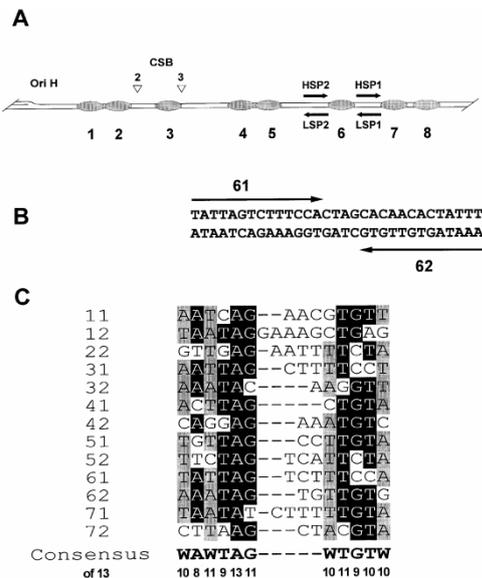


Fig_B5. Conservación entre secuencias de unión específica por TFAM, representado en forma de LOGO (Crooks, Hon et al. 2004). (A) Alineamiento de 5 secuencias de unión de LSP en diferentes organismos (humana, gorila, chimpancé común, chimpancé pigmeo, orangután), donde las cajas grises muestran las regiones de unión para

HMG1 y HMG2 de TFAM en hLSP. Las flechas indican los sitios de intercalación de Leu58 y Leu182. Finalmente subrayado en negro se encuentran los motivos AAC. (B) Alineamiento de las secuencias de unión específica en hTFAM: LSP, HSP, X e Y, donde subrayado en negro se muestran los motivos AAC.

2.4. Ejemplo de motivo en secuencias de unión de xl-TFAM

La presencia de motivos en los sitios de unión específica de TFAM ha sido también descrito para su homólogo en *Xenopus laevis*, *xl-TFAM*. El hecho de que *xl-TFAM* se une de forma tetramérica sugiere que el sitio de unión debe estar compuesto por pequeñas secuencias repetidas que son reconocidas por HMGbox, ensambladas de forma simétrica (Antoshechkin, Bogenhagen et al. 1997). Antoshechkin, *et al.* alinearon todas las secuencias de unión de *xl-TFAM*, previamente divididas por la mitad y alineadas en sentido contrario, lo cual permitió definir el siguiente motivo: WAWTAG (W: A o T) separado por dos o tres bases del siguiente motivo WTGTW. Las mitades se encuentran separadas aproximadamente por 3 o 7 nt y son invertidas una respecto a la otra, formando entonces un lugar o sitio de unión con simetría diada (Fig_B6). Ambos motivos no siempre se encuentran presentes y su combinación puede diferir, esto es consistente con la hipótesis de que cada uno de estos sitios es reconocido por un dominio HMGbox.



Fig_B6: Análisis de las secuencias de unión de *xl-TFAM* a mtADN. (A) Diagrama de la región de control de mtADN. Los sitios de unión de *xl-TFAM* están indicados mediante los óvalos sombreados. (B) Ejemplo de la orientación de la diada para el sitio 6 de unión. (C) Alineamiento de los sitios de unión a TFAM. (Antoshechkin *et al.*)

Si lo comparamos con los resultados y el conocimiento de la unión de hTFAM a LSP22, podríamos concluir que la secuencia definida como 5-AA'C-10pb-C'AA-3', con este tipo de simetría diada para el motivo AAC, conforma un requisito para la unión específica de secuencia de TFAM. Y de forma más imprescindible el primer motivo 5'-AAC-3', que realiza el contacto con el dominio HMG1 de TFAM y se encuentra más extensamente conservado entre secuencias de LSP y las diferentes secuencias humanas de unión específica de secuencia. Esto estaría relacionado con el hecho que HMG1 presenta una mayor afinidad por ADN que HMG2. Por lo tanto, podría sugerirse que el motor o quien dirige la unión específica de secuencia sería HMG1 a través del reconocimiento del motivo AAC.

3. Implicación energética AAC y pre-curvatura inicial

Análisis estructurales y de mutagénesis de los complejos formados por dominios HMGbox (ya sea de unión específica como inespecífica) indican que una parte significativa del doblamiento del ADN es mediante la intercalación de cadena laterales de residuos específicos entre las bases del ADN, implicando un aumento de los valores de *roll* en estos puntos. Sin embargo, hay que destacar que la inserción o intercalación de residuos entre las bases es un proceso termodinámicamente desfavorable. Entonces, la cadena lateral puede situarse entre las bases debido a la acción acumulativa de varias interacciones proteína-ADN. Estas interacciones pueden ser electrostáticas o no-electrostáticas, dependiendo de la interfaz formada (Dragan, Read et al. 2004).

Además del papel de la proteína en la torsión del ADN, cabe destacar el papel de la secuencia nucleotídica en el caso de las HMGbox ss. Es ampliamente conocido que las propiedades elásticas del dúplex de ADN son fuertemente dependientes de la secuencia nucleotídica (Kahn, Yun et al. 1994). Por ejemplo, las secuencias que incluyen repeticiones en fase de *A-tracts* son sinónimo de un doblamiento intrínseco global del ADN (Haran and Mohanty 2009).

Los estudios de la flexibilidad intrínseca del ADN mediante el cálculo de sus características fisico-químicas permiten diferenciar regiones con mínimos energéticos. Uno de estos mínimos se encuentra en el motivo invertido 3'-AA'C-5', exactamente en el punto donde se da la intercalación de Leu182 para HMG2. Para Leu58, el mínimo energético se encuentra desplazado dos pares de base previos y la caena lateral intercala precisamente en un paso energéticamente favorable para su torsión. Por lo tanto, si tenemos en cuenta que los mínimos energéticos en el ADN están relacionados con puntos de flexibilidad o de cierta curvatura, puede ser que favorezca energéticamente la intercalación de las cadenas laterales, y ello explicaría porqué se mantiene el patrón de reconocimiento de TFAM en las dos estructuras descritas, TFAM-LSP22 y TFAM-LSP28. Además ello ayudaría a entender porqué es el primer motivo AAC tan imprescindible para la unión de TFAM, mediante HMG1, y su función de activador transcripcional

4. Región central (*Nxlinker*)

Como se comentó anteriormente (ver discusión, B.1.) en la región de unión específica de TFAM a LSP22, encontramos las dos regiones flanqueantes de unión a HMGbox (5-AA'C-10pb-C'AA-3) y entre estas dos regiones existe una región central, *Nxlinker*. Cuando se alinean las diferentes estructuras de LSP y las secuencias de unión específica de TFAM (Fig_B5) por los motivos AAC, se evidencia que esta zona central presenta una longitud conservada de 10pb. Además esta longitud de la región central se mantiene conservada entre las secuencias LSP y HSP1 para humano y ratón (Dairaghi, Shadel et al. 1995).

Esta conservación en la longitud de la extensión nucleotídica en la región central puede mostrar una cierta correlación con la región *linker* de la TFAM. Cuando se analiza la longitud del *linker* entre mamíferos y metazoos se observa una clara conservación en la longitud de este *linker* (~27aa). Por lo tanto, es posible sugerir que existe una cierta correlación entre el patrón de reconocimiento de AAC y la longitud del *linker*.

4.1. Conservació *cluster* de guaninas central

Otra característica sorprendente de esta región central es la presencia de un grupo de guaninas (en concreto 6 guaninas) en la secuencia LSP22. La importancia de este grupo de guaninas en la cadena no-codificante queda en relevancia cuando se realizan estudios comparativos de secuencia entre LSP de primates. Este grupo se encuentra altamente conservado entre LSP de hominoides; sin embargo, en ratón la secuencia entre los dos motivos es altamente rica en A-T (Fisher, Parisi et al. 1989) (Foran, Hixson et al. 1988) (Fig_A5). Además, cabe destacar que este grupo de guaninas no se encuentra conservado entre las diferentes secuencias de unión específica de TFAM en humano (LSP, HSP, X e Y), sugiriendo una funcionalidad de este grupo de guaninas en la activación de la transcripción en LSP (Fig_B5).

La presencia de grupos de guaninas en el ADN está relacionada con la posibilidad de formación de súper estructuras llamadas *Gquadruplex*. Se trata de estructuras muy estables debido a los enlaces de hidrógeno entre las guaninas que se encuentran formando un *stacking* planar. Es posible que la existencia de este grupo de guaninas pueda regular en cierta forma la afinidad de unión y función de TFAM dentro del complejo de iniciación de la transcripción en la zona promotora de la cadena ligera.

C. EL DOGMA DE LA COLA C-TERMINAL

1. Actividad en reconocimiento específico

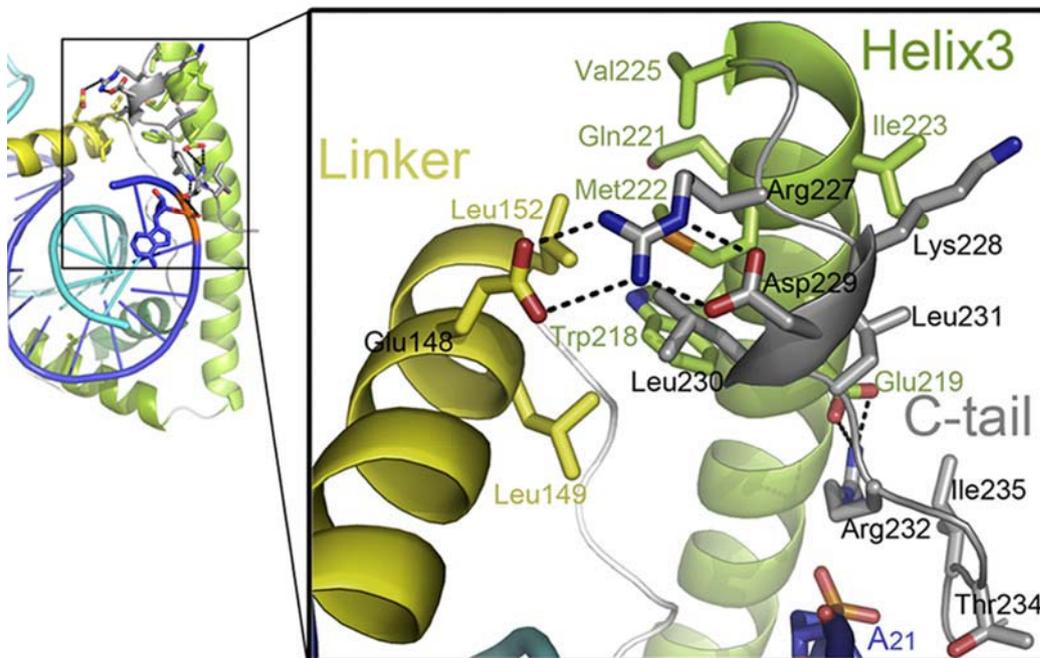
Como se ha comentado hasta el momento se han asociado muchas de las funciones a la cola C-terminal: dimerización de TFAM sobre el ADN, desenrollamiento de éste, interacción con otros factores de transcripción, etc. Pero una de las funciones que más se ha asociado a la cola C-terminal es la capacidad de unión específica con las secuencias de ADN y su actividad en el inicio de transcripción.

En mamíferos la cola C-terminal se encuentra altamente conservada hasta el residuo Lys236. En el cristal, desde la posición Arg232 hasta el último residuo trazado, Lys237, la cola C-terminal se sitúa por encima de los fosfatos del ADN sin realizar ningún contacto específico con las bases.

Experimentos de mutagénesis definieron las regiones críticas de la cola C-terminal para la interacción de TFAM de forma específica al ADN (Dairaghi, Shadel et al. 1995). Mientras que la delección desde el residuo Lys236 hasta el final no provocaba ninguna disminución de la afinidad de unión específica al ADN, la delección de residuos anteriores a Lys236 provocaba una disminución de la unión específica alterando la actividad de inicio de la transcripción. Además, la mutación puntual de Arg232 provocaba el mismo efecto que deleccionar los 15 residuos finales de la cola C-terminal (Gly226). Los autores concluían que los residuos de Gly226 a Lys194 y la carga positiva en la cola C-terminal eran importantes para la unión específica y la función en transcripción.

Ahora bien, analizando la estructura TFAM-LSP22 o LSP28 (ambas finalizan en Lys237) es posible observar que las regiones caracterizadas como imprescindibles para la unión específica (de Gly226 a Lys236) no realizan ningún contacto específico con las bases del ADN. Sin embargo, es posible definir la existencia de un núcleo hidrofóbico formado entre la parte final del linker, la hélice 3 de HMG2 y la cola C-terminal (L/HMG2/C) (Fig_C1), donde la cola C-terminal se encuentra empaquetada hacia la hélice 3 de HMG2 en dirección opuesta al sitio de inicio de la transcripción y realizando distintos contactos con linker y la hélice 3. Por lo tanto, la delección de los residuos de la cola C-terminal que forman este núcleo hidrofóbico provoca una desestabilización del núcleo hidrofóbico L/HMG2/C probablemente desestructurando HMG2. Ello implicaría una disminución de la capacidad de unión al ADN de forma específica y afectaría la capacidad de inducir curvatura en el ADN, por lo tanto, influyendo en la actividad de promotor de la transcripción.

Aún así, el papel de la cola C-terminal no sería tanto de interacción específica con LSP sino de estabilización de HMG2 junto con el linker, para permitir la unión de éste al ADN e inducir la curvatura final de 180° que permite la ubicación de la cola C-terminal cercana al inicio de transcripción para la interacción de esta con TFB2M. Ya en experimentos de interferencia por metilación se observaron tres regiones de contacto de TFAM con LSP (Fisher, Parisi et al. 1989), dos de ellas encajaban con las secuencias predichas de interacción por parte de HMGbox ss y la otra región se situaba en el grupo central de guaninas, región de contacto para el linker en la estructura. Por lo tanto, no existen experimentos bioquímicos que demuestren contacto específico de la cola C-terminal con el ADN.



Fig_C1. Representación del *core* hidrofóbico formado entre la parte final del linker (amarillo), la hélice 3 de HMG2 (en verde) y la cola C-terminal (en gris) (L/HMG2/C), donde con barras se encuentran representados los diferentes residuos que participan en el núcleo mediante diferente tipo de interacciones.

2. Cola C-terminal y flexibilidad intrínseca

El último residuo trazado, Lys237, se encuentra expuesto al solvente, y los residuos Gln238-Cys246 no pudieron ser trazados tanto en la estructura TFAM-LSP22 como en TFAM-LSP28. Posiblemente la incapacidad de trazar los últimos residuos (9 residuos finales) es debida a su flexibilidad intrínseca, característica demostrada mediante experimentos de SAXS para la proteína en su forma no unida al ADN. Además los experimentos de SAXS con TFAM-LSP22 mostraban un mejor ajuste cuando la cola C-terminal era flexible. La flexibilidad intrínseca sería compatible con la habilidad de este último segmento a interactuar con otras proteínas como TFB2M.

La longitud de la cola C-terminal se encuentra más o menos conservada en mamíferos (Fig.C2). Aunque los últimos residuos muestran una menor conservación, posiblemente reflejando una adaptación especie-específica de la molécula diana y la capacidad para modular la interacción con factores de transcripción de cada especie.

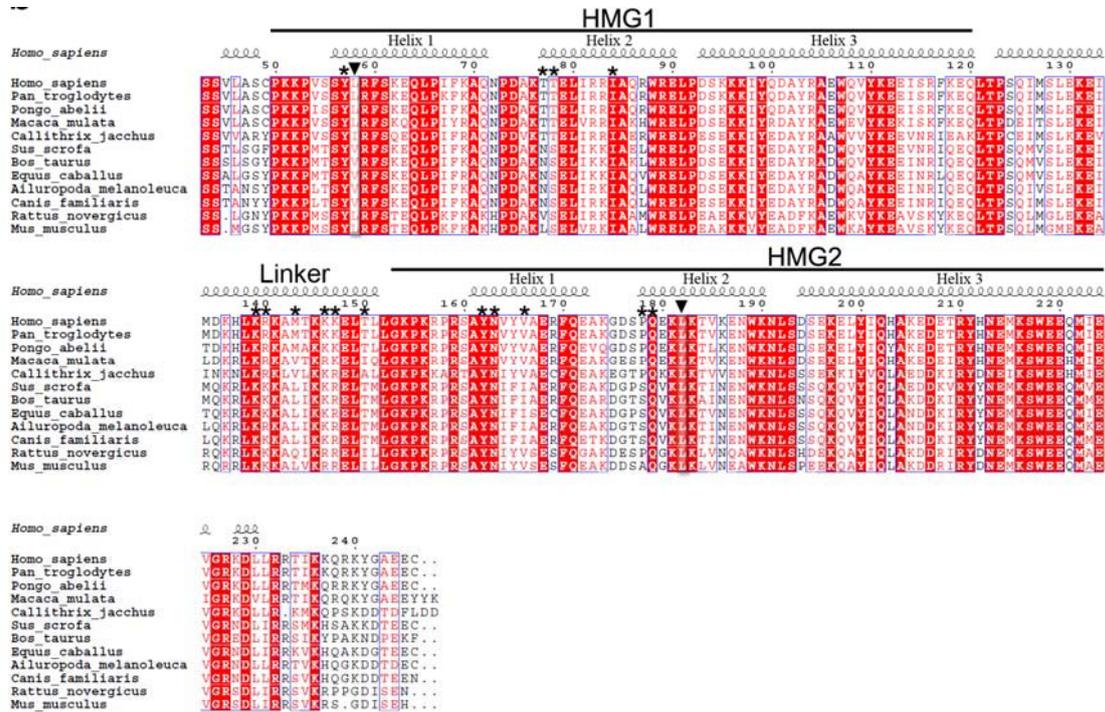
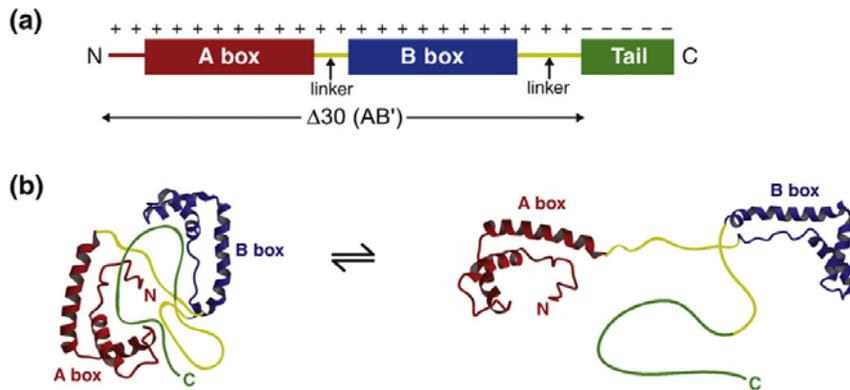


Fig.C2. Alineamiento de TFAM en diferentes mamíferos, donde es posible observar que la parte final de la cola C-terminal muestra un menor grado de conservación.

La flexibilidad intrínseca no es una característica desconocida o nueva para las proteínas de la familia HMGB. Por ejemplo en el caso de HMGB1, la cola C-terminal ácida de 30 residuos presenta una flexibilidad intrínseca que permite regular la interacción de HMGB1 con el ADN, ya que cuando existe en su forma compacta (alojada entre los dos dominios HMGBbox como un sandwich) no es posible la unión con el ADN (Stott, Watson et al.)(Fig.C3). Además mediante estudios de dicroísmo circular e ITC (*Isothermal Titration Calorimetry*) se observó que incluso a bajas temperaturas los dominios HMGBbox no están completamente plegados, concretamente los de unión específica. Parte de este estado de *unfolding* se adjudica a las colas C-terminales básicas que se encuentran desplegadas en la forma no unida a ADN y constituyen un 22-17% de estos dominios. En el caso de LEF86,

cuya cola C-terminal de 19 residuos (representa el 22% de la proteína) aparece en forma extendida a lo largo del surco mayor del ADN, donde realiza una interacción con el ADN puramente electrostática (Dragan, Read et al. 2004).

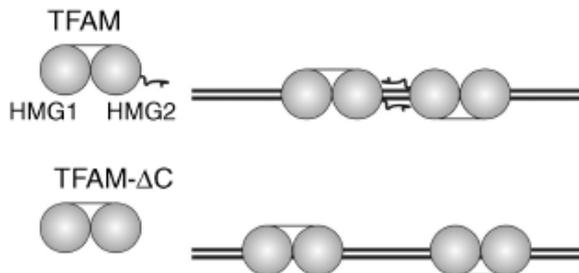


Fig_C3. Representación de los diferentes dominios de HMGB1. (A) Dominios de HMGB1 y su carga. (B) Representación esquemática del equilibrio existente entre las dos formas de HMGB1, donde la cola C-terminal participa activamente.

Por lo tanto, para el caso de TFAM podríamos entender un comportamiento flexible para la cola C-terminal favoreciendo así la posibilidad de interacción proteína-proteína, interacción con ADN o modificaciones pos-traduccionales que regularían la función de la proteína en determinados sitios y necesidades celulares.

3. Cola C-terminal y cooperatividad

El efecto cooperativo positivo que existe en TFAM se veía reducido después de eliminar la cola C-terminal en ensayos de anisotropía de fluorescencia (Wong, Rajagopalan et al. 2009). Wong, *et al.* postularon que la dimerización de TFAM sobre el ADN ocurre a través de la cola C-terminal mediante procesos de cooperatividad positiva. Ensayos descritos anteriormente por Kaufman, *et al.* proponían el siguiente mecanismo de dimerización (Fig_C4), donde la unión de la primera molécula de TFAM sobre ADN permite la interacción de C-terminal con la cola C-terminal de la segunda molécula de TFAM, donde la dimerización sobre ADN presentaba un carácter funcional en la activación de la transcripción.



Fig_C4. Representación esquemática de los modelos de unión de TFAM y TFAM- ΔC a ADN y su dimerización sobre el ADN. (Wong *et al.*)

Pero este mecanismo de dimerización no concuerda con los experimentos de protección previos sobre la zona LSP, donde TFAM cubría una región de 23 pb. Las estructuras

cristalográficas para TFAM-LSP22 y TFAM-LSP28 muestran como TFAM una de forma monomérica a longitudes de ADN de 22 y 28pb. Por lo tanto, ello dificulta el modelo de dimerización propuesto por Wong, *et al.*

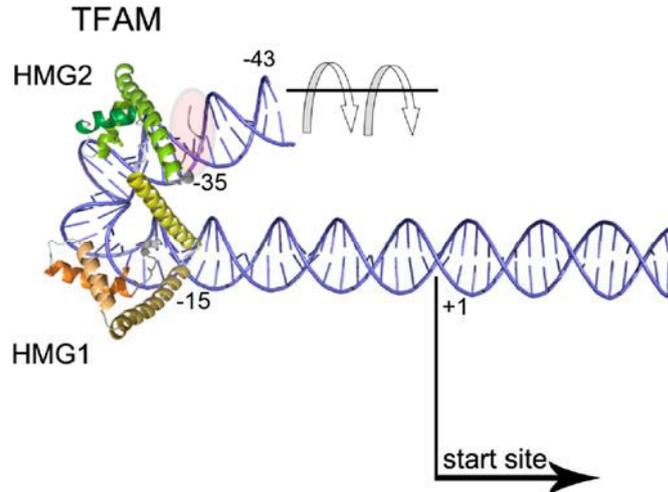
Además el cálculo de las superficies de interacción muestra que las posibles superficies de contacto se encuentran en los dominios HMGbox y en una pequeña región del linker. La interacción proteína-proteína a través del linker entre las dos moléculas de la unidad asimétrica implica la interacción de residuos alojados en linker y la cola C-terminal. Por lo tanto, la delección de la cola C-terminal en el residuo Glu219 o Gln221 podría provocar una desestabilización del núcleo hidrofóbico L/HMG2/C, afectando la posible superficie de contacto proteína-proteína. Por lo tanto, la reducción del efecto cooperativo al deleccionar la cola C-terminal se encontraría más vinculada a la desestructuración del núcleo hidrofóbico y la dificultad para ofrecer una superficie estable de contacto.

Finalmente, destacar que ha sido descrito que la cola C-terminal de TFAM interacciona con el factor TFB2M en el PIC (complejo pre-iniciación de la transcripción). Por lo tanto, este tipo de dimerización seguramente no sería funcional para la activación de la transcripción.

4. Orientación opuesta al inicio de transcripción

Un detalle importante para la interpretación de la estructura en el contexto del inicio de la transcripción en LSP es que la cola C-terminal yace en el ADN en el sentido opuesto al sitio de inicio de la transcripción. Sin embargo, debido a la curvatura en forma de U inducida por TFAM, la cola C-terminal acaba situándose próxima a la secuencia de ADN adyacente al sitio de inicio de la transcripción (Fig_C5). Por lo tanto, gracias a la curvatura del ADN es posible la localización de la cola C-terminal cercana al sitio de inicio de transcripción para LSP, y así interactuar con los factores de transcripción (como TFB2M) y promover la activación de la transcripción.

LSP y HSP1 son funcionalmente diferentes (ver Introducción, apartado B.1.) y presentan un lugar de unión para TFAM en orientación opuesta, por lo que en el caso de HSP1 la unión de TFAM a su secuencia implica la localización próxima de TFAM al sitio de inicio de transcripción, sin necesidad de una curvatura. De hecho estudios de curvatura de TFAM en HSP1 mediante FRET demostraron que la perturbación estructural del ADN es menor en HSP1 (Ngo, Kaiser et al.).



Fig_C5. Representación de la estructura cristalográfica de TFAM-LSP en el entorno de la región promotora LSP. TFAM se une a la secuencia -35 a -15 pb, donde la cola C-terminal yace en dirección opuesta al sitio de inicio de la transcripción (*start site*), indicada con una sombra rojiza. La curvatura del ADN permite localizar la cola C-terminal cercana al sitio de inicio de la transcripción y, por lo tanto, cercana a la maquinaria de inicio de transcripción.

D. IMPORTANCIA DEL LINKER

1. La zona del linker en la torsión del ADN

Durante la realización de los ensayos de permutación circular se obtuvo un ángulo de curvatura de 68° para TFAM en complejo con el ADN, mientras que el dominio D1 presentaba una unión específica por el ADN, pero una curvatura no significativa; y D2C no mostraba ni una curvatura ni unión al ADN. Ello sugirió que la capacidad de torsión del ADN por parte de TFAM no está sólo gobernada por los dominios HMGbox, sino que es necesaria una acción cooperativa de los demás segmentos.

Posteriormente los ensayos de spFRET demostraron el papel del linker en la perturbación global del ADN. El dominio H1 (HMG1+linker) presentaba un mayor valor para E_{FRET} asociada a una menor distancia entre fluoróforos y, por lo tanto, una mayor curvatura del ADN. En concreto D1 provoca una curvatura de 90° ($R \sim 85 \text{ \AA}$), frente a 113° para H1 ($R \sim 78 \text{ \AA}$). El papel de linker en la curvatura final del ADN también fue caracterizada por Ngo, *et al.* que mediante el uso de un mutante de linker para seis residuos cargados positivamente substituidos por alanina, obtuvieron una disminución de $> 30\%$ del fenómeno FRET (Ngo, Kaiser *et al.*).

2. Papel del linker en la unión al ADN y función en la activación de la transcripción

Los experimentos de SAXS para TFAM en su forma no unida a ADN en solución constatan una enorme flexibilidad para la zona del linker con una conformación extendida, mientras que en la estructura cristalográfica TFAM-LSP22 y TFAM-LSP28, el linker presenta una conformación helicoidal donde estabiliza la torsión inducida en el ADN por parte de los

dos dominios HMGbox pero sin realizar contactos específicos con las bases del ADN. Por lo tanto, de un estado a otro es necesaria una reorganización de linker.

Además los dominios D1 y H1 presentan diferente afinidad por LSP, donde el dominio H1 presenta una mayor afinidad por ADN que D1. Ello sugiere la importancia del linker no tanto en la curvatura del ADN, sino también en la afinidad al ADN. De hecho estudios previos de anisotropía de fluorescencia, técnica usada para detectar la interacción de macromoléculas en solución, demostraban la contribución del linker en la unión del ADN (Wong, Rajagopalan et al. 2009). En estos estudios la débil afinidad de HMG2 por el ADN, se veía claramente estimulada mediante la adición de un determinado número de residuos del linker (que incluían Lys146 y Lys147).

Cabe destacar que este linker se encuentra ampliamente conservado sugiriendo su implicación funcional. Estudios de actividad del homólogo de TFAM en levadura, ABF2p, permitieron observar que cuando se añadía el linker de TFAM humano, se confería capacidad para realizar unión específica y activación de la transcripción (Dairaghi, Shadel et al. 1995).

Por lo tanto, la importancia del linker en la unión y torsión del ADN presentan una conexión funcional con el papel de TFAM en el inicio de la transcripción mitocondrial, revelando así un papel de capital importancia para el linker de TFAM y desplazando la atención de la cola C-terminal al linker.

3. El linker participa en el núcleo hidrofóbico que estabiliza HMG2

Tal y como se ha descrito en el apartado de discusión para la cola C-terminal, la estructura cristalográfica de TFAM en complejo con LSP permite dilucidar la existencia de un núcleo hidrofóbico entre el extremo C-terminal del linker, la hélice 3 de HMG2 y la cola C-terminal. Además, estudios de anisotropía de fluorescencia (Wong, Rajagopalan et al. 2009) demostraron que la capacidad de unión del dominio HMGbox2 incrementaba cuando se añadía la zona del linker, aunque la afinidad de unión era 10 veces menor que para HMGbox1.

Este núcleo hidrofóbico permite la estabilización y estructuración correcta de HMG2, y la unión de HMG2 se encuentra modulada por la región final del linker y la parte inicial de la cola C-terminal.

4. Reorientación de los dominios por parte del linker

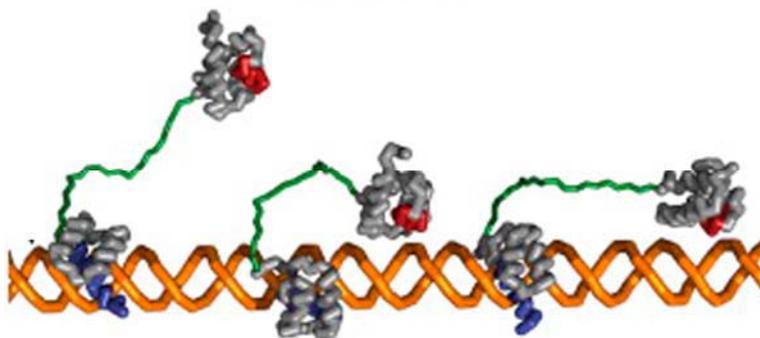
La comparación de las dos estructuras TFAM-LSP22 y TFAM-LSP28 permite ratificar la curvatura en forma de U con 180° que promueve la unión de TFAM al ADN. Aunque cabe destacar que el hecho más sorprendente es que una estructura respecto a la otra muestra un ligero desplazamiento del linker en la región N-terminal, que provoca una reorientación ligeramente diferente para HMG2. Ello sugiere el posible papel del linker en la orientación respectiva de los dominios HMGbox.

Los experimentos de SAXS también sugieren una función reorientativa de los dominios HMGbox para el linker. TFAM en su forma no unida muestra una gran flexibilidad para el linker, que da lugar a una gran variedad de conformaciones en el espacio y las diferentes

posiciones relativas que adoptan HMG1 y HMG2 debido a la flexibilidad del linker. Mientras que en el caso de TFAM en complejo con ADN, la curva experimental de TFAM-LSP22 no acaba de ajustar con la curva teórica de la estructura cristalográfica TFAM-LSP22 o TFAM-LSP28. Por lo tanto, es probable que en solución existan sutiles cambios conformacionales que impliquen diferentes reorganizaciones entre HMG1 y HMG2 mediante ligeros desplazamientos del linker en su región N-terminal.

Este posible papel del linker no sería un hecho novedoso en proteínas multidominio de unión al ADN con presencia de dominios de unión no específica de secuencia (*ns*) y dominios de unión específica de secuencia (*ss*). La presencia de un linker flexible entre los dos dominios con diferente afinidad para ADN permite la comunicación entre los dos dominios y la capacidad a la proteína para realizar una búsqueda óptima de la secuencia de ADN mediante deslizamiento a lo largo del esqueleto de fosfatos del ADN (mecanismo conocido como *sliding*). Este mecanismo es frecuente en factores de transcripción de eucariotas como Oct-1, perteneciente a la familia POU con dos dominios que se estructuran de forma independiente e interaccionan con diferentes afinidades directamente con ADN, y ambos son necesarios para la unión efectiva al ADN (Vuzman, Azia et al.) (Fig_D1). En el caso de TFAM nos encontramos con dos dominios con características fisicoquímicas (cómo el pI, longitud de hélices), afinidades por ADN y patrón de reconocimiento del ADN muy distintos. Para HMG2 ha sido ampliamente descrito que se trata de un dominio más inestable, con una temperatura de fusión inferior a HMG1 (Gangelhoff, Mungalachetty et al. 2009), así como una menor capacidad de unión a ADN (Wong, Rajagopalan et al. 2009).

Todo ello sugiere un mecanismo en que el linker flexible de TFAM permite la cooperatividad de los dos dominios (HMG1, HMG2) y permite acelerar la búsqueda del sitio diana de unión en el ADN. Donde primero existe la unión de HMG1 (con mayor afinidad a ADN) anclándose de forma relativa al ADN. En esta situación HMG2 alternaría sutiles conformaciones con diferentes posiciones relativas en el espacio debido a la flexibilidad del linker, hasta establecer una unión al ADN estable debido a un correcto plegamiento de HMG2 por parte de la formación del núcleo hidrofóbico donde el linker se reorganiza y ofrece su superficie para la interacción con HMG2. La unión final de los dos dominios presenta una constante de afinidad menor, tratándose de una unión mucho más estable (Ngo, Kaiser et al.).



Fig_D1. Modelo de representación del sistema de comunicación o cooperatividad en una proteína de unión al ADN multidominio (Oct-1) con un dominio con mayor afinidad por ADN (en azul) y de menor afinidad (en rojo), y la presencia de un linker flexible (en verde) (Vuzman, et al.).

5. Longitud del linker y correlación con la secuencia de unión de TFAM

El análisis de la conservación del linker de TFAM deja entrever que se trata de un linker ampliamente conservado en cuanto a su carga positiva (con un gran número de Lys y Arg), así como en su longitud (~27 aa). Además cuando se alinean las secuencia de unión específica de TFAM en humanos: LSP, HSP1, X e Y, es posible observar que el número de nucleótidos entre las repeticiones de trinucleótidos invertidas definidas (5'-AA'C-Nxlinker-C'AA-3') es sistemáticamente 10pb (ver Discusión, B.2.3), es decir, topológicamente significaría una vuelta de la hélice de ADN. Lo mismo sucede cuando comparamos las secuencias LSP y HSP1 de humano y ratón (Dairaghi_1995). Ello sugiere que el modo de unión de TFAM a estas secuencias podría ser topológicamente similar, existiendo una relación entre el patrón de reconocimiento y la longitud de linker.

Por lo tanto, parece ser que existe una correlación entre longitud del linker, organización de los dominios y patrón secuencial de unión al ADN, y torsión final del ADN o efecto topológico sobre ADN.

En la literatura encontramos un ejemplo de correlación entre patrón de reconocimiento y longitud del ADN para el caso de la proteína híbrida SRY.B (ver resultados) (Stott, Tang et al. 2006). La curvatura secuencial en el complejo SRY.B/DNA es función de la posición relativa entre los dos dominios, y concretamente de la separación entre los residuos intercalantes. El residuo isoleucina en SRY y fenilalanina en el dominio B, se encuentran separados por 9 pb, número muy cercano a la periodicidad en un B ADN, con lo que los puntos de curvatura se encuentran en fase. Observaciones similares se han descrito para los monómeros de UBF en el *enhacesome*, donde se cree que el espació de 20pb entre los dominios es necesario para mantener la curvatura secuencial del ADN (Stefanovsky, Pelletier et al. 2001). Parece ser que tanto para HMGB1 como en UBF, la longitud del linker (región entre dominios HMGBbox) es crítica para la posición relativa de los dominios y, así asegurar la curvatura secuencial.

Ello sugiere que en el caso de TFAM existe un linker de longitud 27 aa que permite recolocar o alojar los dos dominios HMGBbox en el ADN dando lugar a una curvatura del ADN. Por lo tanto, podría ser el linker quien determinase el patrón secuencial de unión de TFAM, definido con una longitud entre 22 pb y 30 pb según la literatura, permitiendo así el cambio topológico en el ADN traducido en forma de curvatura del ADN.

A continuación se intentará relacionar las funciones bioquímicas conocidas para TFAM con los resultados obtenidos durante la realización de esta tesis doctoral. En concreto, nos centraremos en dos funciones: el papel en la activación de la transcripción en LSP y la implicación de TFAM en el empaquetamiento del mtADN.

E. MODELO DE INICIO DE LA TRANSCRIPCIÓN EN LSP

A partir de los resultados obtenidos para el complejo TFAM-LSP es posible plantear el siguiente modelo de unión y actuación de TFAM en el complejo de pre-iniciación de la transcripción (PIC) formado por TFAM, TFB2M, mtRNAP (Fig_E1).

La unión de TFAM a LSP22 puede plantearse como un modelo de unión progresivo, ya que mediante spFRET y SAXS es posible observar un modelo dinámico de unión, a diferencia de los sistemas *lock-key* descritos para algunos casos.

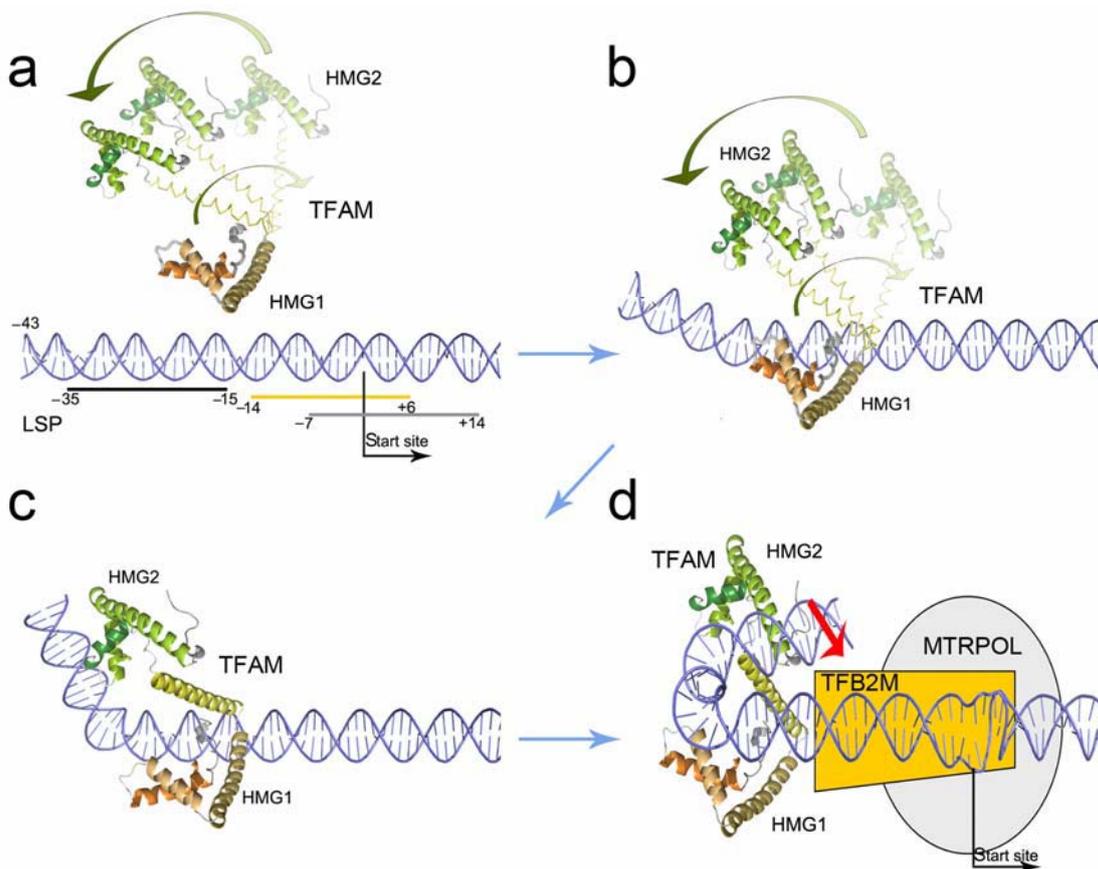
En solución encontramos los dominios HMGbox de TFAM bien plegados de forma independiente y con diferentes posiciones relativas uno respecto al otro, debido a la gran flexibilidad que presenta el linker (Fig_E1 (a)). El dominio HMG1 con mayor afinidad por ADN y un patrón de reconocimiento del ADN típico de HMGbox de unión específica de secuencia, reconoce la región 1 de LSP22 (-15 a -20 pb en LSP) y promueve un primer punto de curvatura mediante la interacción de diversos residuos con las bases del ADN, en particular el residuo Leu58 que se intercala entre el motivo 5-AA'C-3 (Fig_E1 (b)). La interacción de HMG1 con ADN provoca una torsión del ADN de $\sim 90^\circ$, como muestran los estudios de FRET y la estructura cristalográfica TFAM-LSP22. La unión del primer dominio HMG1 permite la reorganización del linker que pasa de una conformación extendida a una conformación helicoidal manteniendo un mayor contacto con el ADN que incrementa el ángulo de torsión del ADN ($\sim 113^\circ$) (Fig_E1 (c)). La reorganización del linker junto con el incremento de torsión del ADN favorece la estabilización de HMG2, mediante la formación del núcleo hidrofóbico L/HMG2/C, y su unión con el ADN en la región 2. Por lo tanto, la curvatura del ADN creada por HMG1 y linker permite la organización de HMG2 de tal forma que existe la intercalación del residuo Leu182 en el motivo 5-C'AA-3, favorecida por los demás residuos que interaccionan con las bases del ADN en la región 2 LSP22 (-28 a -35 pb en LSP). La unión de este segundo dominio induce un nuevo punto de curvatura de 90° en el ADN, imponiéndose una torsión global en forma de U de 180° (Fig_E1 (b)).

La curvatura de 180° en LSP permite que la cola C-terminal, que emerge en sentido opuesto al sitio de inicio de transcripción, se sitúe cercana a este punto. Tal y como se había descrito anteriormente, la cola C-terminal de TFAM interacciona de forma directa con TFB2M. Por lo tanto, la localización de la cola C-terminal después de la curvatura permite su función de reclutamiento de TFB2M para llevar a cabo la apertura y estabilización del substrato y producto de la transcripción que ayuda a la acción de mtRNAP. Sería un caso similar a lo que sucede en el caso de TBP-TATAbox-TFII-RNAPol.

La transcripción mitocondrial por parte de mtRNAP implica, como en otros sistemas de transcripción (levadura, procariotas, eucariotas), la formación de una burbuja en la región de inicio de transcripción rica en AT de 12-14 pb mediante el desenrollamiento del ADN. Este proceso es conocido como formación del *open complex*. La apertura de esta región y la

estabilización del *open complex* depende del estado de torsión del ADN (Revyakin, Ebright et al. 2004). Cabe imaginar el ADN como un sistema donde existe un flujo de tensión con olas de torsión a lo largo del ADN, donde la introducción de perturbaciones estructurales en el ADN, en forma de puntos de curvatura, puede actuar como barrera para esta difusión de tensiones (conocido como “twist barrier”) (Roca). Esta barrera de difusión provoca un desenrollamiento local sucesivo a su bloqueo. Por lo tanto, el hecho que TFAM presente su lugar de unión a una distancia tan corta del sitio definido para *open complex* en LSP y que provoque una curvatura extrema en ADN, puede actuar como *twist barrier* facilitando la apertura en la región rica en AT y la estabilización de la burbuja de transcripción. Además, ha sido descrito que la distancia entre sitio de unión de TFAM y sitio de inicio de transcripción está muy conservada en longitud, y es imprescindible para el inicio de la transcripción.

Ello sugiere que TFAM no actuaría solo como factor de reclutamiento de TFB2M, sino también como elemento estructural importante para la promoción del *open complex* en el PIC.



Fig_E1. Mecanismo de unión propuesto para TFAM en LSP y su participación en la transcripción. (a) Situación donde TFAM se encuentra en solución libre, en el ADN se encuentran las regiones de unión para TFAM en negro, para TFB2M en naranja y en gris para mTRNAP. (b) Unión de HMG1 al ADN y primer punto de curvatura. (c) Reorganización del linker debido a la unión de HMG1, que permite la interacción de HMG2 con su secuencia de unión. (d) la unión de HMG2 da lugar a una curvatura final de 180°, donde la cola C-terminal puede reclutar TFB2M para la formación de PIC (flecha roja) y existe la burbuja del *open complex* que permite la función de mTRNAP. Además, la curvatura de TFAM facilita la estabilización del *open complex*, mediante la formación de un desenrollamiento local después de su unión.

F. MODELO DE EMPAQUETAMIENTO DEL mtADN

A parte de participar en la activación de la transcripción mitocondrial, TFAM es una de las proteínas mayoritarias involucradas en el empaquetamiento del mtADN y su organización. Estudios previos a la resolución de la estructura de TFAM, habían propuesto un mecanismo de organización del mtADN similar al mecanismo de compactación propuesto para ABF2p, homólogo de TFAM en levadura. Este mecanismo consiste en la introducción de puntos de curvatura en el ADN que promueven la formación de lazos (ver Introducción). En este apartado se tratará de recoger las diferentes características de unión de TFAM para acabar ofreciendo un modelo de empaquetamiento y organización del mtADN.

La estructura cristalográfica de TFAM en complejo con ADN revela como los dos dominios HMGbox inciden dos puntos de **curvatura** dando lugar a una torsión global del ADN de 180°. En este complejo la proteína se entrelazan con el ADN, es decir, rodeando el ADN; a diferencia de lo que ocurre en la organización del ADN nuclear eucariota por parte de las histonas, donde es el ADN quien envuelve los nucleosomas.

Estudios recientes de *single molecule* (Farge, Laurens et al.), han demostrado que el monómero de TFAM se desliza rápidamente alrededor del mtADN, fenómeno conocido como **sliding**. Mediante experimentos de difusión observaron que TFAM se desliza a lo largo del esqueleto fosfato del ADN hasta que agrega en un cierto punto del ADN y existe la formación de un clúster mediante la interacción cooperativa entre moléculas de TFAM. Este fenómeno de *sliding* está normalmente asociado a proteínas bidominio con diferentes afinidades por el ADN y un linker flexible (Vuzman, Polonsky et al.), características que coinciden con los resultados hallados para TFAM.

Cuando comparamos el ángulo de torsión creado por TFAM mediante diferentes técnicas es posible observar que existe cierta discrepancia. Mientras que el modelo cristalográfico define un sistema rígido, los experimentos de SAXS y spFRET permiten incluir una cierta **dinámica en el complejo**. Cabe destacar que en el caso de spFRET, la población correspondiente con un ángulo de $\sim 170^\circ$ presenta un perfil con una gran anchura, indicando que existe una gran heterogeneidad dentro de la población. Por lo tanto, en solución la curvatura del ADN por parte de TFAM no es un sistema rígido, sino que existe cierta flexibilidad en el ADN.

Esta flexibilidad en el ADN se puede relacionar con los experimentos con pinzas ópticas y TFAM (Farge, Laurens et al.). Farge, *et al.* evidenciaron que el sistema de compactación del ADN por parte de TFAM no es mediante la introducción de loops, sino debido al **incremento de la flexibilidad en el ADN**. Definían que el incremento en la flexibilidad del ADN era debido a un desenrollamiento de 2-3 pb en los extremos del ADN. La estructura cristalográfica ya mostraba un ligero desenrollamiento en la secuencia de unión (~ 1 pb por cada 22 pb).

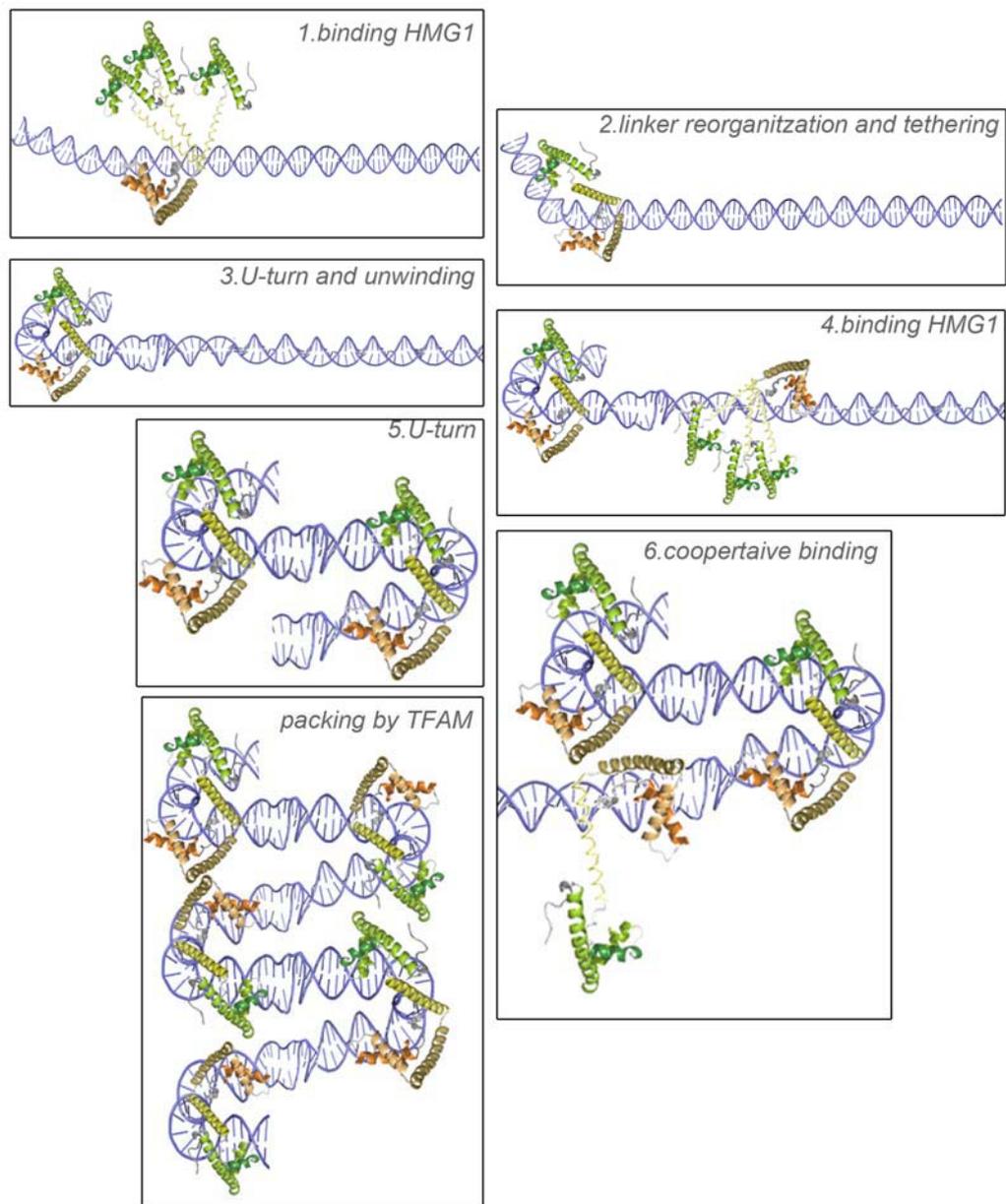
TFAM une de forma **monomérica** cuando se encuentra formando complejos con ADN de longitudes de 22 pb, según la estructura cristalográfica. De hecho mediante experimentos de *footprint single-molecule* se ha definido que la región de unión de tipo no específico de

secuencia (*nss*) para TFAM es $\sim 30 \pm 3$ pb de longitud, es decir, existe una unión de TFAM en fase cada 30 pb. En concordancia con estudios de cuantificación de TFAM por nucleoides, que proponían 1 TFAM: 16 pb – 25 pb (Kukat, Wurm et al.) (Kaufman, Durisic et al. 2007).

TFAM presenta un tipo de **unión cooperativa** ampliamente demostrada (Kaufman, Durisic et al. 2007; Wong, Rajagopalan et al. 2009) y presente en los geles nativos de retardo (EMSA). Mediante el análisis del empaquetamiento cristalino se observó que el dominio HMG1, en concreto su Hélice 3, puede formar una interfaz estable para la unión con otra Hélice 3 del dominio HMG1 de la estructura simétrica. Este tipo de empaquetamiento se daba en ambas estructuras resueltas para TFAM (TFAM-LSP22, TFAM-LSP28). Además, el análisis de posibles superficies de interacción también resaltaba la posible formación de ensamblajes mediante la **interacción proteína-proteína por parte de la Hélice 3 de HMG1**.

La suma de estos resultados permite realizar un modelo de empaquetamiento de mtADN por parte de TFAM (Fig_F2). Imaginemos un sistema donde TFAM se encuentra deslizándose a lo largo del mtADN y el linker flexible permite la comunicación entre los dominios HMGbox con diferente afinidad para el ADN. El bloqueo debido a una torsión inesperada del ADN o por parte de la unión de un macrocomplejo al ADN, provoca una disminución del deslizamiento de TFAM y su unión al ADN. La unión al ADN de TFAM provoca la adquisición de una curvatura en el ADN, pero con presencia de cierto dinamismo en la unión donde ligeros desplazamientos del linker pueden provocar una diferente reorganización para HMG1 respecto HMG2. Además esta unión de TFAM provoca un desenrollamiento de las regiones alojadas en los extremos de su unión, ofreciendo una mayor flexibilidad del ADN. Esta mayor flexibilidad en el ADN y la reorganización de HMG1 favorece la interacción proteína-proteína, posiblemente llevada a cabo por la disposición antiparalela de la hélice 3 de HMG1.

Con ello, finalmente se observaría la unión de TFAM en fase cada 30 pb promovida por la presencia de interacciones proteína-proteína, con la capacidad de inducir curvaturas y desenrollamiento en el ADN. Esta unión en fase permitiría la compactación del mtADN con presencia de cierta flexibilidad. La flexibilidad en el ADN permite resistir a diferentes tensiones existente en el ADN y adaptarse a las necesidades de la mitocondria con un ADN que cabalga entre la replicación, transcripción, empaquetamiento, daño por radiación, etc. es decir, permitiendo un buen mantenimiento del mtADN.



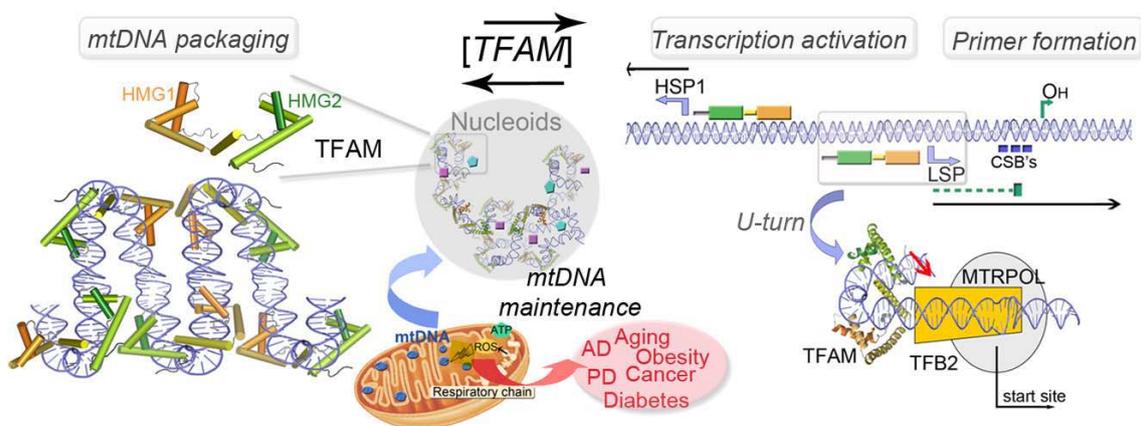
Fig_F2. Modelo de empaquetamiento para TFAM secuencial. (1) TFAM difunde por el mtADN. El dominio HMG1 con mayor afinidad por ADN, se une y provoca una ligera distorsión del ADN. (2) La presencia de un linker flexible permite acompañar los movimientos de HMG2 y permitir la unión de éste al ADN, fenómeno conocido como “*tethering*”. (3) La unión de HMG2 permite la formación de un “*U-turn*” en el ADN, es decir, provoca una curvatura en el ADN en forma de U. Esta curvatura no se puede entender como estado rígido, sino que existe cierta dinámica o flexibilidad en el ADN. Además la introducción de esta torsión en el ADN provoca un desenrollamiento de las bases situadas en los extremos de la región de unión de TFAM. (4) La unión de TFAM favorece la unión de otras proteínas, siguiendo el mismo modelo de unión (5) con introducción de curvaturas en el ADN. (6) La suma de incremento en la flexibilidad de TFAM junto con la capacidad de unión cooperativa, y el hecho que TFAM une en fase cada 30 pb; permite pensar que existe un empaquetamiento del mtADN favorecido por interacciones proteína-proteína en el ADN.

G. TFAM Y BIOGÉNESIS MITOCONDRIAL: zoom-out

La biogénesis mitocondrial es un proceso dinámico y regulado de forma muy precisa. Actualmente se ha sugerido el papel de la concentración de TFAM en el estado de actividad del nucleoide y, en consecuencia, de la mitocondria (Furukawa_Yamada_2012_FEBS). La resolución de la estructura de TFAM en complejo con ADN, y su caracterización bioquímica y biofísica permiten entender mejor el papel de TFAM en la biogénesis mitocondrial (Fig_G1).

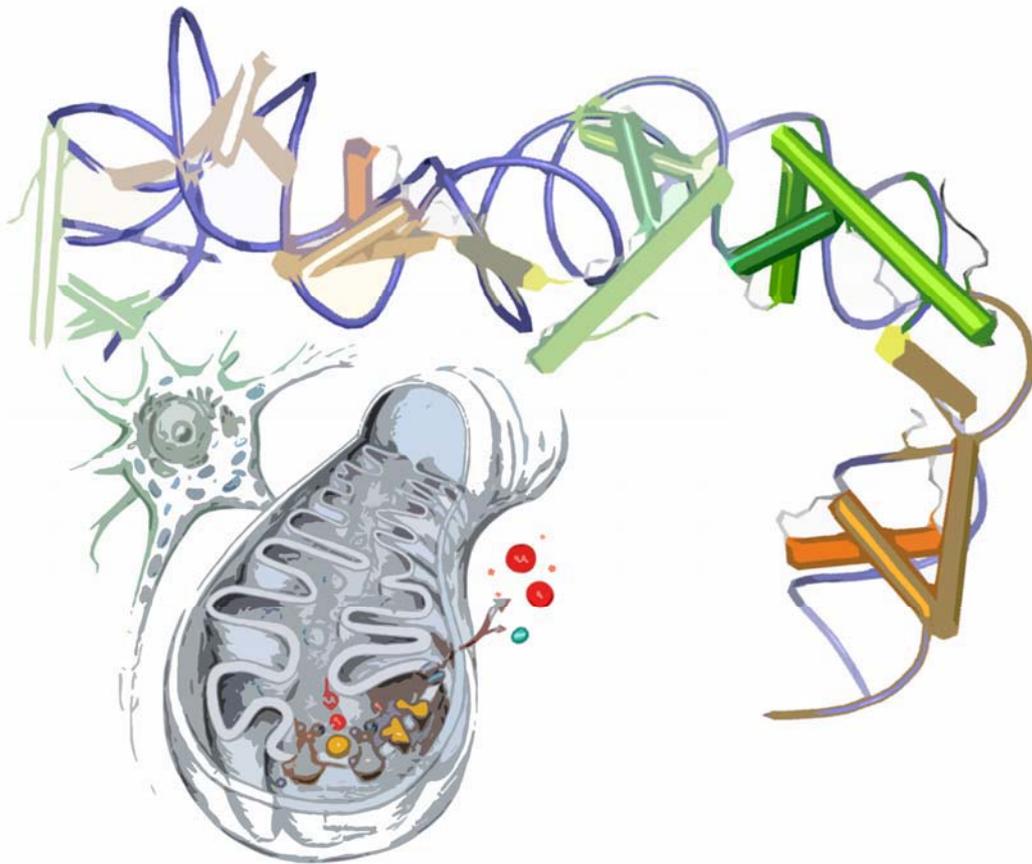
Concentraciones de TFAM óptimas permiten compactación del mtADN en nucleoides mediante su unión con el ADN formando complejos dinámicos donde TFAM se encuentra abrazando el ADN e introduciendo torsiones con ángulos $\sim 180^\circ$, que permiten protegerlo del efecto de ROS. Además, existe una cierta flexibilidad interna en el mtADN que permite su transcripción y replicación. Se ha descrito que en función de la concentración de TFAM existe un desplazamiento del equilibrio transcripción:replicación debido a la diferente afinidad de TFAM por los promotores LSP y HSP1. El trabajo realizado durante la tesis doctoral ha permitido entender el mecanismo de activación de la transcripción realizado por TFAM en la región promotora de LSP, que permite la formación de encebadores para la replicación a partir de O_H y, por lo tanto, está directamente relacionado con el aumento de copias de mtADN.

Concentraciones de TFAM altas implicarían un mayor grado de compactación, inhibiendo cualquier actividad en el ADN y dando lugar a nucleoides en estado semi-senescente.



Fig_G1. Esquema ilustrativo de la implicación de TFAM en la biogénesis de mitocondria. La mitocondria presenta su ADN compactado en estructuras llamadas nucleoides donde la proteína mayoritaria es TFAM, aunque también se encuentran otras proteínas involucradas (en la figura, indicadas con rectángulos, rombos y cuadrados). La compactación en nucleoides del mtADN permite protegerlo del daño por especímenes ROS y su rápida transferencia en procesos de fusión y fisión. La cuantificación de TFAM en nucleoides depende en función del estado de actividad del nucleoide, donde a concentraciones de TFAM inferiores (situación del panel derecho) existe una activación de la transcripción, mientras que en situaciones donde la concentración de TFAM es mayor (situación para el panel izquierdo) existe un desplazamiento del sistema hacia una mayor compactación de mtADN.

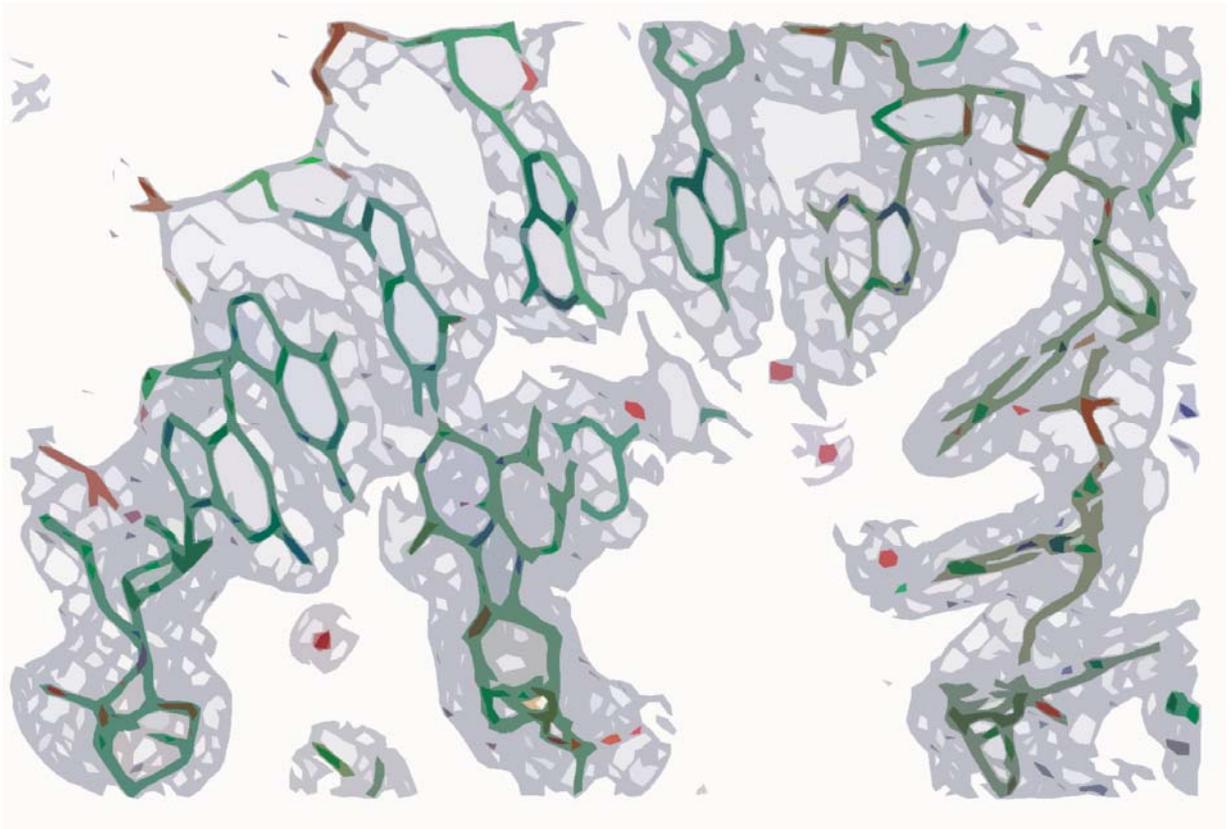
CONCLUSIONES



- TFAM impone una torsión estructural en LSP mediante la introducción de dos puntos de curvatura por parte de los dominios HMGbox que dan lugar a una curvatura en forma de U, observada mediante la estructura cristalográfica y ensayos de spFRET.
- Existe una dinámica en solución del complejo TFAM-LSP22 observada mediante SAXS y spFRET. Los resultados sugieren un sistema heterogéneo con múltiples conformaciones que presentan movimientos moleculares sutiles.
- Los dominios HMGbox de TFAM presentan el mecanismo típico de unión de HMGbox mediante su unión al surco menor del ADN provocando una distorsión local del ADN. La curvatura del ADN implica la introducción de residuos hidrofóbicos que se intercalan entre las bases del ADN (Leu58 para HMG1 y Leu182 para HMG2).
- Los dominios HMGbox de TFAM muestran un patrón de interacción y afinidad a ADN diferente, influyendo en la curvatura final. HMG1 presenta un patrón de interacción específico de secuencia y una mayor afinidad por ADN. La promoción de la curvatura por parte de HMG2 es un proceso cooperativo donde es el efecto conjunto de todos los residuos el que permite obtener la curvatura final.
- La unión de TFAM a LSP22 permite definir un motivo invertido relativamente conservado: 5'AA'C-(10pb)-C'AA-3', ' indica los puntos de intercalación donde existe una variación en los parámetros físico-químicos del ADN. Estos motivos presentan una relación con las secuencias consenso de HMGbox de unión específica de secuencia.
- El punto de intercalación en el ADN definido para Leu182 presenta cierta flexibilidad intrínseca del ADN, favoreciendo energéticamente el coste de la intercalación más torsión.
- La distancia entre los dos motivos invertidos (10 pb, una vuelta de hélice) y la longitud del linker (27 aa) están conservados, sugiriendo la existencia de una correlación entre el patrón de reconocimiento y la longitud del linker.
- La cola C-terminal parece tener una función estabilizadora del complejo TFAM-LSP22, en concreto del dominio HMG2, mediante la formación del núcleo hidrofóbico, más que dirigir la unión específica de secuencia de TFAM o promover la dimerización como se había descrito anteriormente.
- La cola C-terminal muestra una gran flexibilidad en solución, compatible con la habilidad para interactuar con otras proteínas como TFB2M.
- En el complejo TFAM-LSP, la cola C-terminal emerge en dirección opuesta al sitio de inicio de transcripción. Pero la curvatura inducida permite recolocar la cola C-terminal cercano al complejo de pre-iniciación de la transcripción.

- La región del linker, en concreto la zona C-terminal, presenta un papel en la estabilización del complejo y la curvatura final inducida en el ADN mediante la realización de diversos contactos con el esqueleto fosfato del ADN y con la proteína formando parte del núcleo hidrofóbico.
- La región del linker presenta una gran flexibilidad en solución permitiendo un elevado número de conformaciones de TFAM en su forma no unida en solución, debido a la reorganización espacial de los dos dominios HMGbox en el espacio.
- La región N-terminal del linker parece mostrar una cierta flexibilidad en el complejo, cuando se comparan las dos estructuras existentes. Esta flexibilidad en el linker permite una diferente organización y reorientación de los dominios HMGbox y una curvatura del ADN variada.

MATERIALES, MÉTODOS Y FUNDAMENTOS



CLONACIÓN

1. Clonación estándar

La clonación de TFAM, los constructos con la cola C-terminal delecionada (CTT's, CTT2: Ser44-Gln238, CTT3: Ser44-Arg233, CTT4: Ser44-Arg227) y los dominios HMGbox (D1: Ser44-Gln125, D2: Leu149-Arg227, D2C: Leu179-Cys246), se realizó mediante el sistema de clonación estándar, que implica la amplificación del inserto con los primers diseñados con la secuencia de reconocimiento para las enzimas de restricción (Tabla1), y digestión del plásmidos e inserto con las mismas enzimas de restricción. Posteriormente se realiza la ligación de los extremos cohesivos creados.

Los plásmidos usados fueron:

- pET28-b(+): vector comercial (Novagen, Madison, WI, USA) usado para la expresión de proteínas recombinantes en *E.coli* que añade una diana de 6 histidinas en el extremo N- o C-terminal. Este vector presenta 5.4 kb, resistencia a kanamicina.
- Petm40: vector comercial usado para la expresión de proteínas recombinantes en *E.coli* que añade una diana MBP en el extremo N-terminal. Este vector presenta 6.5 kb, resistencia a kanamicina y un sitio de corte para la proteasa TEV entre la proteína de fusión MBP y la proteína interés.

Proteína	plásmido		primer 5'-->3'	Enzima
TFAM	Pet28-b(+)	fw	cat gcc atg gga tca tct gtc tta gca agt tg	Nco
		rv	cc ctc gag aca ctc ctc agc acc ata ttt tc	Xho I
TFAM	Petm40	fw	cat gcc atg gga tca tct gtc tta gca agt tg	Nco
		rv	cgg gat ccc tta aca ctc ctc agc ac	Bam HI
CTT1	Pet28-b(+)	fw	cat gcc atg gga tca tct gtc tta gca agt tg	Nco
		rv	cc ctc gag acc ata ttt tcg ttg ttt ctt tat tg	Xho I
CTT2	Pet28-b(+)	fw	cat gcc atg gga tca tct gtc tta gca agt tg	Nco
		rv	cc ctc gag ttg ttt ctt tat tgt gcg acg tag	Xho I
CTT3	Pet28-b(+)	fw	cat gcc atg gga tca tct gtc tta gca agt tg	Nco
		rv	cc ctc gag gcg acg tag aag atc ctt tc	Xho I
CTT4	Pet28-b(+)	fw	cat gcc atg gga tca tct gtc tta gca agt tg	Nco
		rv	cc ctc gag tcg tcc aac ttc aat cat ttg ttc	Xho I
D1	Pet28-b(+)	fw	cat gcc atg gga tca tct gt	Nco
		rv	ctc gag act tgg agt tag	Xho I
D2	Pet28-b(+)	fw	cat gcc atg gga cag cta act	Nco

		rv	cc ctc gag aca ctc ctc agc acc ata ttt tc	Xho I
D2C	Pet28-b(+)	fw	cat gcc atg gga cag tta aca c	Nco
		rv	cc ctc gag aca ctc ctc agc acc ata ttt tc	Xho I

Tabla1. Listado de los primers usados para la amplificación de los diferentes insertos, se indica la diana presente para la enzima de restricción, el plásmido para llevar a cabo la ligación y el nombre del constructo.

2. Clonación independiente de ligación: In-Fusion™

Los constructos de TFAM con el linker (H1: Ser43-Leu152, H2S: Ser115-Glu219, H2L: Ser115-Cys246) se originaron mediante un sistema versátil de clonación independiente de ligación basado en el sistema In-Fusion™ (*Clontech-Takara Bio Europe*) aplicado en la Plataforma de Expresión en el IRB. En esta plataforma disponen de una serie de vectores, llamado pOPIN, basados en el sistema de clonación vía In-Fusion™, que permiten la expresión de proteína recombinantes en una amplia gama de huéspedes (*E.coli*, células de mamífero y células de insecto) y su purificación mediante columna de níquel, ya que añaden 6 histidinas en el extremo N-terminal (pOPINF) o C-terminal (pOPINE). El sistema In-Fusion™ está basado en la presencia de secuencia homóloga en los insertos amplificados por PCR y los extremos del vector linealizado. La longitud óptima de homología sería de 15 pb. Debido a que la secuencia homóloga en el inserto se añade mediante PCR, no es necesario ningún paso adicional.

Los plásmidos usados fueron:

- pOPINE: Vector que proviene de pTriEx2, con resistencia a ampicilina. Permite añadir una diana de 6 histidinas en el extremo C-terminal, que puede ser delecionada mediante digestión con Carboxypeptidasa. Presenta los promotores: T7lacO, CMV y B-actina, promotor p10/lef-2 y el elemento baculo 1629.
- pOPINF: Vector que proviene de pTriEx2, con resistencia a ampicilina. Permite añadir una diana de 6 histidinas en el extremo N-terminal, que puede ser delecionada mediante digestión con 3C proteasa. Presenta los promotores: T7lacO, CMV y B-actina, promotor p10/lef-2 y el elemento baculo 1629.

3. Mutagénesis dirigida

Una vez resuelta la estructura y analizada, se pudieron clasificar aquellos residuos de cada HMGbox que realizaban contactos específicos con las bases del ADN. Es por ello que se decidió crear mutantes puntuales de estos residuos por alanina, para realizar ensayos de afinidad al ADN.

Para realizar los mutantes puntuales se usó el kit de mutagénesis dirigida comercializado por *Stratagene (QuickChange™ Site-Directed Mutagenesis Kit)* (Fig1). La mutagénesis dirigida consiste en diseñar oligos para la región donde se hallan las bases a modificar, estos oligos deben presentar una extensión de 10-15 pb a cada lado del punto de mutación (Tabla2). Mediante PCR se realiza la amplificación del ADN con la mutación (que puede ser una inserción o deleción) y posteriormente mediante la enzima DpnI se procede a la digestión del ADN parental, ya que éste se encuentra metilado. El ADN restante se transforma en células DH5α o XL21-Blue.

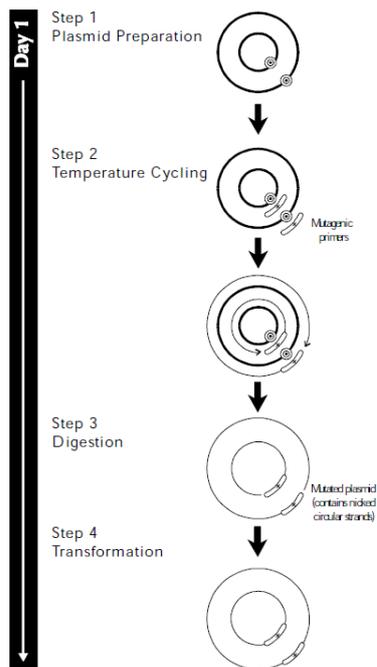


Fig1. Esquema para el proceso de mutagénesis dirigida. El paso 1 consiste en localizar las regiones del ADN amutar y el diseño de los primers. En el paso 2, se lleva a cabo la reacción de amplificación mediante el uso de la enzima *Pfu*, que permite amplificar con elevada fidelidad y velocidad. El paso 3 consiste en la digestión del ADN parental mediante *DpnI*, que reconoce el ADN parental debido a que se encuentra metilado. La digestión se realiza a 37°C durante 1h. En el paso 4, se tranforma el ADN obtenido mediante choque térmico en células competentes DH5α o XL21-Blue.

mutante	oligo	primer 5'--> 3'
L58	L58 FW	CCA AAG AAA CCT GTA AGT TCT TAC GCT CGA TTT TCT AAA GAA C
	L58 RV	G TTC TTT AGA AAA TCG AGC GTA AGA ACT TAC AGG TTT CTT TGG
Y57	Y57 FW	CCA AAG AAA CCT GTA AGT TCT GCC CTT CGA TTT TCT AAA GAA C
	Y57 RV	G TTC TTT AGA AAA TCG AAG GGC AGA ACT TAC AGG TTT CTT TGG
T77	T77 FW	CAG AAC CCA GAT GCA AAA GCT ACA GAA CTA ATT AGA AGA
	T77 RV	TCT TCT ATT TAG TTC TGT AGC TTT TGC ATC TGG GTT CTG
I81	I81 FW	AT GCA AAA ACT ACA GAA CTA GCT AGA AGA ATT GCC CAG CGT TG
	I81 RV	CA ACG CTG GGC AAT TCT TCT AGC TAG TTC TGT AGT TTT TGC AT
Y162	Y162 FW	AAA AGA CCT CGT TCA GCT GCT AAC GTT TAT GTA GCT GAA AGA
	Y162 RV	TCT TTC AGC TAC ATA AAC GTT AGC AGC TGA ACG AGG TCT TTT
N163	N 163 FW	AGA CCT CGT TCA GCT TAT GCC GTT TAT GTA GCT GAA AGA TTC
	N 163 RV	GAA TCT TTC AGC TAC ATA AAC CGC ATA AGC TGA ACG AGG TCT

Q179	Q179 RV	TA GCT AAG GGT GAT TCA CCG GCG GAA AAG CTG AAG ACT GTA AA
	Q179 FW	TT TAC AGT CTT CAG CTT TTC GCG CGG TGA ATC ACC CTT AGC TA
L182	L182 FW	GAT TCA CCG CAG GAA AAG GCG AAG ACT GTA AAG GAA AAC
	L182 RV	GTT TTC CTT TAC AGT CTT GCG CTT TTC CTG CGG TGA ATC

Tabla2. Listado de los primers utilizados para crear los mutantes puntuales a alanina, con las bases rojas señalando el punto de mutación.

PRUEBAS DE EXPRESIÓN

Para obtener un buen rendimiento de expresión de la proteína se realizaron pruebas de expresión, que presentan múltiples variables para optimizar el proceso: cepa bacteriana, temperatura, tipo de medio de cultivo, concentración del agente inductor (IPTG), tiempo de expresión.

Las cepas bacterianas usadas para la expresión fueron:

- **BL21(DE3)** B F- *dcm ompT hsdS*(rB- mB)gal λ (DE3). Proviene de la cepa BL21, pero contienen una copia cromosómica del gen para T7 RNA polimerasa.
- **BL21(DE3)[pLysS]** B F- *dcm ompT hsdS*(rB- mB)gal λ (DE3)[pLysS Cmr]. Similar a BL21 (DE3), pero contiene el plásmido pLysS resistente a cloramfenicol y que codifica para la lisozima de fago T7, un inhibidor de T7 polimerasa que reduce la expresión a partir del promotor de T7 cuando no existe inducción.
- **Rossetta (DE3)** B Δ *lacYZ lon- ompT- λ (DE3)* [pRARE Cmr]. Presentan la capacidad de expresión de genes con código de uso peculiares.
- **Rossetta (DE3) [pLysS]** B Δ *lacYZ lon- ompT- λ (DE3)* [pRARE Cmr]. Presentan la capacidad de expresión de genes con código de uso peculiares y el plásmido pLys, para evitar la expresión prematura.
- **Origami (DE3)** K12-derived. Cmr Tcr. Promueve la formación de puentes disulfuro en el citoplasma, ya que presenta mutación en los genes *trB* (*thioredoxina reductasa*) y *gor* (*glutación reductasa*).

A partir de una colonia única para las células *E.coli* crecidas en una placa de agar, se inocula un medio LB de 4ml, permitiendo el crecimiento durante toda la noche hasta llegar a supersaturación del cultivo. Después de este pre-cultivo se inocula medio fresco LB (aplicando un factor de dilución 1/200) y se realiza un crecimiento a 37 °C con agitación de 250 rpm hasta alcanzar una OD₆₀₀ de 0.6 a 0.8. Las células son luego atemperadas a la temperatura de inducción e inducidas con IPTG en rangos de concentración de 0.1 a 1 mM IPTG. El tiempo de expresión es de 3 a 16 h, dependiendo de la temperatura final. Todos los cultivos usados presentan la concentración de antibióticos óptima para evitar la contaminación de éstos.

SOLUBILIZACIÓN

El siguiente paso después de la expresión, es la solubilización de la proteína y para ello es clave la elección del tampón de solubilización en función de las características físico-químicas de la proteína. Mediante el servidor *ProtParam* (Bairoch, Apweiler et al. 2005) es

posible conocer algunos de los parámetros fundamentales para decidir el tampón de solubilización como son el pI teórico de la proteína que determinará el pH del tampón o el número de Cys que determinará la presencia de agentes reductores. Además, es sabido que las proteínas de unión al ADN, como TFAM, requieren una gran concentración de sal para mimetizar la carga negativa del ADN.

Típicamente el tampón de solubilización o lisis presenta inhibidores de proteasas (como por ejemplo, fluoruro de fenilmetilsulfonilo, PMSF, inhibidor de proteasas de serina o cócteles comerciales de inhibidores de proteasas); así como DNasa para evitar la presencia de una elevada cantidad de ADN en la muestra, frecuente en el caso de proteínas de unión al ADN.

El pelet celular obtenido de la expresión es resuspendido en un volumen indicado de tampón de lisis (factor de dilución 1/10, 1 ml de tampón de lisis por cada 10ml de cultivo). Una vez resuspendido el pelet hasta una solución homogénea, es sonificado durante 15 minutos, con espacios de tiempo de 5 segundos. El proceso de sonicación permite la rotura celular mediante la aplicación de vibración de ultrasonidos al homogenado. Posteriormente, se realiza una centrifugación a 10.000-15.000 rpm durante 20 minutos que permite obtener el sobrenadante con la proteína soluble en él. Este sobrenadante es filtrado mediante el uso de filtros con poros de diámetro entre 0.22-0.45 μm .

Es importante que durante el proceso de solubilización y posterior purificación las muestras se encuentren a una temperatura de $\sim 4^{\circ}\text{C}$, para evitar una posible degradación de la proteína interés.

PURIFICACIÓN

Para la purificación de la proteína se emplean métodos cromatográficos en columna, donde existe una fase estacionaria, material sólido con las características químicas adecuadas que rellena la columna, y una fase móvil, formada por la solución tamponada que se hace pasar a través de la columna. Generalmente estas cromatografías se realizan en sistemas cromatográficos FPLC, en el que se hace uso de bombas de alta presión (100-400 bares) y columnas con materiales que soportan altas presiones. Este sistema reduce notablemente los tiempos de purificación e incrementan el rendimiento.

Existen diferentes tipos de columnas: columna de afinidad, de exclusión molecular, de intercambio iónico, hidrofóbica.

1. Columna de afinidad

Muchas proteínas tienen la capacidad de unirse específicamente a ciertas moléculas. En esta técnica la fase estacionaria de la columna está embebida con una molécula con capacidad de unión específica con la proteína interés, conocida como ligando. Cuando se aplica el extracto, la proteína realiza una unión con el ligando quedando inmovilizada en la columna. Seguidamente mediante el uso de un tampón con características similares al ligando es posible la elusión de la proteína diana en fracciones posteriores.

1.1. Columna de afinidad por Níquel

La columna de afinidad por Níquel está basada en el hecho que el níquel se encuentra quelado en la fase estacionaria de la columna y permite la purificación de proteínas

recombinantes con colas de histidina. La proteína retenida se eluye mediante el uso de un gradiente de imidazol, que actúa como competidor con las proteínas para la unión al níquel. Las columnas de níquel usadas para la purificación de los constructos con cola de Histidinas eran del tipo HisTrap HP 5 ml (*GE Healthcare*).

1.2. Columna de afinidad por MBP (Proteína de unión a maltosa)

Las proteínas recombinantes con dianas MBP son purificadas mediante columnas con una fase estacionaria de dextrin sefarosa, que permite la purificación de la proteína en condiciones fisiológicas. Para eluir la proteína se aplica un tampón al anterior pero con 10 mM maltosa. La forma en que se eluye la proteína en estas columna permite incluso purificar complejo, ya que se mantienen intactos. Además, la proteína de fusión MBP permite incrementar los niveles de expresión y ayuda a la solubilización de la proteína.

Las columnas de MBP usadas para la purificación de los constructos con la proteína de fusión MBP eran del tipo MBP Trap HP 5 ml (*GE Healthcare*).

1.3. Columna de afinidad por Heparina

La heparina es un polímero de tipo glicosaminoglicano que imita el esqueleto azúcar-fosfato del ADN y permite la unión con alta afinidad de proteínas de unión a ácidos nucleicos. Además debido a su carga negativa, crea un efecto de intercambio catiónico y se unen también todo tipo de proteínas básicas. Para llevar a cabo una columna de afinidad por heparina es necesario que la muestra o extracto se encuentre a baja concentración de sal, ya que posterior a la unión se aplica un aumento lineal de la fuerza iónica para eluir la proteína.

Las columnas de Heparina usadas para la purificación de los TFAM y sus diferentes constructos eran del tipo Heparin Trap HP 5 ml (*GE Healthcare*).

2. Cromatografía de exclusión por peso molecular

La cromatografía de exclusión por peso molecular o gel filtración es una técnica que permite separar moléculas en función de su tamaño molecular. La fase estacionaria está constituida por partículas de polímeros con diferente porosidad. La separación se basa en el tamaño de las partículas. Las proteínas más grandes que no pueden penetrar en los poros de las partículas de la matriz de filtración son eluidas con más rapidez que las proteínas más pequeñas, que penetran por los poros de las partículas y siguen un camino más largo.

La capacidad separadora reside fundamentalmente en el gel cuya matriz consta de un gran número de esferas porosas microscópicas. Cada gel se caracteriza por un rango de fraccionamiento que depende del tamaño de sus poros (Fig1). Superdex, sepharyl y superosa son geles de elevada resolución con una amplia variedad de fraccionación.

Para la proteína interés, TFAM, y sus diversos mutantes y dominios, se usaron columna de exclusión molecular de tipo superdex 75 (Superdex 75 HL 26/60), excepto para el caso del constructo MBP-TFAM que se usó una columna superdex 200.

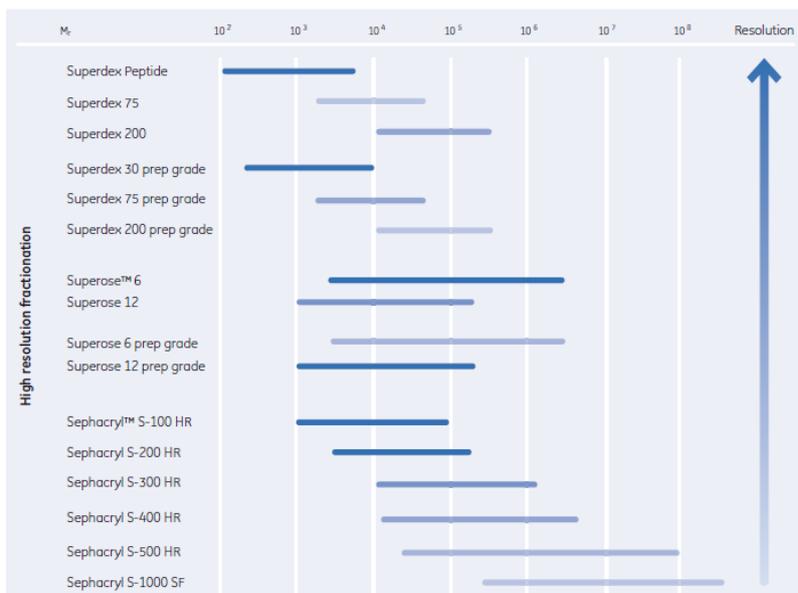


Fig1. Resolución del fraccionamiento de los diferentes tipos de geles en función del peso molecular de la macromolecular a purificar.

3. Columna de intercambio iónico

Las columna de intercambio iónico presentan una fase estacionaria formado por grupos cargados positivamente (intercambio aniónico) o negativamente (intercambio catiónico). Estos grupos cargado normalmente están neutralizados por iones del tampón. Estos iones son reversiblemente reemplazados por las proteínas con más tendencia a unirse al soporte. Las proteínas cargadas pueden unirse a intercambiadores catiónicos o aniónicos dependiendo de su carga neta. Las proteínas más cargadas se unirán más fuertemente al intercambiador y, por lo tanto, serán más difíciles de eluir. La afinidad con la que una proteína se une a un intercambiador iónico depende de la fuerza iónica del medio. Para eluir las proteínas se suele aumentar la fuerza iónica de la fase móvil (por ejemplo, aumentar la concentración de NaCl).

En el caso de TFAM se usaron columnas de intercambio catiónico, MonoS 5/50 GL GE Healthcare.

3.1. Digestión con la proteasa TEV

El constructo de TFAM con la proteína de fusión MBP en el extremo N-terminal, presentaba un sitio de corte para la proteasa TEV entre la proteína y la diana MBP. Durante la purificación de este constructo se llevo a cabo la digestión con TEV después de la columna de afinidad por MBP.

La proteasa TEV es una proteasa específica para cisteína de la familia de peptidasas C4 con un peso molecular de 27 kDa, que se encuentra en el virus del tabaco Etch (*Tobacco Etch Virus*, TEV). El sitio de reconocimiento óptimo es ENLYFQ'GS, donde el sitio de corte ocurre entre Gln y los residuos Gly/Ser. Algunas de las ventajas de esta proteasa es que es extremadamente específica y presenta una elevada eficiencia de corte.

Algunas consideraciones se deben tener en cuenta para llevar a cabo la digestión con TEV. La duración de la reacción es de 24 h, aunque la mayoría de los cortes ocurren en las

primeras horas. La proteasa TEV presenta la mayor eficiencia de corte a 34 °C, aunque en nuestro caso se realizaba la digestión a 20 °C durante 24 h. El ratio de proteasa usado era 1 OD₂₈₀ de TEV por 50 OD₂₈₀ de la proteína, gracias a que se trata de una proteasa altamente específica es posible usar concentraciones elevadas de proteína. Para la digestión era necesario preparar un saco de diálisis con la proteína diluída en condiciones de 750 mM NaCl, 50 mM HEPES pH 7.5, 5 mM DTT, 10 Mm Maltosa, frente a un tampón de diálisis con las siguientes características: 300 mM NaCl, 50 Mm HEPES pH 7.5, 1 Mm DTT. Para la digestión fue necesario aplicar este tampón con una menor concentración sal y una concentración inferior de DTT, ya que se trata de un agente quelante de Zn.

Después de la digestión con TEV proteasa la muestra era cargada en una gel filtración, que permite separar por peso molecular TFAM, la proteasa TEV y MBP.

IDENTIFICACIÓN DE PROTEÍNAS

1. Electroforesis de proteínas en condiciones desnaturalizantes (SDS-PAGE)

La electroforesis en geles de poliacrilamida es uno de los métodos más utilizados para la purificación, análisis y caracterización de proteínas. La técnica permite separar moléculas cargadas y explota diferencias en movilidad cuando se les somete a la acción de un campo eléctrico. Se utilizan agentes desnaturalizantes de proteínas, como pueden ser: detergentes (p.e. SDS), caótopos (p.e. urea) y agentes reductores (2- mercaptoetanol, DTT). La separación de proteínas es en función del tamaño (masa molecular), lo que permite determinar el peso molecular de las proteínas. Las muestras son desnaturalizadas antes de ser cargadas en el gel a 97°C durante 5 min en presencia de 5 % (w/v) SDS y 0.025 % (w/v) de azul de bromofenol. Se usó el marcador de peso molecular *SeeBlue pre-satined protein* (Invitrogen).

Para la visualización de las proteínas se utilizan diferentes métodos de tinción, siendo el más habitual el azul de coomassie.

2. Digestión tríptica de proteínas

La técnica de digestión tríptica de proteínas o *protein mass fingerprinting* sirve para identificar proteínas mediante la digestión de la proteína con enzimas como tripsina o quimiotripsina en pequeños péptidos, cuya masa absoluta se obtiene mediante un espectrómetro de masas. La masa absoluta de cada péptido es luego comparada con una base de datos con secuencia proteicas y el tamaño de los posibles péptidos generados a partir de las proteínas en la base de datos, que permite identificar la proteína digerida.

3. Degradación Edman

La degradación por Edman es un método que permite la secuenciación del extremo N-terminal de péptido. Se trata de un proceso en el que el extremo N-terminal del residuo reacciona con fenilisotiocinato a pH básico y, posteriormente en un pH ácido se corta este residuo derivado sin afectar a los demás enlaces de la cadena peptídica. El residuo compuesto puede ser luego identificado mediante cromatografía o electroforesis. El hecho limitante de este proceso es que la longitud del péptido no puede ser superior a 50 residuos.

CARACTERIZACIÓN DE LA ESTABILIDAD DE LA PROTEÍNA EN SOLUCIÓN

1. *Dynamic Light scattering (DLS)*

DLS es una herramienta comúnmente usada para evaluar la monodispersidad de proteínas en solución, en concreto, en el campo de la cristalografía nos permite valorar la homogeneidad y estabilidad de la solución a cristalizar. DLS es una técnica usada para determinar el radio hidrodinámico de las proteínas a partir de la ecuación *Stokes-Einstein*, $R_h = K_B T / 6\pi\eta D_T$, donde K_B es la constante de Boltzmann, T la temperatura absoluta, η es la viscosidad del solvente a la misma temperatura, D_T el coeficiente de difusión y R_h es el radio hidrodinámico.

La técnica consiste en una cubeta con la solución proteica que es iluminada mediante luz laser. Las moléculas en solución presentan un movimiento browniano, que provoca fluctuaciones en la intensidad de la luz dispersada. Este cambio en la intensidad de la luz es medida a través de un detector situado a 90° del haz incidente. El coeficiente de difusión, D_T , se puede extraer mediante la aplicación de una función de autocorrelación. En general las partículas pequeñas difunden más rápido que las grandes. A partir de D_T es posible obtener el radio hidrodinámico, R_h , de la molécula en solución. El peso molecular o MW puede ser calculado a través del radio hidrodinámico, si se define la forma de la molécula (alargada, globular), aunque no es muy recomendable.

El análisis de datos permite obtener la distribución del % de intensidad o cantidad de luz dispersada versus el radio hidrodinámico de las partículas en solución. A partir de la anchura de los picos de distribución, es decir, la desviación estándar, se puede extraer el valor de la polidispersidad (aunque por defecto se presenta en forma de porcentaje, % Pd). El porcentaje de polidispersidad se calcula dividiendo la polidispersidad por el radio hidrodinámico, multiplicado por 100. Las soluciones con un %Pd <15 son monodispersas y con tendencia a cristalizar.

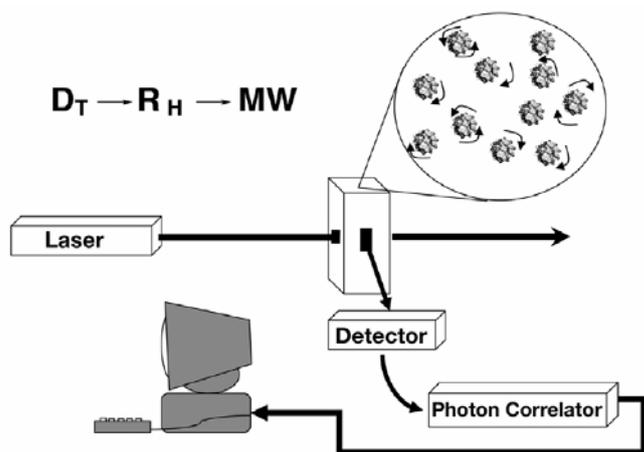


Fig1. Esquema del experimento DLS, donde el haz láser incidente dispersa la solución

Las medidas de DLS se realizaron bajo diferentes condiciones de tampón usando el instrumento *Malvern Zetasizer Nano* (Worcestershire, UK), para ello se usó una cubeta de cuarzo de 1cm de paso de luz. Cada medida se realizó por triplicado.

2. Thermofluor

La técnica de Thermofluor sirve para determinar el tampón en el que la proteína es más estable térmicamente (Ericsson, Hallberg et al. 2006). Para monitorizar la estabilidad térmica de una proteína y testar qué factores (tampones, aditivos, ligandos,...) afectan a dicha estabilidad, este ensayo (*fluorescence-based thermal shift*) utiliza un colorante sensible a la hidrofobicidad del entorno, como Sypro orange (*Molecular probes*). Durante el proceso de desnaturalización, las superficies hidrofóbicas van quedando expuestas al medio incrementado y, por lo tanto, aumenta la fluorescencia debida al colorante. La desnaturalización térmica de una proteína es un proceso que sigue un modelo típico de dos estados con una marcada transición entre los estados nativos y los desnaturalizados, donde la temperatura de fusión (T_m) se define como la temperatura en el punto medio de la transición en la desnaturalización de la proteína que corresponde al máximo en la primera derivada de la curva de fusión. Cuanto mayor es la T_m , mayor es la estabilidad de la proteína. Por lo tanto, este pequeño ensayo permite seleccionar los tampones óptimos para la estabilización de la proteína.

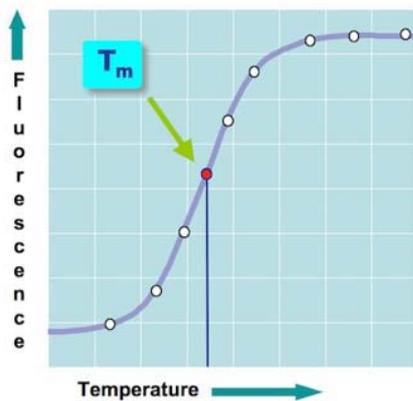


Fig2. Diagrama para un ensayo de Thermofluor. Mediante incremento de temperatura con el tiempo es posible observar un incremento de la fluorescencia hasta alcanzar una fase plateau. La temperatura en el punto medio de la transición en la desnaturalización de la proteína se conoce como temperatura de fusión (T_m), y nos permite validar la calidad del tampón.

Las medidas se realizaron en una máquina de PCR real-time y los cambios de fluorescencia en las muestras se monitorizan simultáneamente con una cámara CCD. Las longitudes de onda de excitación y emisión de la sonda Sypro son 490 y 575 nm, respectivamente.

FORMACIÓN DEL COMPLEJO PROTEÍNA-ADN

Para la formación del complejo proteína-ADN es importante determinar la secuencia de ADN, mediante la búsqueda de secuencias consenso o información sobre análisis footprinting. Una vez escogida la secuencia se realizan ensayos para conocer la capacidad de unión y afinidad mediante geles nativos de retardo o cromatografías de exclusión molecular.

1. Geles nativos de retardo (EMSA)

Los ensayos de retardo en la movilidad electroforética (*Electrophoretic Mobility Shift Ssays*, EMSA) o más comúnmente llamados geles nativos de retardo, se basan en la

observación de que complejos estables proteína-ADN migran más lentamente que los fragmentos de ADN libre en geles de poliacrilamida en condiciones nativas bajo la acción de un campo eléctrico.

Frecuentemente, se usaron los ensayos EMSA para conocer el comportamiento de la muestra cuando se utiliza una concentración constante de ADN (0.1 nmol) frente a concentraciones crecientes de proteína de 0.025 nmol a 0.6 nmol. También eran usados para conocer el estado del complejo proteína/ADN y para los ensayos de FRET.

1.1. Preparación de la muestra, botones de diálisis

En el caso del análisis de la afinidad de la proteína por el ADN. Se preparaban los complejos con concentraciones creciente de proteína y una concentración constante de ADN con volumen final de 20 μ l. Debido a que la proteína se encuentra a una elevada concentración de sal era necesario el uso de botones de diálisis de 15 μ l (*Hampton Research*) para disminuir la concentración de sal (Fig1). El tamaño de poro de la membrana usada era de 3.500 Da.

Las muestras se incubaron 20 min a 4°C, se dializaron con gradiente escalonado de sal hasta 20-100 mM NaCl, cada paso de diálisis presenta una duración mínima de 2 horas con un paso final de 24 horas en la condición del tampón final.

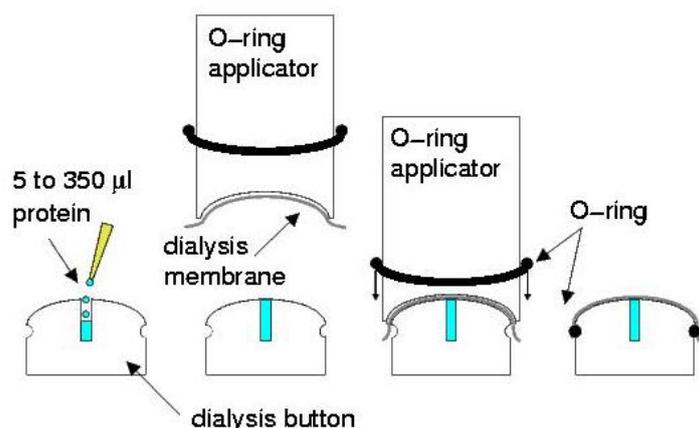


Fig1. Esquema de la preparación de los botones de diálisis de izquierda a derecha. El complejo se coloca en el hueco del botón. Seguidamente se coloca la membrana (previamente sumergida en el tampón del complejo) en la superficie del aplicador y éste se coloca en el botón. Mediante una goma se fija la membrana al botón y se deposita en el tampón de diálisis.

1.2. Preparación de los geles

Los geles nativos de poliacrilamida para los ensayos de formación de complejo presentaban unas dimensiones 8 cm x 10cm x 1.5mm. En el caso de los geles para la permutación circular las dimensiones del gel eran 18 cm x 20 cm x 1 mm.

En función de la longitud del ADN se preparaban geles con diferente porcentaje de poliacrilamida. Para ensayos con los oligos de cristalización (~ 22 pb), se usaban geles al 5.5 % o 7%, mientras que para los geles de permutación circular con ADNs de 143 pb, se usaron geles entre 8% y 12%.

	8%	12%	14%	20%
Archilamida 40% (29:1)	10 ml	15 ml	17.5 ml	25 ml
H ₂ O	34 ml	29 ml	26.5 ml	19 ml
PSA 3%	1 ml			
TEMED	15 µl			
10x TBE o TAE	5 ml			

1.3. Observación de los geles

Las muestras para cargar en el gel se preparan mezclando el complejo con *DNA dye loading buffer* 6x. En el caso de los ADN marcado con fluoróforos se usó un tampón de carga sin bromofenol azul (tampón de carga 5x: 25 mM TrisHCl pH7, 150 mM EDTA, 25 % glicerol).

Una vez cargadas las muestras, se usaba un voltaje de 100V durante 40 min (aunque el tiempo podía variar en función del % de archilamida), evitando la migración de la primera banda azul. Seguidamente se realiza la tinción del gel mediante 20 min con *SyBr Green (Invitrogen)*, aplicando una dilución 1/10.000. El gel teñido se analiza con una luz UV, ya que el complejo resultante ADN-SYBR Green presenta el pico de absorción a $\lambda = 498$ nm y el pico de emisión a $\lambda = 522$ nm (correspondiente a la zona verde del espectro, de ahí su nombre).

Para el caso de los ADNs marcados con fluoróforos se usaba un fluorímetro de escanéo conocido como *Typhon (Molecular dynamics)* que permite obtener una imagen resultante multicanal, producto de la excitación y emisión de cada fluoróforo por separado debido a que presenta diversos filtros. La imagen final de los experimentos FRET era producto de la combinación de tres canales:

- Canal 1 con excitación a 488 nm, y filtro de emisión a 520 nm (para la visualización de Alexa494)
- Canal 2 con excitación a 488 nm, y filtro de emisión a 610 nm (para la visualización de Alexa590)
- Canal 3 con excitación a 488 nm, y filtro de excitación a 532 nm (para visualizar el fenómeno FRET)

CRISTALOGRAFIA DE PROTEÍNAS

1. Introducción

La cristalografía de rayos X es una de las técnicas físicas más potentes para la resolución de proteínas a partir de estructuras de materia sólida ordenada o cristales.

1.1. Cristales y simetría

La estructura cristalina está caracterizada microscópicamente por la agrupación de sus moléculas en un modelo de repetición periódica tridimensional. El conjunto mínimo que se repite ordenadamente en las tres direcciones del espacio se conoce con el nombre de *celdilla elemental* y la *unidad asimétrica* se refiere al conjunto mínimo de moléculas cuya repetición (mediante la aplicación de elementos de simetría) genera la totalidad del contenido de la celdilla elemental. La combinación de los elementos de simetría presentes en los cristales y los distintos tipos de celdilla posibles dan lugar a los denominados *grupos espaciales*.

2. Cristalización

La obtención de cristales únicos de elevada calidad es la base de la cristalografía de rayos X y, frecuentemente, suele ser el punto de embudo durante el proceso de determinación de la estructura cristalográfica de una proteína.

La cristalización de macromoléculas es un proceso donde influyen múltiples parámetros, consiste en aportar una solución altamente concentrada en presencia de un estado de supersaturación, que conducirá la macromolécula hacia un estado sólido, el cristal. Existen numerosos métodos para la cristalización de proteínas y ácidos nucleicos, uno de los más utilizados es el sistema de vapor de difusión.

Este método se basa en la difusión de agua (u otro agente volátil) entre la gota (sentada o colgante) de la solución de la macromolécula con el agente precipitante apropiado y la solución de reservorio en un sistema cerrado. La diferencia de concentración de sales y precipitantes entre la gota y el depósito fuerza la difusión de las especies más volátiles y da lugar a una nueva situación de equilibrio. Esta difusión de solventes modifica el volumen de la gota, modificando la concentración de la proteína en la gota. Mediante un diagrama de fase (Fig1) es posible controlar el grado de saturación al que exponemos la solución y el resultado final del equilibrio que da lugar a nucleación, precipitación amorfa, precipitación en forma cristalina,...

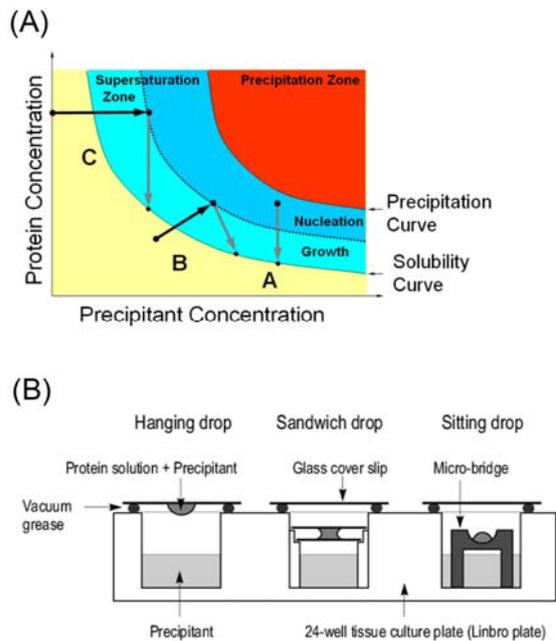


Fig1. (A) Diagrama de fases para el proceso de cristalización, donde es posible distinguir la curva de precipitación y la curva de solubilidad en función de la concentración de proteína y agente precipitante. La zona entre estas dos curvas se conoce como zona de supersaturación (en azul) y puede distinguirse en dos regiones, la región donde sucede la formación de núcleos cristalinos y la región donde sucede el crecimiento cristalino. (B) Método de cristalización por vapor de difusión, con tres tipos de sistema de gota: gota colgante (hanging drop), gota sandwich (sandwich drop), gota sentada (sitting drop).

El tamaño de la gota y la cantidad de precipitante que existen en la solución de cristalización influirán en el número de centros de nucleación que se formara, así como en la velocidad de equilibrio, directamente relacionada con la velocidad a la que aparecen los cristales. Resulta importante para el experimento controlar estos dos parámetros, ya que si forzamos la evaporación pueden surgir un elevado número de núcleos que luego compitan por la proteína restante y den lugar a cristales pobres. No obstante, una difusión limitada puede dar lugar a la dificultad para que se de la formación de núcleos.

Para la cristalización de la proteína y del complejo proteína-ADN se usó el método de difusión de vapor tanto en su versión de gota sentada como colgante) a 20 y 4°C empleando *screenings comerciales sparse-matrix* (Cudney, Patel et al. 1994). En estos *screenings* se realizan múltiples combinaciones de la proteína con mezclas de diferentes agentes precipitantes, como sales (fosfatos, citratos, sulfato de amonio,...), precipitantes (polietilenglicol (PEG), glicerol,...), alcoholes (metilpentendiol, isopropanol,...) y tampones ajustados a diferentes pHs (HEPES, Tris,...).

Estos *screenings* se realizaron en placas de cristalización de 96 pocillos dispersadas con un robot automático (*Cartesian*) de dispersión líquida de nanovolúmenes (100nl:100nl) de la Plataforma Automatizada de Cristalografía (PAC) en el Parque Científico de Barcelona (PCB). Éstas se almacenaron en armarios termoestabilizados que incorporan un sistema automático de visualización mediante imágenes digitales de las nanogotas de cristalización (Crystal Farm, *Oxford Cryosystem 700*) de la PAC-PCB.

La cristalización de complejos proteína-ADN es similar a la cristalización de cualquier otra macromolécula, depende del precipitante, la fuerza iónica, el pH, la presencia de aditivos,... De forma interesante se observó que el pH óptimo para la cristalización de complejo proteína-ADN es neutro, con tendencia a ácido. Esto es probablemente debido a las interacciones polares entre las cargas negativas del esqueleto de ADN y las cargas positivas de las cadenas laterales de la proteína. Debido a estas interacciones de tipo polar (Nadassy, Wodak et al. 1999), es preferible usar precipitantes como PEG o MPD, en lugar de sales. Dentro de los *screenings* comerciales existen dos *screenings* de capital importancia para complejos proteína-ADN, *PEG-Ion screen* y *nucleic acid screen*. Además, algunos aditivos frecuentes en la cristalización de complejos proteína-ADN son *spemine*, iones divalentes, *cobalt hexamine*. Otro factor importante en los ensayos de cristalización de complejos proteína-ADN es la longitud del ADN usado. Normalmente el ADN usado presenta 1 o 5 pb adyacentes a la región central, con variedad de extremos, romos o cohesivos. El uso de extremos cohesivos puede favorecer la formación de fibras lineares de ADN.

2.1. Procesos de optimización

Una vez obtenidos los primeros cristales se procedió a la optimización mediante el uso de placas de cristalización de 24 pocillos, donde las gotas se dispersaban a mano con volúmenes de gota oscilante entre 1 a 6 μ l y jugando con el ratio de mezcla entre la solución de proteína y la solución precipitante. Las pruebas de optimización de la condición consistieron en hacer un *screenings* variando pH y la concentración de agentes precipitantes en los dos sentidos de la placa (vertical u horizontal), así como el uso de aditivos. También se aplicaron métodos previos a la cristalización, en concreto técnicas de **seeding**, que consisten en añadir núcleos preformados a una gota metaestable para limitar el número de núcleos y favorecer el crecimiento del cristal.

Además, se aplicaron métodos post-cristalización para mejorar la calidad de los cristales como la **deshidratación química**, para reducir el contenido de solvente y ayudar a un mejor empaquetamiento cristalino. Estos métodos de deshidratación química de cristales consisten en transferir los cristales a soluciones deshidratantes, que presentan las mismas características que el reservorio, pero con una mayor concentración de agente precipitante o en presencia de agentes crioprotectores (como por ejemplo, PEG400, MPD, glicerol). **Cross-linking del cristal**, que permiten incrementar la consistencia del cristal frente a estrés mecánico y reduce la solubilidad, mediante el uso del compuesto químico glutaraldehído. Este proceso es dependiente del pH, el número y la posición de las lisinas. **Annealing del cristal**, que consiste en el proceso rápido de descongelación y congelación, frecuentemente este proceso se realiza en la línea de sincrotrón.

2.2. Congelación de cristales

Para evitar el daño por radiación de los cristales difractados se usan técnicas de criocristalografía, las cuales permiten proteger el cristal de la formación de hielo sin alterar la estabilidad de la señal. La criocristalografía consiste en sumergir los cristales en soluciones, que una vez congeladas, forman una capa de agua vitrificada, en lugar de formar hielo cristalino que podría provocar la rotura del cristal. Esta solución se conoce como solución crioprotectora, y frecuentemente consiste en una solución de composición similar a la solución del reservorio a la cual se añade un agente crioprotector con glicerol,

etilenglicol o otros alcoholes de bajo peso molecular. Posteriormente, los cristales son difractados a temperaturas cercanas a 100 K (-173°C).

Los cristales de complejos proteína-ADN suelen ser muy frágiles, con lo que parece una buena idea usar *screenings* con características crioprotectantes para evitar posteriores manipulaciones.

3. Rayos X

Para la recogida de la difracción de rayos X es necesario disponer de rayos X de alta energía. En nuestro caso la recogida de datos cristalográficos se realizó la radiación de sincrotrón en ERSF (*European Synchrotron Radiation Facility*) de Grenoble, Francia.

Para la generación de rayos X se parte de electrones que son acelerados a elevadas energías. Mediante imanes electromagnéticos se puede variar la trayectoria de los electrones y al cambiar la dirección de éstos se emite parte de su energía en forma de rayos X, que salen de forma tangencial al giro de los electrones. Los haces de rayos X del sincrotrón presentan una elevada intensidad, son monocromáticos (de una única longitud de onda) y está colimado (todas las ondas son paralelas y con una misma dirección).

3.1. Difracción de rayos X

Cuando un haz de rayos X incide sobre un cristal, éste interacciona con los electrones de los átomos que forman el cristal y los hace vibrar de manera acoplada a las variaciones periódicas de su campo eléctrico. Los electrones oscilantes se convierten así en un *focus* de una nueva radiación X que se emite de forma esférica. Este fenómeno se conoce como *dispersión*, y normalmente se da de forma elástica (la longitud de onda de la radiación emitida por los electrones será similar a la radiación que impacta con el cristal).

Dentro de la dispersión, puede suceder que los rayos dispersados interfieran unos con otros rayos dispersados. Como los átomos del cristal se encuentran ordenados de manera regular y periódica, la geometría del cristal marcará la interacción entre los rayos dispersados ofreciendo la posibilidad de que estos se combinen y refuercen, proceso conocido como interferencia constructiva que da lugar a la denominada dispersión cooperativa o *difracción*.

El fenómeno de difracción puede ser descrito cómo si de una reflexión óptica se tratara mediante la Ley de Bragg. Esta reflexión ocurriría en unos planos de la red cristalina, y así, la difracción se produce entre planos paralelos distanciados entre sí el valor d , para un ángulo de incidencia θ , tal que la diferencia del camino recorrido por los dos trenes de onda sea igual al múltiplo de la longitud de onda λ . Para que se produzcan todos los haces difractados posibles, el cristal debe ser rotado de tal modo que todos lo planos virtuales se coloquen en la posición adecuada para cumplir la Ley de Bragg.

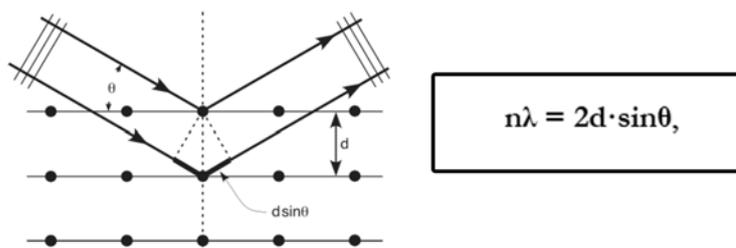


Fig2. Ley de Bragg. Dos haces paralelos que inciden sobre una red cristalina serán difractados siguiendo el ángulo θ . Siempre que la separación entre los dos planos cristalinos sea $2d \cdot \sin\theta$ coincide con un múltiplo de la longitud de onda, $n\lambda = 2d \cdot \sin\theta$. La onda que incide en el plano inferior saldrá en fase con la onda del plano superior. Así, las ondas interfieren constructivamente y la señal se verá reforzada.

Los rayos dispersados por todos los átomos de la molécula irradiada se combinarán entre ellos, por lo que cada una de las reflexiones registradas contendrá información de los átomos irradiados.

4. Recogida de datos

La recogida de datos se realiza típicamente mediante el uso de detectores tipo CDD. Al final de la recogida las intensidades registradas se digitalizan, cosa que permite obtener el difractograma para cada ángulo de rotación del cristal, que resulta en una colección de múltiples reflexiones cada uno de ellos con una determinada intensidad. Según la simetría de la macromolécula y el grupo espacial se deberá recoger un determinado número de imágenes para completar la información.

Para obtener una buena calidad de datos es importante tener en cuenta diversos factores durante la recogida de datos:

- Temperatura de la recogida. Actualmente la recogida se realiza a temperaturas criogénicas (100 K), que permiten reducir el daño por radiación. Este tipo de recogida de datos exige la congelación de los cristales, proceso que puede aumentar la mosaicidad del cristal.
- Ruido por la dispersión del solvente. Para evitar el ruido en la medida y mejorar el ratio señal vs ruido, es necesario minimizar el contenido de reservorio en el lazo.
- Tiempo de exposición y atenuación de la línea. Debe existir un compromiso entre estos dos parámetros. Un tiempo de exposición excesivo puede dar lugar a sobreintensidad de las reflexiones (*overload*) o daño por radiación.
- Ángulo de oscilación, no debe ser mayor a la mosaicidad para evitar sobrexposición de puntos (*overlap*).
- Otros parámetros podrían ser: el tamaño del haz de rayos X, que debe optimizarse con el tamaño del cristal. Métodos de deshidratación acoplados al difractómetro, recogida de datos helicoidales, uso del sistema kappa para adquirir nuevos ángulos de recogida...

5. Procesamiento de datos

Los parámetros que se evalúan a partir del patrón de difracción son la posición del haz difractado, que depende de la red cristalina, y la intensidad de cada reflexión, que depende de la distribución de los átomos en la celdilla. Por lo tanto, los diferentes difractogramas se

procesan e indexan, hecho que permite deducir la simetría del cristal y el grupo espacial al que pertenece. Los programas de indexado analizan la disposición de las reflexiones en el difractograma y calculan la orientación del cristal. Teniendo en cuenta la longitud de onda de la radiación y la distancia entre el detector y el cristal, también es posible calcular las medidas de la celdilla unidad y predecir el grupo espacial. Una vez determinados estos parámetros, se pueden utilizar para predecir la situación de las reflexiones en las demás imágenes. Ésto permite saber si el indexado ha sido exitoso y medir la intensidad registrada en cada una de las predicciones para las reflexiones. Finalmente, los valores obtenidos para cada imagen son escalados entre sí generando para cada reflexión, sus tres coordenadas (índices de Miller, h, k y l) y su intensidad (I_{hkl}).

Los datos obtenidos fueron indexados e integrados mediante MOSFLM (Battye, Kontogiannis et al.) y/o XDS (Kabsch), y escalados posteriormente mediante SCALA.

5.1. Análisis de la calidad de datos

Para evaluar la calidad de los datos una vez procesados existen diversos indicadores: R_{merge} y R_{sym} , la señal de intensidad $I/\sigma(I)$, completitud de los datos (*completeness*) y la gráfica *Wilson-plot* (Glusker 1994).

- R_{merge} y R_{sym} : son parámetros estadísticos que valoran la precisión con la que se han hecho las predicciones de señal en el procesado y como correlacionan con la intensidades experimentales observadas. Se considera que el procesado es correcto cuando los valores no superan el 10%.
- $I/\sigma(I)$: medida de la relevancia de las señales de intensidad observadas respecto a la desviación debida al ruido y los errores experimentales. Cuanto más grande es el valor, más precisos son nuestros datos. Generalmente, se aceptan valores a partir de 2.
- *Completeness*: muestra la completitud de nuestros datos. Cuánto más cercano a 100%, más completitud de datos.
- *Wilson plot*: es la distribución de la media de las intensidades en función de la resolución. Esta distribución permite calcular el factor de temperatura y el factor necesario para poner las intensidades en una escala absoluta.

6. Resolución problema de la fase

El objetivo de un estudio estructural es la determinación de la densidad electrónica, $\rho(xyz)$, para cada punto de la celdilla elemental cristalina. Para ello partimos de la información del espacio recíproco a partir de los datos de difracción. La transformación de espacio recíproco al espacio cristalino (función de densidad electrónica) implica la resolución de la *transformada de Fourier* (Fig3). En ella encontramos los factores de estructura, $F(hkl)$, definidos en cada punto del espacio recíproco (hkl). Los factores de estructura son ondas, y como tales se pueden describir por sus módulos, $|F(hkl)|$, y fases, $\Phi(hkl)$, respecto de un origen común de fases.

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| \cos 2\pi(hx + ky + lz - \phi(hkl))$$

Fig3. Función de densidad electrónica definida en cada punto de la celdilla elemental cristalina de coordenadas x, y, z. Esta ecuación representa la *transformada de Fourier* en el espacio cristalino.

A partir de las medidas del espectro de difracción se obtienen las intensidades $I(hkl)$. Estas intensidades nos proporcionan las amplitudes $|F(hkl)|$ de los factores de estructura (Fig4), pero no sus fases $\Phi(hkl)$. Las fases son pues magnitudes que no pueden medirse experimentalmente, hecho conocido como *problema de la fase*.

$$I(hkl) = K.A.L.p. |F(hkl)|^2$$

Fig4. Relación entre el módulo de los factores de estructura $|F(hkl)|$ y la intensidad $I(hkl)$ de los puntos del espectro de difracción. **K** es un factor que lleva los factores de estructura experimentales a la escala absoluta, es decir, a la escala de los factores de estructura calculados (teóricos). Este factor se puede determinar, de modo aproximado, usando los datos experimentales mediante el denominado plot de Wilson. **A** es un factor de absorción, que igualmente puede estimarse conociendo la composición del cristal y sus dimensiones. **L** es el denominado factor de Lorentz, responsable de corregir la distinta velocidad angular por la que pasan los puntos recíprocos de la esfera de Ewald. **P** es el factor de polarización, que corrige el efecto de la polarización del haz incidente sobre el cristal.

Para la resolución de este problema se aplicaron dos técnicas: método de dispersión anómala (MAD, *Multiple Anomalous Scattering*) y método de remplazamiento molecular (MR, *Molecular Replacement*).

6.1. Método de dispersión anómala múltiple (MAD)

La técnica MAD se basa en variar la longitud de onda alrededor de la longitud de onda de absorción de los átomos pesados introducidos en los cristales de proteína, por ejemplo, mediante la sustitución de los residuos de metionina en la proteína por Seleno-metionina. Ello da como resultado a una dispersión anómala que modifica el factor de dispersión atómicos (f), de tal modo que su expresión se ve modificada con dos términos adicionales f' (real) y f'' (imaginaria) (Fig5).

$$f = f^0 + f' + i f''$$

Fig5. En presencia de dispersión anómala, el factor de dispersión atómico, f^0 , se ve modificado por dos términos adicionales que dan cuenta de las partes real e imaginaria de la dispersión anómala.

La componente real tiene en cuenta las amplitudes difractadas medidas a diferentes longitudes de onda. La componente imaginaria o anómala se obtiene midiendo las llamadas diferencias de Bijovet entre los *pares de Friedel* (h, k, l y $-h, -k, -l$), parejas de reflexiones que en condiciones normales deberían tener la misma amplitud y fase, pero con signos opuestos. El cambio de intensidad detectable entre *parejas de Friedel* recibe el nombre de difracción anómala. Analizando las diferencias en la difracción es posible calcular la distribución de amplitudes y fases que generan los dispersores anómalos. Las fases generadas por estos dispersores anómalos, como una primera aproximación a las fases globales, permite calcular la densidad electrónica para toda la proteína.

El método MAD, desarrollado por Hendrickson y Kahn, implica la medida de los datos de difracción de un cristal de la proteína (que contenga un dispersor anómalo fuerte) usando

radiaciones de distintas energías: la que maximiza f'' (*pico*), la que minimiza f' (*inflexión*) y una energía lejana de estas dos (*remota*). Combinando estos conjuntos de datos de difracción, y en concreto analizando las diferencias entre ellos, es posible calcular la distribución de amplitudes y fases que generan los dispersores anómalos. En el caso de método SAD, método de dispersión anómala única se usa radiación de una sola energía para difractar los cristales.

Para la resolución del complejo TFAM con ADN, primero se derivó la proteína con selenometionina y se optó por un experimento SAD, lo cual implicó realizar un escaneo de fluorescencia del cristal y obtener el espectro de absorción para el átomo de Se. La longitud de onda escogida para el experimento de difracción anómala de longitud de onda única (SAD) fue donde el valor de la segunda derivada (f'' , componente imaginaria) de los datos del espectro de absorción presentaba su máximo, longitud de onda conocida como pico.

Una vez los datos de la difracción fueron recogidos, se indexaron, mezclaron y esclaron. La identificación de las posiciones de los átomos pesados (Se) se realizó mediante el programa SHELXD (Uson and Sheldrick 1999). Posteriormente se realizó una corrección del solvente y una corrección anisotrópica hasta obtener un mapa de densidad electrónica donde era posible distinguir proteína y ADN.

En el caso del complejo con el ADN brominado se optó por una recogida a una longitud de onda remota, un valor más alejado del espectro de absorción. Las amplitudes experimentales de este experimento fueron posteriormente usadas para llevar a cabo un experimento de remplazamiento molecular mediante el uso del modelo estructural obtenido del complejo TFAM-LSP22 con la proteína derivada con seleno metionina.

6.2. Método de remplazamiento molecular (MR)

El método de remplazamiento molecular consiste en asignar fases de una estructura modelo como fases iniciales para nuestra estructura. Para hacerlo primero se deben calcular parámetros de rotación y traslación necesarios para posicionar la proteína modelo en la celdilla unidad de la proteína a resolver.

La calidad de las soluciones de estas funciones se expresa mediante los coeficientes de correlación entre las funciones de Patterson experimental y la calculada con la proteína conocida. Un coeficiente de correlación alto entre dichas funciones equivale a un buen acuerdo entre los datos de difracción experimental y el calculado con la proteína conocida. Una vez orientada y trasladada convenientemente la molécula conocida, se calcula un mapa de densidad electrónica usando los factores de estructura experimentales.

En nuestro caso el método de reemplazo molecular fue usado para combinar la información de difracción del complejo TFAM-LSP22Br con el modelo o estructura obtenida mediante experimento SAD con TFAMSeMet-LSP22, realizando una búsqueda de 2 moléculas por unidad asimétrica mediante el programa MolRep (Vagin and Isupov 2001).

7. Optimización de la fase

Una vez obtenido el mapa de densidad electrónica frecuentemente es necesario aplicar método de optimización de la fase mediante un proceso cíclico hasta conseguir una convergencia entre el mapa experimental y las fases experimentales. Estos métodos están integrados en programas como DM (Cowtan and Zhang 1999), CNS (Abrahams and Leslie 1996) y REFMAC (Murshudov, Vagin et al. 1997).

Dentro de los mecanismos para la optimización de fase encontramos métodos de *corrección de solvente*, que asignan un valor de densidad electrónica al solvente y eliminan la densidad electrónica negativa; y aplicación de *simetría no cristalográfica (NCS)*, que permite atribuir un promedio de densidad entre moléculas idénticas dentro de la misma unidad simétrica.

8. Modelado, refinamiento y validación de la estructura

Una vez optimizado el mapa de densidad electrónica se puede empezar el proceso de construcción de la cadena polipeptídica y el ADN. Este proceso puede realizarse a mano cuando existe una previa experiencia o puede realizarse mediante un proceso automatizado con programa como ARP/wARP. En el caso de TFAMSeMet-LSP22 se realizó una construcción a mano mediante el programa COOT.

La construcción del modelo se realizó en base a mapas de densidad electrónica o mapas de Fourier calculados con coeficientes $2F_{\text{obs}}-F_{\text{calc}}$ y $F_{\text{obs}}-F_{\text{calc}}$ (éstos últimos para comprobar deficiencias negativas debidas a partes del modelo no observadas en los datos experimentales, o positivas, debidas a partes de los datos experimentales que no pueden ser explicadas por el modelo).

8.1. Refinamiento estructural

El modelo estructural viene definido por las coordenadas atómicas (x, y, z) y un factor de temperatura (B) para cada átomo, que expresa la movilidad térmica del átomo alrededor de su posición de equilibrio. En el proceso de refinamiento estructural se ajustan los parámetros (x, y, z, B) para cada átomo para optimizar el acuerdo entre las observaciones y la predicción. El proceso de refinamiento es un proceso iterativo que combina ciclos de cálculo con construcción y ajuste del modelo al mapa de densidad electrónica. Por lo tanto, el número mínimo de datos son cuatro veces (x, y, z, B), el número de átomos de la estructura. Si no hay suficientes datos, hay que imponer restricciones. El incremento en la resolución de la difracción da lugar a un incremento exponencial del número de datos. El objetivo del refinamiento es lograr que el modelo se ajuste lo mejor posible a las amplitudes de los factores de estructura experimentalmente medidas.

En algunos casos el refinamiento debe incluir diversas restricciones:

- *Refinamiento de cuerpo rígido (Rigid Body)*: se suele usar en los primeros periodos del refinado. Se trata de mantener la estructura como un todo, que se modifica sólo en términos de translaciones y giros.
- *TLS*: El uso de TLS permite asumir la existencia de grupos tratados como un colectivo asociado a un mismo valor de desplazamiento atómico (*Atomic Displacement Parameters, ADPs*) o factores de temperatura (*B-factors*). TLS hace

referencia a Translation/Libration/Screw, movimiento vibracional que se aplica a cada cuerpo rígido definido para explicar los valores de B-factor observados.

- *Restricciones estereoquímicas*: dado el conocimiento previo de la estereoquímica presente en proteínas y ADN, es posible aplicar ciertas restricciones en cuanto a ángulos de enlaces, distancias de enlace, quiralidad, etc. Estas limitaciones se impusieron manualmente para refinar las moléculas de ADN en la estructura, imponiendo las restricciones Watson-Crick entre las bases. Para el resto de enlaces, los ángulos y distancias están restringidos a unos valores tabulados en diccionarios que los programas de refinamiento emplean.

El proceso de refinamiento se lleva a cabo a caballo entre el programa REFMAC5 (Murshudov, Vagin et al. 1997) y AUTOBUSTER (Blanc, Roversi et al. 2004), y la inspección visual mediante el programa COOT.

8.2. Validación de la estructura

8.2.1. Parámetros estadísticos

La validación del proceso de refinamiento se lleva a cabo mediante la inspección de dos parámetros estadísticos:

- R_{factor} : medida de la precisión de los resultados del refinamiento, que se obtiene comparando el modelo obtenido con los datos experimentales. Es decir, es un indicador que estima el desacuerdo entre F_{obs} y F_{calc} , de manera que indica lo bien que el modelo predice las amplitudes medidas por las reflexiones. Aunque depende de la resolución final del modelo, valores por debajo del 20% son indicativos de la buena calidad de un modelo.

$$R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|}$$

- R_{free} : mide de forma objetiva el grado en el que el modelo atómico construido predice las amplitudes observadas al ser calculado, empleando el mismo algoritmo que el R_{factor} a partir de un pequeño conjunto de reflexiones tomadas aleatoriamente y no utilizadas en el refinamiento (entre 5-10% del total de reflexiones).

8.2.2. Parámetros estereoquímicos

En la estructura final es necesario observar ausencia de choques entre átomos y que las distancias y ángulos sean adecuados y los parámetros térmicos razonables. Además de estos criterios es importante contrastar que las conformaciones que adoptan los enlaces N-C α y C α -C en el enlace peptídico no den lugar a orientaciones interatómicas energéticamente improbables. Las torsiones del enlace peptídico (Φ , phi y ψ , psi) presentan, en principio, libertad de giro. Sin embargo, cada uno de estos ángulos está energéticamente limitado a ciertos intervalos definidos en el diagrama de *Ramachandran*, que muestra las regiones permitidas de estos ángulos para cada uno de los tipos de estructura secundaria.

PERMUTACIÓN CIRCULAR

1. Descripción de la técnica

La permutación circular está basada en el conocimiento que la migración electroforética en geles de poliacrilamida de fragmentos de ADN con una distorsión central es diferente comparado con fragmentos que presentan la distorsión cerca de los extremos (Wu and Crothers 1984). La medida de la movilidad relativa del complejo proteína/ADN (con el ADN distorsionado, μM) vs la movilidad del ADN libre (μE), ($\mu M/\mu E$), en relación con la localización relativa de la secuencia de unión dentro del fragmento, permite estimar la distorsión introducida en el ADN, mediante los algoritmos generados por Thompson, *et al.* (Thompson and Landy 1988), más tarde ampliados por Ferrari, *et al.* Thompson, *et al.* aplicaban la simplificación final $\mu M/\mu E = \cos \alpha/2$, donde α hace referencia al ángulo de desviación que se produce en un ADN lineal referente a la movilidad relativa ($\mu M/\mu E$) para los fragmentos con el sitio de curvatura central. Esta forma de calcular el ángulo era óptima para ángulos de distorsión entre 0° y 140° , es decir, θ entre 180° y 40° . Debido a que el método descrito por Ferrari, *et al.* consideraba más parámetros se optó por él.

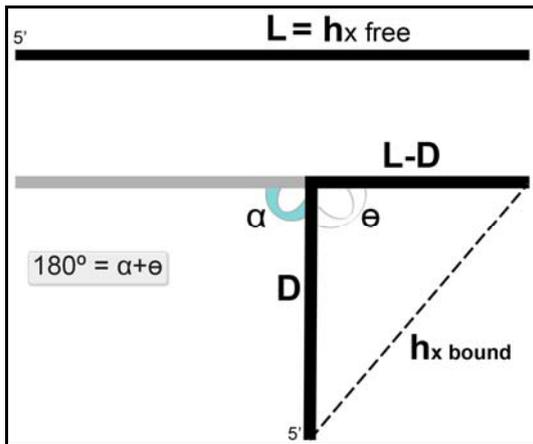
En el caso de complejo proteína/ADN, la técnica consiste en generar una batería de fragmentos de la misma longitud con la secuencia de unión específica situada en una localización diferente para cada fragmento. Posteriormente se observa la movilidad electroforética de estos fragmentos en complejo con la proteína interés. La movilidad electroforética relativa ($\mu M/\mu E$) fue definida por Levene&Zimm (Levene and Zimm 1989) y es proporcional al centro de velocidad de masa (V_{cm}), definido por la longitud del fragmento (L) y la distancia entre los extremos (h_x), y dividido por la intensidad del campo magnético aplicado (E), éste definido por la carga total del ADN, Q , y su fricción constante, ζ :

$$\mu = \langle V_{cm} \rangle / E = \langle h_x^2 / L^2 \rangle Q / \zeta$$

mediante la aplicación de constantes definidas experimentalmente para Q/ζ es posible encontrar la siguiente ecuación:

$$\mu M / \mu E = k_{bound} (h_{x\ bound}^2 / L^2) / k_{free} (h_{x\ free}^2 / L^2) = K (h_{x\ bound}^2 / L^2) \quad (\text{ecuación 1})$$

donde $h_{x\ bound}$ es la distancia entre los dos extremos del ADN debido a la curvatura. Si trasladamos esta incógnita a los vértices de un triángulo y el ángulo que los relaciona:



$$h_{x\text{ bound}}^2 = D^2 + (L-D)^2 - 2D(L-D)\cos\theta \quad (\text{ecuación 2})$$

Cuando se substituye la ecuación 2 en la ecuación 1:

$$\mu\mathbf{M}/\mu\mathbf{E} = K[D^2 + (L-D)^2 - 2D(L-D)\cos\theta] / L^2$$

$$\mu\mathbf{M}/\mu\mathbf{E} = 2K(1 + \cos\theta)(\mathbf{D}/\mathbf{L})^2 - 2K(1 + \cos\theta)(\mathbf{D}/\mathbf{L}) + K \quad (\text{ecuación 3})$$

En esta ecuación, los valores experimentales serían $\mu\mathbf{M}/\mu\mathbf{E}$ definido por la relación entre la movilidad electroforética del complejo y la movilidad del ADN libre; y (\mathbf{D}/\mathbf{L}) o lugar de flexibilidad está definido por la distancia del extremo 5' al sitio donde se encuentra la región de unión respecto a la distancia total. Al representar gráficamente los valores de $\mu\mathbf{M}/\mu\mathbf{E}$ en el eje Y y (\mathbf{D}/\mathbf{L}) en el eje X para los diferentes fragmentos, se obtiene una parábola a la cual se aplica un ecuación polinómica de segundo grado y a partir del elemento de segundo grado $2K(1 + \cos\theta)(\mathbf{D}/\mathbf{L})$ es posible obtener el valor del ángulo θ . A partir de este valor es posible hallar el ángulo final de curvatura ($\alpha = 180^\circ - \theta$).

2. Realización del ensayo de permutación circular de TFAM a LSP

La secuencia del oligo LSP22 utilizado para la cristalización fue clonada en el vector *pBend2* entre el sitio de digestión para la enzima *Xba I* y *Sal I* continuos. El vector *pBend2* mediante su especial diseño del sitio de policlونación o MCS (multi-cloning site) permite la digestión del plásmido creando fragmentos de 143 pb con el sitio LSP22 situado a lo largo de esta extensión en diferentes posiciones (Fig1). Se produjo la digestión con 6 enzimas de restricción: *Bgl II*, *Spe I*, *EcoRV*, *Stu I*, *Asp 718*, *Bam HI* (*NE Biolabs*).

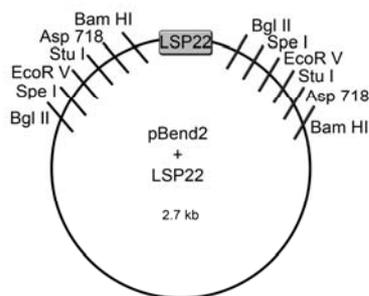


Fig1. Esquema del vector *pBend2* con la secuencia LSP22 clonada. La digestión con las enzimas indicadas da lugar a fragmentos de 143 pb con la secuencia LSP22 situada en puntos diferentes del fragmento.

Fue necesaria la producción de grandes cantidades de pBend2+LSP22, ya que en ese caso no fue usada radiactividad para marcar los fragmentos de ADN. Antes de llevar a cabo el experimento, se realizó una titración para saber el ratio proteína:ADN ideal. Para 50 ng de ADN, se usaba 0.05 $\mu\text{g}/\mu\text{l}$ de proteína.

Los productos de la digestión fueron incubados con TFAM a temperatura ambiente en las siguientes condiciones de tampón: 20mM/100mM NaCl; 50mM Hepes pH 7.5/20mM Bis-Tris pH6.5; 5mM DTT. También se añadía 10%glicerol, para favorecer la formación del complejo y 1 μg BSA para evitar unión inespecífica.

Posteriormente los complejos fueron cargados en geles nativos PAGE al 8% y se aplicó un voltaje de 100V, tanto a temperatura ambiente como a 4°C. La movilidad relativa de los complejos de TFAM con los fragmentos de ADN frente al ADN libre ($\mu\text{M}/\mu\text{E}$) se representa en función la posición relativa de LSP22 dentro del fragmento de 143pb. A la representación se le aplica una función polinómica de segundo grado. El mínimo de la representación coincide con la posición de LSP22 en el centro del fragmento de 143pb, y a partir de este valor es posible extraer el ángulo de curvatura θ ($180^\circ-\alpha$) de acuerdo con el método descrito por Ferrari, *et al.* (Ferrari, Harley et al. 1992).

DISPERSIÓN DE BAJO ÁNGULO EN SOLUCIÓN (SAXS)

1. Dispersión de rayos X a bajo ángulo (SAXS)

El uso de la metodología de dispersión de rayos X a bajo ángulo (SAXS: *Small-angle X-ray scattering*) está siendo mejorada y muy difundida ya que permite la caracterización de estructuras e interacciones de complejos macromoleculares en solución.

SAXS es una técnica que permite estudiar estructuras e interacciones entre complejos biológicos. SAXS puede ser usado para muestras proteicas, ácidos nucleicos, y sus complejos en una variedad amplia de condiciones, desde cercanas a las condiciones fisiológicas a altamente desnaturalizantes. Para ello no es necesario cristalizar la muestra ni existen limitaciones de tamaño, a diferencia de RMN. Históricamente, SAXS ha sido utilizado para obtener parámetros claves como el peso molecular aparente (MW), el radio de giro (R_g), y la distancia máxima intramolecular (D_{max}). Aunque se trata de una técnica de baja resolución (no llegando a resoluciones inferiores a $\sim 10\text{\AA}$), la existencia de fuentes de sincrotrón de mayor poder y el desarrollo de algoritmos que permiten la reconstrucción de mapas de densidad electrónica 3D a partir de perfiles de dispersión 1D, han permitido generar modelos ab-initio de baja resolución a partir de datos de SAXS, lo que está popularizando el uso de esta metodología. Además, SAXS ha sido ampliamente utilizado para el estudio de desplegamiento o plegamiento parcial de sistemas en función del tiempo o de las condiciones de la solución.

Al igual que en el caso de un experimento cristalográfico de difracción de rayos X, en SAXS es necesario colocar la muestra, en este caso en solución, bajo un haz de rayos X altamente colimado y posteriormente se recoge la dispersión de rayos X (Fig1). El ángulo entre el haz de rayos X y la dispersión se define como 2θ .

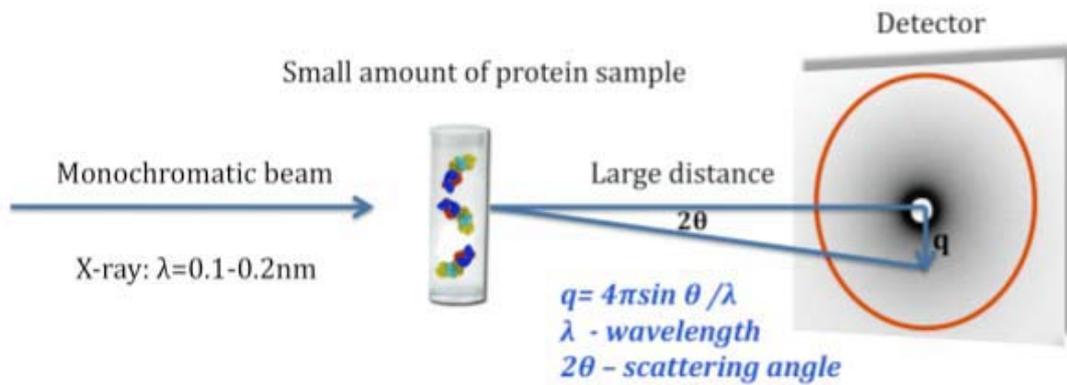


Fig1. Esquema de dispersión de rayos X a bajo ángulo de una solución con muestra de proteína. *Svergun et al, 2002*.

SAXS es un método de contraste donde la señal de dispersión deriva de la diferencia media de densidad electrónica, $\Delta\rho(r)$, entre la densidad de la molécula en solución de interés ($\rho(r)$) y el solvente, ρ_s :

$$\Delta\rho(r) = \rho(r) - \rho_s$$

En la práctica, el experimento de SAXS implica medir la dispersión en el tampón, y posteriormente la dispersión con la muestra. La sustracción de los perfiles de dispersión da lugar a la señal de dispersión final observada debido a la presencia de la macromolécula en solución (Fig2).

La curva de dispersión obtenida a partir de la sustracción del tampón frente a la muestra, $I(q)$, es isotrópica debido a la orientación aleatoria de las partículas en solución. Esta función se representa como logaritmo de la intensidad, $\log(I)$, vs *momentum transfer* (q), donde $q = (4\pi \sin \theta) / \lambda$, 2θ es el ángulo de dispersión y λ es la longitud de onda del haz de rayos X incidente. El parámetro *momentum transfer*, q , también se puede definir como s , relacionado por la siguiente ecuación $q = 2\pi s$. Las unidades de q son inversas a las de la longitud de onda, es decir, \AA^{-1} o nm^{-1} .

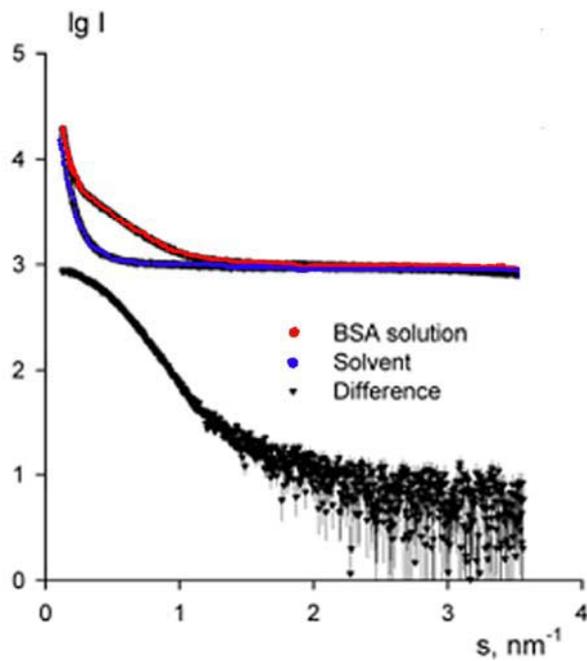


Fig2. Perfil de dispersión típico de una solución de BSA (en rojo) y solvente (en azul), con la curva de sustracción final (en negro) ($Lg(I)$ vs s), representado en escala semi-logarítmica. (Svergun, et al. 2003)

La curva experimental es posible dividirla en tres partes, en función de la información que se puede extraer. Un ejemplo lo tenemos en la Fig3, donde se comparan las curvas de dispersión experimentales para diferentes proteínas. A bajos ángulos o q (2-3 nm de resolución), la curva experimental difiere notablemente entre diferentes proteínas, en esta primera zona q presenta una rápida decadencia dependiente de la forma de la partícula. En resoluciones medias (2-0.5 nm), las diferencias son mucho menos pronunciadas y, en resoluciones por debajo de 0.5 nm, todas las curvas presentan una forma similar.

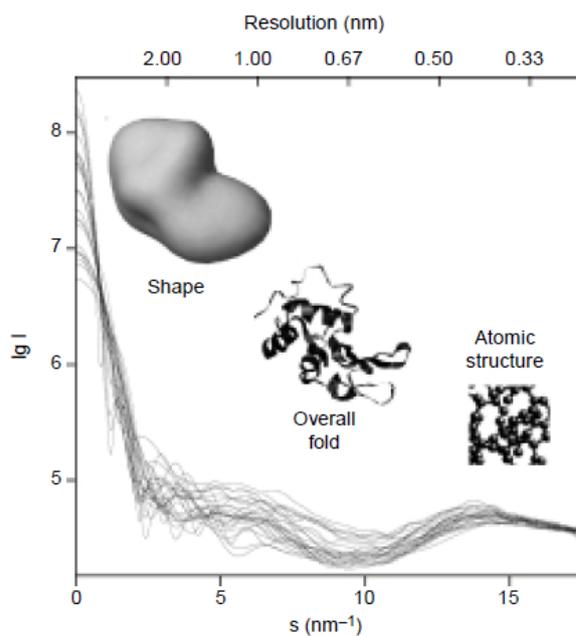


Fig3. Perfil de dispersión de SAXS para 25 proteínas diferentes con pesos moleculares entre 10 a 300kDa. La intensidades de dispersión se muestran en escala logarítmica, el eje x superior muestra la resolución ($\Delta=2\pi/s$)

y el inferior, *s*. Además se encuentra indicado los niveles de organización estructural observados en función de la resolución. *Svergun, et al. 2002*

2. Análisis de los datos

El análisis de los datos se realizó con el programa PRIMUS (*Svergun, et al.*).

Los parámetros más frecuentes extraídos a partir de un perfil de SAXS de biomoléculas en solución son el radio de giro (R_g) y la intensidad de dispersión inicial ($I(0)$), obtenidos a partir de la fórmula de Guinier:

$$I(q) \approx I(0) \exp(-q^2 R_g^2/3)$$

Para *momentum transfer* (q) inferiores, la representación de $\ln(I(q))$ vs q^2 permite obtener la conocida como recta de Guinier (Fig4). A partir de la recta Guinier es posible extraer el R_g , como la pendiente de la recta, y de la extrapolación con el eje y obtenemos $I(0)$. Si la tendencia de la recta no es demasiado lineal indicaría posibles estados de agregación.

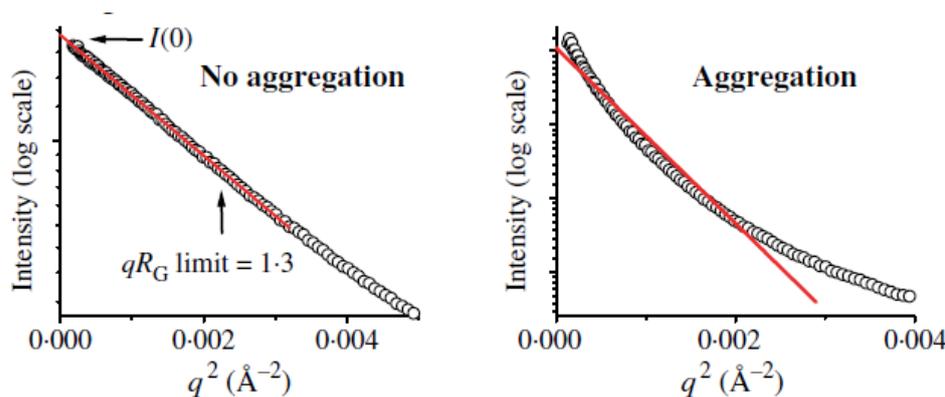


Fig4. Representación Guinier con valores de R_g y $I(0)$, donde una dependencia no lineal de Guinier indica la presencia de agregación. *Putnam, et al.2007.*

R_g es un parámetro que sirve para caracterizar el volumen o tamaño molecular, mientras que $I(0)$ presenta una relación con el peso molecular de la muestra medida, ya que depende del cuadrado del número de electrones:

$$I(0) = \kappa c (\Delta\rho)^2 (MW)^2$$

Donde κ es una constante que puede ser determinada a partir de la mediada de un patrón estándar, es decir, de una proteína con concentración y peso molecular conocido como BSA. c es la concentración de la macromolécula, $\Delta\rho$ es el contraste de densidad electrónica media de la molécula, y MW es el peso molecular. Este MW puede ser usado para determinar posibles estados oligoméricos.

El perfil de dispersión más común es la función de distribución de las distancias atómicas intramoleculares $p(r)$ (Fig5), donde D_{max} es la distancia intramolecular máxima:

$$I(q) = \int_0^{D_{max}} dr p(r) (\sin(qr))/(qr)$$

$P(r)$ se obtiene a partir de la transformada de Fourier indirecta del perfil de dispersión, por ejemplo, mediante el programa GNOM creado por Svergun, D. (Schmidt, König et al. 1995). La función $p(r)$ es un histograma de las distancias interatómicas de la partícula en solución que son promediadas para las diferentes orientaciones. Por lo tanto, a partir de esta función de distribución es posible extraer la longitud máxima en la partícula, D_{\max} , y además ofrece información sobre la forma de la proteína (Fig5). D_{\max} puede ser difícil de caracterizar cuando en las estructuras existen zonas extendidas o con desorden. Sin embargo, en los últimos años el uso de $p(r)$ para visualizar la forma molecular ha sido substituida por algoritmos que proporcionan mapas de densidad electrónica 3D a baja resolución a partir del perfil de SAXS 1D. Esta reconstrucción 3D presenta diversas limitaciones asociadas a la carencia de fase e información respecto a la orientación angular de las moléculas.

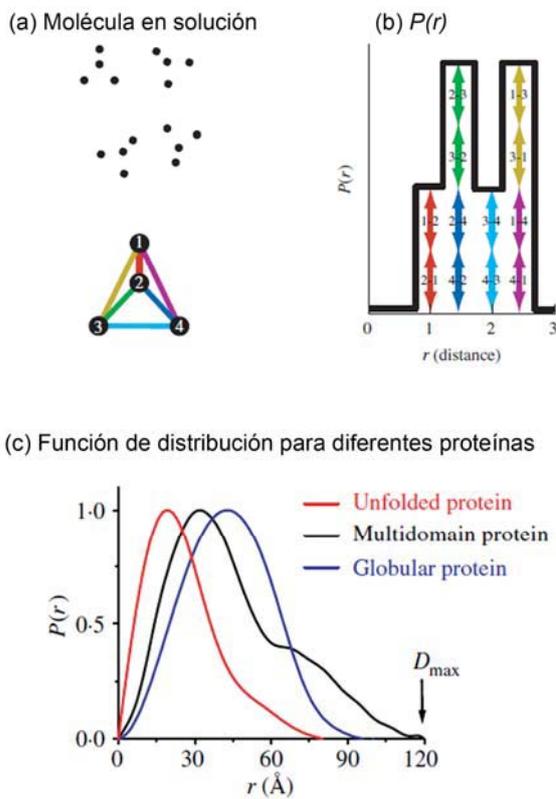


Fig5. La función $p(r)$, resuelve las distancias pero no las direcciones en las moléculas dispersadas. (a) Molécula en solución dispersada con diferentes orientaciones. (b) Función de distribución en forma de histograma con el promedio de las distancias interatómicas. (c) Perfil de la función de distribución para diferentes proteínas. La proteínas globulares suelen mostrar un único pico, mientras que proteínas elongadas muestran una cola alargada

Además, como se ha comentado anteriormente SAXS es un método ideal para identificar y caracterizar moléculas con dominios sin plegamiento o estructura secundaria definida. La representación *Kratky* ($q^2 I(q)$ vs q) (Fig6), que puede ser calculada directamente a partir de la curva de dispersión, es una excelente herramienta para evaluar el plegamiento de las proteínas. Por ejemplo, para dominios plegados, la representación *Kratky* ofrece un pico bien definido similar a una parábola, donde la posición del pico ofrece alguna información

sobre el tamaño global. Mientras que proteínas parcialmente desplegadas muestran una campana de gauss inacabada.

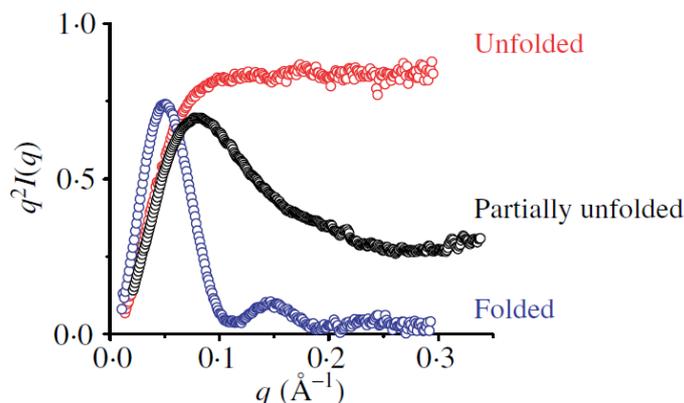


Fig6. Representación Kratky. Esta representación permite identificar la presencia de muestra desplegada. Una molécula globular presentará un pico perfecto y simétrico, mientras una molécula extendida presentará un perfil sin pico y un *plateau* a lo largo del eje q . Putnam, et al. 2007

3. Combinación de datos cristalográficos y SAXS para modelar estructuras

Así como existe una gran variedad de herramientas y técnicas que permiten modelar los datos de SAXS con modelos atómicos, cabe destacar que un punto crítico y nada trivial en este proceso es comparar las curvas de dispersión experimentales con curvas teóricas. Además existen diferentes métodos *ab initio* que permiten generar mallas tridimensionales a baja resolución, similares a las reconstrucciones de microscopía electrónica, que pueden ser usadas para acoplar estructuras atómicas. Análisis mucho más complicados pueden realizarse cuando la macromolécula de interés presenta flexibilidad intrínseca y, por lo tanto, la curva de SAXS es producto de una combinación de diferentes conformaciones. El uso de SAXS para estos diferentes casos es extremadamente útil, ya que muchas veces los modelos con resolución atómica imponen limitaciones.

3.1. Cálculo de curvas de dispersión teóricas a partir de modelos atómicos: CRY SOL

La curva de dispersión experimental es fruto de la diferencia de dispersión entre la molécula diana con su capa de solvatación ordenada y el volumen excluido, que tiene en cuenta la falta de dispersión del solvente debido a la presencia del soluto. Por lo tanto, el volumen excluido puede ser determinado al definirse la forma de la molécula y calculando la dispersión a partir de este volumen imaginado como un volumen lleno de densidad.

Es posible calcular curvas teóricas a partir de modelos atómicos, como sería el caso de una estructura cristalográfica, mediante el uso del programa CRY SOL. Este programa es capaz de crear mallas o envolturas a partir de *spherical harmonics (multipole expansion)*, dependiente del número de átomos, que cubren el modelo y a partir de aquí es posible calcular la intensidad asociada al volumen. Este programa es óptimo para ajustar curvas teóricas a experimentales a baja resolución, pero a mayor resolución ($q > 0.4 \text{\AA}^{-1}$) es más problemático debido a que las reconstrucciones tipo "*spherical reconstructions*" no tienen en cuenta la estructura interna de las partículas dispersadas.

3.1.1. Validación

Ajustar la curva del modelo teórico o cristalográfico a la curva experimental de SAXS requiere un parámetro que correlacione el grado de concordancia entre las dos curvas, al igual que en cristalografía existe la medida *Rfactor*. Una versión de *Rfactor* en SAXS también existe descrita por *Svergun, et al.* como función χ^2 , hace referencia a la discrepancia entre los dos modelos:

$$\chi^2 = 1 / (N_p - 1) \sum_i [(I(q_i)_{exp} - cI(q_i)_{calc}) / \sigma(q_i)]$$

Donde *c* es un factor de escala y $\sigma(q_i)$ es el error experimental. Esta medida claramente impone un mayor peso sobre los datos a baja resolución y es comúnmente usada para comparar curvas experimentales con teóricas mediante el programa CRY SOL (Schmidt, Konig et al. 1995).

3.2. Modelado *Rigid Body* combinado con datos de SAXS: pre_BUNCH

La gama de conformaciones diferentes para proteínas multi-dominio con regiones connectoras o extremos N- o C-terminal se pueden generar a partir del programa pre_BUNCH. Este programa trata los dominios como cuerpos rígidos y los conecta mediante linkers, donde los ángulos diédrico entre C α -C α son seleccionados de forma aleatoria pero cumpliendo con la estereoquímica de Ramachandran y evitando problemas estéricos. Normalmente se generan un gamma de conformaciones capaz de cubrir todo el espacio posible, suele ser un número N=10.000 conformaciones. Los perfiles de dispersión para estas conformaciones son luego calculados mediante el programa CRY SOL en un rango de 0 a 0.5 Å⁻¹.

3.3. Estudio de sistemas flexibles, con múltiples conformaciones: EOM

Muchas proteínas o complejos macromoleculares presentan regiones flexibles que les permiten adoptar un gran número de conformaciones. Para este tipo de muestras (como es el caso de TFAM) se debe olvidar el protocolo de fijar un único modelo que presente el mejor ajuste con los datos experimentales.

Ensemble Optimization Method (EOM) es un método que permite analizar la presencia de múltiples conformaciones de proteínas contribuyendo a la curva experimental de dispersión (Bernado, Mylonas et al. 2007). EOM permite a partir de una mezcla de posibles conformaciones (N>1000), generadas de forma totalmente aleatorias para cubrir todas las posibles conformaciones en el espacio, seleccionar mediante el uso de un algoritmo genético subpoblaciones (N=50) de conformaciones que ajustan con la curva experimental. Las mejores subpoblaciones son sucesivamente escogidas para una posterior evolución. Normalmente se realizan 200 ciclos de algoritmo genético. El número de conformaciones dentro de la subpoblación (N=50) es reflejo de la flexibilidad del sistema. Una vez obtenidas las 50 mejores conformaciones mediante EOM se analiza su ajuste con la curva experimental, la distribución del R_g y la distribución de la distancia entre dominios entre la población inicial de 10.000 y la subpoblación final.

Mediante el uso de datos experimentales y teóricos, EOM es capaz de distinguir entre cuerpos rígidos y regiones flexibles, y evaluar contactos entre dominios.

4. Estudio de sistema dinámicos en SAXS con TFAM

A su vez, SAXS también permite estudiar sistemas macromoleculares dinámicos, concretamente es posible diferenciar los porcentajes de población de conformaciones diferentes en la población global. En el caso del complejo TFAM-LSP22 se trató de combinar modelos creados a partir de las coordenadas cristalográficas existentes y modelos computacionales con los datos experimentales para el complejo. Ello implicó la definición de diferentes escenarios ofreciendo flexibilidad a la proteína y ángulos de curvatura diferentes al ADN. El ADN fue creado a partir del servidor 3D-DART (van Dijk and Bonvin 2009). Para crear los diferentes escenarios, que involucraban dar flexibilidad a TFAM, se usó el programa PRE_BUNCH generando 2000 estructuras; seguidamente se usó el programa CLASH para fusionar la proteína con el ADN evitando impedimentos estéricos entre la proteína y el ADN. Finalmente, mediante CRY SOL y *averaging curves* se procedió al ajuste de las curvas promediadas teóricas con la curva experimental.

Los escenarios creados se definen a continuación:

- **Escenario 1:** el ADN presentaba una curvatura global de 180°, como en la estructura cristalográfica:
 - Escenario **1a:** presentaba un cuerpo rígido definido por casi totalmente todo el complejo, con flexibilidad en la cola C-terminal en su totalidad (res 225-246)
 - Escenario **1b:** presentaba un cuerpo rígido definido por casi totalmente todo el complejo, pero ofreciendo flexibilidad en la zona final de la cola C-terminal (en concreto los 9 aminoácidos que faltaban en la estructura cristalográfica (res 238-246)
- **Escenario 2:** Se definieron tres cuerpos rígidos o dominios (dominio1:HMG1, dominio2: LINKER, dominio3:HMG2) mediante el programa PRE_BUNCH, y luego se formaba el complejo con el ADN definiendo su unión con éste hasta el final del dominio 1 con el programa CLASH.
 - Escenario **2a:** se utilizaba un ADN con una primera curvatura de 90° debido a la unión de HMG1
 - Escenario **2b:** se utilizaba un ADN con una primera curvatura de 90° y una segunda curvatura más ligera de 40°.
- **Escenario 2 c/d:** Difería de 2a y 2b, en que la unión con el ADN implicaba HMG1 y el linker y HMG2 en su totalidad, es decir, del residuo 44-220. Para c se usó el mismo ADN que en escenario 2a, y para 2d el mismo que 2b.
- **Escenario 3:** Se definieron dos cuerpos rígidos (HMG1 y HMG2, por separado), con un linker no helicoidal. Además se imponía la unión con el ADN de HMG1.
 - Escenario **3a:** se utilizó un ADN lineal ideal. Este escenario fue eliminado, ya que existen impedimentos estéricos al intentar modelizar la unión con

un dominio HMGbox. Cabe destacar que está ampliamente descrito y caracterizado, que HMGbox después de su unión imponen una curvatura sobre el ADN.

- Escenario **3b**: se utilizó un ADN con una primera curvatura debido a la unión de HMG1.
- **Escenario 4**: Para este escenario no se creó complejo con el ADN, se utilizaron las 50 mejores conformaciones que definen TFAM en solución.
- **Escenario 5**: Trabajamos con un ADN libre, ideal y linear.
- **Escenario 6**: Se definieron dos cuerpos rígidos (dominio 1: HMG1+linker, dominio2: HMG2), la unión con el ADN comprende hasta HMG1+linker (residuo 44-148).
 - Escenario 6a: se utilizó un ADN con una curvatura de aproximadamente 90°.
 - Escenario 6b: se utilizó un ADN con una primera curvatura de 90° y una segunda curvatura más ligera de 40°.

5. Recogida de datos

5.1. TFAM en su forma no unida al ADN

Para TFAM primero se realizó una primera recogida de muestras en la línea ID14.3 de ESRF, así como para los constructos con la cola C-terminal recortada (CTT2, CTT3). Las concentraciones medidas fueron las siguientes: 0.5 mg/ml, 2.15 mg/ml, 3.5 mg/ml, 5.54 mg/ml, 7.9 mg/ml, 12.82 mg/ml en condiciones de tampón de purificación: 200 mM NaCl, 50 mM HEPES pH 7.5, 5 mM DTT, 10% glicerol. Antes de la recogida de datos es necesario la medida de BSA para calibrar la línea y poder extrapolar luego el peso molecular de la muestra de interés. En el caso de la recogida en ESRF, en la *beamline* ID14-3, fue necesario aplicar una estrategia para evitar el daño por radiación que consistía en hacer una primera medida de 3s/imagen y posteriormente una segunda medida de 30s/imagen, cada medida de 10 imágenes. Además la línea no era suficientemente estable, el detector era demasiado pequeño para captar todos los ángulos y la cantidad de muestra tenía un volumen relativamente grande de 200 µl. Por ello las siguientes medidas de TFAM se realizaron en el sincrotrón DESY (Hamburgo), *beamline* X33, línea mucho más estable y con un aplicador de muestra automático y un volumen necesario de ~ 50 µl. En este caso se realizaron medidas a dos concentraciones de NaCl diferentes: 200 mM y 500 mM. Finalmente se usaron las curvas a una concentración de NaCl de 500 mM, ya que presentaban una menor presencia de comportamientos típico de interacciones entre partículas. Las curvas para las diferentes concentraciones fueron fusionadas mediante el programa PRIMUS, utilizando las medidas de baja concentración para la baja resolución y las de mayor concentración para mayor resolución.

5.2. TFAM en complejo con LSP22

Una vez resuelta la estructura cristalográfica de TFAM en complejo con LSP22, se creyó interesante analizar el comportamiento en solución del complejo. Para ello se recogieron datos de SAXS en ESRF, línea ID14-3, de muestras a diferentes concentraciones que oscilaban entre 0.5 mg/ml a 12 mg/ml. Para estas fechas la línea de BioSAXS del ESRF había sido ampliamente renovada, incluyendo aplicador automático de muestra, un nuevo detector, métodos de colección de muestra en continuo para evitar daño por radiación, señal estable y con atenuación óptima,... Al igual que para TFAM en su forma no unida, las curvas fueron fusionadas y analizadas mediante el programa PRIMUS.

FÖRSTER RESONANCE ENERGY TRANSFER (FRET)

1. Introducción a FRET

La técnica de FRET (*Förster Resonance Energy Transfer*) se ha desarrollado como un mecanismo basado en fluorescencia, potente para el estudio de procesos biológicos debido a que es posible usarlo en células vivas y presenta resolución a escala nanométrica, por ejemplo, es posible observar biomoléculas que adoptan conformaciones estructurales diversas para llevar a cabo su función.

FRET es un fenómeno físico en el que existe una transferencia de energía entre un fluoróforo donador y una molécula aceptora entre un rango de distancia de 1 nm a 10 nm. Esta técnica ofrece grandes ventajas por su elevada sensibilidad, observación a pequeña escala de tiempo (nanosegundos), y el rango de distancias para la técnica coincide con el rango de distancias donde ocurren la mayoría de procesos biológicos. Además, requiere de una simple base experimental y se puede aplicar la técnica a muestras diluidas. Recientemente, se ha desarrollado una metodología que permite la combinación de FRET con experimentos *single-molecule* o *single particle*, smFRET o spFRET. spFRET es una herramienta potente y sensible para determinar cambios conformacionales y interacciones moleculares.

1.1. Principio físico

El principio de FRET se encuentra descrito por una transferencia de energía no-radiativa donde participa un donador, fluoróforo excitado energéticamente, que transfiere energía a otra molécula situada a cierta distancia en el espacio, aceptor, mediante acople dipolo-dipolo (Fig1 (a)). El aceptor excitado vuelve a su estado basal mediante la pérdida de energía a través de la emisión de fotones, por ejemplo, emitiendo fluorescencia. En el caso de FRET, el estado basal de donador y el aceptor se encuentran energéticamente acoplados. Por lo tanto, FRET es un fenómeno donde dos procesos ocurren simultáneamente: en el donador existe una excitación de fotones del estado basal al excitado, y una transferencia de energía de donador al aceptor a través del acople bipolar entre la emisión del donador y la excitación del aceptor (Fig1 (b)). La eficiencia de FRET (E_{FRET}) es dependiente de la distancia entre fluoróforos (R), de forma inversamente proporcional:

$$E_{\text{FRET}} = R_0^6 / (R_0^6 + R^6)$$

donde R_0 es el radio de Förster o la distancia crítica a la que la eficiencia de FRET es del 50% y varia en función de los pares de fluoróforos escogidos. Para distancias menores a R_0 , la eficiencia de FRET es mayor; mientras que distancias mayores a R_0 , la eficiencia es mucho menor a 50%. (Fig1 (c)).

R_0 en solución acuosa está determinado por los siguientes parámetros:

$$R_0 = [8.79 \times 10^{-5} (\kappa^2 \eta^{-4} Q_D J(\lambda))]^{1/6} \text{ \AA}$$

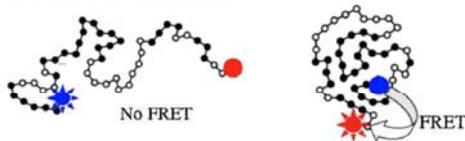
Donde κ^2 representa el ángulo entre los dipolos de los fluoróforos, Q_D es el rendimiento cuántico del donador y η es el índice de refracción del medio. $J(\lambda)$ es el solapamiento en el espectro entre la emisión del donador ($F_D(\lambda)$) y la absorción del aceptor ($\epsilon_A(\lambda)$) (Fig1 (d)), determinado por la siguiente ecuación:

$$J(\lambda) = \int F_D(\lambda) \epsilon_A(\lambda) \lambda^4 d\lambda / (\int F_D(\lambda) d\lambda) \text{ mol}^{-1} \text{ cm}^{-1} \text{ nm}^4$$

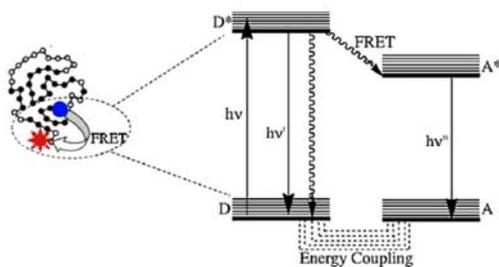
E_{FRET} se encuentra también determinado por la intensidad de la fluorescencia: I_{DA} , intensidad del donador en presencia del aceptor y I_D , intensidad de donador solo.

$$E_{\text{FRET}} = I_{DA} / (\alpha I_{DA} + I_D)$$

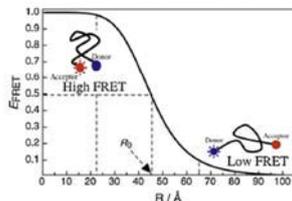
(a) Mecanismo de FRET



(b) Diagrama de Jablonski



(c) Eficiencia FRET



(d) Definición $J(\lambda)$

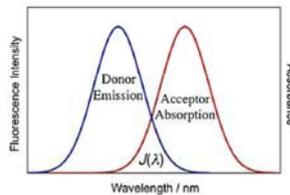


Fig1. (a) Mecanismo de funcionamiento básico de FRET, donde existe una molécula donador (en azul) y un aceptor (en rojo), cuando las dos moléculas se encuentran suficientemente próximas existe una transferencia de energía provocando la excitación del aceptor. (b) Diagrama de *Jablonski* donde se muestra la transferencia de energía existente y los diferentes estados electrónicos de las dos moléculas, aceptor y donador. Es posible observar el acoplamiento entre espines orbitales diferentes entre donador y aceptor. (c) Representación de la Eficiencia FRET en función de la distancia entre fluoróforos. (d) Representación del solapamiento en el espectro necesario para observar el fenómeno FRET. (Sahoo, et al.2011)

1.1.1. Limitaciones en FRET

La eficiencia del fenómeno FRET tiene diversos factores limitantes:

- $J(\lambda)$, el espectro de emisión del donador debe sobreponerse con el espectro de absorción del aceptor (Fig.1(d))
- Distancia entre fluoróforos, R : la distancia entre donador y aceptor debe estar entre un rango de distancia de ~ 1 a 10 nm. La eficiencia de FRET es inversamente proporcional a la distancia entre las sonda elevada a 6, por lo tanto, una pequeña modificación de R implica cambios grandes en E_{FRET} .
- Interacción dipolo-dipolo, κ^2 : el momento dipolo de emisión del donador, el momento dipolo de absorción del aceptor y sus vectores de separación deben encontrarse en óptima orientación. κ^2 depende del ángulo entre los momentos dipolo entre donador y aceptor, de la misma forma que una antena de radio puede afectar la señal de recepción. Si el donador y el aceptor se encuentra orientados de forma paralela uno respecto al otro, la eficiencia de FRET va a ser mayor que si se encuentran orientados de forma perpendicular. El grado de alineamiento define el valor κ^2 , que varía entre 0 (los dipolos se encuentran orientado de forma perpendicular) y 4 (los dipolos están orientados de forma paralela). Normalmente este valor es $2/3$, que es un promedio de todos los valores para los posibles ángulos cuando los fluoróforos se encuentran unidos a la molécula permitiendo su rotación libre.

Existen diversos métodos para medir FRET, como el cambio en la fluorescencia del donador, cambio en la fluorescencia en el aceptor, cálculo de la anisotropía, cálculo de la destrucción fotoquímica del aceptor, etc. Uno de los métodos más usados y comunes es la medida de FRET que involucra la monitorización de la emisión del aceptor después de la excitación del donador.

Cálculo de la anisotropía

Cuando una población de fluoróforos orientada de forma aleatoria es excitada con una luz polarizada, las moléculas cuyos dipolos de absorción/excitación se encuentran orientados de forma paralela al eje de polarización son preferiblemente excitados. La anisotropía resultante (r), es una medida del grado de orientación, determinada a partir de la medida de la intensidad de la emisión mediante haces de luz polarizados de forma vertical (I_v) y horizontal (I_H).

$$r = (I_v - I_H) / (I_v + 2I_H)$$

La depolarización ocurre como resultado de la emisión desde el aceptor durante FRET, por ejemplo, cuando el donador transfiere su energía de excitación al aceptor con una orientación diferente al plano de fotoexcitación original (Fig2).

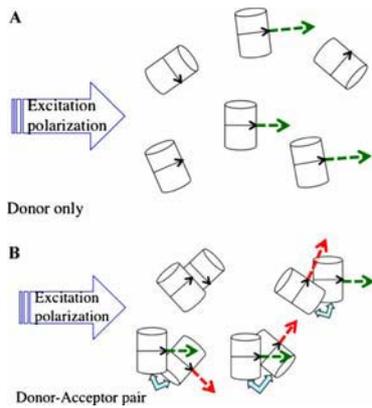


Fig2. Representación de la anisotropía existente dentro de la metodología FRET, cuando los dos momentos dipolo de los fluoróforos no se encuentran orientados de forma paralela. (A) En azul se representa el haz de excitación polarizado, con la molécula donadora y su emisión de excitación. (B) Haz de polarización original con las moléculas aceptora y su emisión (en verde), y la molécula donadora y su posterior emisión en rojo.

1.2. Aplicación de FRET en complejos proteína-ADN

En numerosos complejos formados por proteína y ADN, el ADN se encuentra deformado por la acción de la proteína. Esta deformación muchas veces se traduce en un doblamiento del ADN, donde es posible calcular el ángulo de curvatura mediante experimentos cristalográficos con resolución atómica. Recientemente, se está aplicando la metodología FRET para obtener el grado de curvatura del ADN en solución. Algunos ejemplos de complejos proteína-ADN caracterizados mediante FRET son el caso del estudio de IHF (*Integration Host Factor*) que mostró un ángulo de $\sim 160^\circ$ en concordancia con los resultados cristalográficos o el estudio del efecto de TFIIA en el complejo TBP/TATA, que implica una disminución en el ángulo de curvatura de 20° (TBP/TATA = 104° , TBP/TATA/TFIIA = 80°).

2. Single Molecule FRET (*spFRET*)

La técnica *single molecule fluorescence* puede ser fusionada con la metodología FRET. Ello ofrece ventajas respecto a la medida estándar de FRET. Por ejemplo, en la medida estándar de FRET el resultado final es un promedio de las propiedades de la muestra; mientras que los experimentos de molécula única permiten construir un histograma, de la distribución de las moléculas individuales en función del tiempo. Además, el uso de un volumen pequeño reduce el ruido de la señal debido a una menor dispersión del medio.

2.1. Diagrama del sistema

Para realizar un experimento de *spFRET* usando la *fluorescencia single molecule* se pueden necesitar dos equipamientos claves: un microscopio confocal y el sistema óptico formado por espejos, láser, filtros y amplificadores de señal. La emisión de la muestra es colectada con el mismo plano de excitación y reflejada hacia una *beam splitter* (*dichroic mirror*) para llegar a los detectores (APD). Antes de los detectores, existe un *slip mirror* para desviar la emisión a través de la cámara (ej, EMCCD) para colectar las imágenes de donador y aceptor (Fig3).

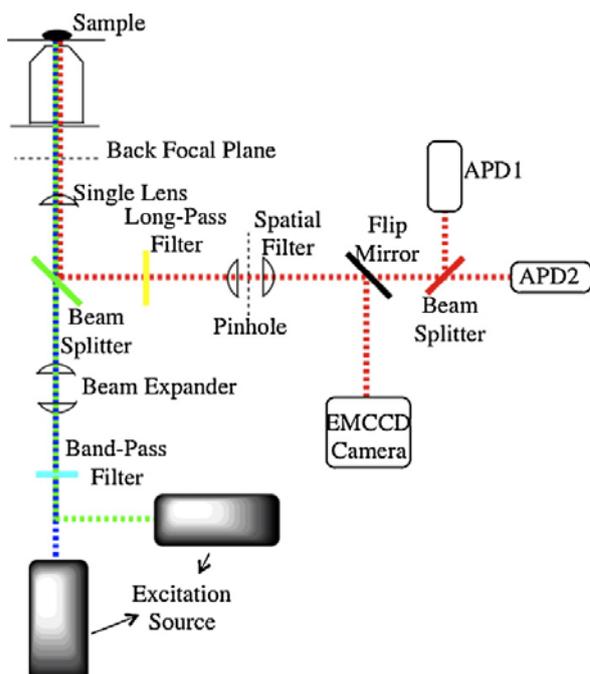


Fig3. Esquema del diagrama de spFRET. El interlineado verde y azul representa la luz de excitación con diferentes longitudes de onda y el interlineado rojo representa la trayectoria de la emisión. El haz excitado es después redireccionado mediante el uso de beam splitter y luego es reflejado a los detectores (APD, amplificadores de la señal) y la cámara (EMCCD). (Sahoo.H., et al.2011).

2.2. Limitaciones del sistema confocal de fluorescencia para spFRET

El perfil de excitación está definido por la intensidad en función de la distribución de fotones y presenta un perfil típico gaussiano. Dentro del sistema confocal y la detección de la fluorescencia existen varios factores limitantes a tener en cuenta cuando se realizan experimento spFRET.

2.2.1. Detección de la fluorescencia

Los fotones fluorescentes son emitidos de forma isotrópica en todas las direcciones, pero los sistemas ópticos o lentes no pueden coleccionar datos en todas las direcciones, factor limitante conocido como *numerical aperture*. Sólo un 25% de los fotones son emitidos a través del objetivo.

La función de detección molecular o MDE (*Molecular Detection Function*) describe la probabilidad de que un fotón generado sea capturado y depende del límite de saturación (es decir, de la difusión), intensidad del fotón y volumen efectivo. Este parámetro MDE depende directamente de diámetro del objetivo. Si el diámetro es demasiado grande, el ruido es mayor que la señal; mientras que si el diámetro es demasiado pequeño, la señal disminuye (Fig4).

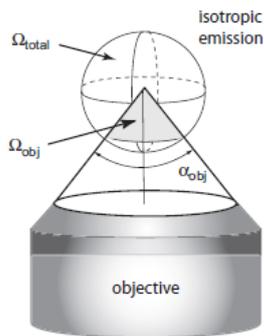


Fig4. Definición de la eficiencia de colección geométrica de la lente del objetivo, Ω_{obj} , frente a la colección total, Ω_{total} ; en un sistema de emisión isotrópica. (*Dissertation_Gansen A_2008*)

2.2.2. Burst size distribution

La distribución de la suma de fotones detectados (*burst size*) en función del tiempo depende de la variación espacial de MDE y de la distribución de trayectorias de difusión: una molécula que difunde en la periferia del *focus* emitirá menos fotones que una molécula pasando a través del centro del *focus*.

2.2.3. Background o ruido de fondo

Mientras que el fluoróforo presenta una irradiación exponencial con un plateau, el ruido presenta una escala lineal con la intensidad. El ruido es proporcional al tamaño del volumen de excitación.

2.2.4. Presencia de eventos múltiples

En experimentos confocales es difícil excluir la presencia de más de una partícula en el *focus*. Para reducir la presencia de eventos multi-partícula se debe reducir el número de partículas en el experimento (N) mediante diluciones, aunque ello implique aumentar en tiempo de análisis.

La probabilidad de encontrar n moléculas en el *focus* en un periodo de tiempo depende del tiempo de difusión de las partículas. Si consideramos un tiempo de difusión demasiado largo, los eventos consecutivos pueden ser considerados como eventos de una sola partícula. Para evitar problemas de eventos multi-partícula pueden ser prevenidos mediante el cálculo del tiempo de difusión medio (T_m) dependiente de:

- Tiempo difusión
- Tamaño del volumen
- Concentración de la partícula

Se deben escoger las condiciones experimentales idóneas para reducir T_m .

2.3. Cálculo de la Eficiencia de FRET en spFRET

Para experimentos de *single molecule*, la eficiencia de FRET se calcula a través del número de fotones detectados en el canal del donador y aceptor. El número de fotones bruto está formado a partir de diversas contribuciones que deben ser corregidas en el proceso conocido como *crosstalk correction*, donde se eliminan las contribuciones debidas a la excitación directa de aceptor y al ruido. La larga cola en la longitud de onda del espectro de emisión del donador resulta en un *crosstalk* del donador en el canal de detección del

aceptor. El ruido no es negligible bajo condiciones *single molecule*, donde sólo una fracción de los fotones detectados dan lugar a fluorescencia.

Por lo tanto, bajo condiciones de saturación, una señal de ráfaga única (*single burst signal*) puede ser definida en ambos canales mediante:

1. Señal en el canal del donador:
señal de emisión sobrante del donador + ruido + *crosstalk* del receptor en el canal de donador

$$I_D^{det} = \eta_D \sigma_D \Phi_D I_0 (1 - E) + B_D + \eta_{AD} \Phi_A \sigma_A I_0$$

2. Señal en el canal del receptor:
emisión del receptor debido a FRET + emisión debida a la excitación directa del receptor + ruido + *crosstalk* del donador en el canal del receptor

$$I_A^{det} = \eta_A \sigma_D \Phi_A I_0 E + \eta_A \Phi_A \sigma_A I_0 + B_A + \eta_{DA} \sigma_D \Phi_D I_0 (1 - E)$$

η_D o η_A = eficiencia de detección del donador o receptor
 η_{AD} o η_{DA} = eficiencia de detección para *crosstalk* del receptor en el donador o el inverso
 $\Phi_{D/A}$ = rendimiento cuántico de fluorescencia para donador y receptor
 $\sigma_{D/A}$ = absorción cruzada del donador y receptor

Si suponemos que E_{FRET} es función de las intensidades en el canal del donador y el receptor. Y el hecho que el *crosstalk* del receptor en el donador es casi 0 y la absorción cruzada del donador es mínima. Es posible simplificar ecuaciones:

$$E = \frac{N_A}{N_A + \gamma N_D}$$

intensidad en el canal de receptor $N_A = (I_A^{det} - B_A) - \frac{\eta_{DA}}{\eta_D} \cdot (I_D^{det} - B_D)$
 intensidad en el canal del donador $N_D = (I_D^{det} - B_D)$
 factor de detección del sistema $\gamma = \eta_A \Phi_A / \eta_D \Phi_D$

Proximity ratio vs E_{FRET}

Existe un parámetro que permite simplificar la ecuación que es *proximity ratio*, P , que mantiene una relación matemática estrecha con E_{FRET} .

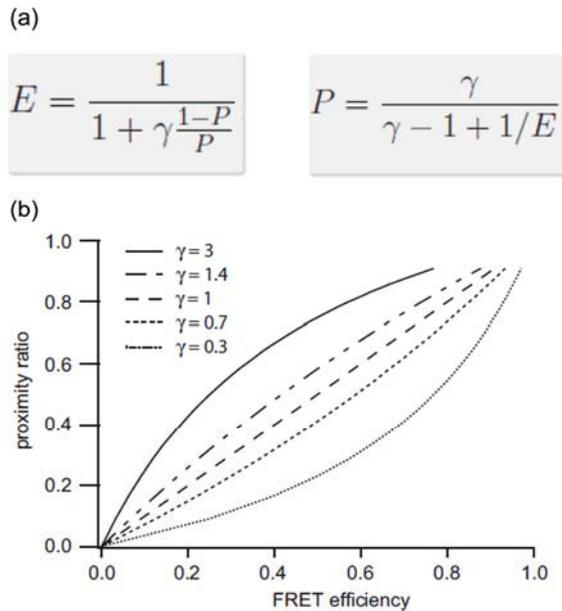


Fig5. (a) Relación matemática entre *proximity ratio* y E_{FRET} (b) Implicación del valor γ entre *proximity ratio* y E_{FRET} . (Dissertation_Gansen A._2008)

La relación entre P y E_{FRET} , depende del factor de detección del sistema (γ), definido por la eficiencia de detección y el rendimiento cuántico de donador y aceptor:

- Para valores $\gamma < 1$, valores bajo de E_{FRET} implican intervalos pequeños de P .
- Para valores $\gamma > 1$, valores bajo de E_{FRET} implica un intervalo más separado en P .
- Para valores $\gamma = 1$, $E_{\text{FRET}} = P$

SINGLE PARTICLE FRET (spFRET) para TFAM en complejo con LSP

1. Diseño del experimento

Para llevar a cabo los experimentos de FRET se diseñaron dos oligos con diferentes longitudes, 30 y 50 pb, en base a modelos realizados a partir de la estructura cristalográfica que muestra una curvatura de 180° . Por lo tanto, a partir de la secuencia de 22 pb usada para la cristalización situada en la zona central se fueron añadiendo bases flanqueante usando la secuencia nucleotídica original de mtADN. El ADN de 30 pb permitirá llevar a cabo los experimentos para dilucidar la E_{FRET} , permitiendo así obtener la distancia entre los fluoróforos y el cálculo final de curvatura. Mientras que, el ADN de 50 pb nos permitirá dilucidar si se trata de un sistema con dos puntos de curvatura como describe la estructura cristalográfica (forma de U) manteniéndose un valor de E_{FRET} similar entre los dos ADNs de diferente longitud (30 y 50 pb) o un sistema con un solo punto de curvatura (forma de V), con valores de E_{FRET} distintos para cada ADN.

Como control positivo se usaron dos ADN ideales de 30 pb con los fluoróforos Alexa situados a 10 o 21 pb para cada uno (más o menos 1 o 2 vueltas de hélice). Como control negativo se usaron ADN de 50 y 30 pb marcados solo con aceptor o sólo con donador, así como ADN de 42 pb marcado con molécula aceptora y donador en los dos extremos.

Los fluoróforos escogidos para el experimento de FRET fueron: Alexa 488, con absorción en el canal verde del espectro ($\lambda = 400-550\text{nm}$) (pico absorción 492 nm, pico emisión 517 nm) y Alexa594, con absorción en el canal rojo ($\lambda = 450-640\text{nm}$) (pico absorción 584 nm, pico emisión 616 nm). Las sondas Alexa presentan características optimizadas respecto a otras sondas fluorescentes como TAMRA, Cianina, rodamina,... Los fluoróforos Alexa presentan un elevado rendimiento cuántico, largo tiempo de vida media y valores elevados de eficiencia de FRET (Fig6).

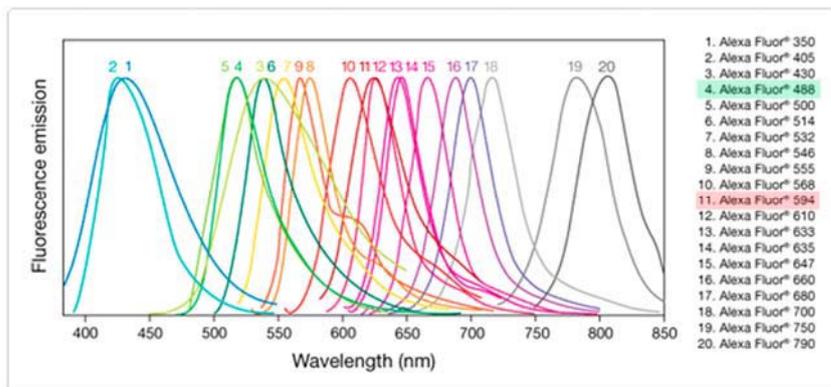


Fig6. Espectro de absorción de los fluoróforos comerciales Alexa, donde se indica los fluoróforos usados para este estudio: Alexa 488 (sombreado en verde, número 4) y Alexa 594 (sombreado en rojo, número 11).

Normalmente es mejor situar los fluoróforos en la cara externa de la curvatura del ADN. El fenómeno FRET puede ser medido de 20 a 100 Å. Para permitir la completa libertad de la sonda, es decir, la flexibilidad en la rotación de las sondas, los fluoróforos son anclados al ADN mediante regiones conectoras o polyCarbonos, normalmente del tipo C3 o C6 con longitudes de 6 o 13 Å. La construcción de modelos de ADN a partir de la estructura cristalográfica con la curvatura de 180°, permite observar que la distancia entre las bases de los extremos, situadas en la cara interna del ADN, es entre 20 a 30 Å. Por lo tanto, colocar las sondas en la cara interna del ADN provocaría que las sondas estuvieran demasiado cercanas, siendo posibles fenómenos ópticos que perturbaran las medidas. Finalmente, se optó por colocar las sondas en la cara externa del ADN, pese al peligro de posibles contactos con la proteína. Previamente se analizó el efecto de apantallamiento de los residuos aromáticos de la proteína en los fluoróforos del ADN. Sólo en el caso de Tyr103 y Trp107 existe una distancia crítica entre los residuos y la sonda Alexa 488.

Para simplificar el ambiente de la sonda y su comportamiento, los fluoróforos fueron conjugados siempre en Timinas, por lo que fue necesario mutar la secuencia de ADN de A a T en los extremos (Fig7). El polylinker o conector usado para añadir la sonda fue C6.

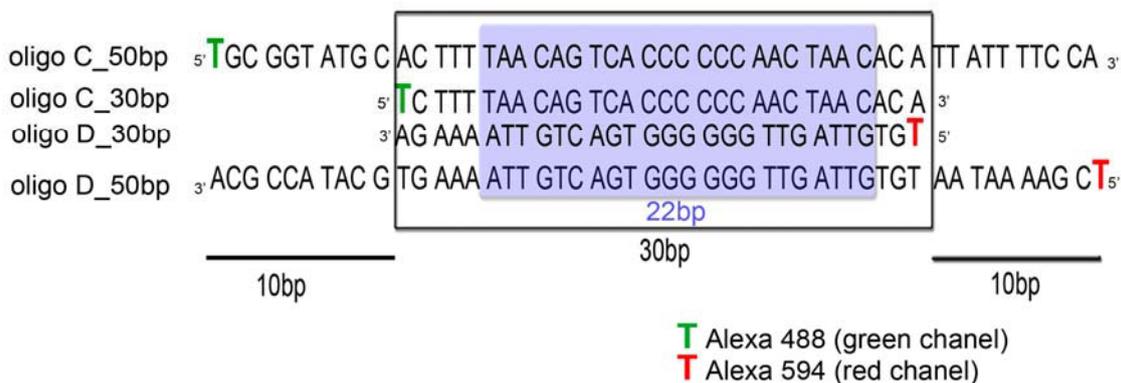


Fig7. Esquema de los ADN de 30 pb y 50 pb usados para el estudio, donde se indica el oligo C (rico en citosinas) y la hebra complementaria, oligo D (rico en guaninas). Sombreado en azul se encuentra la región de 22 pb usada para la cristalización, LSP22. El oligo C fue marcado con el fluoróforo Alexa488 en el extremo 5' y el oligo D marcado con el fluoróforo Alexa 594 en el extremo 3'.

Los oligos fueron obtenidos a partir de la casa comercial IBA (*IBA Naps*, Göttingen, Germany), purificados mediante dos pasos consecutivos de HPLC que permiten obtener una gran calidad de oligo con 100% marcado con las sondas fluorescentes a una concentración final de 0.2 μ M.

2. Características de los oligos y purificación HPLC del dúplex LSP

Los oligos C (cadena rica en citosinas) y oligo D (cadena rica en guaninas) fueron resuspendidos en tampón TE. Posteriormente se chequeó su espectro de absorción y la migración en un gel nativo PAGE 12%. La migración de los dos oligos era muy diversa, en el caso del oligo D existe una migración de bandas múltiples (Fig8). En un principio se planteó la posibilidad que el oligo D presentara impurezas, pero análisis posteriores demostraron que la presencia de múltiples bandas era debida a la presencia de múltiples conformaciones. Las diferentes conformaciones para el oligo D eran debidas a la presencia de un grupo de 6 guaninas seguidas que dan lugar a formación de supraestructuras de ADN, conocidas como *G quadruplex*.

El análisis de la capacidad de unión de TFAM a las cadenas únicas u oligos, permitió observar que TFAM puede formar complejo con las bandas superiores presentes en el oligo D, mientras que en el caso del oligo C, TFAM no presenta unión. Por lo tanto, fue necesario purificar los oligos una vez apareados en forma de cadena doble, para evitar la presencia de *G quadruplex* (Fig8).

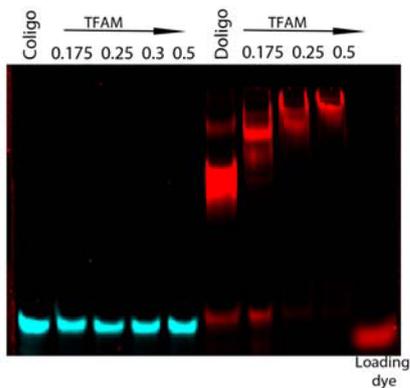
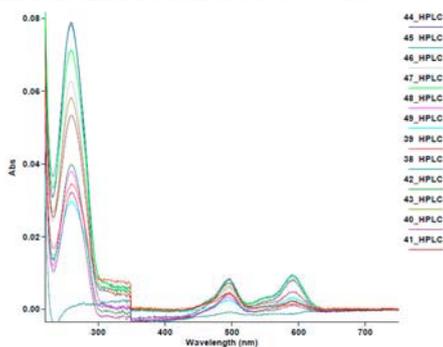


Fig8. Análisis de la migración de los oligos C y D, y la capacidad de unión de TFAM a los oligos por separado mediante concentraciones crecientes de TFAM. Es posible observar que para oligo C no existe unión de TFAM, mientras que para el oligo D existe unión que da lugar a bandas superiores. Para ello se usó un gel nativo del 12% PAGE y el equipo Typhoon (Amersham Biolabs) que permite usar diferentes longitudes de onda para la visualización, que coinciden con las longitudes de onda de excitación y emisión para oligo C y oligo D.

La purificación se realizó mediante un equipo de HPLC usando un gradiente lineal de NaCl (de 20 mM a 1000 mM). Posteriormente las fracciones eran analizadas en el espectrofotómetro para ratificar la presencia de los fluoróforos y se controlaba la migración en gel nativo de las diferentes fracciones (Fig9). Una vez seleccionadas las diferentes fracciones se purificaron mediante una columna de *desalting*, para reducir la concentración de NaCl.

(a) Espectro de las fracciones para el duplex C'D'_30pb



(b) Gel nativo 12% de los diferentes duplex y oligos

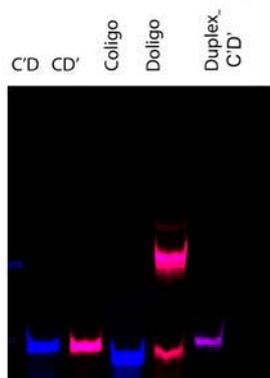


Fig9. (a) Espectro de las diferentes fracciones eluidas en HPLC, donde es posible ver los dos picos de emisión para Alexa 488 a 488 nm y Alexa 594 a 594 nm. (b) Gel nativo al 12% con los diferentes dúplex formados: C'D,

dúplex con oligo C marcado y oligo D simple; **CD'**, dúplex con oligo C simple y oligo D marcado; **C'D'**, dúplex con oligo C y D marcados.

3. Formación del complejo y preparación de la muestra TFAM-LSP para spFRET

Para la adquisición de datos es necesario obtener un complejo lo más homogéneo posible, u para ello lo primero es realizar experimentos de titración para optimizar el ratio de proteína y ADN. Además es necesario ajustar y estabilizar el complejo para las medidas de spFRET.

Para la formación del complejo se usó la misma metodología que para la formación del complejo para la cristalización mediante diálisis por gradiente de 750 mM NaCl a 100 mM NaCl. La concentración final era de 100 mM NaCl para intentar evitar la formación de bandas superiores de complejo, posiblemente asociadas a formas oligoméricas. Para los ensayo de FRET fue necesario modificar dos componentes del tampón original de purificación de la proteína: HEPES pH 7.5 fue modificado por TRIS pH 7.5, debido a que HEPES presenta absorción en el espectro. El otro componente modificado fue la concentración de DTT de 5 mM a 1 mM, ya que se ha descrito que DTT a elevadas concentraciones puede provocar la destrucción fotoquímica del fluoróforo.

Para la preparación de la muestra para el experimento spFRET fue necesario utilizar un tampón especial usado como blanco y para la dilución de las diferentes muestras, ya que los experimentos de spFRET requieren el uso de concentraciones del orden pM (~20-50 pM). Este tampón contenía: 50 mM Tris pH 7.5, 100 mM NaCl, 1 mM DTT, Vitamina B₁₂ (5µl por 1ml final de tampón) y NP-40. La vitamina B₁₂ permite evitar el posible daño fotoquímico de los fluoróforos durante el experimento; mientras que NP-40 permite estabilizar el complejo que se encuentra a concentraciones tan inferiores. El tampón se prepara justo antes de las medidas y es filtrado usando filtros con poro de diámetro ~ 0.1µm.

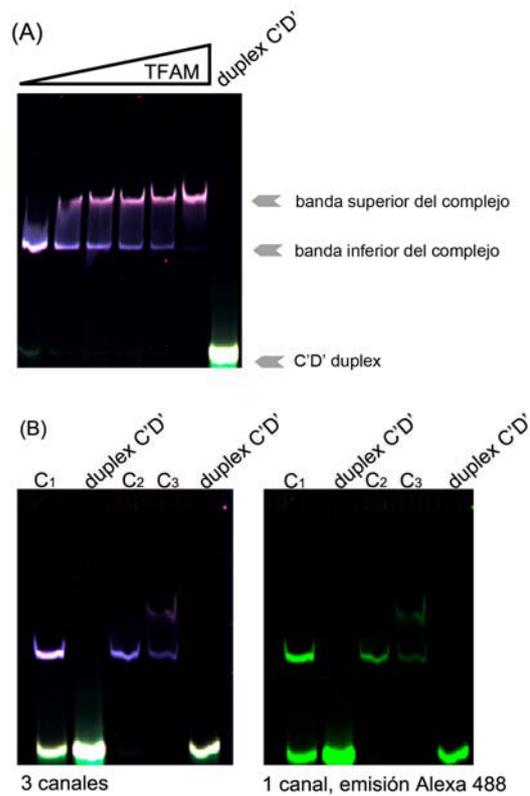


Fig10. Migración de complejos TFAM-dúplex 30pb en geles EMSA al 12% acrilamida y visualización mediante *Typhon* en los tres canales: canal de emisión directa de Alexa 488 (en verde), Alexa 594 (en rojo) y emisión de Alexa 594 por FRET (en morado). (A) Titración del complejo TFAM-30pb manteniendo una concentración fija de ADN frente a una concentración creciente de TFAM, donde es posible observar que existe una primera banda de retardo que desaparece en base a añadir TFAM, y la formación de una banda superior debido a la presencia de estados oligoméricos. (B) (C) Imagen del mismo gel con revelados en los tres canales o sólo en el canal de la emisión directa de Alexa 488 (en verde), respectivamente. Los complejos C1, C2 y C3 muestran una concentración creciente de TFAM donde es posible observar que para C1 existe dúplex sobrante, en el caso C2 el complejo parece perfecto y para C3 existen múltiples bandas de complejo.

Mediante las primeras adquisiciones de datos se pudo constatar que existía FRET pero con una señal de intensidad baja, por ello se decidió usar un complejo mixto que consistían en mantener el ratio de proteína y ADN marcado con fluoróforos, y añadir complejo proteína:ADN sin marcaje. Ello permitía estabilizar el complejo y mejorar la señal de FRET.

El mismo procedimiento se aplicó para los mutantes de TFAM: T77, L182, Q179, N163, Y162; y los dominios D1 y D2.

4. Adquisición de los datos

La adquisición de datos se realizó a través de un complejo sistema de detectores, espejos y amplificadores de señal. A partir de los impulsos detectados es posible obtener un voltaje que dará lugar a una señal digital en código binario (descrito como *.t3r). Cada fotón queda registrado con la siguiente información:

- Un número router que identifica los fotones para el canal donador y el canal aceptor
- El tiempo de recogida, ~ 100 ns

- El tiempo de llegada de la pulsación

Previo a la adquisición de datos es necesaria la calibración del sistema mediante el uso de una solución de Alexa 488 20nM, que permite calibrar el sistema hasta obtener valores de función de autocorrelación y *count rate* correctos. Para experimento de spFRET con celdillas de $1\mu\text{m} \times 1\mu\text{m}$, el tiempo de difusión debe ser alrededor de 52 sec y N (número de partículas) similar a 10. Para realizar medidas de tiempo total de 600 sec.

5. Procesamiento de datos

Los datos de un experimento de spFRET se recogen en forma de fichero binario, que contiene información de los millones de fotones medidos en cada experimento. De forma consecutiva, los datos tienen que ser reducidos para que solo sean procesados los fotones que pertenecen a eventos de partícula única, *single molecule*. El primer paso de reducción y procesamiento de datos se realizó con el programa FRETtchen, desarrollado por el grupo del Prof. Jörg Lagowski en Heidelberg. Posteriormente los datos fueron analizados y visualizados con los programas IGOR PRO (Wavemetrics, OR, USA) y ORIGIN, descritos a continuación.

5.1. FRETtchen

El programa de análisis FRETtchen permite la reducción de datos y la identificación de *single burst* o ráfagas únicas. Por lo tanto, mediante el uso de este programa es posible pasar de fotones individuales colectados en un fichero binario (*.ter) a mostrar distribuciones del número de eventos y su duración en forma de *burst rate*, *burst duration* y la distribución de proximity ratio en forma de fichero (*.txt). Un experimento de *single molecule event* se caracteriza por tres parámetros a definir:

- *Burst duration*, es calculado como la diferencia de tiempo entre el primero y último fotón del evento. Permite analizar cambios grandes en las propiedades de difusión. Generalmente *burst duration* es considerablemente mayor que el tiempo de difusión medio de la molécula.
- *Burst size*, suma de todos los fotones en una ráfaga. La mayoría de los eventos presentan un bajo número de fotones ya que muchas moléculas transitan el *focus* en las regiones periféricas emitiendo menos fluorescencia. La distribución de *burst size* sigue una curva exponencial.
- *Photon rate per burst* es determinada por *burst size* dividido por *burst duration*. Depende de la cantidad de emisión i de la calidad del fluoróforo.

En un experimento de *single molecule*, los parámetros *burst duration* y *burst site* se encuentran correlacionadas (Fig11). En una representación bidimensional de *burst duration* respecto *burst size*, la mayoría de moléculas o eventos se sitúan agrupados y paralelos al eje Y. Los eventos multi-partícula normalmente presentan *burst size* y *burst duration* peculiar, y pueden ser eliminados mediante la modificación del parámetro *burst rate* dentro del programa FRETtchen.

Tanto la E_{FRET} como P (*proximity ratio*) son calculados a través de *burstwise* para analizar la distribución de la distancia entre fluoróforos. De esta forma el programa FRETtchen construye histogramas que son una función de distribución del *proximity ratio* vs el número de eventos, es decir, el número de eventos que se presentan a una distancia determinada entre fluoróforos. Estos histogramas pueden revelar la presencia de

subpoblaciones dentro de la muestra, dando lugar a distribuciones heterogéneas con múltiples picos o máximos de curvas gaussianas. A partir de la distribución con forma de curva gaussiana es posible obtener diversa información (Fig11): la **posición central** de la campana de gauss hace referencia al valor de E_{FRET} o P para la subpoblación, es decir a la distancia entre donador y aceptor. El **área de la campana** es la medida del porcentaje representativo de esta conformación en la muestra. Finalmente, la **anchura** ofrece información sobre la heterogeneidad de la subpoblación, es decir, si se trata de una conformación dinámica o más bien estática.

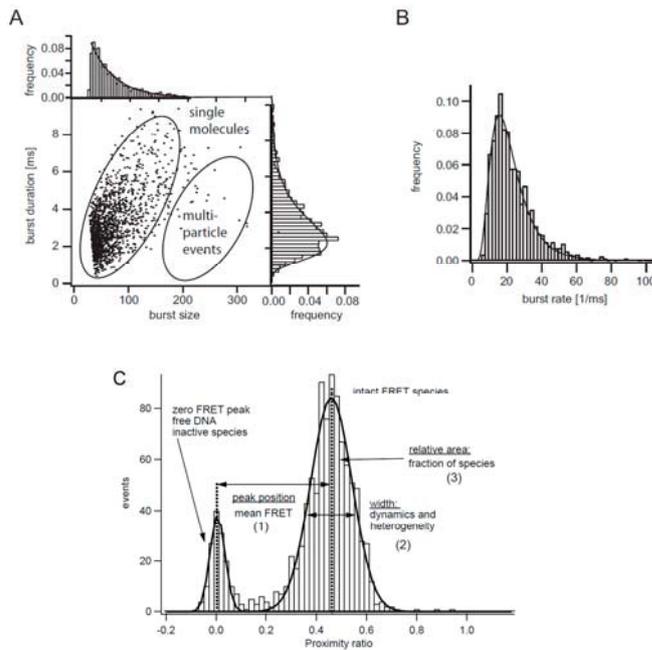


Fig11. Análisis de datos por parte del programa FRETtchen. (A) Diagrama 2D mostrando la relación entre *burst size* y *burst duration*. Para especies únicas o eventos de partícula única ambos parámetros presentan una correlación. Mientras que los eventos de multi-partícula suelen mostrar un *burst size* mucho mayor. También se muestra la distribución de estos parámetros por separado. (B) Distribución de *burst rate*, que es calculado como el ratio entre *burst size* y *burst duration*. (C) Típica distribución para el valor *proximity ratio* y la información que caracteriza estas distribuciones. (*Dissertation_Gansen A._2008*)

La segunda fase después de la obtención de la distribución de *proximity ratio* es realizar dos correcciones en los datos: una de ella es corregir el ruido de la distribución de forma bastante automatizada, y la segunda corrección es llamada *crosscorrelation* o *crossstalk correction*. *Crosstalk correction* consiste en analizar la medida de la muestra sin FRET pero con presencia de los dos fluoróforos, es decir, en nuestro caso la medida del ADN marcado sin presencia de proteína. Será necesario tomar el valor de *proximity ratio* para esta medida y establecerla como el blanco o zero para editar los demás datos.

5.2. Programa IGOR y ORIGIN

El programa IGOR y ORIGIN permiten realizar un análisis gráfico y estadístico de los histogramas.

AGRADECIMIENTOS

L'ocasió de poder treballar en aquest fascinant projecte ha estat meravellosa i de vegades massa difícil. Primer de tot, m'agradaria agrair a la Jasmin, la noia que va començar amb el món TFAM i em va deixar un munt de llibretes i notes que m'han ajudat durante aquests anys. Sense ella no hagués estat capaç d'avançar en aquest projecte. En següent lloc m'agradaria agrair a la meva directora de tesi, Maria Solà, per donar-me la possibilitat de poder treballar amb aquest projecte.

Científicament també m'agradaria donar les gràcies a en Miquel Coll per la disponibilitat d'espai i els group meetings plagats d'idees. Al Jordi Bernués pel seu suport i la Dra. Katy Tóth per l'acollida a Heidelberg i oferir-me la possibilitat de treballar al seu grup. Finalment, m'agradaria donar les gràcies a en SAXSman, Pau Bernadó, per esenyar-me i oferir-me el teu temps davant de qualsevol dubte. Encara et recordo dins del despatx explicant con en Pauet feia "el campinyooooonssss". A la Jenny Colom i la Isabel per l'ajuda amb els equips de purificació. A la Sònia, Yuyu i l'equip de la PAC per facilitar-nos la vida cristal·logràfica.

I ara vull enumerar a tots aquells que dia a dia m'heu fet riure i feliç. Al vostre costat he madurat tots aquets anys i he disfrutat molt amb tots vosaltres. La petita família dels CRIS.

Als antics CRI1, quan vaig arribar em vaig trobar amb un grup de gent excepcional. M'agradaria enumerar-vos un per un. Leo, *la divina proteosomal*, gracias por tus bailes locos y tus sabios consejos. La llegada de Elsa ha cambiado tu mirada. Yuyu, gracias por estar siempre disponible para tomarnos unas cañas en Paco, cool-off,...por las confesiones y por nuestros momentos de cotilleos locos. Mailys, la francesa más viajera y hippie. Porque vivir contigo fue una experiencia corta, pero única. Lionel, a comerrrr! Gracias por tu ayuda científica y las cervezas por Gràcia. Robert-Roberto, gracias por tus bailes y nuestras conversaciones de todo y sobre todo. Esther, per ser la noia amb més ganes de música al labo i els cafetonets. Tom, por nuestras risas y búsquedas de google locas. Solete, per ser la persona més divertida després d'una birra. Esther Peña per la teva ironia, perque ets una gran científica. A la Naybe i en Friki-Dani pels riures mirant vídeos frikis. A la Vane pels concerts de Pedro Guerra. A la Silvia per l'ajuda en cada sincrotronada. A l'Albert per la seva disponibilitat. A en Radek i Roeland per les converses de camí al sincrotró.

A CRI2 per l'alegria que em dóna pujar les escales i xerrar amb vosaltres. A la Maria i la Rosa, per ser les "rubies" més divertides i per estar sempre disponibles per escoltar els meus problemilles. Mari!! no t'oblidis que tu lo vales!. A la Babi, la deessa més divertida i a la vegada responsable. Gràcies pel teu arrós amb llet i la teva complicitat. Espero que el nou camí professional sigui fantàstic. A en Xavi, per ser el príncep Beckelar més eco-hippie i enrollat, gràcies per aguantar les meves bromes. A la Query per ser la persona més positiva y preparar la millor empanada gallega. A en David-pringao, per ser la persona més tranquila i pacificadora. A la Mercè, per mantenir-me en forma i en bon ritme. Lukiiii, mi italiano, tu no podías faltar. Gracias por los buon giorno y por el viaje a Warwick que me permitió conocerte.

A CRI6 pels últims anys compartits. A vosaltres mil gràcies per oferir-me aire fresc i riures cada cop que m'he sentit atrapada. A la Cris, la mare més moderna i enrotllada, i la Rubia que més controla al labo. Ah! I la que porta millor modelets. Al Pablo-Nuevo, me encanta la forma en que cada dia apareces por el labo con tu sonrisa, sobretodo después de comer. A la Celia Lujaaannn, wathever! Por tu alegría y la bocanada de aire fresco. A en Jordi, ataque pokemón, gràcies per les teves abraçades i les teves mirades de complicitat. Espero poder celebrar una festa de disfresses ben aviat. A la Laia per ser una noia valenta i plena d'alegria, perquè sempre tens un somriure preparat. Ah! I ens espera un viatge fantàstic. A en Damià, sempre t'ho he dit...hauria d'haver arribat abans al laboratori! Gràcies per un tot molt global, per les converses causals d'una hora i les idees per fer dolenteries. Espero que els nous projectes et vagin igual de bé que l'experiència de ser pare.

Als quasi CRI'S, Miguelote por unas fiestas de Huesca únicas (qué le estará pasando al pobre Miguel). A en Josep de la Yuyu per ser un crack duracel i la teva vitalitat. Al coronel Pepov, per al més divertit de tots i un gran trekking al Nepal. A l'Oriol per ser un frikifísic. A en Bladi por tu dulzura. A en Paco-Paco, pels divendres, dijous, dimecres...vaja per tots els dies que ens has rebut amb cançons de Sabina i acudits. A en Tiago, Tiby, Núria cerdà i Sergio per ser el CRIS hèlix més molnguis.

Als meus companys de labo. Quan vam començar erem un petit grup de *mitonenes*, que ara ha anat augmentant. Pablo gracias por las risas y gracias por tu ayuda científica. Estoy muy segura que vas a llegar muy lejos, vales mucho. Oskar, tu estancia fue breve pero nuestras guerras de narices únicas. Cuppi, Cuppi ánimo en el camino que queda y haz el favor de ir a las clases de danza los viernes. Reicy, por ser la más dulce del grupo. Arko, thanks for the chance to know you and I hope to enjoy still more jazz concerts with you. Seb-Duomo por ayudarnos a crecer científicamente cada día. I finalment a la meva Cris-paleta, gracias por tu ayuda día a día, por tus ganas de facilitarme los días difíciles y las miradas de complicitad. Sin tí los dos últimos años hubieran sido muy diferentes.

I dins de l'ambient del labo no podia deixar d'agrair a la meva gran amiga i companya. Nere, trobar-te m'ha fet créixer i madurar molt. Gràcies pels kit-kats, les sincrotronades, els balls al labo i les ganes de cuidar de tots. El peixet serà un privilegiat amb aquests dos pedasso de pares.

A les nenes de Heidelberg per ser genials i oferir-me la possibilitat de disfrutar de tres mesos a Deutschland. A l'Oriol per la teva generositat, gràcies per compartir els teus amics amb mi, i sobretot gràcies per les llargues converses. Ah! I perque encara ens queden moltes jams de swing per millorar els quicks.

A en Juanma, per ser el fidel amic que escolta i explica el què sent. Crec que ets una de les persones que millor em conceix. A l'Arnau, perquè m'estàs ajudant a créixer. A la Clara per ser un budha espiritual. Als escaladors pels caps de semana de riures i escalada. A la gent d'Erasmus per les serates de vi. Als amics de Sant Pol, pels estius de la infantesa més divertits! cada estiu era una nova aventura. I el veïns de Sant Andreu per les converses a través del pati i els berenars.

A les nenes del cole: Mireia, Anna, Maite i Anni. Per ser incondicionals, per ser-hi sempre encara que els anys i les experiències ens distanciïn en forma de ser. A la Mireia per la teva

alegria, a l'Anni per ser la fidel confident, a la Maite per la seva tendresa i protecció, i a l'Anna per ser tan única. Han quedat enrera els dies de colònies i posteriors viatgets; anem madurant i espero que la nova etapa en la que us endinseu sigui la més feliç i radiant. Perquè heu trobat uns companys de camí que estic segura us cuidaran com us mereixeu (Dani, Roger, Lauri i Oriol).

A les lobes, Núria i Natàlia per les quedades on solucionem el món i un viatge a Eivissa inolvidable. A tu Nú per fer-me sempre costat, perquè encara ens queden tantes coses per aprendre i espero sempre tenir-te ben aprop.

Als de la Uni (Dele, Aida, Montxe, Aaró, Cris, Gemma, Núria, Larry, David, Bea, Albertot, Sandra, Sara) per ser el grup més dinàmic, per necessitar mil mails i cadenes per acabar organitzant un sopar. Pels viatges, sopars i Setmanes Santes temàtiques, les hores al bar de facultat i ferrocarrils. Perquè sou una part molt important de la meua vida i m'agrada pensar que en uns mesos tindrem una petitona Alba entre nosaltres i així podrem deixar de parlar de gats. En concret, a l'Aida, Marc, Montse, Aaró i Cris pels kebabs improvisats i els temps que passem junts. Dono gràcies per haver-me decidit a estudiar bioquímica i recordar quan xutava papareres al entrar a classe o practicava operacions a ratolins en estats... ejem ejem. A tu Aida, per la teva sinceritat. Seràs una mare brutal!. A la Montxe per ser l'organitzadora de les mogudes. A la Cris per la teva comprensió i tendresa. Al Marc i l'Aaró per ser els tio/tia de les nostres quedades de bruixes. Ah! I a la Gemmota per ser la Sherpa més molongui i més bruixeta de tots. Gràcies per les nits sevillanes i per ser cop d'aire fresc en un moment difícil.

Finalmente, a la meua família. A la Maribel i en Valentí, per l'acollida i per sentir-me tan estimada entre vosaltres tots els dissabtes durant uns anys. Als hermostissssmos, por ser tan acogedores. Tete, per ser el meu segon pare, per cuidar-me ja des de que era un cigronet i per fer-me riure. Crec que gràcies a tu he après a fer bromes i fugir dels cops que de vegades ens ofereix la vida. Al papa perquè setmana darrera setamana em soprén amb la força per seguir endavant. Papa, es por ti que esta profesión me fascina. Porque la razón de mi pasión por la ciencia es el estímulo de encontrar solución a las batallas que nos encontramos durante el camino llamado vida. A la mare per ser-hi SEMPRE. Per la teua vitalitat, posar-te unes raquetes i caminar per la Vall de Núria en plena nevada, ballar a casa mentre estàs plaxant roba o les nits de lluna plena al balcó de Sant Pol. Ets la millor persona que conec.

BIBLIOGRAFÍA

- Akhmetov, II, D. V. Popov, et al. "[Association of the mitochondrial transcription factor (TFAM) gene polymorphism with physical performance of athletes]." *Fiziol Cheloveka* **36**(2): 121-5.
- Albring, M., J. Griffith, et al. (1977). "Association of a protein structure of probable membrane derivation with HeLa cell mitochondrial DNA near its origin of replication." *Proc Natl Acad Sci U S A* **74**(4): 1348-52.
- Alonso-Montes, C., M. G. Castro, et al. (2008). "Mitochondrial transcription factors TFA, TFB1 and TFB2: a search for DNA variants/haplotypes and the risk of cardiac hypertrophy." *Dis Markers* **25**(3): 131-9.
- Alvarez, V., A. I. Corao, et al. (2008). "Mitochondrial transcription factor A (TFAM) gene variation and risk of late-onset Alzheimer's disease." *J Alzheimers Dis* **13**(3): 275-80.
- Antoshechkin, I. and D. F. Bogenhagen (1995). "Distinct roles for two purified factors in transcription of *Xenopus* mitochondrial DNA." *Mol Cell Biol* **15**(12): 7032-42.
- Antoshechkin, I., D. F. Bogenhagen, et al. (1997). "The HMG-box mitochondrial transcription factor xl-mtTFA binds DNA as a tetramer to activate bidirectional transcription." *EMBO J* **16**(11): 3198-206.
- Asin-Cayuela, J., M. Helm, et al. (2004). "A monomer-to-trimer transition of the human mitochondrial transcription termination factor (mTERF) is associated with a loss of in vitro activity." *J Biol Chem* **279**(15): 15670-7.
- Asin-Cayuela, J., T. Schwend, et al. (2005). "The human mitochondrial transcription termination factor (mTERF) is fully active in vitro in the non-phosphorylated form." *J Biol Chem* **280**(27): 25499-505.
- Attardi, G. and G. Schatz (1988). "Biogenesis of mitochondria." *Annu Rev Cell Biol* **4**: 289-333.
- Attardi, G., M. Yoneda, et al. (1995). "Complementation and segregation behavior of disease-causing mitochondrial DNA mutations in cellular model systems." *Biochim Biophys Acta* **1271**(1): 241-8.
- Barrientos, A., V. Volpini, et al. (1996). "A nuclear defect in the 4p16 region predisposes to multiple mitochondrial DNA deletions in families with Wolfram syndrome." *J Clin Invest* **97**(7): 1570-6.
- Belin, A. C., B. F. Bjork, et al. (2007). "Association study of two genetic variants in mitochondrial transcription factor A (TFAM) in Alzheimer's and Parkinson's disease." *Neurosci Lett* **420**(3): 257-62.
- Bianchi, M. E., M. Beltrame, et al. (1989). "Specific recognition of cruciform DNA by nuclear protein HMG1." *Science* **243**(4894 Pt 1): 1056-9.
- Boesch, P., F. Weber-Lotfi, et al. "DNA repair in organelles: Pathways, organization, regulation, relevance in disease and aging." *Biochim Biophys Acta* **1813**(1): 186-200.
- Bogenhagen, D. F., D. Rousseau, et al. (2008). "The layered structure of human mitochondrial DNA nucleoids." *J Biol Chem* **283**(6): 3665-75.
- Bonawitz, N. D., D. A. Clayton, et al. (2006). "Initiation and beyond: multiple functions of the human mitochondrial transcription machinery." *Mol Cell* **24**(6): 813-25.
- Brown, W. M., J. Shine, et al. (1978). "Human mitochondrial DNA: analysis of 7S DNA from the origin of replication." *Proc Natl Acad Sci U S A* **75**(2): 735-9.
- Camasamudram, V., J. K. Fang, et al. (2003). "Transcription termination at the mouse mitochondrial H-strand promoter distal site requires an A/T rich sequence motif and sequence specific DNA binding proteins." *Eur J Biochem* **270**(6): 1128-40.
- Canugovi, C., S. Maynard, et al. "The mitochondrial transcription factor A functions in mitochondrial base excision repair." *DNA Repair (Amst)* **9**(10): 1080-9.
- Cotney, J. and G. S. Shadel (2006). "Evidence for an early gene duplication event in the evolution of the mitochondrial transcription factor B family and maintenance of rRNA methyltransferase activity in human mtTFB1 and mtTFB2." *J Mol Evol* **63**(5): 707-17.
- Croteau, D. L., R. H. Stierum, et al. (1999). "Mitochondrial DNA repair pathways." *Mutat Res* **434**(3): 137-48.
- Cullen, K. J., Z. Yang, et al. (2007). "Mitochondria as a critical target of the chemotherapeutic agent cisplatin in head and neck cancer." *J Bioenerg Biomembr* **39**(1): 43-50.
- Chang, D. D. and D. A. Clayton (1984). "Precise identification of individual promoters for transcription of each strand of human mitochondrial DNA." *Cell* **36**(3): 635-43.
- Chang, D. D., J. E. Hixson, et al. (1986). "Minor transcription initiation events indicate that both human mitochondrial promoters function bidirectionally." *Mol Cell Biol* **6**(1): 294-301.
- Cho, J. H., S. J. Ha, et al. (1998). "A novel DNA-binding protein bound to the mitochondrial inner membrane restores the null mutation of mitochondrial histone Abf2p in *Saccharomyces cerevisiae*." *Mol Cell Biol* **18**(10): 5712-23.
- Choi, Y. S., S. Kim, et al. (2001). "Mitochondrial transcription factor A (mtTFA) and diabetes." *Diabetes Res Clin Pract* **54 Suppl 2**: S3-9.
- Dairaghi, D. J., G. S. Shadel, et al. (1995). "Addition of a 29 residue carboxyl-terminal tail converts a simple HMG box-containing protein into a transcriptional activator." *J Mol Biol* **249**(1): 11-28.
- Dairaghi, D. J., G. S. Shadel, et al. (1995). "Human mitochondrial transcription factor A and promoter spacing integrity are required for transcription initiation." *Biochim Biophys Acta* **1271**(1): 127-34.
- Diffley, J. F. and B. Stillman (1991). "A close relative of the nuclear, chromosomal high-mobility group protein HMG1 in yeast mitochondria." *Proc Natl Acad Sci U S A* **88**(17): 7864-8.

- Dinardo, M. M., C. Musicco, et al. (2003). "Acetylation and level of mitochondrial transcription factor A in several organs of young and old rats." *Biochem Biophys Res Commun* **301**(1): 187-91.
- Ekstrand, M. I., M. Falkenberg, et al. (2004). "Mitochondrial transcription factor A regulates mtDNA copy number in mammals." *Hum Mol Genet* **13**(9): 935-44.
- Ekstrand, M. I., M. Terzioglu, et al. (2007). "Progressive parkinsonism in mice with respiratory-chain-deficient dopamine neurons." *Proc Natl Acad Sci U S A* **104**(4): 1325-30.
- Eynon, N., M. Moran, et al. "The champions' mitochondria: is it genetically determined? A review on mitochondrial DNA and elite athletic performance." *Physiol Genomics* **43**(13): 789-98.
- Falkenberg, M., M. Gaspari, et al. (2002). "Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA." *Nat Genet* **31**(3): 289-94.
- Fernandez-Silva, P., F. Martinez-Azorin, et al. (1997). "The human mitochondrial transcription termination factor (mTERF) is a multizipper protein but binds to DNA as a monomer, with evidence pointing to intramolecular leucine zipper interactions." *EMBO J* **16**(5): 1066-79.
- Fisher, R. P. and D. A. Clayton (1985). "A transcription factor required for promoter recognition by human mitochondrial RNA polymerase. Accurate initiation at the heavy- and light-strand promoters dissected and reconstituted in vitro." *J Biol Chem* **260**(20): 11330-8.
- Fisher, R. P. and D. A. Clayton (1988). "Purification and characterization of human mitochondrial transcription factor 1." *Mol Cell Biol* **8**(8): 3496-509.
- Fisher, R. P., T. Lisowsky, et al. (1992). "DNA wrapping and bending by a mitochondrial high mobility group-like transcriptional activator protein." *J Biol Chem* **267**(5): 3358-67.
- Fisher, R. P., M. A. Parisi, et al. (1989). "Flexible recognition of rapidly evolving promoter sequences by mitochondrial transcription factor 1." *Genes Dev* **3**(12B): 2202-17.
- Fisher, R. P., J. N. Topper, et al. (1987). "Promoter selection in human mitochondria involves binding of a transcription factor to orientation-independent upstream regulatory elements." *Cell* **50**(2): 247-58.
- Foran, D. R., J. E. Hixson, et al. (1988). "Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis." *Nucleic Acids Res* **16**(13): 5841-61.
- Friddle, R. W., J. E. Klare, et al. (2004). "Mechanism of DNA compaction by yeast mitochondrial protein Abf2p." *Biophys J* **86**(3): 1632-9.
- Garrido, N., L. Griparic, et al. (2003). "Composition and dynamics of human mitochondrial nucleoids." *Mol Biol Cell* **14**(4): 1583-96.
- Gaspari, M., M. Falkenberg, et al. (2004). "The mitochondrial RNA polymerase contributes critically to promoter specificity in mammalian cells." *EMBO J* **23**(23): 4606-14.
- Gaweda-Walerych, K., K. Safranow, et al. "Mitochondrial transcription factor A variants and the risk of Parkinson's disease." *Neurosci Lett* **469**(1): 24-9.
- Ghivizzani, S. C., C. S. Madsen, et al. (1994). "In organello footprint analysis of human mitochondrial DNA: human mitochondrial transcription factor A interactions at the origin of replication." *Mol Cell Biol* **14**(12): 7717-30.
- Giese, K., J. Pagel, et al. (1994). "Distinct DNA-binding properties of the high mobility group domain of murine and human SRY sex-determining factors." *Proc Natl Acad Sci U S A* **91**(8): 3368-72.
- Giles, R. E., H. Blanc, et al. (1980). "Maternal inheritance of human mitochondrial DNA." *Proc Natl Acad Sci U S A* **77**(11): 6715-9.
- Goodell, S. and G. Cortopassi (1998). "Analysis of oxygen consumption and mitochondrial permeability with age in mice." *Mech Ageing Dev* **101**(3): 245-56.
- Goto, A., Y. Matsushima, et al. (2001). "Drosophila mitochondrial transcription factor A (d-TFAM) is dispensable for the transcription of mitochondrial DNA in Kc167 cells." *Biochem J* **354**(Pt 2): 243-8.
- Gray, M. W., G. Burger, et al. (1999). "Mitochondrial evolution." *Science* **283**(5407): 1476-81.
- Grosschedl, R., K. Giese, et al. (1994). "HMG domain proteins: architectural elements in the assembly of nucleoprotein structures." *Trends Genet* **10**(3): 94-100.
- Guo, J., L. Zheng, et al. "Frequent truncating mutation of TFAM induces mitochondrial DNA depletion and apoptotic resistance in microsatellite-unstable colorectal cancer." *Cancer Res* **71**(8): 2978-87.
- Gur, E. and R. T. Sauer (2009). "Degrons in protein substrates program the speed and operating efficiency of the AAA+ Lon proteolytic machine." *Proc Natl Acad Sci U S A* **106**(44): 18503-8.
- Ha, P. K., B. C. Tong, et al. (2002). "Mitochondrial C-tract alteration in premalignant lesions of the head and neck: a marker for progression and clonal proliferation." *Clin Cancer Res* **8**(7): 2260-5.
- Hayashi, Y., M. Yoshida, et al. (2008). "Reverse of age-dependent memory impairment and mitochondrial DNA damage in microglia by an overexpression of human mitochondrial transcription factor A in mice." *J Neurosci* **28**(34): 8624-34.
- He, J., C. C. Mao, et al. (2007). "The AAA+ protein ATAD3 has displacement loop binding properties and is involved in mitochondrial nucleoid organization." *J Cell Biol* **176**(2): 141-6.
- Hixson, J. E. and D. A. Clayton (1985). "Initiation of transcription from each of the two human mitochondrial promoters requires unique nucleotides at the transcriptional start sites." *Proc Natl Acad Sci U S A* **82**(9): 2660-4.
- Hobbs, A. E., M. Srinivasan, et al. (2001). "Mmm1p, a mitochondrial outer membrane protein, is connected to mitochondrial DNA (mtDNA) nucleoids and required for mtDNA stability." *J Cell Biol* **152**(2): 401-10.

- Huang, J. C., D. B. Zamble, et al. (1994). "HMG-domain proteins specifically inhibit the repair of the major DNA adduct of the anticancer drug cisplatin by human excision nuclease." *Proc Natl Acad Sci U S A* **91**(22): 10394-8.
- Huppert, J. L. and S. Balasubramanian (2007). "G-quadruplexes in promoters throughout the human genome." *Nucleic Acids Res* **35**(2): 406-13.
- Iborra, F. J., H. Kimura, et al. (2004). "The functional organization of mitochondrial genomes in human cells." *BMC Biol* **2**: 9.
- Im, S. H. and J. Lee (2003). "Identification of HMG-5 as a double-stranded telomeric DNA-binding protein in the nematode *Caenorhabditis elegans*." *FEBS Lett* **554**(3): 455-61.
- Jimenez-Menendez, N., P. Fernandez-Millan, et al. "Human mitochondrial mTERF wraps around DNA through a left-handed superhelical tandem repeat." *Nat Struct Mol Biol* **17**(7): 891-3.
- Kajander, O. A., A. T. Rovio, et al. (2000). "Human mtDNA sublimons resemble rearranged mitochondrial genomes found in pathological states." *Hum Mol Genet* **9**(19): 2821-35.
- Kanki, T., K. Ohgaki, et al. (2004). "Architectural role of mitochondrial transcription factor A in maintenance of human mitochondrial DNA." *Mol Cell Biol* **24**(22): 9823-34.
- Kasai, H., P. F. Crain, et al. (1986). "Formation of 8-hydroxyguanine moiety in cellular DNA by agents producing oxygen radicals and evidence for its repair." *Carcinogenesis* **7**(11): 1849-51.
- Kaufman, B. A., N. Durisic, et al. (2007). "The mitochondrial transcription factor TFAM coordinates the assembly of multiple DNA molecules into nucleoid-like structures." *Mol Biol Cell* **18**(9): 3225-36.
- Keeney, P. M., C. K. Quigley, et al. (2009). "Mitochondrial gene therapy augments mitochondrial physiology in a Parkinson's disease cell model." *Hum Gene Ther* **20**(8): 897-907.
- Kelly, D. P. and R. C. Scarpulla (2004). "Transcriptional regulatory circuits controlling mitochondrial biogenesis and function." *Genes Dev* **18**(4): 357-68.
- Kidani, A., H. Izumi, et al. (2009). "Thioredoxin2 enhances the damaged DNA binding activity of mtTFA through direct interaction." *Int J Oncol* **35**(6): 1435-40.
- Klass, J., F. V. t. Murphy, et al. (2003). "The role of intercalating residues in chromosomal high-mobility-group protein DNA binding, bending and specificity." *Nucleic Acids Res* **31**(11): 2852-64.
- Koster, J. C., B. A. Marshall, et al. (2000). "Targeted overactivity of beta cell K(ATP) channels induces profound neonatal diabetes." *Cell* **100**(6): 645-54.
- Kruse, B., N. Narasimhan, et al. (1989). "Termination of transcription in human mitochondria: identification and purification of a DNA binding protein factor that promotes termination." *Cell* **58**(2): 391-7.
- Kucej, M. and R. A. Butow (2007). "Evolutionary tinkering with mitochondrial nucleoids." *Trends Cell Biol* **17**(12): 586-92.
- Kucej, M., B. Kucejova, et al. (2008). "Mitochondrial nucleoids undergo remodeling in response to metabolic cues." *J Cell Sci* **121**(Pt 11): 1861-8.
- Kuhn, A., R. Voit, et al. (1994). "Functional differences between the two splice variants of the nucleolar transcription factor UBF: the second HMG box determines specificity of DNA binding and transcriptional activity." *EMBO J* **13**(2): 416-24.
- Kukat, C., C. A. Wurm, et al. "Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and frequently contain a single copy of mtDNA." *Proc Natl Acad Sci U S A* **108**(33): 13534-9.
- Larsson, N. G., J. Wang, et al. (1998). "Mitochondrial transcription factor A is necessary for mtDNA maintenance and embryogenesis in mice." *Nat Genet* **18**(3): 231-6.
- Laudet, V., D. Stehelin, et al. (1993). "Ancestry and diversity of the HMG box superfamily." *Nucleic Acids Res* **21**(10): 2493-501.
- Lu, B., S. Yadav, et al. (2007). "Roles for the human ATP-dependent Lon protease in mitochondrial DNA maintenance." *J Biol Chem* **282**(24): 17363-74.
- Mambo, E., X. Gao, et al. (2003). "Electrophile and oxidant damage of mitochondrial DNA leading to rapid evolution of homoplasmic mutations." *Proc Natl Acad Sci U S A* **100**(4): 1838-43.
- Mangus, D. A., S. H. Jang, et al. (1994). "Release of the yeast mitochondrial RNA polymerase specificity factor from transcription complexes." *J Biol Chem* **269**(42): 26568-74.
- Marchetti, P., T. Hirsch, et al. (1996). "Mitochondrial permeability transition triggers lymphocyte apoptosis." *J Immunol* **157**(11): 4830-6.
- Martin, M., J. Cho, et al. (2005). "Termination factor-mediated DNA loop between termination and initiation sites drives mitochondrial rRNA synthesis." *Cell* **123**(7): 1227-40.
- Matsunaga, M. and J. A. Jaehning (2004). "Intrinsic promoter recognition by a "core" RNA polymerase." *J Biol Chem* **279**(43): 44239-42.
- Matsusaka, H., M. Ikeuchi, et al. (2005). "Selective disruption of MMP-2 gene exacerbates myocardial inflammation and dysfunction in mice with cytokine-induced cardiomyopathy." *Am J Physiol Heart Circ Physiol* **289**(5): H1858-64.
- Matsushima, Y., K. Matsumura, et al. (2003). "Functional domains of chicken mitochondrial transcription factor A for the maintenance of mitochondrial DNA copy number in lymphoma cell line DT40." *J Biol Chem* **278**(33): 31149-58.

- McCauley, M. J., J. Zimmerman, et al. (2007). "HMGB binding to DNA: single and double box motifs." *J Mol Biol* **374**(4): 993-1004.
- McCulloch, V., B. L. Seidel-Rogol, et al. (2002). "A human mitochondrial transcription factor is related to RNA adenine methyltransferases and binds S-adenosylmethionine." *Mol Cell Biol* **22**(4): 1116-25.
- McCulloch, V. and G. S. Shadel (2003). "Human mitochondrial transcription factor B1 interacts with the C-terminal activation region of h-mtTFA and stimulates transcription independently of its RNA methyltransferase activity." *Mol Cell Biol* **23**(16): 5816-24.
- Megraw, T. L. and C. B. Chae (1993). "Functional complementarity between the HMG1-like yeast mitochondrial histone HM and the bacterial histone-like protein HU." *J Biol Chem* **268**(17): 12758-63.
- Mercer, T. R., S. Neph, et al. "The human mitochondrial transcriptome." *Cell* **146**(4): 645-58.
- Metodiev, M. D., N. Lesko, et al. (2009). "Methylation of 12S rRNA is necessary for in vivo stability of the small subunit of the mammalian mitochondrial ribosome." *Cell Metab* **9**(4): 386-97.
- Miyakawa, I., N. Sando, et al. (1987). "Isolation of morphologically intact mitochondrial nucleoids from the yeast, *Saccharomyces cerevisiae*." *J Cell Sci* **88 (Pt 4)**: 431-9.
- Montoya, J., T. Christianson, et al. (1982). "Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA." *Proc Natl Acad Sci U S A* **79**(23): 7195-9.
- Nass, M. M. (1969). "Reversible generation of circular dimer and higher multiple forms of mitochondrial DNA." *Nature* **223**(5211): 1124-9.
- Nass, S. and M. M. Nass (1963). "Intramitochondrial Fibers with DNA Characteristics. Ii. Enzymatic and Other Hydrolytic Treatments." *J Cell Biol* **19**: 613-29.
- Nemajerova, A., S. Erster, et al. (2005). "The post-translational phosphorylation and acetylation modification profile is not the determining factor in targeting endogenous stress-induced p53 to mitochondria." *Cell Death Differ* **12**(2): 197-200.
- Noack, H., T. Bednarek, et al. (2006). "TFAM-dependent and independent dynamics of mtDNA levels in C2C12 myoblasts caused by redox stress." *Biochim Biophys Acta* **1760**(2): 141-50.
- O'Mahoney, J. V., M. R. Brandon, et al. (1991). "Developmental and tissue-specific regulation of ovine insulin-like growth factor II (IGF-II) mRNA expression." *Mol Cell Endocrinol* **78**(1-2): 87-96.
- Ohgaki, K., T. Kanki, et al. (2007). "The C-terminal tail of mitochondrial transcription factor a markedly strengthens its general binding to DNA." *J Biochem* **141**(2): 201-11.
- Ohno, T., S. Umeda, et al. (2000). "Binding of human mitochondrial transcription factor A, an HMG box protein, to a four-way DNA junction." *Biochem Biophys Res Commun* **271**(2): 492-8.
- Parisi, M. A. and D. A. Clayton (1991). "Similarity of human mitochondrial transcription factor 1 to high mobility group proteins." *Science* **252**(5008): 965-9.
- Parisi, M. A., B. Xu, et al. (1993). "A human mitochondrial transcriptional activator can functionally replace a yeast mitochondrial HMG-box protein both in vivo and in vitro." *Mol Cell Biol* **13**(3): 1951-61.
- Parra, V., H. Verdejo, et al. "The complex interplay between mitochondrial dynamics and cardiac metabolism." *J Bioenerg Biomembr* **43**(1): 47-51.
- Pastukh, V., I. Shokolenko, et al. (2007). "Human mitochondrial transcription factor A possesses multiple subcellular targeting signals." *FEBS J* **274**(24): 6488-99.
- Paull, T. T., M. J. Haykinson, et al. (1993). "The nonspecific DNA-binding and -bending proteins HMG1 and HMG2 promote the assembly of complex nucleoprotein structures." *Genes Dev* **7**(8): 1521-34.
- Pfannschmidt, T., A. Nilsson, et al. (1999). "Direct transcriptional control of the chloroplast genes psbA and psaAB adjusts photosynthesis to light energy distribution in plants." *IUBMB Life* **48**(3): 271-6.
- Pham, X. H., G. Farge, et al. (2006). "Conserved sequence box II directs transcription termination and primer formation in mitochondria." *J Biol Chem* **281**(34): 24647-52.
- Reyes, A., M. Mezzina, et al. (2002). "Human mitochondrial transcription factor A (mtTFA): gene structure and characterization of related pseudogenes." *Gene* **291**(1-2): 223-32.
- Ringel, R., M. Sologub, et al. "Structure of human mitochondrial RNA polymerase." *Nature* **478**(7368): 269-73.
- Seidel-Rogol, B. L., V. McCulloch, et al. (2003). "Human mitochondrial transcription factor B1 methylates ribosomal RNA at a conserved stem-loop." *Nat Genet* **33**(1): 23-4.
- Seidel-Rogol, B. L. and G. S. Shadel (2002). "Modulation of mitochondrial transcription in response to mtDNA depletion and repletion in HeLa cells." *Nucleic Acids Res* **30**(9): 1929-34.
- Shadel, G. S. and D. A. Clayton (1997). "Mitochondrial DNA maintenance in vertebrates." *Annu Rev Biochem* **66**: 409-35.
- Shoubridge, E. A. (2009). "Something old, something new, something borrowed." *Cell Metab* **9**(4): 307-8.
- Shutt, T. E., M. Bestwick, et al. "The core human mitochondrial transcription initiation complex: It only takes two to tango." *Transcription* **2**(2): 55-59.
- Silva, J. P., M. Kohler, et al. (2000). "Impaired insulin secretion and beta-cell loss in tissue-specific knockout mice with mitochondrial diabetes." *Nat Genet* **26**(3): 336-40.
- Singh, G., N. Neckelmann, et al. (1987). "Conformational mutations in human mitochondrial DNA." *Nature* **329**(6136): 270-2.
- Smeets, A., C. Evrard, et al. (2005). "Crystal structures of oxidized and reduced forms of human mitochondrial thioredoxin 2." *Protein Sci* **14**(10): 2610-21.

- Sologub, M., D. Litonin, et al. (2009). "TFB2 is a transient component of the catalytic site of the human mitochondrial RNA polymerase." *Cell* **139**(5): 934-44.
- Sterner, R., G. Vidali, et al. (1978). "Postsynthetic modification of high mobility group proteins. Evidence that high mobility group proteins are acetylated." *J Biol Chem* **253**(21): 7601-4.
- Stros, M., J. Reich, et al. (1994). "Calcium binding to HMG1 protein induces DNA looping by the HMG-box domains." *FEBS Lett* **344**(2-3): 201-6.
- Sumitani, M., K. Kasashima, et al. "Biochemical properties of *Caenorhabditis elegans* HMG-5, a regulator of mitochondrial DNA." *J Biochem* **149**(5): 581-9.
- Sutovsky, P., R. D. Moreno, et al. (2000). "Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos." *Biol Reprod* **63**(2): 582-90.
- Thomas, J. O. and A. A. Travers (2001). "HMG1 and 2, and related 'architectural' DNA-binding proteins." *Trends Biochem Sci* **26**(3): 167-74.
- Tiranti, V., A. Savoia, et al. (1997). "Identification of the gene encoding the human mitochondrial RNA polymerase (h-mtRPO1) by cyberscreening of the Expressed Sequence Tags database." *Hum Mol Genet* **6**(4): 615-25.
- Topper, J. N. and D. A. Clayton (1989). "Identification of transcriptional regulatory elements in human mitochondrial DNA by linker substitution analysis." *Mol Cell Biol* **9**(3): 1200-11.
- Uchiumi, T. and D. Kang "The role of TFAM-associated proteins in mitochondrial RNA metabolism." *Biochim Biophys Acta* **1820**(5): 565-70.
- Ugrinova, I., E. A. Pasheva, et al. (2001). "In vivo acetylation of HMG1 protein enhances its binding affinity to distorted DNA structures." *Biochemistry* **40**(48): 14655-60.
- Van Remmen, H., Y. Ikeno, et al. (2003). "Life-long reduction in MnSOD activity results in increased DNA damage and higher incidence of cancer but does not accelerate aging." *Physiol Genomics* **16**(1): 29-37.
- Wai, T., D. Teoli, et al. (2008). "The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes." *Nat Genet* **40**(12): 1484-8.
- Wallace, D. C. (2005). "Mitochondria and cancer: Warburg addressed." *Cold Spring Harb Symp Quant Biol* **70**: 363-74.
- Wallace, D. C. (2005). "A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine." *Annu Rev Genet* **39**: 359-407.
- Wallace, K. B. (2008). "Mitochondrial off targets of drug therapy." *Trends Pharmacol Sci* **29**(7): 361-6.
- Wanrooij, P. H., J. P. Uhler, et al. "G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation." *Proc Natl Acad Sci U S A* **107**(37): 16072-7.
- Weir, H. M., P. J. Kraulis, et al. (1993). "Structure of the HMG box motif in the B-domain of HMG1." *EMBO J* **12**(4): 1311-9.
- Winkley, C. S., M. J. Keller, et al. (1985). "A multicomponent mitochondrial RNA polymerase from *Saccharomyces cerevisiae*." *J Biol Chem* **260**(26): 14214-23.
- Wong, T. S., S. Rajagopalan, et al. (2009). "Biophysical characterizations of human mitochondrial transcription factor A and its binding to tumor suppressor p53." *Nucleic Acids Res* **37**(20): 6765-83.
- Xu, B. and D. A. Clayton (1996). "RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA-DNA hybrids serving as primers." *EMBO J* **15**(12): 3135-43.
- Yoshida, Y., H. Izumi, et al. (2002). "Human mitochondrial transcription factor A binds preferentially to oxidatively damaged DNA." *Biochem Biophys Res Commun* **295**(4): 945-51.
- Zhang, J., M. J. McCauley, et al. (2009). "Mechanism of DNA flexibility enhancement by HMGB proteins." *Nucleic Acids Res* **37**(4): 1107-14.
- Zhang, Q., J. T. Yu, et al. "Mitochondrial transcription factor A (TFAM) polymorphisms and risk of late-onset Alzheimer's disease in Han Chinese." *Brain Res* **1368**: 355-60.
- Zhang, X., Y. Zheng, et al. "Disruption of the mitochondrial thioredoxin system as a cell death mechanism of cationic triphenylmethanes." *Free Radic Biol Med* **50**(7): 811-20.
- Bairoch, A., R. Apweiler, et al. (2005). "The Universal Protein Resource (UniProt)." *Nucleic Acids Res* **33**(Database issue): D154-9.
- Coll, M., C. A. Frederick, et al. (1987). "A bifurcated hydrogen-bonded conformation in the d(A.T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin." *Proc Natl Acad Sci U S A* **84**(23): 8385-9.
- Cowtan, K. D. and P. Main (1996). "Phase combination and cross validation in iterated density-modification calculations." *Acta Crystallogr D Biol Crystallogr* **52**(Pt 1): 43-8.
- Dairaghi, D. J., G. S. Shadel, et al. (1995). "Addition of a 29 residue carboxyl-terminal tail converts a simple HMG box-containing protein into a transcriptional activator." *J Mol Biol* **249**(1): 11-28.

- Dickerson, R. E. and H. R. Drew (1981). "Kinematic model for B-DNA." Proc Natl Acad Sci U S A **78**(12): 7318-22.
- Dragan, A. I., C. M. Read, et al. (2004). "DNA binding and bending by HMG boxes: energetic determinants of specificity." J Mol Biol **343**(2): 371-93.
- Eisenberg, D., R. M. Weiss, et al. (1982). "The helical hydrophobic moment: a measure of the amphiphilicity of a helix." Nature **299**(5881): 371-4.
- Evans, P. (2006). "Scaling and assessment of data quality." Acta Crystallogr D Biol Crystallogr **62**(Pt 1): 72-82.
- Fisher, R. P., T. Lisowsky, et al. (1992). "DNA wrapping and bending by a mitochondrial high mobility group-like transcriptional activator protein." J Biol Chem **267**(5): 3358-67.
- Fisher, R. P., J. N. Topper, et al. (1987). "Promoter selection in human mitochondria involves binding of a transcription factor to orientation-independent upstream regulatory elements." Cell **50**(2): 247-58.
- Gabadinho, J., A. Beteva, et al. "MxCuBE: a synchrotron beamline control environment customized for macromolecular crystallography experiments." J Synchrotron Radiat **17**(5): 700-7.
- Gangelhoff, T. A., P. S. Mungalachetty, et al. (2009). "Structural analysis and DNA binding of the HMG domains of the human mitochondrial transcription factor A." Nucleic Acids Res **37**(10): 3153-64.
- Kaufman, B. A., N. Durisic, et al. (2007). "The mitochondrial transcription factor TFAM coordinates the assembly of multiple DNA molecules into nucleoid-like structures." Mol Biol Cell **18**(9): 3225-36.
- Laudet, V., D. Stehelin, et al. (1993). "Ancestry and diversity of the HMG box superfamily." Nucleic Acids Res **21**(10): 2493-501.
- Love, J. J., X. Li, et al. (1995). "Structural basis for DNA bending by the architectural transcription factor LEF-1." Nature **376**(6543): 791-5.
- McCulloch, V. and G. S. Shadel (2003). "Human mitochondrial transcription factor B1 interacts with the C-terminal activation region of h-mtTFA and stimulates transcription independently of its RNA methyltransferase activity." Mol Cell Biol **23**(16): 5816-24.
- McGuffin, L. J., K. Bryson, et al. (2000). "The PSIPRED protein structure prediction server." Bioinformatics **16**(4): 404-5.
- Murphy, E. C., V. B. Zhurkin, et al. (2001). "Structural basis for SRY-dependent 46-X,Y sex reversal: modulation of DNA bending by a naturally occurring point mutation." J Mol Biol **312**(3): 481-99.
- Murphy, F. V. t., R. M. Sweet, et al. (1999). "The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition." EMBO J **18**(23): 6610-8.
- Murshudov, G. N., A. A. Vagin, et al. (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallogr D Biol Crystallogr **53**(Pt 3): 240-55.
- Ngo, H. B., J. T. Kaiser, et al. "The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA." Nat Struct Mol Biol **18**(11): 1290-6.
- Parkhurst, K. M., R. M. Richards, et al. (1999). "Intermediate species possessing bent DNA are present along the pathway to formation of a final TBP-TATA complex." J Mol Biol **289**(5): 1327-41.
- Potterton, L., S. McNicholas, et al. (2004). "Developments in the CCP4 molecular-graphics project." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2288-94.
- Sheldrick, G. M. (2008). "A short history of SHELX." Acta Crystallogr A **64**(Pt 1): 112-22.

- Smart, O. S., T. O. Womack, et al. "Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER." *Acta Crystallogr D Biol Crystallogr* **68**(Pt 4): 368-80.
- Stott, K., G. S. Tang, et al. (2006). "Structure of a complex of tandem HMG boxes and DNA." *J Mol Biol* **360**(1): 90-104.
- Thompson, J. F. and A. Landy (1988). "Empirical estimation of protein-induced DNA bending angles: applications to lambda site-specific recombination complexes." *Nucleic Acids Res* **16**(20): 9687-705.
- Wong, T. S., S. Rajagopalan, et al. (2009). "Biophysical characterizations of human mitochondrial transcription factor A and its binding to tumor suppressor p53." *Nucleic Acids Res* **37**(20): 6765-83.
- Antoshechkin, I., D. F. Bogenhagen, et al. (1997). "The HMG-box mitochondrial transcription factor xl-mTFA binds DNA as a tetramer to activate bidirectional transcription." *EMBO J* **16**(11): 3198-206.
- Crooks, G. E., G. Hon, et al. (2004). "WebLogo: a sequence logo generator." *Genome Res* **14**(6): 1188-90.
- Dairaghi, D. J., G. S. Shadel, et al. (1995). "Addition of a 29 residue carboxyl-terminal tail converts a simple HMG box-containing protein into a transcriptional activator." *J Mol Biol* **249**(1): 11-28.
- Dairaghi, D. J., G. S. Shadel, et al. (1995). "Human mitochondrial transcription factor A and promoter spacing integrity are required for transcription initiation." *Biochim Biophys Acta* **1271**(1): 127-34.
- Dragan, A. I., C. M. Read, et al. (2004). "DNA binding and bending by HMG boxes: energetic determinants of specificity." *J Mol Biol* **343**(2): 371-93.
- Farge, G., N. Laurens, et al. "Protein sliding and DNA denaturation are essential for DNA organization by human mitochondrial transcription factor A." *Nat Commun* **3**: 1013.
- Fisher, R. P., M. A. Parisi, et al. (1989). "Flexible recognition of rapidly evolving promoter sequences by mitochondrial transcription factor 1." *Genes Dev* **3**(12B): 2202-17.
- Fisher, R. P., J. N. Topper, et al. (1987). "Promoter selection in human mitochondria involves binding of a transcription factor to orientation-independent upstream regulatory elements." *Cell* **50**(2): 247-58.
- Foran, D. R., J. E. Hixson, et al. (1988). "Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis." *Nucleic Acids Res* **16**(13): 5841-61.
- Gangelhoff, T. A., P. S. Mungalachetty, et al. (2009). "Structural analysis and DNA binding of the HMG domains of the human mitochondrial transcription factor A." *Nucleic Acids Res* **37**(10): 3153-64.
- Gaspari, M., M. Falkenberg, et al. (2004). "The mitochondrial RNA polymerase contributes critically to promoter specificity in mammalian cells." *EMBO J* **23**(23): 4606-14.
- Haran, T. E. and U. Mohanty (2009). "The unique structure of A-tracts and intrinsic DNA bending." *Q Rev Biophys* **42**(1): 41-81.
- Kaufman, B. A., N. Durisic, et al. (2007). "The mitochondrial transcription factor TFAM coordinates the assembly of multiple DNA molecules into nucleoid-like structures." *Mol Biol Cell* **18**(9): 3225-36.
- Kukat, C., C. A. Wurm, et al. "Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and frequently contain a single copy of mtDNA." *Proc Natl Acad Sci U S A* **108**(33): 13534-9.
- Kuznetsov, S. V., S. Sugimura, et al. (2006). "Direct observation of DNA bending/unbending kinetics in complex with DNA-bending protein IHF." *Proc Natl Acad Sci U S A* **103**(49): 18515-20.
- Laudet, V., D. Stehelin, et al. (1993). "Ancestry and diversity of the HMG box superfamily." *Nucleic Acids Res* **21**(10): 2493-501.
- Ngo, H. B., J. T. Kaiser, et al. "The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA." *Nat Struct Mol Biol* **18**(11): 1290-6.
- Revyakin, A., R. H. Ebright, et al. (2004). "Promoter unwinding and promoter clearance by RNA polymerase: detection by single-molecule DNA nanomanipulation." *Proc Natl Acad Sci U S A* **101**(14): 4776-80.
- Roca, J. "Transcriptional inhibition by DNA torsional stress." *Transcription* **2**(2): 82-85.
- Stefanovsky, V. Y., G. Pelletier, et al. (2001). "DNA looping in the RNA polymerase I enhancer is the result of non-cooperative in-phase bending by two UBF molecules." *Nucleic Acids Res* **29**(15): 3241-7.
- Stott, K., G. S. Tang, et al. (2006). "Structure of a complex of tandem HMG boxes and DNA." *J Mol Biol* **360**(1): 90-104.
- Stott, K., M. Watson, et al. "Tail-mediated collapse of HMGB1 is dynamic and occurs via differential binding of the acidic tail to the A and B domains." *J Mol Biol* **403**(5): 706-22.
- Vuzman, D., A. Azia, et al. "Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails." *J Mol Biol* **396**(3): 674-84.
- Vuzman, D., M. Polonsky, et al. "Facilitated DNA search by multidomain transcription factors: cross talk via a flexible linker." *Biophys J* **99**(4): 1202-11.
- Wong, T. S., S. Rajagopalan, et al. (2009). "Biophysical characterizations of human mitochondrial transcription factor A and its binding to tumor suppressor p53." *Nucleic Acids Res* **37**(20): 6765-83.

Abrahams, J. P. and A. G. Leslie (1996). "Methods used in the structure determination of bovine mitochondrial F1 ATPase." *Acta Crystallogr D Biol Crystallogr* **52**(Pt 1): 30-42.

Albring, M., J. Griffith, et al. (1977). "Association of a protein structure of probable membrane derivation with HeLa cell mitochondrial DNA near its origin of replication." *Proc Natl Acad Sci U S A* **74**(4): 1348-52.

Alonso-Montes, C., M. G. Castro, et al. (2008). "Mitochondrial transcription factors TFA, TFB1 and TFB2: a search for DNA variants/haplotypes and the risk of cardiac hypertrophy." *Dis Markers* **25**(3): 131-9.

Alvarez, V., A. I. Corao, et al. (2008). "Mitochondrial transcription factor A (TFAM) gene variation and risk of late-onset Alzheimer's disease." *J Alzheimers Dis* **13**(3): 275-80.

Anderson, S., A. T. Bankier, et al. (1981). "Sequence and organization of the human mitochondrial genome." *Nature* **290**(5806): 457-65.

Antoshechkin, I. and D. F. Bogenhagen (1995). "Distinct roles for two purified factors in transcription of *Xenopus* mitochondrial DNA." *Mol Cell Biol* **15**(12): 7032-42.

Antoshechkin, I., D. F. Bogenhagen, et al. (1997). "The HMG-box mitochondrial transcription factor xl-mtTFA binds DNA as a tetramer to activate bidirectional transcription." *EMBO J* **16**(11): 3198-206.

Asin-Cayuela, J., M. Helm, et al. (2004). "A monomer-to-trimer transition of the human mitochondrial transcription termination factor (mTERF) is associated with a loss of in vitro activity." *J Biol Chem* **279**(15): 15670-7.

Asin-Cayuela, J., T. Schwend, et al. (2005). "The human mitochondrial transcription termination factor (mTERF) is fully active in vitro in the non-phosphorylated form." *J Biol Chem* **280**(27): 25499-505.

Attardi, G. and G. Schatz (1988). "Biogenesis of mitochondria." *Annu Rev Cell Biol* **4**: 289-333.

Bairoch, A., R. Apweiler, et al. (2005). "The Universal Protein Resource (UniProt)." *Nucleic Acids Res* **33**(Database issue): D154-9.

Barrientos, A., V. Volpini, et al. (1996). "A nuclear defect in the 4p16 region predisposes to multiple mitochondrial DNA deletions in families with Wolfram syndrome." *J Clin Invest* **97**(7): 1570-6.

Battye, T. G., L. Kontogiannis, et al. "iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM." *Acta Crystallogr D Biol Crystallogr* **67**(Pt 4): 271-81.

Belin, A. C., B. F. Bjork, et al. (2007). "Association study of two genetic variants in mitochondrial transcription factor A (TFAM) in Alzheimer's and Parkinson's disease." *Neurosci Lett* **420**(3): 257-62.

Bernado, P., E. Mylonas, et al. (2007). "Structural characterization of flexible proteins using small-angle X-ray scattering." *J Am Chem Soc* **129**(17): 5656-64.

Bianchi, M. E., M. Beltrame, et al. (1989). "Specific recognition of cruciform DNA by nuclear protein HMG1." *Science* **243**(4894 Pt 1): 1056-9.

Blanc, E., P. Roversi, et al. (2004). "Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT." *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1): 2210-21.

Boesch, P., F. Weber-Lotfi, et al. "DNA repair in organelles: Pathways, organization, regulation, relevance in disease and aging." *Biochim Biophys Acta* **1813**(1): 186-200.

Bogenhagen, D. F., D. Rousseau, et al. (2008). "The layered structure of human mitochondrial DNA nucleoids." *J Biol Chem* **283**(6): 3665-75.

Bonawitz, N. D., D. A. Clayton, et al. (2006). "Initiation and beyond: multiple functions of the human mitochondrial transcription machinery." *Mol Cell* **24**(6): 813-25.

Brown, W. M., J. Shine, et al. (1978). "Human mitochondrial DNA: analysis of 7S DNA from the origin of replication." *Proc Natl Acad Sci U S A* **75**(2): 735-9.

Camasamudram, V., J. K. Fang, et al. (2003). "Transcription termination at the mouse mitochondrial H-strand promoter distal site requires an A/T rich sequence motif and sequence specific DNA binding proteins." *Eur J Biochem* **270**(6): 1128-40.

Canugovi, C., S. Maynard, et al. "The mitochondrial transcription factor A functions in mitochondrial base excision repair." *DNA Repair (Amst)* **9**(10): 1080-9.

Coll, M., C. A. Frederick, et al. (1987). "A bifurcated hydrogen-bonded conformation in the d(A.T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin." *Proc Natl Acad Sci U S A* **84**(23): 8385-9.

Cotney, J. and G. S. Shadel (2006). "Evidence for an early gene duplication event in the evolution of the mitochondrial transcription factor B family and maintenance of rRNA methyltransferase activity in human mtTFB1 and mtTFB2." *J Mol Evol* **63**(5): 707-17.

Cowtan, K. D. and P. Main (1996). "Phase combination and cross validation in iterated density-modification calculations." *Acta Crystallogr D Biol Crystallogr* **52**(Pt 1): 43-8.

Cowtan, K. D. and K. Y. Zhang (1999). "Density modification for macromolecular phase improvement." *Prog Biophys Mol Biol* **72**(3): 245-70.

Crooks, G. E., G. Hon, et al. (2004). "WebLogo: a sequence logo generator." *Genome Res* **14**(6): 1188-90.

Croteau, D. L., R. H. Stierum, et al. (1999). "Mitochondrial DNA repair pathways." *Mutat Res* **434**(3): 137-48.

Cudney, R., S. Patel, et al. (1994). "Screening and optimization strategies for macromolecular crystal growth." *Acta Crystallogr D Biol Crystallogr* **50**(Pt 4): 414-23.

Cullen, K. J., Z. Yang, et al. (2007). "Mitochondria as a critical target of the chemotherapeutic agent cisplatin in head and neck cancer." *J Bioenerg Biomembr* **39**(1): 43-50.

- Chang, D. D. and D. A. Clayton (1984). "Precise identification of individual promoters for transcription of each strand of human mitochondrial DNA." *Cell* **36**(3): 635-43.
- Chang, D. D., J. E. Hixson, et al. (1986). "Minor transcription initiation events indicate that both human mitochondrial promoters function bidirectionally." *Mol Cell Biol* **6**(1): 294-301.
- Chen, V. B., W. B. Arendall, 3rd, et al. "MolProbity: all-atom structure validation for macromolecular crystallography." *Acta Crystallogr D Biol Crystallogr* **66**(Pt 1): 12-21.
- Cho, J. H., S. J. Ha, et al. (1998). "A novel DNA-binding protein bound to the mitochondrial inner membrane restores the null mutation of mitochondrial histone Abf2p in *Saccharomyces cerevisiae*." *Mol Cell Biol* **18**(10): 5712-23.
- Choi, Y. S., S. Kim, et al. (2001). "Mitochondrial transcription factor A (mtTFA) and diabetes." *Diabetes Res Clin Pract* **54 Suppl 2**: S3-9.
- Dairaghi, D. J., G. S. Shadel, et al. (1995). "Addition of a 29 residue carboxyl-terminal tail converts a simple HMG box-containing protein into a transcriptional activator." *J Mol Biol* **249**(1): 11-28.
- Dairaghi, D. J., G. S. Shadel, et al. (1995). "Human mitochondrial transcription factor A and promoter spacing integrity are required for transcription initiation." *Biochim Biophys Acta* **1271**(1): 127-34.
- Dickerson, R. E. and H. R. Drew (1981). "Kinematic model for B-DNA." *Proc Natl Acad Sci U S A* **78**(12): 7318-22.
- Diffley, J. F. and B. Stillman (1991). "A close relative of the nuclear, chromosomal high-mobility group protein HMG1 in yeast mitochondria." *Proc Natl Acad Sci U S A* **88**(17): 7864-8.
- Dinardo, M. M., C. Musicco, et al. (2003). "Acetylation and level of mitochondrial transcription factor A in several organs of young and old rats." *Biochem Biophys Res Commun* **301**(1): 187-91.
- Dragan, A. I., C. M. Read, et al. (2004). "DNA binding and bending by HMG boxes: energetic determinants of specificity." *J Mol Biol* **343**(2): 371-93.
- Eisenberg, D., R. M. Weiss, et al. (1982). "The helical hydrophobic moment: a measure of the amphiphilicity of a helix." *Nature* **299**(5881): 371-4.
- Ekstrand, M. I., M. Falkenberg, et al. (2004). "Mitochondrial transcription factor A regulates mtDNA copy number in mammals." *Hum Mol Genet* **13**(9): 935-44.
- Ekstrand, M. I., M. Terzioglu, et al. (2007). "Progressive parkinsonism in mice with respiratory-chain-deficient dopamine neurons." *Proc Natl Acad Sci U S A* **104**(4): 1325-30.
- Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1): 2126-32.
- Ericsson, U. B., B. M. Hallberg, et al. (2006). "Thermofluor-based high-throughput stability optimization of proteins for structural studies." *Anal Biochem* **357**(2): 289-98.
- Evans, P. (2006). "Scaling and assessment of data quality." *Acta Crystallogr D Biol Crystallogr* **62**(Pt 1): 72-82.
- Falkenberg, M., M. Gaspari, et al. (2002). "Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA." *Nat Genet* **31**(3): 289-94.
- Farge, G., N. Laurens, et al. "Protein sliding and DNA denaturation are essential for DNA organization by human mitochondrial transcription factor A." *Nat Commun* **3**: 1013.
- Fernandez-Silva, P., F. Martinez-Azorin, et al. (1997). "The human mitochondrial transcription termination factor (mTERF) is a multizipper protein but binds to DNA as a monomer, with evidence pointing to intramolecular leucine zipper interactions." *EMBO J* **16**(5): 1066-79.
- Ferrari, S., V. R. Harley, et al. (1992). "SRY, like HMG1, recognizes sharp angles in DNA." *EMBO J* **11**(12): 4497-506.
- Fisher, R. P. and D. A. Clayton (1985). "A transcription factor required for promoter recognition by human mitochondrial RNA polymerase. Accurate initiation at the heavy- and light-strand promoters dissected and reconstituted in vitro." *J Biol Chem* **260**(20): 11330-8.
- Fisher, R. P. and D. A. Clayton (1988). "Purification and characterization of human mitochondrial transcription factor 1." *Mol Cell Biol* **8**(8): 3496-509.
- Fisher, R. P., T. Lisowsky, et al. (1992). "DNA wrapping and bending by a mitochondrial high mobility group-like transcriptional activator protein." *J Biol Chem* **267**(5): 3358-67.
- Fisher, R. P., M. A. Parisi, et al. (1989). "Flexible recognition of rapidly evolving promoter sequences by mitochondrial transcription factor 1." *Genes Dev* **3**(12B): 2202-17.
- Fisher, R. P., J. N. Topper, et al. (1987). "Promoter selection in human mitochondria involves binding of a transcription factor to orientation-independent upstream regulatory elements." *Cell* **50**(2): 247-58.
- Foran, D. R., J. E. Hixson, et al. (1988). "Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis." *Nucleic Acids Res* **16**(13): 5841-61.
- Fridde, R. W., J. E. Klare, et al. (2004). "Mechanism of DNA compaction by yeast mitochondrial protein Abf2p." *Biophys J* **86**(3): 1632-9.
- Gabadinho, J., A. Beteva, et al. "MxCuBE: a synchrotron beamline control environment customized for macromolecular crystallography experiments." *J Synchrotron Radiat* **17**(5): 700-7.
- Gangelhoff, T. A., P. S. Mungalachetty, et al. (2009). "Structural analysis and DNA binding of the HMG domains of the human mitochondrial transcription factor A." *Nucleic Acids Res* **37**(10): 3153-64.
- Garrido, N., L. Griparic, et al. (2003). "Composition and dynamics of human mitochondrial nucleoids." *Mol Biol Cell* **14**(4): 1583-96.

- Gaspari, M., M. Falkenberg, et al. (2004). "The mitochondrial RNA polymerase contributes critically to promoter specificity in mammalian cells." *EMBO J* **23**(23): 4606-14.
- Gaweda-Walerych, K., K. Safranow, et al. "Mitochondrial transcription factor A variants and the risk of Parkinson's disease." *Neurosci Lett* **469**(1): 24-9.
- Ghivizzani, S. C., C. S. Madsen, et al. (1994). "In organello footprint analysis of human mitochondrial DNA: human mitochondrial transcription factor A interactions at the origin of replication." *Mol Cell Biol* **14**(12): 7717-30.
- Giese, K., J. Pagel, et al. (1994). "Distinct DNA-binding properties of the high mobility group domain of murine and human SRY sex-determining factors." *Proc Natl Acad Sci U S A* **91**(8): 3368-72.
- Giles, R. E., H. Blanc, et al. (1980). "Maternal inheritance of human mitochondrial DNA." *Proc Natl Acad Sci U S A* **77**(11): 6715-9.
- Glusker, J. P. (1994). "X-ray crystallography of proteins." *Methods Biochem Anal* **37**: 1-72.
- Goodell, S. and G. Cortopassi (1998). "Analysis of oxygen consumption and mitochondrial permeability with age in mice." *Mech Ageing Dev* **101**(3): 245-56.
- Gray, M. W., G. Burger, et al. (1999). "Mitochondrial evolution." *Science* **283**(5407): 1476-81.
- Grosschedl, R., K. Giese, et al. (1994). "HMG domain proteins: architectural elements in the assembly of nucleoprotein structures." *Trends Genet* **10**(3): 94-100.
- Guo, J., L. Zheng, et al. "Frequent truncating mutation of TFAM induces mitochondrial DNA depletion and apoptotic resistance in microsatellite-unstable colorectal cancer." *Cancer Res* **71**(8): 2978-87.
- Gur, E. and R. T. Sauer (2009). "Degrons in protein substrates program the speed and operating efficiency of the AAA+ Lon proteolytic machine." *Proc Natl Acad Sci U S A* **106**(44): 18503-8.
- Ha, P. K., B. C. Tong, et al. (2002). "Mitochondrial C-tract alteration in premalignant lesions of the head and neck: a marker for progression and clonal proliferation." *Clin Cancer Res* **8**(7): 2260-5.
- Haran, T. E. and U. Mohanty (2009). "The unique structure of A-tracts and intrinsic DNA bending." *Q Rev Biophys* **42**(1): 41-81.
- Hayashi, Y., M. Yoshida, et al. (2008). "Reverse of age-dependent memory impairment and mitochondrial DNA damage in microglia by an overexpression of human mitochondrial transcription factor a in mice." *J Neurosci* **28**(34): 8624-34.
- He, J., C. C. Mao, et al. (2007). "The AAA+ protein ATAD3 has displacement loop binding properties and is involved in mitochondrial nucleoid organization." *J Cell Biol* **176**(2): 141-6.
- Hixson, J. E. and D. A. Clayton (1985). "Initiation of transcription from each of the two human mitochondrial promoters requires unique nucleotides at the transcriptional start sites." *Proc Natl Acad Sci U S A* **82**(9): 2660-4.
- Hobbs, A. E., M. Srinivasan, et al. (2001). "Mmm1p, a mitochondrial outer membrane protein, is connected to mitochondrial DNA (mtDNA) nucleoids and required for mtDNA stability." *J Cell Biol* **152**(2): 401-10.
- Huang, J. C., D. B. Zamble, et al. (1994). "HMG-domain proteins specifically inhibit the repair of the major DNA adduct of the anticancer drug cisplatin by human excision nuclease." *Proc Natl Acad Sci U S A* **91**(22): 10394-8.
- Huppert, J. L. and S. Balasubramanian (2007). "G-quadruplexes in promoters throughout the human genome." *Nucleic Acids Res* **35**(2): 406-13.
- Iborra, F. J., H. Kimura, et al. (2004). "The functional organization of mitochondrial genomes in human cells." *BMC Biol* **2**: 9.
- Im, S. H. and J. Lee (2003). "Identification of HMG-5 as a double-stranded telomeric DNA-binding protein in the nematode *Caenorhabditis elegans*." *FEBS Lett* **554**(3): 455-61.
- Jimenez-Menendez, N., P. Fernandez-Millan, et al. "Human mitochondrial mTERF wraps around DNA through a left-handed superhelical tandem repeat." *Nat Struct Mol Biol* **17**(7): 891-3.
- Kabsch, W. "Xds." *Acta Crystallogr D Biol Crystallogr* **66**(Pt 2): 125-32.
- Kahn, J. D., E. Yun, et al. (1994). "Detection of localized DNA flexibility." *Nature* **368**(6467): 163-6.
- Kajander, O. A., A. T. Rovio, et al. (2000). "Human mtDNA sublimons resemble rearranged mitochondrial genomes found in pathological states." *Hum Mol Genet* **9**(19): 2821-35.
- Kang, D., S. H. Kim, et al. (2007). "Mitochondrial transcription factor A (TFAM): roles in maintenance of mtDNA and cellular functions." *Mitochondrion* **7**(1-2): 39-44.
- Kanki, T., K. Ohgaki, et al. (2004). "Architectural role of mitochondrial transcription factor A in maintenance of human mitochondrial DNA." *Mol Cell Biol* **24**(22): 9823-34.
- Kasai, H., P. F. Crain, et al. (1986). "Formation of 8-hydroxyguanine moiety in cellular DNA by agents producing oxygen radicals and evidence for its repair." *Carcinogenesis* **7**(11): 1849-51.
- Kaufman, B. A., N. Durisic, et al. (2007). "The mitochondrial transcription factor TFAM coordinates the assembly of multiple DNA molecules into nucleoid-like structures." *Mol Biol Cell* **18**(9): 3225-36.
- Keeney, P. M., C. K. Quigley, et al. (2009). "Mitochondrial gene therapy augments mitochondrial physiology in a Parkinson's disease cell model." *Hum Gene Ther* **20**(8): 897-907.
- Kelly, D. P. and R. C. Scarpulla (2004). "Transcriptional regulatory circuits controlling mitochondrial biogenesis and function." *Genes Dev* **18**(4): 357-68.

- Kidani, A., H. Izumi, et al. (2009). "Thioredoxin2 enhances the damaged DNA binding activity of mtTFA through direct interaction." *Int J Oncol* **35**(6): 1435-40.
- Klass, J., F. V. t. Murphy, et al. (2003). "The role of intercalating residues in chromosomal high-mobility-group protein DNA binding, bending and specificity." *Nucleic Acids Res* **31**(11): 2852-64.
- Koster, J. C., B. A. Marshall, et al. (2000). "Targeted overactivity of beta cell K(ATP) channels induces profound neonatal diabetes." *Cell* **100**(6): 645-54.
- Krissinel, E. and K. Henrick (2007). "Inference of macromolecular assemblies from crystalline state." *J Mol Biol* **372**(3): 774-97.
- Kruse, B., N. Narasimhan, et al. (1989). "Termination of transcription in human mitochondria: identification and purification of a DNA binding protein factor that promotes termination." *Cell* **58**(2): 391-7.
- Kucej, M. and R. A. Butow (2007). "Evolutionary tinkering with mitochondrial nucleoids." *Trends Cell Biol* **17**(12): 586-92.
- Kucej, M., B. Kucejova, et al. (2008). "Mitochondrial nucleoids undergo remodeling in response to metabolic cues." *J Cell Sci* **121**(Pt 11): 1861-8.
- Kuhn, A., R. Voit, et al. (1994). "Functional differences between the two splice variants of the nucleolar transcription factor UBF: the second HMG box determines specificity of DNA binding and transcriptional activity." *EMBO J* **13**(2): 416-24.
- Kukat, C., C. A. Wurm, et al. "Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and frequently contain a single copy of mtDNA." *Proc Natl Acad Sci U S A* **108**(33): 13534-9.
- Kuznetsov, S. V., S. Sugimura, et al. (2006). "Direct observation of DNA bending/unbending kinetics in complex with DNA-bending protein IHF." *Proc Natl Acad Sci U S A* **103**(49): 18515-20.
- Larsson, N. G., J. Wang, et al. (1998). "Mitochondrial transcription factor A is necessary for mtDNA maintenance and embryogenesis in mice." *Nat Genet* **18**(3): 231-6.
- Laudet, V., D. Stehelin, et al. (1993). "Ancestry and diversity of the HMG box superfamily." *Nucleic Acids Res* **21**(10): 2493-501.
- Levene, S. D. and B. H. Zimm (1989). "Understanding the anomalous electrophoresis of bent DNA molecules: a reptation model." *Science* **245**(4916): 396-9.
- Lobley, A., M. I. Sadowski, et al. (2009). "pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination." *Bioinformatics* **25**(14): 1761-7.
- Love, J. J., X. Li, et al. (1995). "Structural basis for DNA bending by the architectural transcription factor LEF-1." *Nature* **376**(6543): 791-5.
- Lu, B., S. Yadav, et al. (2007). "Roles for the human ATP-dependent Lon protease in mitochondrial DNA maintenance." *J Biol Chem* **282**(24): 17363-74.
- Malarkey, C. S., M. Bestwick, et al. "Transcriptional activation by mitochondrial transcription factor A involves preferential distortion of promoter DNA." *Nucleic Acids Res* **40**(2): 614-24.
- Mambo, E., X. Gao, et al. (2003). "Electrophile and oxidant damage of mitochondrial DNA leading to rapid evolution of homoplasmic mutations." *Proc Natl Acad Sci U S A* **100**(4): 1838-43.
- Mangus, D. A., S. H. Jang, et al. (1994). "Release of the yeast mitochondrial RNA polymerase specificity factor from transcription complexes." *J Biol Chem* **269**(42): 26568-74.
- Marchetti, P., T. Hirsch, et al. (1996). "Mitochondrial permeability transition triggers lymphocyte apoptosis." *J Immunol* **157**(11): 4830-6.
- Martin, M., J. Cho, et al. (2005). "Termination factor-mediated DNA loop between termination and initiation sites drives mitochondrial rRNA synthesis." *Cell* **123**(7): 1227-40.
- Matsusaka, H., M. Ikeuchi, et al. (2005). "Selective disruption of MMP-2 gene exacerbates myocardial inflammation and dysfunction in mice with cytokine-induced cardiomyopathy." *Am J Physiol Heart Circ Physiol* **289**(5): H1858-64.
- Matsushima, Y., K. Matsumura, et al. (2003). "Functional domains of chicken mitochondrial transcription factor A for the maintenance of mitochondrial DNA copy number in lymphoma cell line DT40." *J Biol Chem* **278**(33): 31149-58.
- McCulloch, V., B. L. Seidel-Rogol, et al. (2002). "A human mitochondrial transcription factor is related to RNA adenine methyltransferases and binds S-adenosylmethionine." *Mol Cell Biol* **22**(4): 1116-25.
- McCulloch, V. and G. S. Shadel (2003). "Human mitochondrial transcription factor B1 interacts with the C-terminal activation region of h-mtTFA and stimulates transcription independently of its RNA methyltransferase activity." *Mol Cell Biol* **23**(16): 5816-24.
- McGuffin, L. J., K. Bryson, et al. (2000). "The PSIPRED protein structure prediction server." *Bioinformatics* **16**(4): 404-5.
- Megraw, T. L. and C. B. Chae (1993). "Functional complementarity between the HMG1-like yeast mitochondrial histone HM and the bacterial histone-like protein HU." *J Biol Chem* **268**(17): 12758-63.
- Mercer, T. R., S. Neph, et al. "The human mitochondrial transcriptome." *Cell* **146**(4): 645-58.
- Metodiev, M. D., N. Lesko, et al. (2009). "Methylation of 12S rRNA is necessary for in vivo stability of the small subunit of the mammalian mitochondrial ribosome." *Cell Metab* **9**(4): 386-97.

- Miyakawa, I., N. Sando, et al. (1987). "Isolation of morphologically intact mitochondrial nucleoids from the yeast, *Saccharomyces cerevisiae*." *J Cell Sci* **88** (Pt 4): 431-9.
- Montoya, J., T. Christianson, et al. (1982). "Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA." *Proc Natl Acad Sci U S A* **79**(23): 7195-9.
- Mouw, K. W. and P. A. Rice (2007). "Shaping the *Borrelia burgdorferi* genome: crystal structure and binding properties of the DNA-bending protein Hbb." *Mol Microbiol* **63**(5): 1319-30.
- Murphy, E. C., V. B. Zhurkin, et al. (2001). "Structural basis for SRY-dependent 46-X,Y sex reversal: modulation of DNA bending by a naturally occurring point mutation." *J Mol Biol* **312**(3): 481-99.
- Murphy, F. V. t., R. M. Sweet, et al. (1999). "The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition." *EMBO J* **18**(23): 6610-8.
- Murshudov, G. N., A. A. Vagin, et al. (1997). "Refinement of macromolecular structures by the maximum-likelihood method." *Acta Crystallogr D Biol Crystallogr* **53**(Pt 3): 240-55.
- Nadassy, K., S. J. Wodak, et al. (1999). "Structural features of protein-nucleic acid recognition sites." *Biochemistry* **38**(7): 1999-2017.
- Nass, M. M. (1969). "Reversible generation of circular dimer and higher multiple forms of mitochondrial DNA." *Nature* **223**(5211): 1124-9.
- Nass, S. and M. M. Nass (1963). "Intramitochondrial Fibers with DNA Characteristics. II. Enzymatic and Other Hydrolytic Treatments." *J Cell Biol* **19**: 613-29.
- Nemajerova, A., S. Erster, et al. (2005). "The post-translational phosphorylation and acetylation modification profile is not the determining factor in targeting endogenous stress-induced p53 to mitochondria." *Cell Death Differ* **12**(2): 197-200.
- Ngo, H. B., J. T. Kaiser, et al. "The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA." *Nat Struct Mol Biol* **18**(11): 1290-6.
- Nikolov, D. B., H. Chen, et al. (1996). "Crystal structure of a human TATA box-binding protein/TATA element complex." *Proc Natl Acad Sci U S A* **93**(10): 4862-7.
- Noack, H., T. Bednarek, et al. (2006). "TFAM-dependent and independent dynamics of mtDNA levels in C2C12 myoblasts caused by redox stress." *Biochim Biophys Acta* **1760**(2): 141-50.
- O'Mahoney, J. V., M. R. Brandon, et al. (1991). "Developmental and tissue-specific regulation of ovine insulin-like growth factor II (IGF-II) mRNA expression." *Mol Cell Endocrinol* **78**(1-2): 87-96.
- Ohgaki, K., T. Kanki, et al. (2007). "The C-terminal tail of mitochondrial transcription factor a markedly strengthens its general binding to DNA." *J Biochem* **141**(2): 201-11.
- Ohno, T., S. Umeda, et al. (2000). "Binding of human mitochondrial transcription factor A, an HMG box protein, to a four-way DNA junction." *Biochem Biophys Res Commun* **271**(2): 492-8.
- Parisi, M. A. and D. A. Clayton (1991). "Similarity of human mitochondrial transcription factor 1 to high mobility group proteins." *Science* **252**(5008): 965-9.
- Parisi, M. A., B. Xu, et al. (1993). "A human mitochondrial transcriptional activator can functionally replace a yeast mitochondrial HMG-box protein both in vivo and in vitro." *Mol Cell Biol* **13**(3): 1951-61.
- Parkhurst, K. M., R. M. Richards, et al. (1999). "Intermediate species possessing bent DNA are present along the pathway to formation of a final TBP-TATA complex." *J Mol Biol* **289**(5): 1327-41.
- Parra, V., H. Verdejo, et al. "The complex interplay between mitochondrial dynamics and cardiac metabolism." *J Bioenerg Biomembr* **43**(1): 47-51.
- Pastukh, V., I. Shokolenko, et al. (2007). "Human mitochondrial transcription factor A possesses multiple subcellular targeting signals." *FEBS J* **274**(24): 6488-99.
- Paull, T. T., M. J. Haykinson, et al. (1993). "The nonspecific DNA-binding and -bending proteins HMG1 and HMG2 promote the assembly of complex nucleoprotein structures." *Genes Dev* **7**(8): 1521-34.
- Pfannschmidt, T., A. Nilsson, et al. (1999). "Direct transcriptional control of the chloroplast genes *psbA* and *psaAB* adjusts photosynthesis to light energy distribution in plants." *IUBMB Life* **48**(3): 271-6.
- Pham, X. H., G. Farge, et al. (2006). "Conserved sequence box II directs transcription termination and primer formation in mitochondria." *J Biol Chem* **281**(34): 24647-52.
- Revyakin, A., R. H. Ebright, et al. (2004). "Promoter unwinding and promoter clearance by RNA polymerase: detection by single-molecule DNA nanomanipulation." *Proc Natl Acad Sci U S A* **101**(14): 4776-80.
- Reyes, A., M. Mezzina, et al. (2002). "Human mitochondrial transcription factor A (mtTFA): gene structure and characterization of related pseudogenes." *Gene* **291**(1-2): 223-32.
- Rice, P. A., S. Yang, et al. (1996). "Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn." *Cell* **87**(7): 1295-306.
- Ringel, R., M. Sologub, et al. "Structure of human mitochondrial RNA polymerase." *Nature* **478**(7368): 269-73.
- Roca, J. "Transcriptional inhibition by DNA torsional stress." *Transcription* **2**(2): 82-85.
- Rubio-Cosials, A., J. F. Sidow, et al. "Human mitochondrial transcription factor A induces a U-turn structure in the light strand promoter." *Nat Struct Mol Biol* **18**(11): 1281-9.
- Schmidt, B., S. Konig, et al. (1995). "Small-angle X-ray solution scattering study on the dimerization of the FKBP25mem from *Legionella pneumophila*." *FEBS Lett* **372**(2-3): 169-72.

- Seeman, N. C., J. M. Rosenberg, et al. (1976). "Sequence-specific recognition of double helical nucleic acids by proteins." *Proc Natl Acad Sci U S A* **73**(3): 804-8.
- Seidel-Rogol, B. L., V. McCulloch, et al. (2003). "Human mitochondrial transcription factor B1 methylates ribosomal RNA at a conserved stem-loop." *Nat Genet* **33**(1): 23-4.
- Seidel-Rogol, B. L. and G. S. Shadel (2002). "Modulation of mitochondrial transcription in response to mtDNA depletion and repletion in HeLa cells." *Nucleic Acids Res* **30**(9): 1929-34.
- Shadel, G. S. and D. A. Clayton (1997). "Mitochondrial DNA maintenance in vertebrates." *Annu Rev Biochem* **66**: 409-35.
- Sheldrick, G. M. (2008). "A short history of SHELX." *Acta Crystallogr A* **64**(Pt 1): 112-22.
- Shoubridge, E. A. (2009). "Something old, something new, something borrowed." *Cell Metab* **9**(4): 307-8.
- Shutt, T. E., M. Bestwick, et al. "The core human mitochondrial transcription initiation complex: It only takes two to tango." *Transcription* **2**(2): 55-59.
- Silva, J. P., M. Kohler, et al. (2000). "Impaired insulin secretion and beta-cell loss in tissue-specific knockout mice with mitochondrial diabetes." *Nat Genet* **26**(3): 336-40.
- Singh, G., N. Neckelmann, et al. (1987). "Conformational mutations in human mitochondrial DNA." *Nature* **329**(6136): 270-2.
- Smart, O. S., T. O. Womack, et al. "Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER." *Acta Crystallogr D Biol Crystallogr* **68**(Pt 4): 368-80.
- Smeets, A., C. Evrard, et al. (2005). "Crystal structures of oxidized and reduced forms of human mitochondrial thioredoxin 2." *Protein Sci* **14**(10): 2610-21.
- Sologub, M., D. Litonin, et al. (2009). "TFB2 is a transient component of the catalytic site of the human mitochondrial RNA polymerase." *Cell* **139**(5): 934-44.
- Stefanovsky, V. Y., G. Pelletier, et al. (2001). "DNA looping in the RNA polymerase I enhancer is the result of non-cooperative in-phase bending by two UBF molecules." *Nucleic Acids Res* **29**(15): 3241-7.
- Sterner, R., G. Vidali, et al. (1978). "Postsynthetic modification of high mobility group proteins. Evidence that high mobility group proteins are acetylated." *J Biol Chem* **253**(21): 7601-4.
- Stott, K., G. S. Tang, et al. (2006). "Structure of a complex of tandem HMG boxes and DNA." *J Mol Biol* **360**(1): 90-104.
- Stott, K., M. Watson, et al. "Tail-mediated collapse of HMGB1 is dynamic and occurs via differential binding of the acidic tail to the A and B domains." *J Mol Biol* **403**(5): 706-22.
- Sumitani, M., K. Kasashima, et al. "Biochemical properties of *Caenorhabditis elegans* HMG-5, a regulator of mitochondrial DNA." *J Biochem* **149**(5): 581-9.
- Sutovsky, P., R. D. Moreno, et al. (2000). "Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos." *Biol Reprod* **63**(2): 582-90.
- Thomas, J. O. and A. A. Travers (2001). "HMG1 and 2, and related 'architectural' DNA-binding proteins." *Trends Biochem Sci* **26**(3): 167-74.
- Thompson, J. F. and A. Landy (1988). "Empirical estimation of protein-induced DNA bending angles: applications to lambda site-specific recombination complexes." *Nucleic Acids Res* **16**(20): 9687-705.
- Tiranti, V., A. Savoia, et al. (1997). "Identification of the gene encoding the human mitochondrial RNA polymerase (h-mtRPOL) by cyberscreening of the Expressed Sequence Tags database." *Hum Mol Genet* **6**(4): 615-25.
- Tominaga, K., J. Hayashi, et al. (1993). "Smaller isoform of human mitochondrial transcription factor 1: its wide distribution and production by alternative splicing." *Biochem Biophys Res Commun* **194**(1): 544-51.
- Topper, J. N. and D. A. Clayton (1989). "Identification of transcriptional regulatory elements in human mitochondrial DNA by linker substitution analysis." *Mol Cell Biol* **9**(3): 1200-11.
- Travers, A. and H. Drew (1997). "DNA recognition and nucleosome organization." *Biopolymers* **44**(4): 423-33.
- Uchiumi, T., K. Ohgaki, et al. "ERAL1 is associated with mitochondrial ribosome and elimination of ERAL1 leads to mitochondrial dysfunction and growth retardation." *Nucleic Acids Res* **38**(16): 5554-68.
- Ugrinova, I., E. A. Pasheva, et al. (2001). "In vivo acetylation of HMG1 protein enhances its binding affinity to distorted DNA structures." *Biochemistry* **40**(48): 14655-60.
- Uson, I. and G. M. Sheldrick (1999). "Advances in direct methods for protein crystallography." *Curr Opin Struct Biol* **9**(5): 643-8.
- Vagin, A. A. and M. N. Isupov (2001). "Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps." *Acta Crystallogr D Biol Crystallogr* **57**(Pt 10): 1451-6.
- van Dijk, M. and A. M. Bonvin (2009). "3D-DART: a DNA structure modelling server." *Nucleic Acids Res* **37**(Web Server issue): W235-9.
- Van Remmen, H., Y. Ikeno, et al. (2003). "Life-long reduction in MnSOD activity results in increased DNA damage and higher incidence of cancer but does not accelerate aging." *Physiol Genomics* **16**(1): 29-37.
- Van Tuyle, G. C. and M. L. McPherson (1979). "A compact form of rat liver mitochondrial DNA stabilized by bound proteins." *J Biol Chem* **254**(13): 6044-53.
- Vuzman, D., A. Azia, et al. "Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails." *J Mol Biol* **396**(3): 674-84.

- Vuzman, D., M. Polonsky, et al. "Facilitated DNA search by multidomain transcription factors: cross talk via a flexible linker." *Biophys J* **99**(4): 1202-11.
- Wai, T., D. Teoli, et al. (2008). "The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes." *Nat Genet* **40**(12): 1484-8.
- Wallace, D. C. (2005). "Mitochondria and cancer: Warburg addressed." *Cold Spring Harb Symp Quant Biol* **70**: 363-74.
- Wallace, D. C. (2005). "A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine." *Annu Rev Genet* **39**: 359-407.
- Wallace, K. B. (2008). "Mitochondrial off targets of drug therapy." *Trends Pharmacol Sci* **29**(7): 361-6.
- Wanrooij, P. H., J. P. Uhler, et al. "G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation." *Proc Natl Acad Sci U S A* **107**(37): 16072-7.
- Weir, H. M., P. J. Kraulis, et al. (1993). "Structure of the HMG box motif in the B-domain of HMG1." *EMBO J* **12**(4): 1311-9.
- Winkley, C. S., M. J. Keller, et al. (1985). "A multicomponent mitochondrial RNA polymerase from *Saccharomyces cerevisiae*." *J Biol Chem* **260**(26): 14214-23.
- Wong, T. S., S. Rajagopalan, et al. (2009). "Biophysical characterizations of human mitochondrial transcription factor A and its binding to tumor suppressor p53." *Nucleic Acids Res* **37**(20): 6765-83.
- Wu, H. M. and D. M. Crothers (1984). "The locus of sequence-directed and protein-induced DNA bending." *Nature* **308**(5959): 509-13.
- Xu, B. and D. A. Clayton (1996). "RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA-DNA hybrids serving as primers." *EMBO J* **15**(12): 3135-43.
- Yoshida, Y., H. Izumi, et al. (2002). "Human mitochondrial transcription factor A binds preferentially to oxidatively damaged DNA." *Biochem Biophys Res Commun* **295**(4): 945-51.
- Zhang, Q., J. T. Yu, et al. "Mitochondrial transcription factor A (TFAM) polymorphisms and risk of late-onset Alzheimer's disease in Han Chinese." *Brain Res* **1368**: 355-60.
- Zhang, X., Y. Zheng, et al. "Disruption of the mitochondrial thioredoxin system as a cell death mechanism of cationic triphenylmethanes." *Free Radic Biol Med* **50**(7): 811-20.
- Zlatanova, J., J. Yaneva, et al. (1998). "Proteins that specifically recognize cisplatin-damaged DNA: a clue to anticancer activity of cisplatin." *FASEB J* **12**(10): 791-9.