



# Estudio estructural de proteínas implicadas en el metabolismo del genoma mitocondrial: Helicasa y Factor de Transcripción A Mitocondrial, TFAM

Pablo Fernández Millán

**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](http://www.tdx.cat)) i a través del Dipòsit Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) 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 ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (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](http://www.tdx.cat)) y a través del Repositorio Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) 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 o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (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](http://www.tdx.cat)) service and by the UB Digital Repository ([diposit.ub.edu](http://diposit.ub.edu)) 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 nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those 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

DEPARTAMENT

Bioquímica i Biologia Molecular

**Estudio estructural de proteínas  
implicadas en el metabolismo del  
genoma mitocondrial: Helicasa y Factor  
de Transcripción A Mitocondrial, TFAM.**

**Pablo Fernández Millán**

[2014]



UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA

DEPARTAMENT

Bioquímica i Biologia Molecular

PROGRAMA DE DOCTORAT

Biotecnologia

**Estudio estructural de proteínas implicadas en el metabolismo del genoma mitocondrial:  
Helicasa y Factor de Transcripción Mitocondrial A, TFAM.**

Memòria presentada per Pablo Fernández Millán per optar al títol de doctor per la universitat  
de Barcelona

Directora de la tesis Maria Solà i Vilarrubias

Doctorando Pablo Fernández Millán

Tutora de la tesis Josefa Badia Palacín

Pablo Fernández Millán 2014



# 1 ÍNDICE.

|       |   |    |
|-------|---|----|
| 1     | ÍNDICE.....   | 5  |
| 2     | INDICE DE TABLAS. ....  | 9  |
| 3     | ÍNDICE DE FIGURAS. ....   | 9  |
| 4     | OBJETIVOS. ....   | 11 |
| 5     | AGRADECIMIENTOS.....  | 13 |
|       | <b>INTRODUCCIÓN</b> .....   | 17 |
| 6     | INTRODUCCIÓN.....   | 19 |
| 6.1   | Mitocondria y ADN mitocondrial. ....  | 19 |
| 6.1.1 | El genoma mitocondrial humano. ....   | 20 |
| 6.2   | Replicación. ....   | 23 |
| 6.2.1 | Replicación del genoma mitocondrial.....  | 23 |
| 6.3   | Helicasas.....  | 25 |
| 6.3.1 | Clasificación de las helicasas. ....  | 25 |
| 6.3.2 | Mecanismo de hidrólisis de NTP de helicasas hexaméricas. ....                                       | 29 |
| 6.3.3 | Movimiento de traslocación asociado a la actividad NTPasa de las helicasas hexaméricas. ....        | 30 |
| 6.3.4 | Replisoma de bacteriófago T7 y actividad de gp4.....  | 31 |
| 6.4   | Helicasa mitocondrial humana, Twinkle. ....   | 33 |
| 6.4.1 | Mutaciones deletéreas de Twinkle. ....  | 36 |
| 6.5   | Transcripción del genoma mitocondrial. ....   | 38 |
| 6.6   | Factor de transcripción A mitocondrial humano (TFAM).....   | 40 |
| 6.6.1 | TFAM es una proteína con dos dominios HMGbox.....   | 40 |
| 6.6.2 | Interacción de TFAM a las secuencias de ADN LSP y HSP1. ....  | 41 |
| 6.6.3 | TFAM y la compactación del mtADN.....   | 44 |
| 6.6.4 | Características generales de los dominios “ <i>High mobility group</i> ” (HMG). ....                | 45 |
| 6.6.5 | Análisis termodinámico de los complejos HMGbox con ADN.....   | 45 |
|       | <b>MATERIALES, MÉTODOS</b> .....  | 49 |
| 7     | MATERIALES, MÉTODOS Y FUNDAMENTOS.....  | 51 |
| 7.1   | Obtención de proteína recombinante: clonación, expresión y solubilización. Descripción general..... | 51 |
| 7.1.1 | Biología molecular de TFAM. ....  | 53 |
| 7.1.2 | Biología molecular de Twinkle y sus distintos constructos.....                                      | 53 |
| 7.1.3 | Expresión heteróloga de Twinkle. ....   | 56 |
| 7.1.4 | Solubilidad de Twinkle y sus múltiples constructos. ....  | 58 |
| 7.2   | Purificación de proteínas recombinantes. ....   | 59 |
| 7.2.1 | Fundamento.....   | 59 |

|         |  |    |
|---------|--|----|
| 7.2.1.1 | Cromatografía de afinidad. ....  | 59 |
| 7.2.1.2 | Cromatografía de exclusión por tamaño.....   | 60 |
| 7.2.2   | Protocolo de purificación de TFAM.....   | 61 |
| 7.2.3   | Protocolo de purificación de Twinkle y sus constructos.....  | 61 |
| 7.3     | Análisis por electroforesis no desnaturizante de la interacción de proteína con ADN (EMSA).....                            | 63 |
| 7.3.1   | Fundamentos. ....  | 63 |
| 7.3.2   | Análisis de la interacción TFAM/ADN mediante EMSA.....   | 64 |
| 7.3.3   | Análisis de la interacción Twinkle/ADN mediante EMSA.....  | 64 |
| 7.4     | Calorimetría Isotérmica de Titulación (ITC) aplicada a la interacción proteína/ADN. 65                                     |    |
| 7.4.1   | Fundamentos. ....  | 65 |
| 7.4.2   | Estudio de la interacción entre TFAM y distintas secuencias de ADN mediante ITC.. 66                                       |    |
| 7.5     | Ensayo <i>in vitro</i> de actividad helicasa.....  | 67 |
| 7.5.1   | Fundamentos. ....  | 67 |
| 7.5.2   | Ensayo de actividad helicasa con Twinkle.....  | 68 |
| 7.6     | Dispersión dinámica de la luz.....   | 69 |
| 7.6.1   | Fundamentos. ....  | 69 |
| 7.6.2   | Estudios de DLS con Twinkle. ....  | 69 |
| 7.7     | Barrido diferencial de fluorescencia. ....   | 70 |
| 7.7.1   | Fundamentos. ....  | 70 |
| 7.7.2   | Análisis de estabilidad mediante DSF con Twinkle. ....   | 70 |
| 7.8     | Cristalografía de proteínas y difracción de rayos X.....   | 71 |
| 7.8.1   | Principios y métodos de cristalización. ....   | 71 |
| 7.8.2   | Cristalización de TFAM en complejo con SITE-X. ....  | 73 |
| 7.8.3   | Ensayos de cristalización de Twinkle.....  | 74 |
| 7.8.4   | Difracción de rayos X de cristales de macromoléculas. ....   | 75 |
| 7.8.4.1 | Fundamentos. ....  | 75 |
| 7.8.4.2 | Procesado, escalado de los datos de difracción y resolución de la estructura cristalográfica del complejo TFAM/Site-X..... | 79 |
| 7.9     | Dispersión de bajo ángulo de rayos X. ....   | 79 |
| 7.9.1   | Fundamentos. ....  | 79 |
| 7.9.2   | Estudios de SAXS con Twinkle. ....   | 82 |
| 7.10    | Construcción de proteínas por homología: MODELLER.....   | 85 |
| 7.10.1  | Fundamentos. ....  | 85 |
| 7.10.2  | Modelo tridimensional de Twinkle mediante predicción de estructura.....  | 86 |
| 7.11    | Microscopía Electrónica con Twinkle.....   | 87 |

|           |   |            |
|-----------|---|------------|
| 7.11.1    | Fundamentos.....  | 87         |
|           | <b>RESULTADOS.....</b>  | <b>91</b>  |
|           | <b>RESULTADOS Y DISCUSIÓN.....</b>  | <b>93</b>  |
| 8         | <b>RESULTADOS Y DISCUSIÓN.....</b>  | <b>95</b>  |
| 8.1       | Resultado y discusión de TFAM en complejo con Site-X.....   | 95         |
| 8.1.1     | Resultados obtenidos con el complejo TFAM/Site-X.....   | 95         |
| 8.1.1.1   | Clonación, purificación y formación del complejo TFAM/Site-X.....   | 95         |
| 8.1.1.2   | Cristalización y resolución de la estructura TFAM/Site-X.....   | 96         |
| 8.1.1.3   | Cristalización de TFAM/Site-X.....  | 96         |
| 8.1.1.4   | Procesado de datos y resolución de la estructura cristalográfica.....   | 97         |
| 8.1.1.5   | Estructura Tridimensional del complejo TFAM/Site-X.....   | 101        |
| 8.1.1.5.1 | Estructura complejo TFAM/Site-X.....  | 101        |
| 8.1.1.5.2 | Interacción de TFAM con el ADN Site-X.....  | 105        |
| 8.1.1.5.3 | Interacción del “ <i>linker</i> ” con las HMGbox1 y 2.....  | 108        |
| 8.1.1.5.4 | Comparación de la estructura de TFAM/Site-X con las estructuras de otros complejos TFAM/ADN.....                                | 109        |
| 8.1.1.5.5 | Análisis de la estructura del ADN de Site-X.....  | 112        |
| 8.1.1.5.6 | Comparación las interacciones con el ADN entre distintos complejos de TFAM.....   | 115        |
| 8.1.1.6   | Análisis complejo TFAM/DNA mediante EMSA.....   | 116        |
| 8.1.1.7   | Calorimetría de Titulación Isoterma de la interacción de TFAM con el ADN.....   | 118        |
| 8.1.1.8   | Análisis computacional de las propiedades físicas de la secuencia Site-X.....   | 121        |
| 8.1.2     | Discusión de los resultados del estudio de TFAM en complejo con Site-X.....   | 123        |
| 8.1.2.1   | Estructura TFAM/Site-X.....   | 123        |
| 8.1.2.2   | Propiedades mecánicas del ADN Site-X y su reconocimiento por TFAM.....  | 126        |
| 8.1.2.3   | Análisis termodinámico de la formación del complejo TFAM/Site-X.....  | 126        |
|           | <b>RESULTADOS Y DISCUSIÓN.....</b>  | <b>129</b> |
|           | <b>TWINKLE.....</b>   | <b>129</b> |
| 8.2       | Resultados y discusión del estudio estructural de Twinkle.....  | 131        |
| 8.2.1     | Resultados del estudio estructural de Twinkle.....  | 131        |
| 8.2.1.1   | Diseño de constructos genéticos de Twinkle.....   | 131        |
| 8.2.1.2   | Expresión de Twinkle recombinante en E. coli y ensayos de solubilidad.....  | 135        |
| 8.2.1.3   | Purificación y estudios biofísicos del dominio helicasa de Twinkle: (GB1)- $\Delta$ 358Twinkle.....                             | 139        |
| 8.2.1.3.1 | Purificación de (GB1)- $\Delta$ 358Twinkle.....   | 139        |
| 8.2.1.3.2 | Barrido diferencial de fluorescencia (DSF) y dispersión dinámica de la luz (DLS) del constructo (GB1)- $\Delta$ 358Twinkle..... | 141        |
| 8.2.1.3.3 | Microscopía electrónica del constructo (GB1)- $\Delta$ 358TWINKLE.....  | 142        |
| 8.2.1.4   | Purificación y estudios biofísicos del constructo Twinkle.....  | 143        |
| 8.2.1.4.1 | Purificación de Twinkle.....  | 143        |
| 8.2.1.4.2 | Barrido diferencial de fluorescencia (DSF) y dispersión dinámica de la luz (DLS) del constructo Twinkle.....                    | 144        |

|           |  |     |
|-----------|--|-----|
| 8.2.1.4.3 | Análisis del complejo Twinkle-ADN por técnicas electroforéticas: EMSA.             | 146 |
| 8.2.1.4.4 | Análisis de la actividad de Twinkle: ensayo helicasa.....                          | 148 |
| 8.2.1.5   | Cristalización de Twinkle y distintos constructos.....                             | 149 |
| 8.2.1.6   | Generación de un modelo de Twinkle in silico con MODELLER.....                     | 151 |
| 8.2.1.7   | Microscopia electrónica.....   | 155 |
| 8.2.1.8   | Estructura de Twinkle en solución por Dispersión de rayos X de ángulo bajo (SAXS). | 158 |
| 8.2.1.8.1 | Medida experimental de Twinkle mediante SAXS.....                                  | 158 |
| 8.2.1.8.2 | Ajuste de los modelos de Twinkle a las curvas experimentales de SAXS.              | 162 |
| 8.2.2     | Discusión de los resultados de Twinkle.....  | 166 |
|           | <b>CONCLUSIONES</b> .....  | 171 |
| 9         | CONCLUSIONES.....  | 173 |
| 9.1       | Conclusiones del proyecto TFAM en complejo con Site-X.....                         | 173 |
| 9.2       | Conclusiones Twinkle.....  | 173 |
|           | <b>BIBLIOGRAFÍA</b> .....  | 175 |
| 10        | BIBLIOGRAFÍA.....  | 177 |
| 11        | ANEXO.....   | 187 |

## 2 ÍNDICE DE TABLAS.

|  |     |
|--|-----|
| Tabla_1. Lista de enzimas de restricción.....  | 53  |
| Tabla_2. Lista de cebadores. ....  | 54  |
| Tabla_3. Vectores de expresión. ....   | 55  |
| Tabla_4. Lista de tampones de rotura. ....   | 59  |
| Tabla_5. Secuencia de los Oligonucleótidos empleados en ITC. ....                                    | 67  |
| Tabla_6. Lista de tampones empleados en los ensayos de DSF. ....                                     | 71  |
| Tabla_7. Muestras de Twinkle analizadas por SAXS. ....   | 83  |
| Tabla_8. Delimitación de los residuos de las regiones flexibles de Twinkle.....                      | 85  |
| Tabla_9. Valores de las soluciones de PHASER para el complejo TFAM/Site-X.....                       | 98  |
| Tabla_10. Datos cristalográficos de los cristales del complejo TFAM/Site-X.....                      | 100 |
| Tabla_11. Valores de los r.m.s.d de los complejos TFAM/Site-X.....                                   | 103 |
| Tabla_12. Valores de r.m.s.d de los complejos TFAM/ADN.....  | 111 |
| Tabla_13. Resultados ITC. ....   | 121 |
| Tabla_14. Resultado expresión de los constructos de Twinkle en distintas cepas bacterianas.<br>..... | 136 |
| Tabla_15. Resultado de la expresión, solubilidad y purificación. ....                                | 138 |
| Tabla_16. Resultado cristalización de Twinkle y distintos constructos. ....                          | 150 |
| Tabla_17. Experimentos de SAXS con Twinkle.....  | 161 |
| Tabla_18. Resultados de SAXS con Twinkle. ....   | 165 |

## 3 ÍNDICE DE FIGURAS.

|  |     |
|--|-----|
| Figura_1. Cadena de transporte de electrones y la fosforilación oxidativa.....                       | 20  |
| Figura_2. Genoma mitocondrial y el “D-loop”.....   | 22  |
| Figura_3. Replicación mitocondrial.....  | 25  |
| Figura_4. Estructura de dos motivos tipo ASCE, y topología de los dominios ASCE, RecA &<br>AAA+..... | 26  |
| Figura_5. Mecanismo de las helicasas hexaméricas. ....   | 30  |
| Figura_6. Modelo del replisoma de bacteriófago T7. ....  | 33  |
| Figura_7. Mutaciones asociadas a <i>Twinkle</i> . ....   | 37  |
| Figura_8. Origen de la transcripción. ....   | 39  |
| Figura_9. Representación de TFAM en bloques y estructura secundaria. ....                            | 40  |
| Figura_10. Sitios de unión específicos de TFAM.....  | 42  |
| Figura_11. TFAM complejo con las secuencias LSP y HSP1 de ADN. ....                                  | 43  |
| Figura_12. Rango de resolución de las columnas de purificación por exclusión por tamaño. ....        | 60  |
| Figura_13. Esquema de un termograma de ITC.....  | 66  |
| Figura_14. Esquema de un ensayo teórico helicasa. ....   | 68  |
| Figura_15. Diagrama de fases de solubilidad.....   | 72  |
| Figura_16. Clones de TFAM.....   | 95  |
| Figura_17. Purificación TFAM. ....   | 96  |
| Figura_18. Resultado gotas de cristalización del complejo. ....                                      | 97  |
| Figura_19. Mapa de densidad electrónica. ....  | 99  |
| Figura_20. Representación de Ramachandran. ....  | 100 |

|   |     |
|---|-----|
| Figura_21. TFAM/Site-X. ....  | 101 |
| Figura_22. Unidad Asimétrica de TFAM/Site-X.....  | 102 |
| Figura_23. Superposición complejos TFAM/Site-X y sus HMGbox. ....   | 104 |
| Figura_24. Interacción TFAM con Site-X. ....  | 107 |
| Figura_25. Interacciones entre el “ <i>linker</i> ” y las dos HMGbox. ....                                      | 109 |
| Figura_26. Superposición de los complejos TFAM/ADN.....   | 110 |
| Figura_27. Superposición TFAM con todas las secuencias. ....  | 112 |
| Figura_28. Inclínación y Giro helicoidal del ADN Site-X. ....   | 113 |
| Figura_29. Comparación Site-X con Dodecámero. ....  | 115 |
| Figura_30. Geles EMSA de TFAM. ....   | 117 |
| Figura_31. Alineamiento secuencias de ADN que une TFAM. ....  | 119 |
| Figura_32. Resultados ITC. ....   | 120 |
| Figura_33. Propiedades mecánicas de Site-X. ....  | 123 |
| Figura_34. Predicción de estructura secundaria.....   | 131 |
| Figura_35. Constructos de Twinkle. ....   | 132 |
| Figura_36. Resultado de expresión y solubilidad de (GB1)- $\Delta$ 358Twinkle analizado por geles SDS-PAGE..... | 137 |
| Figura_37. Purificación de (GB1)- $\Delta$ 358Twinkle.....  | 140 |
| Figura_38. Termograma de DSF de (GB1)- $\Delta$ 358Twinkle y ensayos por DLS. ....                              | 142 |
| Figura_39. Microscopía electrónica de (GB1)- $\Delta$ 358Twinkle. ....  | 143 |
| Figura_40. Purificación de Twinkle. ....  | 144 |
| Figura_41. Resultado y termograma de DSF del constructo Twinkle. ....   | 145 |
| Figura_42. Ensayos de interacción Twinkle y el ADN mediante EMSA. ....  | 147 |
| Figura_43. Ensayo helicasa de la actividad de Twinkle. ....   | 148 |
| Figura_44. Estructuras empleadas como plantilla para MODELLER. ....   | 152 |
| Figura_45. Alineamiento para la elaboración del modelo atómico 3D. ....   | 153 |
| Figura_46. Modelos generados con MODELLER. ....   | 154 |
| Figura_47. Imágenes de microscopía electrónica y mapa a baja resolución. ....                                   | 156 |
| Figura_48. Mapa de crio-microscopía electrónica.....  | 157 |
| Figura_49. Ajuste de modelos 3D a los mapas de crio-EM. ....  | 158 |
| Figura_50. Curvas de SAXS de Twinkle.....   | 160 |
| Figura_51. Curva SAXS Twinkle. ....   | 162 |
| Figura_52. Ajuste de los modelos de Twinkle a las curvas de SAXS. ....  | 163 |
| Figura_53. Resultado del ajuste a la curva de SAXS y modelos EOM. ....  | 164 |

#### **4 OBJETIVOS.**

- Cristalizar TFAM en complejo con el sitio de unión al ADN mitocondrial Site-X.
- Resolver la estructura cristalográfica del complejo TFAM/Site-X.
- Estudiar bioquímicamente el complejo TFAM/Site-X.
- Clonar Twinkle en un sistema de expresión para su producción en *E. coli*.
- Establecer un protocolo para la solubilización, purificación y estabilización de Twinkle y sus constructos genéticos para su cristalización.
- Cristalizar Twinkle.
- Estudiar la estructura de Twinkle en disolución: SAXS.
- Resolver la estructura tridimensional de Twinkle por microscopía electrónica.



## 5 AGRADECIMIENTOS.

Quiero agradecer a todas aquellas personas que están involucradas en el desarrollo de los proyectos que expongo en esta tesis, primero por supuesto a mi directora de tesis Dra. Solà por darme la oportunidad de realizar mi trabajo de tesis en su grupo de investigación, y otra segunda oportunidad en la investigación, al Dr. Valle y a la Melisa Lazaro por su excelente trabajo y colaboración, aportando todos los datos de microscopía que han sido esenciales para el proyecto de Twinkle, al Dr Bernado por enseñarme la parte SAXS de la ciencia y su indispensable supervisión en mi trabajo en este campo, al Prof. J. Spelbrink, a su grupo y especialmente al Dr. Gerhold y a Sirin Cainz con quienes trabajé codo con codo durante mi estancia en Holanda y con los que hemos peleado mucho por avanzar todo lo posible en el proyecto de Twinkle, al Dr. Prohens por nuestras instructivas discusiones de termodinámica y su supervisión con los análisis calorimétricos, a la Dra Batistini por su aportación al proyecto TFAM/Site-X con los estudios computacionales del ADN y enseñame que el ADN es mucho más que una secuencia. Finalmente, quiero también agradecer al Dr. Gomis por acojerme en su laboratorio durante un año para desarrollar mi proyecto.

Tampoco puedo cerrar este capítulo sin agradecer a otras muchas personas, que aunque su aportación haya sido de un modo menos profesional no por ello ha tenido menor importancia. No puedo olvidarme de aquellos que mis primeros mentores en la ciencia, gracias Danel y Maria Lucas por vuestro esfuerzo en enseñarme lo que es la investigación y vuestro apoyo incondicional en los malos momentos, y de paso gracias a Bea por sus cafés de los lunes, a Borja y los conciertos que siempre me perdí, a Xavi... unos de mis mejores compañeros de piso y por esas noches de "the Young ones", a David, a Oscar M. (ya sabes porque), a Tammo y nuestras excursiones en las pincherías, a Luis y los fines de semana de sofá y maratón de deporte.

Quiero hacer un hueco especial a Elena Rubio, gracias por apollo y ayuda en la universidad, y después de ella, siempre has sido una referencia de quien aprender, y es un privilegio tenerte como amiga.

Gracias a Cri7, cómo ha cambiado desde que llegue, primero un gracias por soportar mis chistes, chascarrillos y comentarios jocosos bastate faltos de sutileza. De modo individual, gracias Nere por darle siempre a todo cordura, gracias Albarito (Oskar para los amigos) tuviste un paso corto pero intenso por el laboratorio, a Reicy por hacernos ver las cosas de un modo más fácil, a Sebs por enseñarnos que EMSAs son algo más que cuatro banda y no me refiero a cuatro-G, a Cristinha por intentar que trabajemos en un laboratorio ordenado y limpio con su

tacto característico, jejeje, a Cuppi por enseñarnos los secretos de la pasta siciliana, a Arka por recordarme que es currar en la investigación pero no hacía falta que trabajes tantas horas para ello ¡!!, a Anna por ser única y una buena amiga.

Aquellos que también habéis compartido laboratorio conmigo en algún momentos como a Damia, a Cristina, a Jordi, a Arnau, a Laia, a Celia por su alegría contagiosa, tantas fiestas de tesis videos, etc... a Jorge y Mireia por esos partidos del Barça, las cervecitas del Yu Nao, Holanda, Valencia, las fiestas de cambrils, tanto buenos momentos... a los compañeros de cri3, y esas cervecitas en el argentino a Nuria, a Soraya, a Raquel, a Tiago, a Sergio por su Beca entre otras muchas cosas, a Tibi, Cynthia, a Pedro, a todos en general. Esta amaneciendo y no me da tiempo... a todo el departamento de cristalografía en general, David, Garreta, Luca, Maria, Simone, Carpena, Merce, Rosa, Roeland, Esther, etc.

No quiero ni podría no dar las gracias a Elena, sin ti esta tesis no hubiera sido posible, y a Rufo por sus entrenidas charlas mientras paseábamos.

Gracias mi hermana y su apoyo incondicional y su eterna confianza en mi, y a mis padres por todo lo que me han enseñado y todo su cariño.

Me dejo a mucha gente y muchas cosas que contar, pero los que me conocéis sabéis lo importantes que habéis sido para esta tesis.

Gracias a todos.

“Sólo para locos”.  
Extracto de: “El lobo estepario”.  
Por Hermann Hesse.



# **INTRODUCCIÓN**



## 6 INTRODUCCIÓN.

### 6.1 Mitocondria y ADN mitocondrial.

La mitocondria es un orgánulo presente en todas las células eucariotas (Hatefi 1985) cuya función principal es la producción de energía química para la célula, en forma de ATP, pero no la única, además participa otras funciones vitales celulares como la regulación homeostática del calcio (Gunter, Gerstner et al. 2012), la síntesis de lípidos (Schenkel and Bakovic 2014), y la apoptosis programada (Barbour and Turner 2014), entre otras. La mitocondria es un orgánulo dinámico que presenta una gran plasticidad y adapta su morfología según el tipo celular, según si es hepática, muscular, etc. Además, tiene capacidad de moverse en el interior de la célula, adaptando su metabolismo a las necesidades energéticas de ésta, lo que conlleva modificaciones de tamaño, forma y número, lo que es regulado mediante la fusión o división del orgánulo (Lowell and Spiegelman 2000) (Hood 2001).

La mitocondria posee un sistema de doble membrana el cual define cinco compartimentos, la membrana externa y la membrana interna en sí mismas, el espacio existente entre ambas membranas denominado “espacio inter-membrana”, la matriz mitocondrial la cual está limitada por la membrana interna, y finalmente el espacio inter-crestas que se produce por invaginaciones de la membrana interna hacia la matriz y que genera compartimentos aislados del espacio inter-membrana y de la matriz mitocondrial. Gracias a la alta impermeabilidad de las membranas, se almacenan gran cantidad de protones en el espacio inter-membrana, generados durante la fosforilación oxidativa, y posteriormente empleados para la síntesis de ATP a través de la ATP sintasa (ver más abajo). Debido a su impermeabilidad, en las membranas existen una gran cantidad de transportadores como por ejemplo el complejo TIM/TOM (complejos “*transporter inner membrane/transporter outer membrane*”) tanto para proteínas o el canal de transporte de calcio (“*sodium-calcium exchanger*”, NCX) para iones.

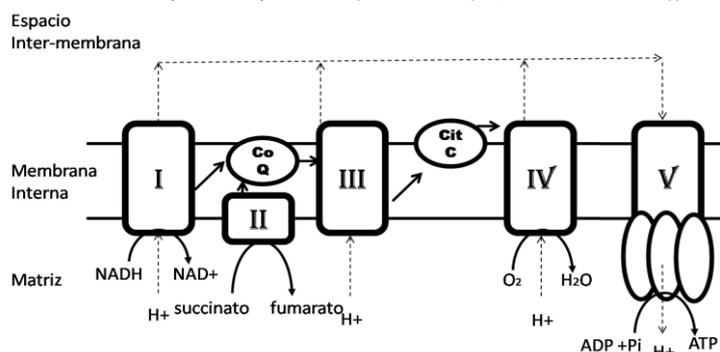
La membrana externa además de contener un gran número de proteínas que conectan el exterior de la mitocondria con la matriz mitocondrial, presenta una estrecha relación con el retículo endoplasmático, permitiendo una correcta homeostasis en el mantenimiento de la concentración celular del calcio y para el transporte de lípidos y de proteínas mitocondriales.

La membrana interna posee mayor superficie que la membrana externa y presenta unas grandes invaginaciones denominadas “crestas”, cuya topología es dinámica como la propia mitocondria, y van a variar en función del estado fisiológico mitocondrial, pudiendo incluso formar un espacio independiente al espacio inter-membrana y a la matriz, con funciones específicas (Jakobs 2006). En la membrana interna se localizan los macrocomplejos que y las

proteínas móviles que los conectan forman en la “cadena de transporte de electrones” (CTE) y que participan en la producción de ATP realizada por la ATP sintasa.

La fosforilación oxidativa es uno de los procesos más importante que se producen en la mitocondria y consiste en la principal vía metabólica de producción de ATP para la célula (figura\_1). En esta reacción participan los cuatro macro-complejos proteicos de la CTE, junto con la ATP sintasa (“Complejo V”), la coenzima Q (CoQ) y el citocromo C (citC), siendo éstos dos últimos los elementos móviles del sistema. Los complejos “Complejo I” (NADH deshidrogenasa) y “Complejo II” (succinato deshidrogenasa) transportan electrones extraídos del NADH y FADH<sub>2</sub> respectivamente, originados durante el ciclo de Krebs en la matriz del orgánulo. Desde los complejos I y II los electrones son transferidos al Complejo III (Q-citocromo C oxidoreductasa) y al Complejo IV (citocromo C oxidasa) mediante CoQ y citC, respectivamente. El complejo IV emplea los electrones que recibe para reducir moléculas de O<sub>2</sub> de la matriz mitocondrial a H<sub>2</sub>O. La transferencia de electrones a lo largo de la CTE genera la liberación de protones por parte de los complejos I, III y IV al espacio inter-membrana. El aumento de protones en el espacio intermembrana crea una diferencia de potencial con la matriz mitocondrial que la ATP sintasa (complejo V) utiliza para transportar los protones hacia esta última, aprovechando este proceso para transformar el gradiente de protones en energía química, mediante la fosforilación de ADP y generando ATP como producto de la reacción (figura\_1).

**Figura\_1. Cadena de transporte de electrones y la fosforilación oxidativa.**  
Se indica de modo esquemático la reacción de los cinco complejos de la cadena de transporte de electrones y la fosforilación oxidativa. Se indica con números romanos el número cada complejo, con flechas continuas el transporte de electrones y en discontinuas el transporte de protones (Basado en (He, Mao et al. 2007)).



### 6.1.1 El genoma mitocondrial humano.

La mitocondria consta de un pequeño genoma, el ADN mitocondrial (mtADN), localizado en la matriz mitocondrial junto con la maquinaria necesaria para la expresión y traducción de proteínas y los ribosomas mitocondriales. Las proteínas implicadas en estos procesos y el mtADN se encuentran en una superestructura denominada nucleóide, que es un

macrocomplejo nucleoprotéico encargado del mantenimiento, replicación, transcripción y empaquetamiento de este genoma (Bogehagen 2012). Se ha propuesto que los nucleoides se encuentran anclados en la membrana interna (Holt, He et al. 2007) y su número por mitocondria es variable, dependiendo del estado fisiológico de la misma (Iborra, Kimura et al. 2004). No se conocen todavía la totalidad de las proteínas que conforman los nucleoides pero sí se han identificado algunas proteínas como la helicasa Twinkle, la polimerasa Pol- $\gamma$ , el factor de transcripción A (TFAM), la proteína de unión a ADN monocatenario mtSSB o la proteína “AAA Domain-containing protein 3” (ATAD3) la cual reconoce la estructura de triple cadena del “D-loop” (He, Mao et al. 2007) (He, Cooper et al. 2012), entre otras proteínas (Garrido, Griparic et al. 2003; Bogehagen 2012).

Según la teoría endosimbiótica de Margulis, hace aproximadamente 2 billones de años que las mitocondrias dejaron de ser  $\alpha$ -proteobacterias independientes y pasaron a vivir en el interior de células eucariotas de modo simbiótico (Thorington and Margulis 1981; Gray 1999). Desde entonces se ha producido una migración de genes de la mitocondria al núcleo, y el resultado es que en la actualidad la longitud y el número de genes que se codifican en el genoma mitocondrial es muy variable entre los distintos organismos eucariotas. El genoma mitocondrial actual es una reminiscencia de su ancestro en  $\alpha$ -proteobacteria.

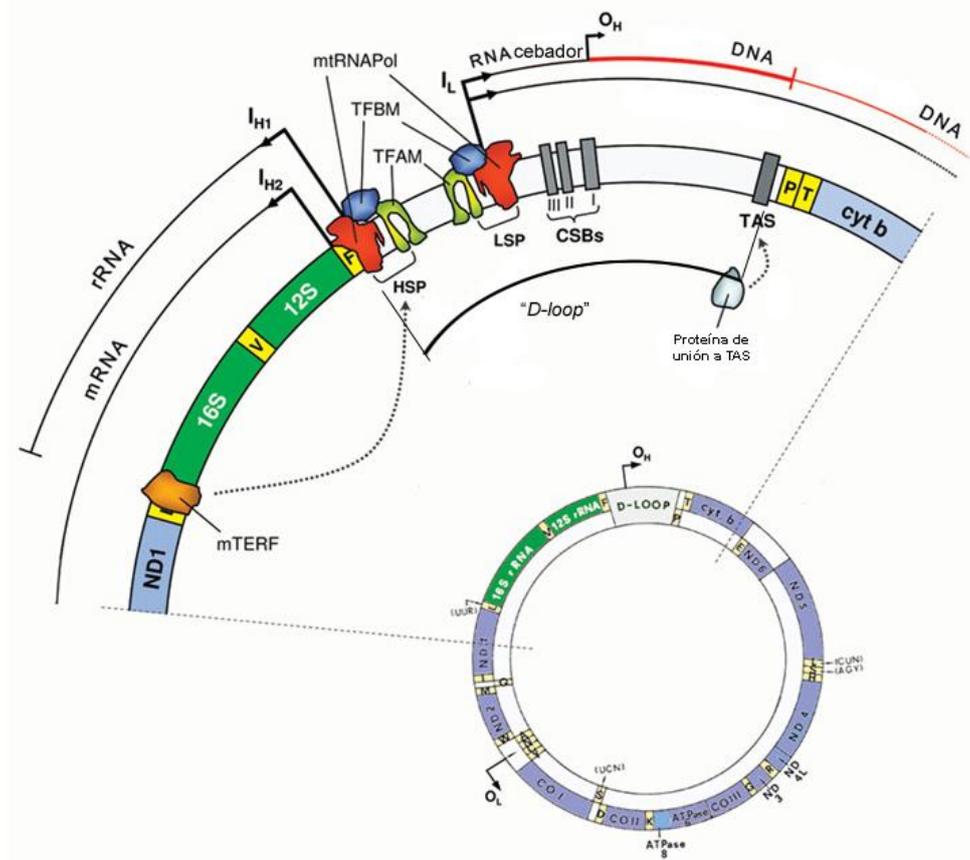
El genoma mitocondrial es circular y posee un tamaño de 16,5 Kb. Ambas hebras tienen una distribución heterogénea en el contenido de guaninas y citosinas, lo que genera un diferente coeficiente de sedimentación al centrifugarlas con un gradiente de CsCl, por lo que una se denominaron “hebra pesada” la que tiene un contenido más alto de guaninas (H, de “*heavy strand*”) y ligera (L, “*light strand*”) a la complementaria, que por tanto contiene un elevado contenido de citosinas (Taanman 1999). La información genética está codificada de un modo muy condensado: hay superposición de genes entre ambas cadenas y no hay intrones (Anderson, Bankier et al. 1981). En el mtADN hay codificados 13 polipéptidos de las aproximadamente 90 proteínas que componen los macrocomplejos de la cadena de transporte de electrones. Además, codifica dos ARN ribosómicos (12S y 16S) y 22 ARN de transferencia (tARN), todo ello suficiente para la traducción de las proteínas codificadas (Ojala, Merkel et al. 1980) (figura\_2).

No todo el ADN mitocondrial es codificante, existe una región no codificante (“*Non Coding Region*”, NCR) donde se encuentran los elementos reguladores del mtADN. En el NCR se alojan el origen de la transcripción y de la replicación de la hebra pesada ( $O_H$ ), el promotor de la transcripción de la hebra pesada (“*heavy strand promoter*” HSP1,  $I_{H1}$ ), el de la hebra ligera (“*light strand promoter*” LSP,  $I_L$ ), además de tres secuencias altamente conservadas (“*Conserved Sequence Block*”) CSB I, II y III, y las secuencias Site-X y Site-Y reconocidas por el

factor de transcripción A (TFAM) (Fisher, Topper et al. 1987) (figura\_2). La NCR también posee una característica peculiar denominada “D-loop”, la cual es una superestructura compuesta por tres hebras: una hebra denominada ADN 7S de unos 650-700 pares de bases que se presenta hibridada con la hebra ligera desplazando la hebra pesada (Fish, Raule et al. 2004).

**Figura\_2. Genoma mitocondrial y el “D-loop”.**

En la ampliación se muestra la región de control en la cual se indica el origen de la replicación de la hebra pesada ( $O_H$ ), los orígenes de la transcripción de la hebra pesada  $I_{H1}$   $I_{H2}$  y de la hebra ligera  $I_L$ ; además las “Conserved Sequence Blocks” CSB I, II y III, y la unión de las proteínas mtRNAPol, TFAM y el TFB2 (TFBM). Abajo el esquema del genoma mitocondrial humano, se indica en verde los rARN 16S y 12S, con cuadrados amarillos los tRNA, color azul los genes codificantes de las subunidades del NADH deshidrogenasa, citocromo B, ATPasa y las subunidades citocromo oxidasa. Imagen adaptada de (Musicco, Roberti et al. 2000).



Actualmente este esquema de organización de la zona control está en cuestión, ya que además se había descrito un promotor adicional para la cadena pesada HSP2 ( $I_{H2}$ ) el cual sólo es considerado por algunos autores como Attardi o Montoya (Attardi and Montoya 1983; Montoya, Gaines et al. 1983; Fernandez-Silva, Enriquez et al. 2003). Según estos autores, este promotor da lugar a un transcripto que comprende a los dos genes ribosomales (12S y 16S) y dos tRNA (Val y Phe) y cuya frecuencia de transcripción es mucho mayor que la originada desde HSP1, cuyo transcripto implica casi la totalidad del mtADN (Attardi and Montoya 1983; Montoya, Gaines et al. 1983; Fernandez-Silva, Enriquez et al. 2003). Por el contrario otros autores sólo reconocen un inicio de la transcripción en la hebra H (HSP) que corresponde con HSP1, y no consideran HSP2 como inicio de la transcripción porque sólo han observado un

transcripto que finaliza justo antes del gen tRNA<sup>PHE</sup> (Falkenberg, Larsson et al. 2007). Según estos autores la longitud del transcripto originado en HSP1 tendría distinta longitud: cuando se une la proteína MTERF llegaría hasta el final de rARN 16S, pero cuando no se une MTERF el transcripto sería de todo el mtADN (Falkenberg, Larsson et al. 2007).

## 6.2 Replicación.

La replicación es la reacción por la cual se realiza la duplicación semiconservativa del ADN o ARN de un genoma. Este proceso es vital para la supervivencia de la célula, sin el cual no se puede realizar la división celular. En el caso de la replicación mitocondrial es necesario no sólo para la división, sino para la conservación del genoma. En general, la replicación es una reacción muy compleja que requiere la participación y coordinación de multitud de macrocomplejos proteicos que participan directamente.

### 6.2.1 Replicación del genoma mitocondrial.

La replicación del genoma mitocondrial es un fenómeno celular que ocurre de modo independiente a la división celular (Clayton, White et al. 1997), y tampoco está relacionado con la división de la mitocondria, tan sólo está determinado por el requerimiento energético de la célula. (Shadel and Clayton 1997). Aunque no se conoce con profundidad el mecanismo subyacente de la replicación mitocondrial, sí es evidente que es un proceso asimétrico de polimerización entre ambas hebras y por tanto es un proceso diferente al que sucede en el núcleo de eucariotas y en la replicación procarionota. Actualmente, la comunidad científica se encuentra en debate entre dos modelos diferentes que intentan explicar el proceso de la replicación mitocondrial: el “modelo de desplazamiento de hebra” (Clayton 2003) y el “modelo de incorporación de ribonucleótidos a través de la hebra retardada” (RITOLS), denominado anteriormente “modelo de hebra acoplada” (Holt and Jacobs 2003);.

- **Modelo de desplazamiento de hebra:**

En el modelo de “desplazamiento de hebra” la replicación se origina en dos puntos concertados, uno para cada hebra (ver figura\_3). En la hebra conductora, que corresponde a la cadena H, el origen de replicación se encuentra en NCR y se denomina O<sub>H</sub>. Para la cadena L, el origen de replicación O<sub>L</sub> se encuentra dos tercios “aguas abajo” del origen O<sub>H</sub>, en la hebra complementaria. Así, la replicación se origina desde el O<sub>H</sub> para la hebra conductora o cadena H, y puede dar dos resultados, uno es que se polimerice sólo el “D-loop”, tras el cual puede que ocurra una pausa o parada, el otro es que continúe para copiar toda la hebra pesada. Cuando prosigue “aguas abajo” del “D-loop”, hay un momento en el que el O<sub>L</sub> queda expuesto, y en ese punto da comienzo la replicación de la hebra retardada (cadena L) en sentido opuesto (Clayton 2003). El hecho que la cadena L se

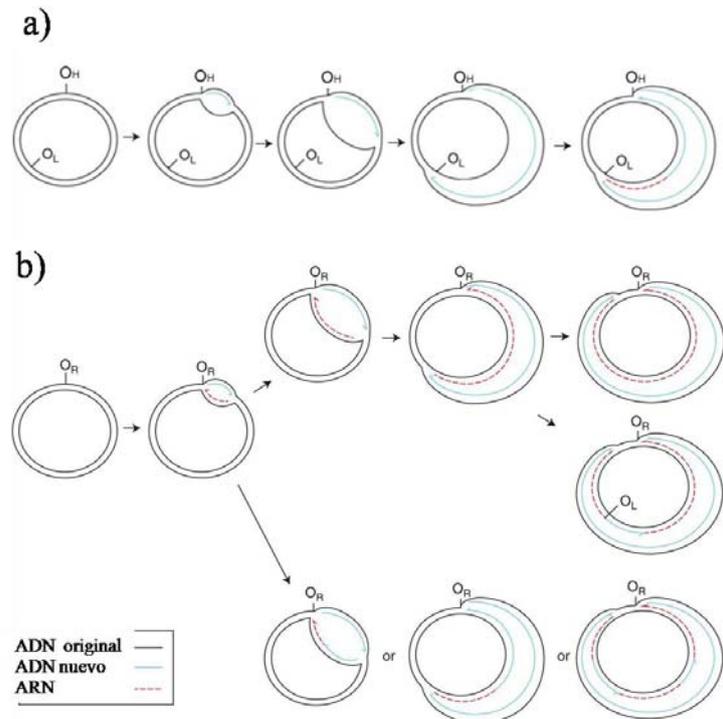
replica una vez iniciada la replicación de la cadena H da lugar a que este modelo se denomine hebra retardada. Durante la replicación no se interrumpe la polimerización de ninguna de las dos hebras nacientes, como consecuencia tanto la hebra complementaria a la conductora como la complementaria a la retardada son continuas. Desde el comienzo de la replicación, la nueva cadena complementaria se hibrida a la cadena H, y forma un ADN de doble cadena estable; sin embargo, la cadena L, retardada, se mantiene en forma monocatenaria hasta que la maquinaria de replicación, que se desplaza por la cadena H, deja expuesto el  $O_L$ . Para estabilizar a la hebra retardada se le unen las proteínas mtSSB, hasta que también se replica. Los cebadores para la replicación son producidos por la ARN polimerasa de mitocondria (POLRNA) (Wanrooij, Fuste et al. 2008; Fuste, Wanrooij et al. 2010).

- **Modelo de hebra acoplada, RITOLS:**

El modelo original de “hebra acoplada” está elaborado a partir de los datos que se han obtenido de geles neutros nativos de agarosa 2D (2DAGE) a partir de extractos purificados de mitocondria. Según los resultados mediante 2DAGE, la polimerización de ambas hebras es simultánea, coordinada, y además se producen intermediarios de la replicación de doble cadena que en un principio se pensaba eran de ADN (Holt 2000). Más tarde se constató que los intermediarios no eran doble cadena de ADN sino híbridos ARN-ADN, lo que dio lugar a denominar a este modelo RITOLS (“*RNA incorporation throughout the lagging strand*”). Además se asignaron dos orígenes de replicación en la cadena pesada, situados en el “*D-loop*” (Yasukawa, Yang et al. 2005),  $O_H$  y  $O_b$ . Mientras se replica la cadena H, la cadena L (retardada) se estabiliza con fragmentos discontinuos de ARN, formando híbridos ARN/ADN. Los híbridos ARN/ADN que se forman inicialmente en la hebra retardada son sustituidos posteriormente por ADN de doble cadena, y entonces es cuando la cadena nueva es completamente continua (Reyes, Kazak et al. 2013). El origen para la sustitución de los híbridos ARN/ADN se produce en múltiples puntos, aunque se produce con más frecuencia en  $O_L$ . La replicación de cada hebra es unidireccional y de modo coordinado, se realiza en ambas hebras simultáneamente, pero la hebra L no es continua durante la polimerización con híbridos de ARN, por lo que se produce el retraso característico de la hebra retardada con respecto a la hebra conductora. Los híbridos ARN/ADN de la hebra ligera no necesita la unión con las proteínas mtSSB para evitar la recombinación de la hebra retardada consigo misma (Holt and Reyes 2012) tal y como se indica en la Figura\_3.

**Figura\_3. Replicación mitocondrial.**

**a)** Modelo de desplazamiento de hebra, se inicia la replicación de la cadena pesada H en el  $O_H$  (en azul) y la cadena L, la cual no comienza a replicarse hasta que H se ha replicado hasta  $O_L$ , a partir de un cebador de ARN (rojo discontinuo). En negro las hebras parentales. **(b)** Modelo de RITOLS, donde comienza la replicación desde  $O_R$  (que incluye a  $O_H$  y  $O_B$  en azul) de la cadena H mientras la cadena L se estabiliza con ARN (rojo); éste se sustituye por ADN (azul) a partir de múltiples puntos cuando se alcanzan dos tercios de longitud del genoma desde  $O_R$ , aunque el más frecuente es  $O_L$  indicado en la figura. En negro se representan las hebras parentales. Basado en (Holt and Reyes 2012).

**6.3 Helicasas.**

En la reacción de replicación una de las proteínas esenciales son las helicasas catalizando el desenrollamiento de las dos cadenas complementarias de nucleótidos, ADN o ARN, cuya reacción esta mediada por la energía de la hidrólisis de trinucleótidos (dNTPs) (Lohman and Bjornson 1996). Aunque su función más importante es su participación en la replicación, algunas helicasas también participan en otros procesos biológicos como la recombinación y reparación del ADN, en coordinación con otras macromoléculas.

**6.3.1 Clasificación de las helicasas.**

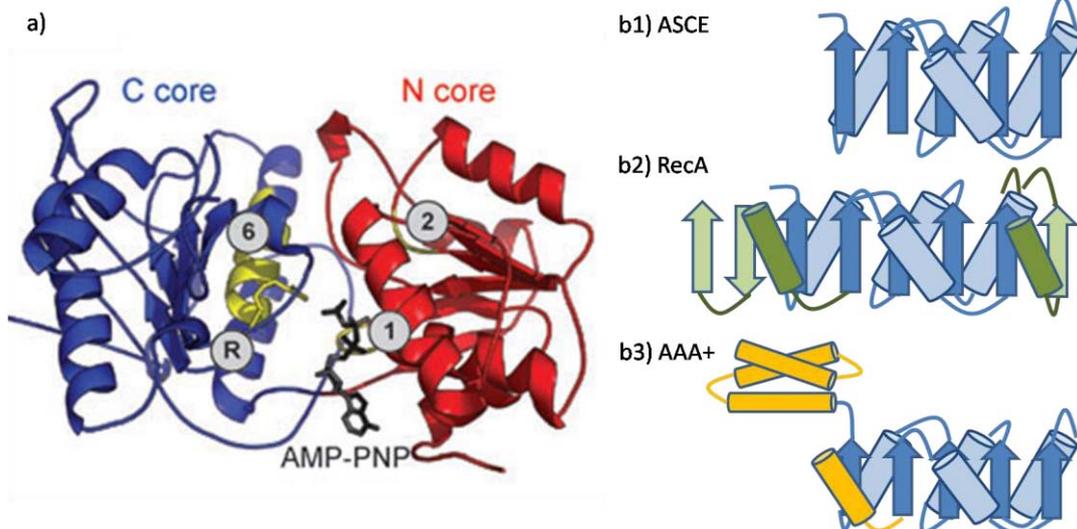
Las helicasas poseen una gran variabilidad a nivel de secuencia y dominios, pero aun así existe un mismo patrón estructural conservado entre todas ellas denominado "Hebra Adicional Conservada E" ("*additional strand conserved E*", ASCE), el cual es a su vez un subgrupo de la familia de proteínas "NTPasas P-loop" (Erzberger, Mott et al. 2006). A nivel topológico el dominio ASCE consiste en un plegamiento de la cadena polipeptídica en forma de sándwich denominado "plegamiento  $\alpha\beta$ ", que se compone por un conjunto de cinco hebras  $\beta$  centrales flanqueadas en cada lado de la hoja  $\beta$  por dos grupos de hélices  $\alpha$  (Koonin 1993). Entre la

hebra  $\beta$ 1, que posee la secuencia altamente conservada "A/GxxxxGKT/S" (donde x es cualquier aminoácido), y la siguiente hélice  $\alpha$ , se encuentra el motivo Walker A. Éste es el responsable de la interacción con el fosfato gamma de los NTPs. Por otra parte la hebra  $\beta$ 3 posee el motivo "zzzzDE" (siendo z cualquier aminoácido hidrofóbico), en el cual se coordina un ion de magnesio y se define como motivo Walker B. Este ion interviene en la unión con los NTPs. La unidad catalítica la componen dos motivos ASCE y entre la superficie de interacción entre ellos se definen el bolsillo en el cual se produce la unión de los nucleótidos (Ye, Osborne et al. 2004) (figura\_4a).

Conservando esta topología común del grupo ASCE, que define el lugar de unión a los NTPs, se construyen las helicasas con diferentes dominios auxiliares, dotando a cada una de ellas de unas propiedades funcionales particulares y diferentes, y que varían también según su distinta interacción con otras proteínas. En función de estos los dominios auxiliares las helicasas se clasifican en dos subgrupos RecA y AAA+. Las helicasas tipo RecA tienen un plegamiento en el que unas hebras  $\beta$  auxiliares y dos hélices  $\alpha$  rodean el motivo principal ASCE. En cambio, las helicasas tipo AAA+ tienen un dominio extra de múltiples hélices  $\alpha$  totalmente independiente del motivo ASCE (figura\_4b1, b2 y b3).

**Figura\_4. Estructura de dos motivos tipo ASCE, y topología de los dominios ASCE, RecA & AAA+.**

En **a)** Se representa cada motivo ASCE con distintos colores azul y rojo, los motivos Walker A y B se indican con los números 1 y 2 respectivamente, la arginina y la hélice  $\alpha$  que interaccionan con los fosfatos de los NTPs se indican con una R el número 6 respectivamente (en color amarillo). En negro se detalla un análogo NTP ubicado en el bolsillo de unión de los nucleótidos. En **b)** se presentan la topología de las helicasas: con flechas se simboliza las hebras  $\beta$  y con cilindros las hélices  $\alpha$ . Se indica en azul el plegamiento tipo ASCE (**b1**); los dominios extra, que se diferencian en los subtipos RecA en **b2** (indicado en verde), y en **b3**) AAA+ (dominio extra indicado en naranja). Figura basada en (Erzberger and Berger 2006).



Como consecuencia de la gran variabilidad de las helicasas tanto desde un punto de vista estructural como bioquímico existen multitud de clasificaciones en función de la característica que se analice, por ejemplo en función de la dirección del desenrollamiento 5'→3' o 3'→5'; o del sustrato que catalizan, sea tipo ARN, ADN o ambos; o por su similitud en la secuencia

primaria; por su estructura cuaternaria u oligomerización; etc. La clasificación más extendida es teniendo en cuenta la secuencia primaria, y se dividen en seis superfamilias: SF1, SF2, SF3, SF4, SF5 y SF6.

Sólo las proteínas que pertenecen a las SF1 y SF2 no poseen una forma hexamérica. Estas dos familias poseen un gran parecido entre ellas en los motivos I, II y IV que definen el bolsillo de unión al ATP, pero el grado de conservación entre motivos es variable, lo que provoca las diferencias entre ambas superfamilias (Leipe, Wolf et al. 2002).

- La SF1 se compone de dos tipos o subgrupos helicasa: SF1A y SF1B. El subgrupo SF1A tiene un plegamiento formado por dos subdominios. Cada subdominio tiene un plegamiento tipo RecA, y se encuentran separados por una hendidura en la cual produce la unión del ATP (Korolev, Hsieh et al. 1997). El subgrupo SF1B también conservan la pareja de dominios RecA divididos en subdominios separados por una hendidura donde alojan el ATP como las SF1A; sin embargo tienen un dominio N-terminal extra que interacciona con la subunidad RecC, siendo ésta una subunidad del macrocomplejo RecBCD. Todas las helicasas SF1 han cristalizado en estado de monómero, y aunque como monómero tienen actividad helicasa, ésta es mayor si se produce cooperación entre múltiples monómeros (Ha, Rasnik et al. 2002; Brendza, Cheng et al. 2005).
- La SF2 está compuesta por helicasas con mecanismos funcionales muy diversos: helicasas ADN/ARN, bidireccionales, etc (Linder and Lasko 2006), pero estructuralmente conservan el plegamiento tipo SF1, con los dos dominios RecA divididos por el bolsillo de unión al ATP.

El resto de familias de helicasas (SF3, SF4, SF5 y SF6) presentan una estructura cuaternaria de hexámero, aunque en muchas de ellas es necesaria la presencia de magnesio, NTPs o ADN/ARN para estabilizar la estructura cuaternaria (Hingorani and Patel 1996). Gracias a la microscopía electrónica se ha podido observar cómo se conserva esta estructura cuaternaria en helicasas de procariotas, eucariotas y virus, aunque haya diferencias en el tamaño del anillo y el diámetro del canal central como consecuencias de movimientos entre monómeros (Barcena, Martin et al. 1998); (Yu, Hingorani et al. 1996).

- La SF3 forman hexámeros, sencillos o dobles, y desenrollan el ADN siguiendo la dirección 5'→3'. Poseen cuatro dominios conservados A, B, B' y C; los dominios A y B corresponde al motivo "Walker A y B" presentes también en las familias SF4, SF5 y SF6. El motivo B' se encuentra orientado al canal central e interacciona con el esqueleto de fosfato del ADN monocatenario (Singleton, Dillingham et al. 2007), mientras que el domino C es específico para las helicasas SF3 (Hickman, Perez et al. 2005). Todas ellas

presentan el motivo AAA+, en el cual se encuentra el lugar de unión al nucleótido, que le caracteriza un patrón estructural específico de hebras  $\beta$  en la superficie de interacción entre dos subunidades. En el hexámero existen seis bolsillos de unión a los nucleótidos pero no todos presentan la misma conformación (Fouts, Yu et al. 1999), debido a movimientos entre monómeros. El ADN de cadena sencilla es estabilizado por “*loops*” de la proteína que interaccionan con el esqueleto de fosfatos del ADN en el interior del canal del hexámero.

- La SF4 son helicasas que presentan una estructura cuaternaria de anillo y una actividad 5'→3'; algunas además presentan actividad cebadora. Los ejemplos más estudiados de esta familia son DnaB de *E.coli* y gp4 del bacteriófago T7 (gp4). Estructuralmente el denominador común de estas enzimas son los motivos H1, H1a, H2, H3 y H4. Los motivos H1 y H2 definen el motivo Walker A y B (Ilyina and Koonin 1992), por lo que DnaB y comparten similitud en la zona C-terminal de gp4. DnaB tiene, además, un dominio N-terminal de reconocimiento a DnaC, responsable de formar la estructura de hexámero. Por el contrario, en gp4 cada subunidad está compuesta por dos dominios N- y C-terminal funcionalmente independientes. El dominio N-terminal que posee una actividad primasa o cebadora, estructuralmente se divide en dos subdominios: dominio de unión al Zinc (ZBD) y dominio de ARN polimerasa (RPD). (Toth, Li et al. 2003);(Kato, Ito et al. 2003; Crampton, Ohi et al. 2006). El C-terminal posee actividad helicasa, es el responsable de la oligomerización, y además se encuentra el bolsillo de unión a dTTP ubicado entre dos subunidades, cuya hidrólisis permite la translocación del ADN. La enzima gp4, mediante este dominio, no sólo forma una estructura modular de anillo de seis sino también de siete subunidades en presencia de ADN y nucleótidos trifosfatos, como se demostró con la resolución de su estructura cristalográfica (Patel and Hingorani 1995; Wigley 2000; Toth, Li et al. 2003). Las helicasas gp4 y Twinkle (helicasa mitocondrial humana) presentan un elevado grado de similitud de secuencia entre ellas (Spelbrink, Li et al. 2001) tanto para el dominio N- como el C-terminal, y Twinkle es también miembro de SF4. Por este motivo gp4 es un buen sistema de referencia para Twinkle, el objeto de estudio de esta tesis.
- En la SF5 se han encontrado ejemplos de proteínas que presentan estructura de hexámero en dos conformaciones diferentes, formando un anillo o bien formando una estructura en espiral, como en el caso de Rho (Skordalakes and Berger 2003). La existencia de estas dos conformaciones se interpreta como que la estructura en espiral es un intermediario durante la carga al ADN, y la estructura de anillo cuando se realiza la translocación del ADN.

- La SF6 la componen helicasas que a pesar de poseer el dominio AAA+ no pertenecen a SF3. Las principales representantes de esta familia son las proteínas de mantenimiento del mini cromosoma (proteínas MCM) de eucariota y la procariota RuvB, cuyo sustrato son estructuras de ADN tipo “*fourway junctions*”. En el caso de las MCM, las cuales pertenecen a las *Arqueas*, son las mejor caracterizadas bioquímicamente y aunque no se conoce la estructura tridimensional cristalográfica sí hay bastante información estructural gracias a la microscopía electrónica, a cuya densidad electrónica se ha podido ajustar un modelo cristalográfico basado en una proteína AAA+. Se ha visto que son hexámeros con un canal suficientemente grande para acomodar un ADN de doble cadena (Fletcher, Bishop et al. 2003).

### 6.3.2 Mecanismo de hidrólisis de NTP de helicasas hexaméricas.

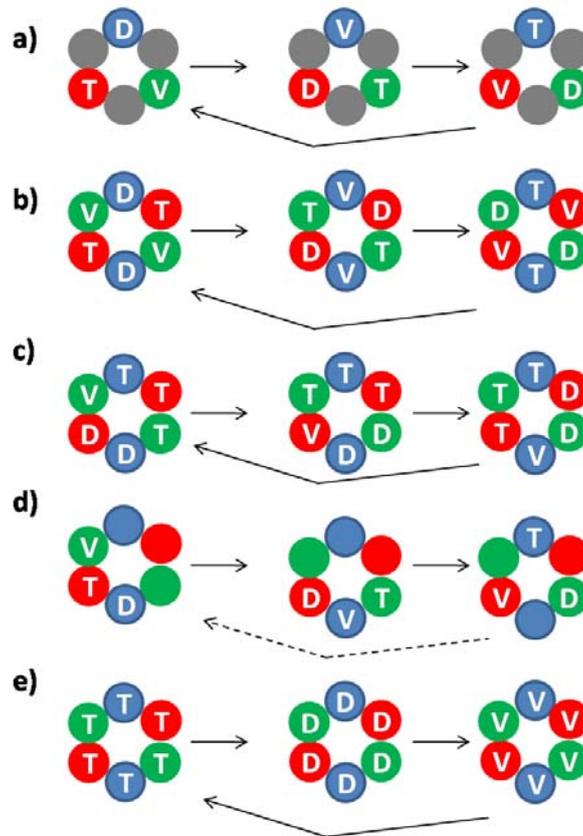
Para las proteínas hexaméricas se proponen cinco modelos diferentes de hidrólisis, los cuales están realizados por diferentes familias de proteínas.

- **Modelo de 3 sitios de unión secuenciales.** En el hexámero, hay seis sitios totales de unión a los trinucleótidos de los cuales participan en la reacción tres simultáneamente, cada uno en una conformación diferente que alternan de modo secuencial. Las tres conformaciones son: a) conformación de carga, donde el bolsillo acoge al nucleótido trifosfato (conformación T); b) unión a nucleótido difosfato como consecuencia de la hidrólisis de un fosfato del trifosfato (conformación D); c) finalmente vacía, que es cuando expulsa el nucleótido (conformación V) (figura\_5a).
- **Modelo de unión secuencial de todos los sitios.** Este modelo alterna también las tres conformaciones descritas en el modelo anterior, pero en este caso participan todas las subunidades de modo simultáneo, siguiendo el patrón T-D-V-T-D-V, no sólo tres como el modelo anterior (figura\_5b).
- **Modelo de unión secuencial de todos los sitios en bloque.** Es una variante del modelo anterior, en la cual hay un bloque de tres monómeros seguidos en la conformación T, dos en la D y uno en la vacía, cambiando en cada ciclo la conformación de cada monómero sin interrumpir el bloque (figura\_5c).
- **Modelo Estocástico.** En los modelos anteriores se producían todas las transiciones de las tres conformaciones de un modo concertado y secuencial. No obstante, se han encontrado helicasas donde los cambios entre las tres conformaciones es aleatoria entre las seis subunidades. Este modelo se ha encontrado para helicasas de tipo AAA+ (figura\_5d).

- Modelo concertado.** El último modelo conocido consiste en que todos los monómeros se encuentran al mismo tiempo en la misma conformación y cambian simultáneamente de conformación T, D o V; por tanto los monómeros, y con ello la helicasa entera, se pueden encontrar en un estado conformacional T, D o V (figura\_5e).

**Figura\_5. Mecanismo de las helicasas hexaméricas.**

Se muestran los distintos mecanismos de las helicasas hexaméricas **a)** modelo de unión a tres sitios secuenciales; **b)** secuenciales de todos los sitios; **c)** secuenciales de todos los sitios en bloque; **d)** modelos estocástico; **e)** modelo concertado. La conformación de cada subunidad con NTP, NDP o vacía se indica con T, D y V respectivamente.



### 6.3.3 Movimiento de traslocación asociado a la actividad NTPasa de las helicasas hexaméricas.

Las helicasas con una estructura hexamérica poseen en el dominio RecA un motivo conservado en forma de “loop” orientado hacia la cara interna del canal central del anillo del hexámero. Este “loop” tiene por función unirse al ADN, más concretamente, interaccionar con el esqueleto de fosfatos (Washington, Rosenberg et al. 1996). La hidrólisis de NTPs produce cambios conformacionales en las subunidades del anillo que provocan desplazamientos de este “loop” cuando está interaccionando con el ADN. En consecuencia, el ADN es desplazado

longitudinalmente a lo largo del canal del anillo por el movimiento de todos los “loops” del hexámero (Singleton, Sawaya et al. 2000; Enemark and Joshua-Tor 2006).

A su vez, la transición entre las conformaciones T, D y V inducidos por la hidrólisis de los NTPs, mencionada anteriormente desplaza el ADN entre un monómero y el adyacente. Por tanto, la hidrólisis de dNTPs genera, a través de pequeños cambios conformacionales tanto de los “loops” como rotacionales entre subunidades, la translocación del ADN. Además, un ciclo completo T, D, V conlleva la hidrólisis de seis NTPs con ella la translocación de seis nucleótidos de ADN de cadena sencilla y la liberación de seis NDPs (Enemark and Joshua-Tor 2008).

#### **6.3.4 Replisoma de bacteriófago T7 y actividad de gp4.**

La maquinaria mínima requerida para reproducir la reacción de replicación mitocondrial *in vitro* es conocida (Korhonen, Pande et al. 2008), sin embargo el mecanismo funcional no está claro. Para entender como interaccionan estas proteínas se emplea como modelo un sistema muy similar muy bien caracterizado, como es el replisoma de bacteriófago T7. El sistema de bacteriófago T7 se ha estudiado en la última década profundamente y además las proteínas implicadas presentan un gran parecido a las mitocondriales: la helicasa gp4 es similar a Twinkle (Spelbrink, Li et al. 2001), la polimerasa gp5 del bacteriófago se parece a Pol- $\gamma$  (Lecrenier, Van Der Bruggen et al. 1997) y la T7 ARN polimerasa es similar a la mitocondrial (“RNA polymerase”- RNAPOL) (Tiranti, Savoia et al. 1997). Por tanto *a priori* es un buen sistema de referencia para entender y elucidar la función que desempeñan las proteínas mitocondriales en el replisoma.

Las proteínas esenciales del replisoma en bacteriófago T7 son cuatro: la polimerasa gp5, la helicasa gp4, las proteínas de unión a ADN de cadena sencilla gp2.5 (SSB) y la tioredoxina (TRX). TRX y gp5 forman un complejo estable gp5/TRX a través de la interacción del lazo de esta última denominado “dominio de unión a Tioredoxina” (TBD) (Doublié and Ellenberger 1998). Funcionalmente se ha comprobado la importancia de la interacción gp5/TRX, que genera un aumento de 20-80 veces la tasa de incorporación de dNTPs en la polimerización de la hebra complementaria (Tabor, Huber et al. 1987). La proteína gp2.5, a través de su dominio C-terminal, también interaccionan con el TBD del heterocomplejo gp5/TRX (Hamdan, Marintcheva et al. 2005) y con una región que componen los residuos K587, K589, R590 y R591 que forman un parche cargado positivamente en la zona frontal de gp5 (Zhang, Lee et al. 2011). Esta interacción gp2.5/gp5/TRX también es esencial para la replicación (Park, Debyser et al. 1998) y además permite la coordinación entre la polimerización de la hebra conductora y la hebra retardada de ADN (Lee, Chastain et al. 1998).

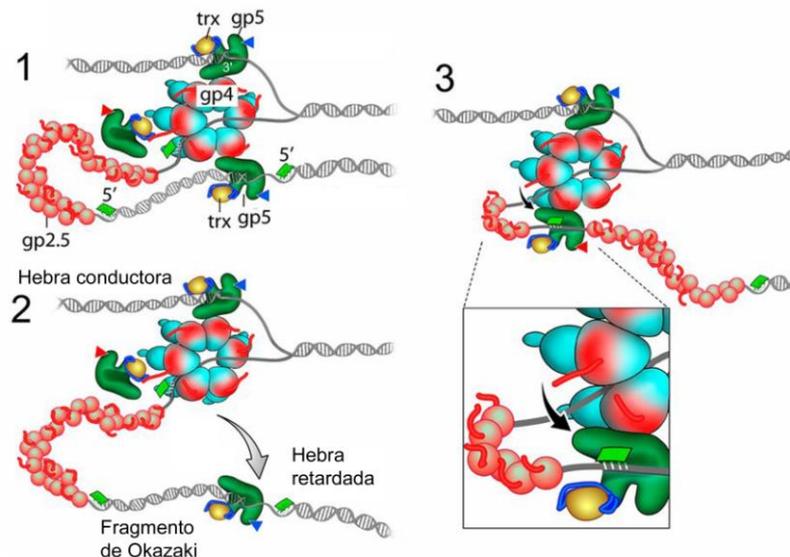
La helicasa gp4 del fago T7 interacciona mediante una cola C-terminal con la polimerasa gp5 en la misma región de ésta que interacciona con gp2.5 y además con el TBD de TRX (Lee and Richardson 2011) (Zhang, Lee et al. 2011), aunque sólo se produce esta interacción cuando gp4 está cargada en la hebra retardada. La helicasa gp4 de T7 presenta un equilibrio entre heptámero y hexámero en solución, sin embargo al interactuar con el ADN sólo se encuentra en forma hexamérica, perdiendo una subunidad (Ahnert, Picha et al. 2000).

La forma hexamérica cargada al ADN está activa, y desenrolla el ADN empleando la energía de hidrólisis de dTTP (Akabayov, Akabayov et al. 2010). Durante la translocación, el anillo hexamérico, formado por los dominios C-terminales, queda posicionado frente las hebras de ADN que se separan, mientras los dominios N-terminales quedan en la cara opuesta al avance. La hebra retardada pasa a través del canal central del hexámero y la hebra conductora pasa por el exterior del hexámero. La helicasa gp4 también tiene una función primasa, realizada por los dos subdominios que se encuentran en el extremo N-terminal: ZBD y RPD (Lee and Richardson 2011). Para realizar la función cebadora durante la replicación se produce una interacción tipo *trans* entre el dominio ZBD de una subunidad con el RPD de la subunidad vecina. El mecanismo consiste en el reconocimiento inicial por parte del ZBD de una secuencia específica de ADN monocatenario, y gracias a un “*linker*” que permite que el movimiento de ZBD puede desplazar el ADN hasta interactuar con el centro activo del RPD de otro monómero adyacente, el cual realiza la reacción de generar el ARN cebador mientras el ZBD se estabiliza la posición del ADN (Lee and Richardson 2002). Este mecanismo de interacción *trans* también se ha observado en otras primasas como DnaG la cual dimeriza en la forma activa (Bailey, Eliason et al. 2007).

Las actividades de la helicasa gp4 y la polimerasa en complejo con tioredoxina se retroalimentan de modo recíproco durante la replicación de la hebra conductora, como consecuencia de su interacción mutua (figura\_6-1). La replicación de la hebra retardada resulta más compleja, el dominio N-terminal con actividad primasa de gp4 reconoce una secuencia específica a partir de la cual se generan los cebadores o fragmentos de Okazaki (figura\_6-2) (Toth, Li et al. 2003) (Lee, Zhu et al. 2010). Estudios de FRET sugieren que durante la translocación del ADN la interacción del dominio N-terminal primasa de gp4 con el cebador y con el ADN de cadena sencilla se mantiene durante el avance de la helicasa dando lugar a un “lazo de cebamiento”, hasta que la interacción de gp5 provoca que el dominio primasa de gp4 libere al ssADN (figura\_6-3) (Pandey, Syed et al. 2009) tal y como se describe en la Figura\_6.

**Figura\_6. Modelo del replisoma de bacteriófago T7.**

La helicasa gp4 desenrolla el ADN bicatenario en dos hebras monocatenarias y rodea a la hebra retardada. El heterocomplejo polimerasa/tioredoxina (gp5/TRX) se une a la helicasa y cataliza la polimerización de nucleótidos sobre la hebra conductora. La cadena sencilla de ADN retardada extruida por la translocación de la helicasa forma un lazo con el dominio primasa de gp4 catalizando la síntesis de tetranucleótidos (verde) que sirven como cebador para gp5/TRX, que también está unido a la helicasa en la hebra retardada. Las proteínas SSB (gp2.5) protegen al ADN monocatenario de estar expuesto (Imagen adaptada de (Geertsema, Kulczyk et al. 2014).

**6.4 Helicasa mitocondrial humana, Twinkle.**

A comienzos de este siglo se descubrió en el genoma nuclear el gen codificante de la helicasa mitocondrial humana, Twinkle. Por microscopía confocal se observó la co-localización de la proteína con los nucleoides mitocondriales (Spelbrink, Li et al. 2001). Esta proteína es un componente esencial en el mantenimiento del genoma mitocondrial (mtADN), la inhibición de la expresión de Twinkle produce un descenso importante de la cantidad de ADN mitocondrial y, por el contrario, la sobreexpresión genera un aumento del ADN mitocondrial (Tynismaa, Sembongi et al. 2004). Este fenómeno también se observó bloqueando la expresión de Twinkle mediante iARN en ratones, en los que se produjo un descenso en la replicación y cantidad de mtADN impidiendo el desarrollo embrionario (Milenkovic, Matic et al. 2013).

Análisis de secuencia primaria demostraron que Twinkle presenta un elevado grado de homología con la helicasa gp4 del bacteriófago T7, y se predijo una composición estructural y funcional de diferentes dominios. La predicción indicó un dominio N-terminal dividido en dos subdominios “Zinc Binding Domain” (ZBD) y el “RNA polymerase Domain” (RPD), y un dominio C-terminal (Spelbrink, Li et al. 2001); (Shutt and Gray 2006). Posteriormente, se comprobó experimentalmente con estudios de proteólisis limitada con tripsina la estructura modular de Twinkle (Ziebarth, Farr et al. 2007), tal y como habían predicho anteriormente los autores citados.

Un análisis más detallado del dominio N-terminal mediante la comparación de múltiples secuencias de Twinkle de distintos metazoos evidenció una gran variabilidad en este dominio. Se destaca el hecho de que no todos los residuos implicados en la actividad primasa del dominio N-terminal de gp4 y de DnaG se encuentran conservados en Twinkle. En concreto, los ZBD de las proteínas gp4 y DnaG poseen cuatro cisteínas que están conservadas en primasas de distintos reinos naturales, también en los ortólogos de Twinkle de todos los metazoos a excepción, curiosamente, de *Homo sapiens*. (Spelbrink, Li et al. 2001). Estudios posteriores *in vitro* realizados por Farge y coautores mostraron por cromatografía por exclusión de tamaño que el dominio N-terminal, definido entre los aminoácidos 43-345, presentaba forma monomérica en solución (Farge, Holmlund et al. 2008).

Al contrario que el dominio N-terminal, el alineamiento de la secuencia de la región C-terminal entre múltiples metazoos evidencia un elevado grado de conservación de este dominio (Shutt and Gray 2006), además se identificaron los motivos Walker A y B, por lo que se predijo que este dominio tendría actividad helicasa (Spelbrink, Li et al. 2001). Tanto por estudios de proteólisis limitada, como estudios *in vitro* de la producción recombinante del dominio C-terminal, comprendida entre los aminoácidos 372 al 684, se demostró que este dominio presenta un estado oligomérico de hexámero, siendo por tanto el responsable de la oligomerización de la proteína, y además se comprobó la actividad ATPasa de este dominio (Farge, Holmlund et al. 2008).

Twinkle, además de tener una estructura modular, también presenta una estructura cuaternaria compuesta por hexámeros y heptámeros, tal y como demostraron Zierbarth y coautores, con la expresión de Twinkle recombinante madura (aa 43-684, sin la secuencia guía de mitocondria) en células de insecto y su posterior análisis mediante las técnicas de cromatografía de exclusión molecular y mediante “*crosslinking*” con glutaraldehído (analizado por SDS-PAGE), siendo este resultado consistente con otros miembros de la misma familia de helicasas, entre ellos gp4 y DnaB, y tal y como se ha mencionado anteriormente con el dominio C-terminal (Ziebarth, Farr et al. 2007). Estudios posteriores de Farge y coautores, mostraron que, además, este hexámero recombinante es estable sin necesitar ningún cofactor como ADN, dNTPs o magnesio, fenómeno que no es común en todas las helicasas hexaméricas de SF4 (Farge, Holmlund et al. 2008).

La actividad helicasa predicha por los alineamientos de secuencia con gp4 se comprobó *in vitro*, donde se demostró que Twinkle tiene la capacidad de desenrollar ADN de doble cadena, que posee una direccionalidad 5'→3'. Estos estudios *in vitro* además demostraron que Twinkle sólo presenta actividad con sustratos tipo horca en los cuales existe un extremo 5' monocatenario con una longitud mínima de 20 bases. Como todas las helicasas, Twinkle

requiere el consumo de dNTPs para catalizar la reacción (Korhonen, Gaspari et al. 2003), y realiza su actividad con trifosfatos diferentes (UTP, ATP, GTP, etc.) (Jemt, Farge et al. 2011) aunque la actividad es mayor con UTP. La actividad helicasa de Twinkle se ve estimulada por la presencia de la proteína de unión a cadena sencilla mitocondrial mtSSB (Korhonen, Gaspari et al. 2003; Jemt, Farge et al. 2011). A pesar de la evidente capacidad helicasa por parte de Twinkle, ésta no puede desenrollar más allá de 55 pb de ADN sin la presencia de Pol- $\gamma$ , cuya interacción produce un aumento de la procesividad de Twinkle, permitiendo desenrollar longitudes de hasta 15000pb (Korhonen, Pham et al. 2004).

Estudios *in vitro* han mostrado que la unidad mínima del replisoma en mitocondria lo componen la Pol- $\gamma$  y Twinkle (nótese que ninguna de estas proteínas puede realizar la reacción de modo independiente una de la otra) y, como se había observado en los ensayos helicasa, la replicación también se ve fuertemente estimulada por la presencia de mtSSB, dando productos de mayor longitud. Se intentó realizar la replicación *in vitro* empleando la polimerasa de fagoT7, la que tiene una gran homología con la polimerasa mitocondrial, y Twinkle, y con la Pol- $\gamma$  con la helicasa gp4, y el resultado fue que no se observó actividad en ninguno de los dos casos siendo indicativo de la existencia de interacciones específicas entre Twinkle y Pol- $\gamma$  (Korhonen, Pham et al. 2004). En concreto, la interacción de Pol- $\gamma$  con Twinkle esta mediada por la subunidad Pol- $\gamma$ B (Farge, Holmlund et al. 2008). Por el contrario, la interacción de Twinkle con mtSSB es inespecífica, ya que si ensayo *in vitro* de replicación se intercambian a mtSSB por un ortólogo SSB de gp5 del fagoT7 no se ve afectado el resultado de la reacción (Sen, Nandakumar et al. 2012). Para la replicación, al igual que para los ensayos helicasa, es necesaria la presencia de dNTPs (Korhonen, Pham et al. 2004). Estudios *in vitro* muestran que un sustrato de ADN de doble cadena que contenga una estructura de burbuja de replicación da lugar a una reacción de replicación, lo que demuestra que Twinkle es capaz de cargarse por sí misma (sin ayuda de una molécula cargadora adicional) en un ADN circular cerrado monocatenario (Jemt, Farge et al. 2011).

Twinkle es capaz de interactuar *in vitro* tanto con ADN de cadena sencilla (ssADN) como doble (dsADN), pero muestra mayor afinidad por este último, y no requiere nucleótidos para la unión (Sen, Nandakumar et al. 2012). Además se ha observado que para la interacción con el ADN la formación de hexámeros es indispensable, lo que implica que el dominio C-terminal es el principal responsable de la unión mientras que la interacción predicha para el del N-terminal es más marginal. Sin embargo, cuando se elimina el dominio N-terminal la actividad helicasa *in vitro* es menor que la proteína completa, tanto si es parcial eliminando sólo el ZBD o si se elimina también el RPD; éste último genera un descenso más acusado de la actividad (Farge, Holmlund et al. 2008). Este efecto es más evidente si se realiza un ensayo de *in vitro*

replicación, donde la sustracción del ZBD produce un ligero descenso en la longitud de los productos procesados, mientras que cuando se elimina totalmente el dominio N-terminal la actividad es casi nula (Farge, Holmlund et al. 2008). Estos resultados no permitieron asignar una función específica para el dominio N-terminal pero si denotan su relevancia en Twinkle (Farge, Holmlund et al. 2008).

Twinkle, como otras helicasas (Machwe, Xiao et al. 2005), además de poder catalizar la reacción de desenrollamiento del ADN bicatenario puede catalizar la reacción de hibridación de dos fragmentos de ADN monocatenario, sin necesitar el consumo de NTPs, al contrario que para la reacción de desenrollamiento (Sen, Nandakumar et al. 2012).

#### **6.4.1 Mutaciones deletéreas de Twinkle.**

La enfermedad de la Oftalmoplegia Externa Progresiva Autosómica Dominante (adPEO) presenta como síntomas una debilidad en los músculos externos del ojo, y como consecuencia los movimientos oculares son muy limitados (Kawashima, Ohta et al. 1994). Esta deficiencia se relaciona con un descenso en la cantidad de mtADN en tejidos oculares originado por una serie de mutaciones en genes nucleares (Zeviani, Servidei et al. 1989). Concretamente, son tres los genes implicados en el desarrollo de esta disfunción: *POLG1*, *Twinkle* y *ANT* los cuales codifican, respectivamente, a la polimerasa Pol- $\gamma$ , la helicasa Twinkle y a “*ADP/ATP translocase 1*”. El análisis a nivel celular del efecto de estas mutaciones en *Twinkle*, tanto en células de mamífero (Wanrooij, Goffart et al. 2007) como en *Drosophila* (Matsushima and Kaguni 2007), ha demostrado que se produce un acusado descenso en el contenido del mtADN.

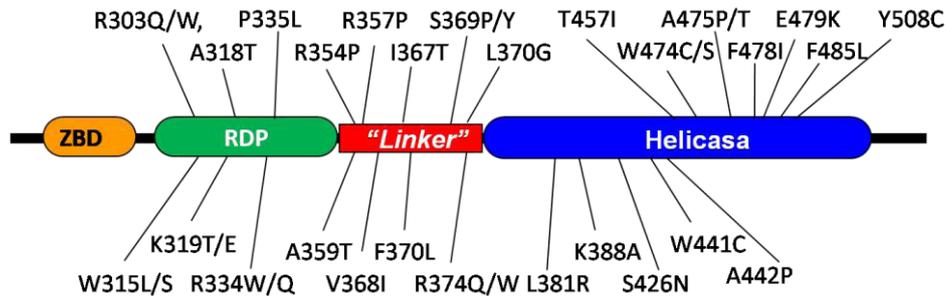
En el caso particular de Twinkle las principales mutaciones se sitúan en el “*linker*” entre los dominios N- y C-terminal (figura\_7) (Spelbrink, Li et al. 2001), aunque, en nuevas poblaciones de estudio, se van descubriendo nuevas mutaciones que no se encuentran en el “*linker*” y también generan el desarrollo de adPEO. Por otra parte, dos mutaciones puntuales Y508C y A318T de Twinkle se han relacionado con la ataxia espino cerebral infantil (IOSCA), una enfermedad neurodegenerativa que provoca muerte prematura (Nikali, Suomalainen et al. 2005) o episodios epilépticos (Lonnqvist, Paetau et al. 2009). Otras mutaciones como R334W/Q asociadas a adPEO también se han relacionado a otras enfermedades como Parkinson (figura\_7) (Vandenberghe, Van Laere et al. 2009).

En total, las mutaciones de Twinkle relacionadas con enfermedades son R334W/Q, P335L, R303Q/W, W315L/S, K319T/E en el dominio N-terminal. En el “*linker*”, las mutaciones son R354P, R357P, A359T, L360G, I367T, V368I, S369P/Y, F370L, R374Q/W, L381R. Finalmente, en el dominio helicasa, son K388A, S426N, W441C, A442P, T457I, W474C/S, A475P/T, F478I,

E479K, F485L, (Lewis, Hutchison et al. 2002) (Arenas, Briem et al. 2003) (Deschauer, Kiefer et al. 2003) (Liu, Ding et al. 2008) (Sanchez-Martinez, Calleja et al. 2012) (Sarzi, Goffart et al. 2007).

**Figura\_7. Mutaciones asociadas a Twinkle.**

Con diagrama de bloques se indican los dominios de Twinkle ZBD (naranja), RPD (verde), "linker" (rojo) y helicasa (azul), indicándose las mutaciones localizadas en Twinkle entre diversas poblaciones humanas.



Con el objetivo de comprender el efecto de estas mutaciones en Twinkle se han realizado distintos estudios tanto *in vitro* como *in vivo* en *Drosophila*, principalmente de aquellas localizadas en el "linker". De los estudios *in vitro* de las mutaciones A359T, I367T, V368I, S369P, R374Q y L381P, se observó que no todas poseen los mismos efectos en la integridad y actividad de Twinkle. Así como el mutante V368I tan sólo presenta un pequeño descenso de la actividad NTPasa, los demás producen grandes alteraciones en la actividad y estabilidad de la proteína, como los mutantes I367T y R374Q, que no conservan la estructura hexamérica y sólo están presentes en forma de monómero, y por tanto son totalmente inactivos. El mutante L381P todavía es capaz de formar hexámero pero no tiene actividad helicasa. Las mutaciones A359T y S369P presentan cierta actividad marginal, la primera presenta un descenso del 50% en la eficiencia de la replicación, mientras que la segunda no es capaz de realizar la reacción de replicación aunque conserve una actividad helicasa parcial (Korhonen, Pande et al. 2008).

Empleando *Drosophila* como sistema de estudio se ha analizado el efecto de múltiples mutaciones, como la sobreexpresión de los mutantes A326T, R341Q y W441C, y se observó que no producían un descenso del d-mtADN, sin embargo los mutantes I334T, A442P, K388A, D483A sí redujeron la cantidad del d-mtADN. En células de insecto todas estas mutaciones son letales (Matsushima, Ide et al. 2006), aunque algunas mutaciones como V368I ya se había observado en humanos que no provoca el desarrollo de adPEO (Arenas, Briem et al. 2003). También se han estudiado el efecto de las mutaciones K574A, R576A, Y577A, D580A, E587A, F588A, K590A e Y595A en *Drosophila*, que afecta a aminoácidos que estarían implicados en la interacción con el dNTP y en su hidrólisis. Estas últimas mutaciones conservan el estado hexamérico, sin embargo en la mayoría produce un descenso en la cantidad de mtADN,

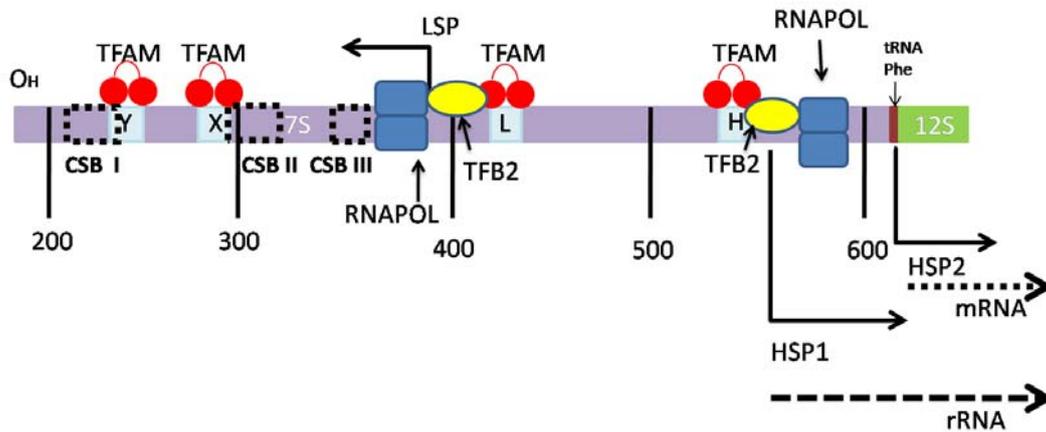
excepto los mutantes D580A, E587A y K590A (Matsushima, Farr et al. 2008; Matsushima and Kaguni 2009).

## 6.5 Transcripción del genoma mitocondrial.

Tal y como se ha explicado al principio de esta introducción, cada hebra del genoma mitocondrial tiene sus propios promotores específicos, “*Heavy Strand Promoter*” (HSP1 y HSP2) para la hebra pesada y “*Light Strand Promoter*” (LSP) para la hebra ligera. Ambos promotores están situados en la región control no codificante, NCR (ver más arriba) (Falkenberg, Larsson et al. 2007). Durante la transcripción mitocondrial de ambas cadenas H y L se genera un precursor de ARN policistrónico que contiene prácticamente toda la información de todos los genes. Este ARN se procesa posteriormente para producir los mRNA, rRNA y tRNA (Montoya, Ojala et al. 1981). Dado que los genes codificantes de las proteínas están separados por tRNAs, la configuración del mtDNA se dice que sigue un “modelo de puntuación de tRNA” (Montoya, Ojala et al. 1981), en el que el ARN transcrito se procesa mediante ARNasas que cortan los extremos de los tARNs, liberándose así los mARNs correspondientes a proteínas. Para la hebra pesada se han descrito dos lugares de inicio de la transcripción HSP1 y HSP2 (Montoya, Christianson et al. 1982). HSP1 se encuentra en el extremo de la NCR, a 16 pb “aguas arriba” del primer gen, que corresponde al del tRNA PHE (figura\_8), y produce un transcripto que incluye a los rARNs 12S y 16S, separados por el tRNA VAL y seguidamente acaba en el tRNA LEU. Este transcripto no llega al final de la cadena H debido a la unión del factor de transcripción MTERF en una parte del tRNA de LEU (Kruse, Narasimhan et al. 1989) (figura\_2). HSP2 se encuentra cerca del fin 3’ del gen del rARN 12S, y genera un transcripto policistrónico para casi toda la hebra pesada. No obstante, como se ha mencionado en el apartado 6.1.1. de la Introducción, la existencia de HSP2 se encuentra bajo discusión en el campo, ya que hay grupos de investigación que postulan que esta segunda transcripción también nace en HSP1 (Falkenberg, Larsson et al. 2007). Por su parte, desde el promotor de transcripción LSP se produce el transcripto de gran parte de la cadena L que da lugar a una molécula de ARN que codifica ocho tRNA y la subunidad ND6 de la NADH deshidrogenasa. Además, desde LSP, se producen algunos productos RNA truncados (Bonawitz, Clayton et al. 2006; Brown and Clayton 2006) que se detienen en CSB II, y que se ha propuesto se emplean como cebadores de la replicación de la hebra complementaria (figura\_8) (Wanrooij, Uhler et al. 2012).

**Figura\_8. Origen de la transcripción.**

Con dos círculos rojos conectados se representa los dos dominios de TFAM sobre las secuencias de unión que se han descrito, indicadas en cuadros azul celeste y marcadas con las letras L para LSP, H para HSP, y X e Y para los sitios de unión a las secuencias X e Y. La secuencia que codifica para el rARN 7S se indica con color violeta y para el rARN 12S en verde. El factor de transcripción TFB2 se indica en círculos amarillos y la RNAPOL con dos cuadrados azules. La secuencia de codificación de tRNA<sup>PHE</sup> se indica en marrón. Con flecha puntuada se indica el transcrito desde HSP2, y con flecha discontinua el transcrito HSP1. Las secuencias conservadas CSB I, II y III se indican con cuadrados de bordes de puntos de color negro.



En el inicio de la transcripción mitocondrial están involucradas la RNAPOL, el factor de transcripción B2 (TFB2) y TFAM. RNAPOL necesita estos dos cofactores para iniciar la transcripción, no como su homóloga de bacteriófago T7 que no necesita a ninguno. Por ello se puede anticipar que RNAPOL tiene sitios de reconocimiento específico para los factores de transcripción. Ello confirmó gracias a la resolución de su estructura cristalográfica del complejo RNAPOL (Ringel, Sologub et al. 2011) donde se observa que RNAPOL conserva el plegamiento común de las “polimerasa tipo A” con la forma de una mano, presentando los subdominios “palma, dedos y pulgar”. RNAPOL tiene un dominio diferente adicional a la polimerasa de T7 denominado “dominio PPR”, rico en prolinas, que se encuentra ubicado en el motivo denominado “pulgar” y se encarga del reconocimiento del ARN. TFB2 pertenece a la familia de las rARN metiltransferasas bacterianas, y forma un heterocomplejo estable con RNAPOL (McCulloch and Shadel 2003). Su función es facilitar la formación de la burbuja inicial para el desenrollamiento del ADN separando ambas hebras en el promotor (Ringel, Sologub et al. 2011). La tercera proteína implicada en la transcripción es TFAM (Gaspari, Falkenberg et al. 2004; Gaspari, Larsson et al. 2004); (Shi, Dierckx et al. 2012), que reconoce específicamente una secuencia de 22pb situada a 15pb aguas arriba de HSP1 y LSP (posición -35 a -15) desde donde contacta el heterocomplejo RNAPOL/TFB2 (McCulloch and Shadel 2003);(Sologub, Litonin et al. 2009; Ngo, Kaiser et al. 2011; Rubio-Cosials, Sidow et al. 2011; Ngo, Lovely et al. 2014; Yakubovskaya, Guja et al. 2014). Además, TFAM cumple otra función esencial en la mitocondria, que es la de compactar el mtADN mediante interacciones inespecíficas (Brown, Tkachuk et al. 2011; Kukat, Wurm et al. 2011).

## 6.6 Factor de transcripción A mitocondrial humano (TFAM).

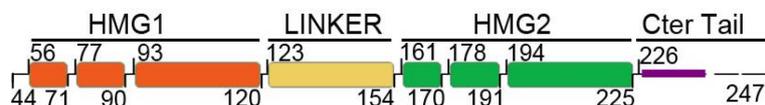
### 6.6.1 TFAM es una proteína con dos dominios HMGbox.

El gen *TFAM* está codificado en el genoma nuclear (Fisher and Clayton 1988). La proteína TFAM consta de 246 residuos de longitud, pero pierde los 42 iniciales como consecuencia de la escisión de la “secuencia guía mitocondrial” durante la importación a la mitocondria. TFAM está compuesta por dos dominios que presentan un plegamiento tipo HMG box, HMGbox1 (desde el residuo 42 al 122) y HMGbox2 (desde el residuo 154 al 225) unidos por un “*linker*” de 32 residuos de longitud (desde el residuo 123 al 153) y seguidos de una cola C-terminal (desde el residuo 226 al 246) (figura\_9a). En general, las HMGbox están constituidas por tres hélices, hélices 1, 2 y 3, precedidas por una secuencia extendida, y el conjunto da lugar a un dominio en forma de L. La secuencia extendida N-terminal se estabiliza con la hélice más larga, hélice 3, de modo antiparalelo generando el brazo largo de la forma L. Las hélices 1 y 2 son de similar longitud y se disponen entre ellas de forma antiparalela, dando lugar al brazo corto de la forma L del dominio (Parisi and Clayton 1991) (figura\_9b).

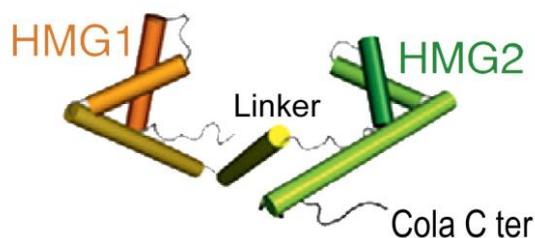
#### Figura\_9. Representación de TFAM en bloques y estructura secundaria.

En **a)** se representa con un esquema en bloques la estructura secundaria, toda hélices- $\alpha$  de TFAM. Se indican el primer y último residuo de cada hélice, los dominios HMGbox 1 y 2, el “*linker*” y la cola C-terminal. Abajo en **b)** el modelo obtenido de la estructura cristalográfica de TFAM/LSP (Rubio-Cosials, Sidow et al. 2011), donde se indica las hélices- $\alpha$  con cilindros. Adaptada de (Rubio-Cosials and Sola 2013).

a)



b)



La mayoría de proteínas con dominios HMG box en tándem unen al ADN de forma inespecífica. TFAM es una excepción a esta norma, ya que posee una actividad dual, específica de unión promotores e inespecífica al ADN en general (Fisher and Clayton 1988) (Alam, Kanki et al. 2003). Según el análisis de secuencia primaria cada dominio HMGbox tendría un papel diferente en el reconocimiento del ADN, el primer dominio tendría reconocimiento específico pero el segundo reconocería el ADN de forma no específica (Masse, Wong et al. 2002). Posteriormente, la resolución de la estructura cristalográfica del complejo de TFAM con la

secuencia de unión al promotor de la cadena ligera LSP (TFAM/LSP) (Ngo, Kaiser et al. 2011; Rubio-Cosials, Sidow et al. 2011) demostró que la interacción de HMGbox1 se produce como se había previsto para la interacción de HMGbox específicas, sin embargo, la interacción de HMGbox2 con el ADN tenía el mismo patrón de reconocimiento estructural que se había asignado a la secuencia de HMGbox específicas (Williams, Cai et al. 2004) (Rubio-Cosials, Sidow et al. 2011). En cuanto al “*linker*”, estudios de dispersión a bajo ángulo (*Small Angle X-ray Scattering, SAXS*) demostraron que presenta una estructura desordenada en disolución pero que adopta conformación de hélice cuando se une al ADN, tal y como muestran sus estructuras cristalográficas (Ngo, Kaiser et al. 2011; Rubio-Cosials, Sidow et al. 2011). Finalmente, la cola C-terminal presenta una estructura desordenada desplegada (figura\_9b) (Rubio-Cosials, Sidow et al. 2011).

### 6.6.2 Interacción de TFAM a las secuencias de ADN LSP y HSP1.

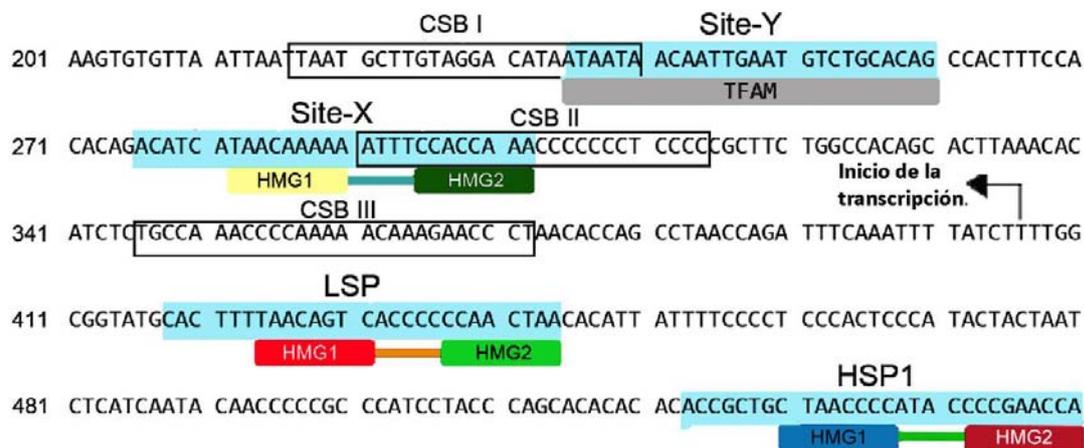
Los dos procesos más conocidos en los que participan TFAM son el inicio de la transcripción y la compactación del ADN, sin embargo se encuentra relacionada con múltiples procesos adicionales: reconoce ADN dañado por cis-platino o por oxidación (Yoshida, Izumi et al. 2002), y cruce de cuatro hebras (“*four-ways junctions*”) (Ohno, Umeda et al. 2000). Por análisis de protección a la metilación mediante dimetil sulfato (DMS) se demostró tanto *in organello* como *in vitro* que TFAM protege cuatro lugares en NCR: LSP, HSP1, Site-X y Site-Y, estos últimos entre los CSB-I y II (Fisher, Topper et al. 1987) (Ghivizzani, Madsen et al. 1994) (figura\_10).

En concreto Site-X, tema principal de estudio en su interacción con TFAM en esta tesis, comprende desde el nucleótido 276 al 303 del mtADN, se encuentra a 104 pb aguas abajo respecto de LSP y se solapa parcialmente con CSB-II (299-315). Se ha constatado que se produce una terminación prematura de la transcripción de la hebra ligera inmediatamente “aguas abajo” de CSB-II, ayudada por una secuencia poli-dT que está incluida en el Site-X; de hecho, si se elimina la región CSB-II no se produce dicha interrupción de la transcripción (Pham, Farge et al. 2006) (Wanrooij, Uhler et al. 2010). La secuencia CSB-II tiene, en la cadena H, dos secuencias consecutivas ricas en guaninas. Durante la transcripción de la cadena L, en esta región se genera un ARN complementario que también contiene la dos secuencias ricas en guaninas. Trabajos de los últimos años han demostrado que las cuatro secuencias ricas en guaninas (dos secuencias en la cadena H y dos más en la hebra de ARN) forman una estructura terciaria del ADN denominada cuádruples de guaninas (“*G-quadruplex*”). En general estas estructuras consisten en la unión de cuatro hebras ricas en guaninas donde cuatro guaninas de distintas hebras contactan entre sí mediante de puentes de hidrógeno tipo Hoogsteen,

formando un plano cuadrado. Por tanto, la interacción de una secuencia con guaninas consecutivas con otras tres hebras también con guaninas consecutivas dando lugar a un apilamiento de “*G-quadruplexes*”, implicando las cuatro hebras. Estudios *in vitro* han demostrado que esta estructura “*G-quadruplex*” está implicada en la terminación prematura del transcripto iniciado en LSP y terminado aguas debajo del CSB-II (Wanrooij, Uhler et al. 2010).

**Figura\_10. Sitios de unión específicos de TFAM.**

Se indica la secuencia del mtADN desde el nucleótido 200 al 553, la coordenada del mtADN se indica a la izquierda de cada fila de la secuencia. Con un rectángulo de borde negro se indica las secuencias correspondientes a CSB I, II y III, y con un rectángulo azul celeste se indica las secuencias identificadas TFAM. Con un modelo de dos rectángulos conectados se simbolizan los distintos dominios de TFAM contactando con la secuencia de ADN de las estructuras cristalográficas de los complejos con LSP, HSP1 y Site-X. La flecha indica la dirección de la transcripción de la cadena ligera.



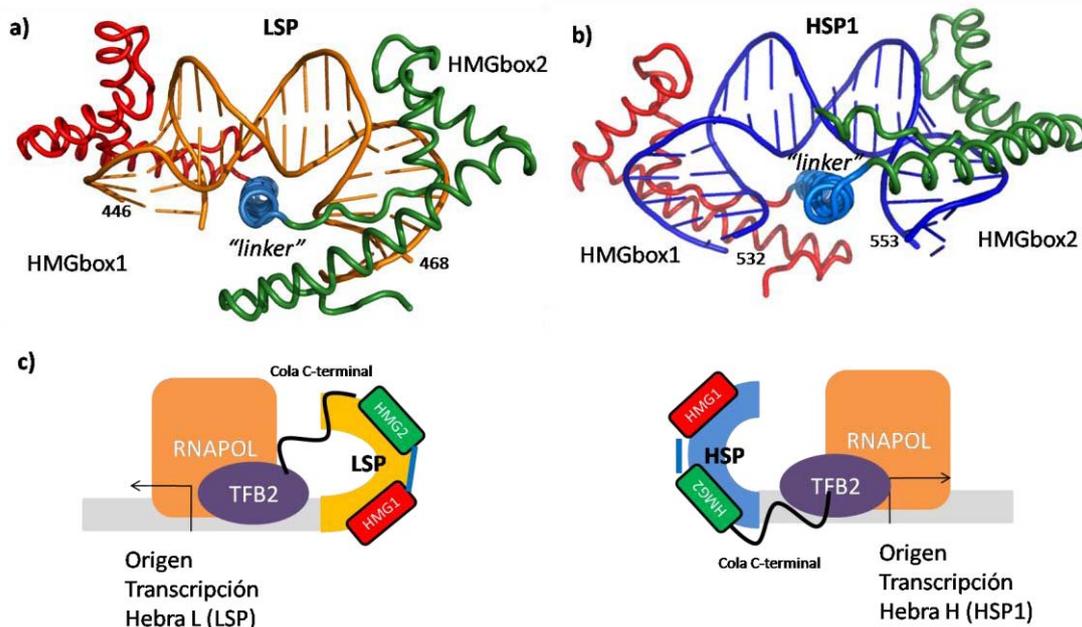
Ambos dominios HMG box de TFAM poseen un parche hidrofóbico en la zona cóncava de la forma de L de la HMGbox, en la cual se encuentran una serie de aminoácidos que insertan entre dos pares de bases del ADN a través del surco menor, además de la intervención de otros residuos polares en este contacto. Esta interacción genera un doblamiento del ADN con un ángulo de 90° por HMGbox, provocando entre HMGbox1 y 2 una curvatura total de 180°, generando una estructura denominada “*U-turn*” en la cual la proteína se enrolla alrededor del ADN (figura\_11). Los dos dominios HMGbox contactan al ADN por caras opuestas mientras que el “*linker*” que los conecta discurre por la cara interna del “*U-turn*”, estabilizando las cargas de los fosfatos que se acercan entre ellos debido al doblamiento del ADN. De todos los residuos de las HMGbox 1 y 2 implicados en la interacción con las bases se destacan los residuos hidrofóbicos Leu58 y Leu182, de HMGbox 1 y 2 respectivamente. Estos dos residuos se insertan entre pares de bases rompiendo el apilamiento entre ellas y contribuyendo a la distorsión, estabilizando simultáneamente la apertura del surco menor en las regiones

dobladas ADN. Alrededor de este parche hidrofóbico mencionado, centrado en Leu58 y Leu182, se encuentran residuos básicos como lisinas y argininas que interactúan inespecíficamente con el esqueleto de fosfatos del ADN.

La unión de TFAM a LSP y HSP1 se ha caracterizado cristalográficamente (Ngo, Kaiser et al. 2011; Rubio-Cosials, Sidow et al. 2011; Ngo, Lovely et al. 2014). En LSP la secuencia de unión de TFAM se encuentra a 40 pb antes del sitio de iniciación de la transcripción, mientras que HSP1 se encuentra a 30 pb. La dirección de la transcripción de las cadenas H y L se produce en sentido opuesto, en cambio, la unión de TFAM con los dos promotores es en la misma orientación. Esto da lugar a que en LSP, el dominio de HMGbox1 quede cerca del sitio de inicio de la transcripción y, por el contrario, en HSP1 es HMGbox2 es la que queda cerca del sitio de inicio de transcripción en la cadena H (figura\_11c) (Ngo, Kaiser et al. 2011; Rubio-Cosials, Sidow et al. 2011; Ngo, Lovely et al. 2014).

**Figura\_11. TFAM complejo con las secuencias LSP y HSP1 de ADN.**

En **a)** se representa el complejo TFAM-LSP (Código del Protein Data Bank, PDB, 3TQ6) y **b)** se indica el complejo TFAM/HSP (PDB 4NOD). En rojo se indica HMGbox 1, en azul el "linker" y en verde a HMGbox 2. El ADN con la secuencia LSP está en color naranja y la secuencia HSP1 en azul. Se indica el número de la primera y última base de la cadena H de cada ADN. En **c)** se esquematiza la unión de TFAM a los promotores LSP y HSP1 en la cual se observa cómo TFAM tiene la misma orientación respecto a las secuencias de ADN mientras que los inicios de interacción tienen orientación contrario en los cuales se unen RNAPOL (rosa) y TFB2 (violeta). Mismo código de colores para TFAM, LSP y HSP1 que a) y b).



La función de TFAM en la unión con HSP1 y LSP es reclutar a la maquinaria de la transcripción, lo que realiza mediante la interacción de su cola C-terminal con el heterocomplejo TFB2/RNAPOL (Dairaghi, Shadel et al. 1995). Las estructuras cristalográficas de TFAM en complejo con LSP y HSP1 muestran la orientación de la cola C-terminal de TFAM respecto a la secuencia de ADN y por tanto su orientación respecto al inicio de la transcripción en la cadena

ligera y en la pesada. En el caso de la transcripción de la cadena ligera, el “*U-turn*” permite acercar y orientar la cola C-terminal hacia el promotor LSP (Ngo, Kaiser et al. 2011; Rubio-Cosials, Sidow et al. 2011; Ngo, Lovely et al. 2014). Por el contrario, en la unión con HSP la cola C-terminal queda directamente orientada hacia el promotor (Ngo, Lovely et al. 2014) (figura\_11c).

### **6.6.3 TFAM y la compactación del mtADN.**

Observaciones por microscopía de fuerza atómica (AFM) y análisis resonancia de superficie de plasmon (SPR) (Kaufman, Durisic et al. 2007) demostraron que TFAM compacta el mtADN uniéndose a lo largo del genoma de un modo cooperativo cada 35pb, doblando el ADN y provocando la compactación del mismo. Recientemente, la estructura cristalográfica de TFAM en complejo con un ADN de secuencia diferente a LSP y HSP1, llamada secuencia no específica (nsADN), ha demostrado que también dobla el ADN en forma de “*U-turn*” aunque la unión sea inespecífica (Ngo, Lovely et al. 2014). La interacción del “*linker*” con el ADN tiene también un papel determinante para la estabilización del “*U-turn*” con esta secuencia inespecífica. Estudios mediante la técnica de pinzas ópticas en los cuales se analizó la compactación que sufre un ADN al ser unido por TFAM, demostraron que tanto HMGbox1 como el “*linker*” participan en dicha compactación. En estos experimentos mutaciones en los residuos que contactan con el esqueleto de ADN reduce la compactación a la mitad, equivalente a la que produce el constructo de la HMGbox1 por sí solo (Ngo, Lovely et al. 2014). Por otra parte, este mismo trabajo demostró que para que la compactación sea efectiva requiere que TFAM forme dímeros al interactuar con el ADN (Ngo, Lovely et al. 2014). Los dímeros formados por TFAM involucra la superficie de HMGbox1 opuesta a la superficie de contacto con el ADN, siendo el contacto antiparalelo entre dos HMGbox1 (el brazo corto de un dominio contacta con el brazo largo del otro dominio). La disrupción de los residuos que contactan entre sí entre dos HMGbox1 dio lugar a la disminución de la capacidad de compactación del ADN.

El mecanismo de cómo la célula regula la doble función de TFAM, la transcripción y compactación del ADN, es todavía desconocido, y las últimas hipótesis apuntan a la idea de que la concentración de TFAM presente en la matriz mitocondrial, regulada mediante el control de la expresión y/o degradación de la misma, induce una reacción u otra (Campbell, Kolesar et al. 2012). En este modelo, a concentraciones elevadas de TFAM se compactaría todo el genoma e impediría el acceso de la maquinaria de transcripción a los promotores; en cambio, a concentraciones inferiores se uniría a sus dianas específicas para el inicio de la transcripción, reclutando así a la maquinaria para ello y promocionando la expresión génica del mtADN ((Pohjoismaki, Wanrooij et al. 2006) .

#### 6.6.4 Características generales de los dominios “*High mobility group*” (HMG).

Las proteínas HMG se localizan en toda la célula, es decir, en núcleo, en citoplasma y en mitocondria, (Goodwin, Sanders et al. 1973) y poseen una gran diversidad funcional. Las proteínas HMG se clasifican en tres familias según su mecanismo de unión al ADN: “*HMG AT-hook*” (HMG-A), “HMG-dominio de unión nucleosomal” de (HMG-N) y “HMG box” (HMG-B) (Bustin 2001). La estructura tridimensional de las tres familias de proteínas son totalmente diferentes entre sí, y su clasificación conjunta se debe a que en general estas proteínas migraban rápidamente en geles tipo SDS-PAGE. Las proteínas HMG-A interaccionan con el ADN por el surco menor mediante una cola extendida sin producir distorsiones en la estructura del ADN (Dragan, Klass et al. 2003); las proteínas HMG-N (Gerlitz and Bustin 2009) interaccionan de modo no específico con el ADN en los nucleosomas y forman parte de la fibras de cromatina (Bianchi and Agresti 2005). La familia HMG-B, a la que pertenece TFAM, es el grupo más amplio y diverso. Las funciones de las proteínas HMG-B en la célula son el reconocimiento y mantenimiento del ADN. En el núcleo celular, las proteínas de esta familia pueden componerse por un dominio HMGbox único o doble (en tándem) o de más dominios, y pueden ser específicas de secuencia de unión al ADN, estando entonces su actividad relacionada a factores de transcripción, o no específicas de unión encontrándose asociadas a los cromosomas (Stros, Launholt et al. 2007).

A pesar de la gran variedad funcional de las proteínas HMG-B, estructuralmente cada dominio HMG-B tiene un plegamiento común en forma de L, compuesto por tres hélices- $\alpha$ , dos más pequeñas (hélice 1 y 2) y de longitud parecidas que forman un ángulo aproximado de  $80^\circ$  con la hélice 3 más larga, que se estabiliza con una región inicial de estructura extendida antiparalela a la hélice 3 (Stros, Launholt et al. 2007) (figura\_9), como se describió anteriormente para TFAM.

Las estructuras cristalográficas de las HMG box, tanto individuales como en tándem, han demostrado la capacidad no solo de unir el ADN a través del surco menor, sino que además induce una dobladura del ADN, tanto si la unión es específica como no específica (Malarkey and Churchill 2012).

#### 6.6.5 Análisis termodinámico de los complejos HMGbox con ADN.

Desde el punto de vista termodinámico la unión de las proteínas HMGbox al ADN es un proceso espontáneo ( $\Delta G < 0$ ) y endotérmico ( $\Delta H > 0$ ) (Dragan, Read et al. 2004). Procesos tipo endotérmico también se ha observado en otras proteínas que interaccionan con el ADN por el surco menor tales como la “proteína de unión a la secuencia TATA” (TBP), la “*Catabolite Activator Protein*” (CAP) (Jen-Jacobson, Engler et al. 2000) y las HU $\alpha\beta$  (Koh, Shkel et al. 2011).

A pesar de tener estructuras tridimensionales completamente diferentes, todas estas proteínas provocan una marcada dobladura del ADN. El caso opuesto son las HMGA1 AT-hook, que contactan por el surco menor sin producir ninguna distorsión del ADN y siendo la reacción exotérmica (Dragan, Liggins et al. 2003).

Cuando se realiza una interacción al ADN por el surco menor induciendo el doblamiento del mismo conlleva un proceso termodinámico endotérmico, *per ende* entálpicamente desfavorable. Por tanto, la formación espontánea del complejo está entrópicamente dirigida ( $\Delta S > 0$ ). De modo contrario, si la unión de las proteínas se produce por el surco mayor del ADN, el proceso espontáneo es exotérmico, la entalpía es positiva y favorable, por tanto entálpicamente dirigido. El coste energético que tiene doblar el ADN por el surco menor podría ser un motivo para justificar que el proceso de interacción sea endotérmico (Dragan, Read et al. 2004). Si se comparan en detalle la naturaleza termodinámica de la formación de complejos de proteínas con el surco menor o mayor del ADN se observa como las interacciones no polares son de mayor magnitud en la unión al surco menor; en cambio, en la unión al surco mayor predominan las interacciones electrostáticas (Privalov, Dragan et al. 2007). De modo general, durante la interacción entre proteínas y ADN, independientemente de si el proceso está dirigido entrópicamente o entálpicamente, se produce el fenómeno denominado “compensación entálpica/entrópica”, por el cual el aumento o descenso de la entalpía produce el efecto opuesto de la entropía, compensándose la variación de energía y manteniéndose la energía total de la reacción (Jen-Jacobson, Engler et al. 2000).

Cuando se produce una interacción entre dos macromoléculas suceden de modo simultáneo una gran cantidad de procesos entálpicos y entrópicos. A nivel entálpico encontramos la formación y/o rotura de enlaces no covalentes (puentes de hidrógeno, interacciones de van der Waals, interacciones hidrofóbicas y/o electrostáticas) entre las dos macromoléculas, y a su vez se rompen los enlaces que existen en la capa de solvatación tanto de moléculas de agua como de iones de cada macromolécula. Además, se producen procesos entrópicos como la liberación de moléculas de agua y/o iones de la capa de solvatación al disolvente, produciendo variaciones del grado de libertad a nivel macromolecular o atómico, etc. A pesar de esa gran complejidad, detallados estudios termodinámicos han permitido constatar que en el caso particular de las interacciones de las HMGbox con el ADN, y posiblemente en la mayoría de proteínas que producen una dobladura del ADN a través del surco menor, el principal factor entrópico (pero no el único) es la liberación de moléculas de agua que se encuentran en el surco menor del ADN como consecuencia de la inserción de los residuos hidrofóbicos entre los pares de bases a través de este surco (Privalov, Dragan et al. 2007; Dragan and Privalov 2008).

Aunque la formación de complejos de las proteínas HMGbox con el ADN está dirigido entrópicamente, existen algunas diferencias entre los complejos de proteínas HMGbox específicas y no-específicas. Las HMGbox específicas son estructuralmente más inestables sin unir el ADN y se estabilizan con la unión, y las interacciones electrostáticas son más determinantes en la formación del complejo siendo éste más compacto. Por el contrario, las proteínas HMGbox no-específicas tienen una estructura más rígida, y la variación de entropía tiene tendencia a ser de mayor magnitud en comparación con las HMGbox específicas (Dragan, Read et al. 2004) (Privalov, Dragan et al. 2011).



# **MATERIALES, MÉTODOS Y FUNDAMENTOS.**



## 7 MATERIALES, MÉTODOS Y FUNDAMENTOS.

### 7.1 Obtención de proteína recombinante: clonación, expresión y solubilización. Descripción general.

En las técnicas de biología estructural en general y en el caso particular de la cristalografía de macromoléculas se necesita una gran cantidad de proteína de gran pureza y homogeneidad. Para conseguirlo en la mayoría de ocasiones es necesaria la expresión heteróloga de las proteínas en sistemas de expresión simplificados que permiten realizar grandes cultivos celulares de los que extraer el material a cristalizar. El primer paso para conseguirlo es clonar el gen que codifica a la proteína diana, por ejemplo una proteína humana, en un vector adecuado para la expresión en función del sistema de expresión, entre otros en bacterias, levaduras, células de insecto o células de mamífero.

El tipo de sistema de expresión es crítico para la obtención de una proteína heteróloga en buen estado, ya sea estructural (es decir, bien plegada) como químico (sin degradación). No obstante hay proteínas que requieren un tratamiento adicional específico, por ejemplo modificaciones post-transduccionales como la glicosilación, la formación de puentes disulfuro, o que para el correcto plegamiento necesita de la participación de alguna chaperona específica. La expresión heteróloga puede ocasionar que el sistema de expresión no posea el sistema adecuado para el correcto plegamiento de la proteína, a pesar de que se exprese en grandes cantidades.

Para una solubilidad óptima durante la etapa de crecimiento del cultivo donde se produce la expresión de la proteína de interés hay que tener en cuenta distintas variables. La primera de ellas es el sistema de expresión empleado, entre otros puede probarse diferentes cepas de bacteria o levadura, o diferentes tipos de células de insecto y/o de mamífero. Las condiciones de crecimiento del cultivo también son críticas, en el caso de expresión en bacterias la solubilidad puede sufrir grandes variaciones en función de si la expresión es a 37 °C, 20 °C o 16 °C, o bien de la cantidad de inductor empleado para la expresión (0.5 o 1 mM) y el tiempo de inducción (3h, 16h, 24 o 36h). Para la expresión bacteriana, las proteínas heterólogas pueden ser inestables e insolubles a pesar de haber variado los parámetros mencionados más arriba. Una alternativa es fusionar en forma de quimera la proteína de interés con otras proteínas que sí sean estables en bacteria para así mejorar su solubilidad. Algunas de las proteínas de fusión más empleadas son: Proteína de unión a la maltosa (MBP), la proteína s-glutación transferasa (GST), tioredoxina (TRX), Protein G-1 (GB1), la proteína "*N utilization substance protein A*" (NusA), a la "*E. coli cell division protein Ftsz*" (Ztag), etc.

Durante la lisis celular, es crítico el tampón empleado en el cual la proteína de interés se verá expuesta. En el tampón de rotura de las células podrá variar la concentración de sal, el pH, se le puede añadir aditivos como reductores, iones, etc., que van a determinar la solubilidad de la proteína a una determinada concentración. Para determinar el tipo de tampón se realiza un barrido más o menos extenso de condiciones diversas para cubrir el mayor número de condiciones químicas diferentes en un mínimo número de ensayos. Para que este tampón sea óptimo se tiene en cuenta el pH, (que tiene que estar alejado 1.5 puntos del punto isoeléctrico de la proteína). También si existen cisteínas en la secuencia de la proteína, ya que entonces es conveniente añadir agentes reductores como el ditioneitol (DTT) o  $\beta$ -mercapto-etanol ( $\beta$ ME)). Finalmente, la actividad de la proteína, por ejemplo para proteínas de unión al ADN es aconsejable emplear concentraciones altas de NaCl entre 0.5 y 1M; para proteínas transmembrana es aconsejable emplear detergentes, etc. En último lugar, hay que tener en cuenta el empleo de inhibidores de proteasas endógenas del sistema de expresión para evitar la proteólisis de la proteína recombinante durante la lisis.

Frecuentemente, se construyen proteínas quimeras para facilitar las etapas de purificación y para facilitar la solubilidad, o ambas si es posible. Actualmente, la construcción de proteínas recombinantes fusionadas a otra proteína o a un fragmento polipeptídico es una práctica rutinaria en los laboratorios de biología estructural para conseguir una mejor eficiencia en las etapas de purificación, ya que se aprovecha la interacción específica entre la proteína y una resina. Como se ha mencionado anteriormente esta estrategia se emplea además para ayudar a la solubilización de las proteínas de interés, ya que la fusión de una proteína que es altamente soluble a la proteína diana genera una proteína quimera más soluble que la proteína diana independiente.

Tal y como se detalla en los siguientes apartados, en esta tesis se han combinado todas estas opciones para poder determinar las condiciones óptimas de solubilidad para cada una de las proteínas estudiadas. Se han realizado clones con diferentes quimeras: colas de histidinas (tanto en el extremo N como C-terminal), con GB1, TRX, MBP y NusA. La expresión de estas fusiones se probaron en nueve cepas diferentes de *E. coli* a dos temperaturas: 37 °C durante 3 horas y 20 °C durante 16 horas. Además se comparó el efecto de romper las bacterias en ocho tampones diferentes para los análisis iniciales de solubilidad. Posteriormente, las condiciones de solubilidad se optimizaron con la ayuda de técnicas como Dispersión dinámica de luz (DLS) y Barrido diferencial de fluorescencia (DSF), que se explicarán más adelante.

### 7.1.1 Biología molecular de TFAM.

El clonaje de la proteína TFAM y el constructo TFAM-CTT3 fue realizado con anterioridad por la Dra. Rubio Cosials, para ver más detalles ver artículo (Rubio-Cosials, Sidow et al. 2011) y tesis de Anna Rubio Cosials “Análisis bioquímico y estructural del factor de transcripción mitocondrial humano A, TFAM, en complejo con la secuencia promotora LSP”.

### 7.1.2 Biología molecular de Twinkle y sus distintos constructos.

Para conseguir los múltiples clones utilizados en el proyecto de Twinkle se siguió el protocolo clásico para generar los constructos de la proteína entera y sus fragmentos, y el sistema “KOD” más eficiente para generar variantes cortas y mutantes puntuales. Excepto para los clones TwinkleS359P,  $\Delta 386$ Twinkle $\Delta 636$ ,  $\Delta 386$ Twinkle $\Delta 647$ ,  $\Delta 386$ Twinkle $\Delta 663$ ,  $\Delta 386$ Twinkle $\Delta 636$ , Twinkle $\Delta 647$  y Twinkle $\Delta 663$  para los que se empleó el sistema “KOD”(explicado brevemente al final de esta sección), para los demás clones se siguió el protocolo clásico, basado en insertar un gen en el vector de expresión mediante una reacción de ligación. En el sistema clásico, inicialmente se amplía el inserto mediante la reacción en cadena de la polimerasa (PCR) con cebadores que introducen en los extremos del fragmento genético a amplificar la diana de corte de una enzima de restricción específica. En este caso se emplearon tres enzimas diferentes: Sfil (Thermo Scientific), Ndel (en 5') (Invitrogen) y XhoI (en 3') (Invitrogen), cuyas secuencias de corte se introdujeron en los extremos de los ADN amplificados (Tabla\_1).

**Tabla\_1. Lista de enzimas de restricción.**

| Enzima de restricción | Secuencia diana | Temperatura reacción |
|-----------------------|-----------------|----------------------|
| Sfil                  | GGCCNNNNNGGCC   | 50 °C                |
| Ndel                  | CATATG          | 37 °C                |
| XhoI                  | CTGGAG          | 37 °C                |

En el diseño de los cebadores, además de incluir la secuencia de la diana de las enzimas de restricción se tuvo en cuenta que la temperatura de hibridación de ambos cebadores fuera parecida en un rango entre 45-60 °C, calculada mediante la ecuación\_1. Aparte, se revisó que los cebadores no formasen estructuras secundarias llamadas horquillas, empleando distintos servidores de la red.

$$\text{Ecuación}_1: T_m = 81,5 + 0,41*(GC\%) - 675/N - \% \text{ base desapareada}$$

Siendo la  $T_m$  la temperatura de hibridación, G el número de guaninas, C el número de citosinas y N el número de bases totales.

En la tabla\_2 se detallan los cebadores utilizados, su secuencia, la temperatura de fusión teórica y la enzima de restricción cuya diana tiene incluida.

**Tabla 2. Lista de cebadores.**

En la tabla se indica el nombre de cada cebador con su secuencia correspondiente, su temperatura de fusión (Tm) calculada y la diana de la enzima de restricción correspondiente.

| Cebador        | Secuencia (5'→3')                                    | Tm (°C) | enzima |
|----------------|--|---------|--------|
| Sfil+TEV       | CCAAC TACGGCCA ACTCGGCCGAGAATTTGTATTTTCAGGGT         | 54.0    | Sfil   |
| T43_for_pHart  | GCCGAGAATTTGTATTTTCAGGGTGCCACTCTCCAAGCCTTGGATATGCCAG | 60.6    | ---    |
| K684_rev_pHart | ACTTAGTGGCCGAGGCGGCCTTACTTTGAACGCTTGGAGGTGCTGG       | 62.4    | Sfil   |
| A386_for_pHart | GCCGAGAATTTGTATTTTCAGGGTGCAGCTGGCCTCCGCTGG           | 61.3    | ---    |
| S75_for_pHart  | GCCGAGAATTTGTATTTTCAGGGTAGTTGCCGCGGCACTGAGC          | 63.2    | ---    |
| A118_for_pHart | GCCGAGAATTTGTATTTTCAGGGTGCAGAAGGGAGCTGGGAAGACTTCC    | 63.3    | ---    |
| S151_for_pHart | GCCGAGAATTTGTATTTTCAGGGTAGCGAGGAGGTCCGGAGGATCTG      | 63.5    | ---    |
| A359_rev_pHart | ACTTAGTGGCCGAGGCGGCCTTAGTTAGAAAGAGCATAAGAAGCATGGCGG  | 62.1    | Sfil   |
| N348_rev_pHart | ACTTAGTGGCCGAGGCGGCCTTAGCAGGGGACCTCCGGGACTTG         | 63.5    | Sfil   |
| T358_for_pHart | GCCGAGAATTTGTATTTTCAGGGTACCGCCCTGCCTGCCTGG           | 62.7    | ---    |
| 663_for_KOD    | CTCAGGCCAGGCCCCCA                                    | 58.0    | ---    |
| 647_G_for      | CAAATCTCAGAGTTCTATGTCGTAGCC                          | 63.0    | ---    |
| 636_K_for      | CCACCAAAGA ACTAGGCCCGGCTCAAGAAG                      | 65.7    | ---    |
| 663_rev_KOD    | CAAATCTCAGAGTTCTATGTCGTAGCC                          | 59.0    | ---    |
| 647_G_rev      | CTTTTGGCCACTGGTCAAGTGCATCCTTG                        | 63.0    | ---    |
| 636_K_rev      | CTTCTTGAGCCGGGCTAGTTCTTTGGTGG                        | 65.7    | ---    |
| S369P_for      | GAGGAGGTGCTAGGAGA ACTGTCA                            | 59.0    | ---    |
| S369P_rev      | CCGAAGCTGCCGAAATGTACGATGGACTTGTG                     | 66.0    | ---    |
| K684_rev_pCri  | TCGCTCGAGTTACTTTGAACGCTTGGAGGTGCTCTGG                | 62.4    | XhoI   |
| T358_for_pCriB | GGAATTCATATGCCGGCCACCGCCCTGCCTGCCTGG                 | 62.7    | NdeI   |
| A118_for_pCriB | GGAATTCATATGCCGGCCGAGAAGGGAGCTGGGAAGACTTC            | 60.5    | NdeI   |
| A386_for_pCriB | GGAATTCATATGGGCAGCTGGCCTCCGCTGG                      | 61.3    | NdeI   |

La amplificación de los genes para el clonaje en una variación del vector pTriEX (vector pHART1201, cedido por Darren Hart, EMBL, Grenoble, ver más abajo), se realizó con dos reacciones de PCR consecutivas ya que el fragmento a introducir era muy largo. Dicho fragmento incluía la secuencia de corte para la enzima de restricción (Sfil), una secuencia codificante para una cola de histidinas y el sitio de corte de la proteasa del virus de tabaco (TEV, producida en el laboratorio) entre la cola de histidinas y la proteína de interés. Esta secuencia se añadió al gen codificante de la proteína mediante dos cebadores encabalgados en el extremo 5' del clon (Tabla\_2, cebador Sfil+TEV y cebador 5' y 3' que definen el origen y final del constructo). Para generar el extremo 5' del constructo con estas secuencias adicionales, una primera PCR se realizó con un cebador en 5' que incluía la secuencia 5' complementaria al gen de interés además de la secuencia que define el lugar de corte de la TEV, y con una

segunda reacción de PCR se añadió la diana de la enzima de restricción Sfil, la secuencia de cola de histidinas y la secuencia complementaria al ADN que codifica la secuencia del lugar de corte de TEV.

Los vectores empleados para la expresión de los constructos fueron pET28, pET40 y pHART (tabla\_3). Los tres vectores de expresión empleados poseen una región denominada policonector (*polylinker*) en la cual se encuentran multitud de dianas de enzimas de restricción permitiendo introducir y ubicar el clon en un lugar específico del vector y poder emplear alguna de las enzimas de restricción que no corten nuestro constructo de interés. Los tres vectores tienen el promotor T7-lac que controla la expresión del fragmento clonado bajo inducción por isopropil- $\beta$ -D-1-tiogalactopiranosido (IPTG) (1992\_Neubauer). El vector pHART también se puede emplear para la expresión de la proteína recombinante en células de insecto y en células de mamífero.

**Tabla\_3. Vectores de expresión.**

| Vector | Longitud (pb) | Resistencia | Tag     | Proteasa corte |
|--------|---------------|-------------|---------|----------------|
| pET28  | 5.369         | Kanamicina  | His-tag | TEV            |
| pET30  | 7.181         | Kanamicina  | MBP     | TEV            |
| pTriEX | 5.934         | Ampicilina  | His-tag | TEV            |
| pET28  | 6.245         | Kanamicina  | GB1     | TEV            |
| pET30  | 7.559         | Kanamicina  | NusA    | TEV            |
| pET28  | 6.401         | Kanamicina  | TRX     | TEV            |

Tras la digestión de cada inserto y de cada vector con la enzima de restricción adecuada (las condiciones de digestión fueron especificadas por la casa comercial de cada enzima) se purificaron el vector y el inserto mediante kit comerciales de purificación (Omega). La unión entre insertos y plásmidos se realizó mediante la reacción de ligación empleando la enzima T4 ligasa (FERMENTAS). El producto de ligación se empleó para transformar la cepa bacteriana *E.coli* DH5- $\alpha$ , siguiendo protocolo estándar. Esta cepa está optimizada para la incorporación de ADN exógeno gracias a las mutaciones que inactivan varias endonucleasas y evitan la recombinación con su genoma, de esta manera mejora la eficiencia de la transformación a partir de una reacción de ligación. Esta cepa permitió obtener una gran cantidad de réplicas del inserto ligado en el vector que, una vez purificado mediante kit comercial (QUIAGEN), se empleó para transformar las cepas bacterianas de expresión.

Para obtener variantes de los constructos genéticos y sus mutantes se empleó el sistema KOD, el cual mediante PCR se amplifica todo el vector con el inserto incluido a partir de cebadores que introducen mutaciones en los sitios de interés o causan deleciones de fragmentos largos,

generándose constructos más cortos. Los oligonucleótidos que introducen la mutación contienen un “*mismatch*” en el triplete a mutar, y su longitud es de 10 pb a cada lado del mismo. Para generar constructos más cortos que la proteína entera, según la posición relativa de los cebadores se genera un constructo más largo o más corto. El producto de la PCR obtenido es lineal, por lo que tras eliminar el ADN parental con la enzima DpnI (Invitrogen), que reconoce el ADN metilado, este producto se fosforila y simultáneamente se liga empleando la ligasa T4 y la quinasa polinucleótidos T4 (PNK) (FERMENTAS) en la misma reacción. Este tipo de productos se emplearon para transformar las bacterias *E.coli* DH5- $\alpha$ .

### 7.1.3 Expresión heteróloga de Twinkle.

Para la expresión se emplearon cepas derivadas del *E.coli* DE3 modificadas para la expresión heteróloga de proteínas. Las cepas DE3 tienen el gen de la RNA polimerasa de T7 bajo control del promotor lacUV5, inducible con IPTG. Estas cepas permiten la producción de proteína recombinante a partir de vectores que poseen el gen de interés clonado bajo control del promotor T7. Por tanto, tras la inducción por IPTG de la expresión de la RNA polimerasa de T7, ésta se une al promotor regulador de la expresión de nuestro gen (1992\_Neubauer).

Las distintas cepas bacterianas derivadas de DE3 que se emplearon en nuestro proyecto fueron:

- **BL21 (DE3)**; genotipo: F- ompT hsdSB (rB-mB-) gal dcm (DE3). Presenta el sistema de expresión del promotor T7.
- **BL21 (DE3) C43**; genotipo: F – ompT hsdSB (rB- mB-) gal dcm (DE3). Esta cepa contiene mutaciones sin caracterizar y genera un crecimiento mayor de la membrana y las vesículas internas siendo una buena cepa para la expresión de proteínas de membrana.
- **BL21 (DE3) star**; genotipo: F- ompT hsdSB (rB-mB-) gal dcm rne131 (DE3). Poseen una mutación en el gen de rne131 que codifica para la enzima “RNasa E” impidiendo que pueda degradar el mRNA, resultando un aumento en la estabilidad del mRNA.
- **BL21 (DE3) pLys-S**; genotipo: F- ompT hsdSB (rB-mB-) gal dcm (DE3) pLys-S (CamR). Deriva de la cepa BL21 (DE3) modificada para que se produzca la expresión de la lisozima T7, que degrada la T7 polimerasa clonada en *E. coli* (DE3). Así, se impide la expresión basal, descontrolada, de la proteína de T7 hasta que no se produzca la inducción por IPTG. De esta manera se evita la toxicidad producida por la expresión basal de proteínas recombinantes tóxicas, las cuales se inducen una vez el cultivo ha crecido suficientemente. Tiene resistencia al cloranfenicol.

- **Rosetta (DE3)**; genotipo F- ompT hsdSB (rB m) –B - gal dcm (DE3) pRARE2 (CamR). El código de los codones de traducción de bacteria es diferente al eucariota y en ocasiones se producen errores de traducción que impiden la expresión. Esta cepa ha sido modificada con siete ARNs de transferencia (tRNA): AGA, AGG, AUA, CUA, GGA, CCC, y CGG, que son diferentes entre bacteria y eucariota. Tiene resistencia al cloranfenicol.
- **BL21 (DE3) Origami 2**; genotipo:  $\Delta(\text{ara-leu})7697 \Delta\text{lacX74} \Delta\text{phoA}$  PvuII phoR araD139 ahpC galE galK rpsL F' [lac+ lacIq pro] gor522::Tn10 TRXB (StrR, TetR). La cepa origami2 tiene mutaciones en la tioredoxina reductasa y la glutatión deshidrogenasa, optimizándolas para que ambas proteínas realicen correctamente la formación de enlaces disulfuro de las proteínas citoplasmáticas de *E.coli*. Tiene resistencia a la tetraciclina y a la estreptomycinina.
- **Rosetta2pLys(DE3)**; genotipo:  $\Delta(\text{ara-leu})7697 \Delta\text{lacX74} \Delta\text{phoA}$  PvuII phoR araD139 ahpC galE galK rpsL (DE3) F' [lac+ lacIq pro] gor522::Tn10 TRXB pLysSRARE2 (CamR, StrR, TetR). Esta cepa bacteriana es una combinación de tres de las cepas anteriormente citadas: BL21 (DE3) Origami 2, BL21 (DE3) pLys-S y Rosetta (DE3). Posee resistencia a tres antibióticos: cloranfenicol, tetraciclina y estreptomycinina.
- **BL21 (DE3) GroEL-GroES**; genotipo: F-, ompT, hsdSB (rB- mB-), gal, dcm. Esta cepa ha sido modificada con el vector pGRO7, que contiene el promotor araB para sobreexpresar las chaperonas GroEL-GroES. Resistente a cloranfenicol.
- **BL21 (DE3) pG-KJE8**; genotipo: F-, ompT, hsdSB (rB- mB-), gal, dcm. Esta cepa ha sido modificada con un vector policistrónico pG-KJE8, el cual permite la expresión de las chaperonas dnaK-dnaJ-grpE mediante el promotor araB, inducible por L-Arabinosa, y de las chaperonas GroES-GroEL bajo control del promotor Pzt-1, inducible por tetraciclina. Resistencia a cloranfenicol.

Para averiguar qué cepa bacteriana ofrecía un resultado óptimo de expresión, se transformaron todas las cepas bacterianas arriba descritas con los constructos obtenidos por ingeniería genética. Se crecieron las cepas a 37 °C en cultivos de 5 ml de LB (medio Luria-Broth) hasta alcanzar una densidad óptica de 0.6, medida a una longitud de onda de 595 nm. Entonces se indujeron con ITPG (1mM) a dos temperaturas diferentes y por diferente tiempo: a 37 °C durante 3 horas, o a 20 °C toda la noche.

El análisis de la expresión de las muestras se realizó comparando una fracción equivalente del cultivo previo a la inducción y del producto al finalizar la inducción, por geles desnaturalizantes SDS-PAGE y tinción por Comassie.

#### **7.1.4 Solubilidad de Twinkle y sus múltiples constructos.**

Para determinar la solubilidad de cada constructo expresado en ocho cepas distintas se empleó un sistema que combinaba el empleo de una pipeta multicanal de ocho puntas con placas de 96 pocillos (Molecular Dimensions) (8 filas por 12 columnas), que permitieron analizar la solubilidad en ocho tampones distintos. Para ello se crecieron cultivo celulares (4 ml en tubos de vidrio de 20 ml) de las cepas que previamente en determinadas condiciones de temperatura habían mostrado una expresión positiva para cada clon. Una vez crecidos, mediante dos centrifugaciones consecutivas a 4500 rpm (Centrífuga 5810/5810, rotor Beckman Coulter A-2-DWP-AT) los *pellets* de 4 ml de los cultivos se acumularon en placas de 96 pocillos (Molecular Dimensions). Se descartó los 96 sobrenadantes con una pipeta de 8 puntas y los pellets se resuspendieron con los ocho tampones que se indican en la tabla 4 (un tampón por fila). La finalidad de emplear estas placas fue poder realizar de modo organizado las etapas de centrifugación, lisis y recuperación de la fracción soluble para analizar la solubilidad.

Los tampones empleados tenían la finalidad de cubrir un amplio espectro de pH de 6.5 a 9.5, y a distintas concentraciones de NaCl de 100, 300, 700 y 1000 mM, aunque la mayoría de las condiciones probadas ya tenían elevadas concentraciones de NaCl porque las proteínas de unión al ADN suelen ser un factor que evita la agregación de estas proteínas. Además se añadió 5 mM de  $\beta$ -ME para mantener el estado de oxidación de las cisteínas en forma reducida, ya que Twinkle posee ocho de estos aminoácidos. Posteriormente se lisaron las bacterias usando dos protocolos:

- En el caso de emplear microtubos en vez de placas de 96 pocillos, las células se lisaron por sonicación: el tubo se sumergía en hielo, y se sonicaba con una sonda del sonicador que tenía con una punta de 4mm de grosor. Se aplicaron pulsos de 1"on, 3"off durante 10" en on.
- Para las placas de 96 pocillos, se alternó congelación (en nitrógeno líquido) durante 30" con descongelación (empleando un baño a 37 °C) durante 90", con agitación fuerte con *vortex* entre descongelación y congelación, repitiéndose 5 veces.

Tras una centrifugación del producto de rotura a 16000g, 20 min a 4 °C se comparó la fracción que se conservaba en el sobrenadante (fracción soluble) y la fracción sedimentada (fracción

insoluble) mediante electroforesis SDS-PAGE con geles de acrilamida al porcentaje adecuado para cada constructo genético, 10 o 12%.

**Tabla\_4. Lista de tampones de rotura.**

Lista de los tampones utilizados en los análisis inicial de solubilidad.

| Número | Tampón  |
|--------|---|
| 1      | Fosfato sódico 100mM pH 7.5 NaCl 700mM BME 5mM      |
| 2      | Tris-HCl 100mM pH 7.5 NaCl 1M BME 5mM               |
| 3      | Fosfato sódico 100mM pH 7.5 NaCl 700mM glicerol 20% |
| 4      | Citrato sódico 100mM pH 6.0 NaCl 100mM BME 5mM      |
| 5      | Bicina 100mM pH 9.5 NaCl 1M BME 5mM                 |
| 6      | Citrato sódico 100mM pH 6.5 NaCl 700mM BME 5mM      |
| 7      | Tris-HCl 100mM pH 8.5 NaCl 700mM BME 5mM            |
| 8      | Fosfato sódico 100mM pH 7.5 NaCl 300mM BME 1mM      |

## 7.2 Purificación de proteínas recombinantes.

### 7.2.1 Fundamento.

Para abordar las técnicas de biología estructural que se desarrollaron en este trabajo fue necesario conseguir elevadas cantidades de proteína, que tuvieran un elevado grado de pureza (>95% pureza evaluada mediante gel SDS-PAGE teñido por Comassie) y que la muestra fuese homogénea. Para lograr una muestra que cumpliera con estos criterios de pureza se emplearon técnicas como cromatografía de afinidad y de exclusión de tamaño. En general, para la preparación de muestras para la cristalografía, la estrategia de purificación se divide en tres fases principales: una primera fase se denomina preparativa en la que se realiza un primer paso de purificación, en general por cromatografía de afinidad, donde normalmente se elimina la mayor parte del extracto crudo celular de la lisis, y se separa, concentra y estabiliza la proteína diana. Según la pureza obtenida se suele incluir una segunda etapa de purificación normalmente por cromatografía de intercambio iónico o de otro tipo donde se eliminan más contaminantes, y finalmente una tercera etapa de pulido en la cual nos aseguramos la homogeneidad y la pureza de la muestra final, normalmente mediante cromatografía por exclusión de tamaño. Pasamos a describir los fundamentos de las técnicas mencionadas.

#### 7.2.1.1 Cromatografía de afinidad.

La cromatografía por afinidad se basa en la unión específica reversible entre la proteína de la fase móvil y un ligando anclado en una matriz estacionaria. Posteriormente, añadiendo un

competidor que presente mayor afinidad por la proteína (en la fase móvil) que el ligando de la fase estacionaria se produce la elución de la proteína. Para conferir afinidad a la proteína diana se generan quimeras con proteínas o fragmentos de unión específica a resinas de purificación específicas, tales como la proteína de unión a la maltosa (MBP), colas de histidina, o la proteína s-glutación transferasa, para las cuales existen unas resinas específicas para cada una de ellas. Teóricamente, el producto de elución debería de ser tan solo la quimera de unión específica, pero experimentalmente se producen unión inespecífica de otras proteínas celulares.

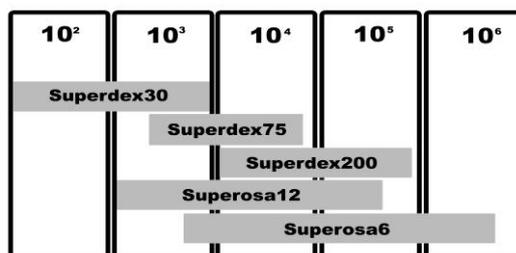
### 7.2.1.2 Cromatografía de exclusión por tamaño.

La cromatografía de exclusión por tamaño se basa en la separación de partículas de la fase móvil en función de diferente coeficiente de migración de dichas partículas de la fase móvil a través de la fase estacionaria. El diferente coeficiente de migración es consecuencia del distinto volumen aparente de cada partícula, que depende del radio de *Stocks* de la misma. La fase estacionaria es una resina que se puede formar con polímeros de distinta composición química, la cual genera unas redes o mallas esféricas con un diferente entramado, y este entramado provoca una diferente retención de las partículas. Las partículas de menor tamaño tienen un mayor tiempo de retención, mientras que las partículas de mayor tamaño poseen una velocidad de migración mayor y eluyen antes.

Existen multitud de resinas en el mercado, definiendo un rango específico de resolución óptima. Para la purificación de macromoléculas las columnas denominadas Superdex (GE HealthCare Life Sciences) fueron las empleadas durante esta tesis, de las que destacaron dos: Superdex 75 y Superose 6. Cada una posee un rango óptimo de purificación diferente, por tanto para cada una de las proteínas estudiadas en esta tesis se emplearon columnas distintas (Figura\_12.)

#### Figura\_12. Rango de resolución de las columnas de purificación por exclusión por tamaño.

En la parte superior se indica el peso molecular (en Dalton) de proteínas globulares y sobre recuadros grises el rango de uso de cada una de las diferentes resinas.



### 7.2.2 Protocolo de purificación de TFAM.

Ver artículo (Rubio-Cosials, Sidow et al. 2011) y tesis de Anna Rubio Cosials “Análisis bioquímico y estructural del factor de transcripción mitocondrial humano A, TFAM, en complejo con la secuencia promotora LSP”.

### 7.2.3 Protocolo de purificación de Twinkle y sus constructos.

A continuación se describe el protocolo de purificación final de Twinkle, para todas las construcciones clonadas: con una cola de histidinas en el extremo N- o C-terminal, y las construcciones (histag)- $\Delta$ 151-Twinkle, (GB1)- $\Delta$ 358-Twinkle, (TRX)- $\Delta$ 358-Twinkle, (GB1)- $\Delta$ 386-Twinkle, (GB1)- $\Delta$ 386-Twinkle $\Delta$ 636, (GB1)- $\Delta$ 386-Twinkle $\Delta$ 647 y (GB1)- $\Delta$ 386-Twinkle $\Delta$ 663. En todos los casos el protocolo empleado se dividió en tres etapas comunes, una primera de preparado de muestra y dos etapas cromatográficas de purificación, aunque hay algunas diferencias en los tampones según el constructo, que se detalla más adelante:

Etapa 1: Rotura de células: a partir de un cultivo de 2,4 L de LB crecido en cuatro Erlenmeyer de 2 L se fraccionaron las células por centrifugación (ThermoScientific Avanti; rotor Beckman-JA8100, 4000 rpm, 20 min, 4°C). Seguidamente se resuspendieron las bacterias en dos tampones según el tipo de constructo: para la proteína entera o con dominios adicionales al dominio helicasa se empleó 150 ml (50 ml/L de cultivo, 3L en total) de 50 mM citrato sódico pH 6.5, NaCl 1.5 M, L-arginina 50 mM, L-glutámico 50 mM, imidazol 50 mM, antiproteasas EDTA-free (Roche, una pastilla por 50ml) y 200  $\mu$ l de DNAsa a 2 mg/ml. Para los constructos del dominio helicasa se empleó como tampón 50 mM Tris-HCl pH7.5, NaCl 1.0 M, L-arginina 50 mM, L-glutámico 50 mM, imidazol 50 mM, antiproteasas EDTA-free (Roche). Se sonicó la muestra en hielo con una punta de 4 mm de grosor durante 20min con pulsos de 1" on 2" off, manteniendo la muestra en hielo. A continuación se centrifugaron las células lisadas a 20000 rpm (rotor JA20 de Beckman) durante 20 min a 4 °C. Se descartó el pellet y el sobrenadante se filtró con unos filtros de tamaño de poro de 0.25  $\mu$ m (MILLEX).

Etapa 2: El producto filtrado se inyectó en una columna de afinidad de 1ml que contenía níquel y era apropiada para proteínas fusionadas a una cola de histidinas. Esta columna estaba conectada a un sistema de purificación ÄKTA (GE HealthCare Life Sciences). Para la elución de los constructos Twinkle y  $\Delta$ 151-Twinkle se realizó un gradiente lineal de 20 volúmenes de columna (cv) desde 50 mM a 500 mM de imidazol, que compitió con el níquel de la resina al cual se habían unido las colas de

histidina. Para el gradiente los tampones utilizados fueron: Tampón A: 50 mM citrato sódico pH6.5, NaCl 1.0 M, L-arginina 50 mM, L-glutámico 50 mM, Imidazol 50 mM. Tampón B: 50 mM citrato sódico pH6.5, NaCl 1.0 M, L-arginina 50 mM, L-glutámico 50 mM, Imidazol 500 mM. Twinkle eluyó alrededor del 35% de tampón B (175 mM de imidazol). Para los constructos que contenían sólo el dominio helicasa (GB1)- $\Delta$ 358-Twinkle, (TRX)- $\Delta$ 358-Twinkle, (GB1)- $\Delta$ 386-Twinkle, (GB1)- $\Delta$ 386-Twinkle $\Delta$ 636, (GB1)- $\Delta$ 386-Twinkle $\Delta$ 647 y (GB1)- $\Delta$ 386-Twinkle $\Delta$ 663 los tampones A' y B' empleados fueron iguales que los A y B mencionados con la excepción que sustituyó el citrato sódico 50 mM pH 6.5 por Tris-HCl 50 mM pH 7.5. Para estos constructos helicasa no se realizó un gradiente lineal, sino con tres escalones a una concentración fija del tampón B a 10, 35 y 70%. En este caso, aunque los constructos del dominio helicasa de Twinkle eluían en los tres escalones, la fracción más soluble y menos heterogénea correspondió a las fracciones eluidas al 35% de tampón B, al 10% presentaba menor pureza por geles de SDS-PAGE y las fracciones al 70% mostraban una agregación mayor del 60% por cromatografía de exclusión de tamaño.

Etapa 3: Las fracciones de la etapa anterior que mostraron la proteína más pura mediante geles de SDS-PAGE y tinción de Comassie se concentraron con un concentrador de tamaño de poro de 30 KDa (Vivaspin 6 ml) hasta un volumen de 1 ml y se realizaron cuatro inyecciones de 250  $\mu$ l (a una concentración variable de 3-6 mg/ml) en una columna Superosa 6 10/300 también conectada a un sistema ÄKTA (GE HealthCare Life Sciences) y previamente equilibrada con el tampón 50 mM citrato sódico pH6.5, NaCl 1.0 M, L-arginina 50 mM, L-glutámico 50mM, para los constructos (GB1)- $\Delta$ 358-Twinkle, (TRX)- $\Delta$ 358-Twinkle, (GB1)- $\Delta$ 386-Twinkle, (GB1)- $\Delta$ 386-Twinkle $\Delta$ 636, (GB1)- $\Delta$ 386-Twinkle $\Delta$ 647 y (GB1)- $\Delta$ 386-Twinkle $\Delta$ 663. Para la elución se empleó el tampón 50 mM Tris-HCl pH7.5, NaCl 1.0 M, L-arginina 50 mM, L-glutámico 50mM. Todos los constructos eluyeron a 12,5 ml (el peso molecular aparente correspondería a 550 KDa, casi ocho monómeros) y su pureza se comprobó por geles de SDS-PAGE al 10% de acrilamida.

Además, se realizó un protocolo de purificación por gravedad con el constructo (GB1)- $\Delta$ 358-Twinkle siguiendo las siguientes etapas:

- El lisado se realizó siguiendo el protocolo descrito anteriormente y empleando como tampón de lisis 50 mM Tris-HCl pH7.5, NaCl 1.0 M, antiproteasa EDTA-free (Roche).
- Se empleó una columna (Bio-Rad) rellena con 1 ml de resina de níquel (Roche) previamente equilibrada con 50 mM Tris-HCl pH7.5, NaCl 1.0 M, 50 mM de imidazol, en la cual se incubó 20 min el sobrenadante obtenido de la lisis y se eluyó por

gravedad. Posteriormente se añadieron en volúmenes de 2 ml el tampón Tris-HCl pH 7.5, NaCl 1.0 M con cantidades crecientes de imidazol 75, 100, 150, 200, 300 y 500 mM. Las fracciones recogidas fueron de 1 ml. La pureza de cada dos fracciones se analizaron mediante geles de SDS-PAGE del 12.5%.

Se compararon las purificaciones de todos los constructos anteriormente mencionados, realizadas a temperatura ambiente y a 4 °C, y no se observó ninguna diferencia, por lo que se tomó la decisión de purificar a temperatura ambiente.

### **7.3 Análisis por electroforesis no desnaturizante de la interacción de proteína con ADN (EMSA).**

#### **7.3.1 Fundamentos.**

Las técnicas electroforéticas son aquellas en las cuales se realiza un desplazamiento de moléculas bajo la acción de un campo eléctrico. En el caso concreto de complejos proteína-ADN se emplea electroforesis en condiciones no desnaturizantes (también llamadas nativas), en la cual la muestra se dispone en un medio sólido. El medio más utilizado son geles, bien de acrilamida o de agarosa. Los geles evitan tanto la convección como la difusión de la muestra. La muestra migra en función de su carga, tamaño y forma, teniendo en cuenta que la carga de la muestra puede variar en función del pH y la fuerza iónica del medio, es decir, el tampón de electroforesis. Para el caso particular de análisis de complejos proteína/ADN se realiza la electroforesis en condiciones nativas, por tanto ni el tampón de electroforesis ni el gel deben contener ningún agente desnaturizante. La resolución de la electroforesis también viene determinada por el porcentaje de acrilamida y/o agarosa, se emplea porcentajes mayores para resolver diferencias entre moléculas pequeñas y viceversa. Para visualizar el resultado se emplean dos métodos diferentes: añadiendo una sustancia fluorescente intercalante del ADN, por ejemplo SyBrSafe (Invitrogen), o bien marcando el ADN radiactivamente con el isótopo de fósforo 32, o con cromóforos como fluoresceína, ya sea añadida en los ADN comerciales (SigmaAldrich) o mediante PCR (a partir de oligonucleótidos comerciales con fluoresceína).

En el caso particular de análisis de complejos de ADN con proteína, se suele emplear la técnica llamada "*electrophoretic mobility shift assay*" (EMSA), en la que se compara la migración de ADN libre con la migración del complejo, normalmente a diferentes ratios entre una cantidad fija de ADN y cantidades crecientes de proteína. Las condiciones electroforéticas de análisis de cada complejo deben establecerse de modo particular y depende estrictamente de sus propiedades físico-químicas. En este trabajo se han establecido las condiciones descritas en el apartado siguiente para cada sistema de estudio.

### 7.3.2 Análisis de la interacción TFAM/ADN mediante EMSA.

Los análisis de los complejos TFAM con distintas secuencias de ADN, todas ellas de 22 pb y doble cadena, Site-X, poly-dA y poly-dA-C se realizaron con el siguiente protocolo: geles de mezcla de acrilamida 30% (w/v) y bis-acrilamida 0.8% (w/v) polimerizada con un porcentaje del 12% en soportes de 10cm de longitud. Como tampón electroforético se empleó Tris-HCl 89 mM a pH 7.5, borato 89 mM, EDTA 2 mM pH 7.5. Las electroforesis se realizaron a temperatura ambiente.

Para este tipo de análisis, los complejos se realizaron mediante dilución serial de la proteína: se diluyó la proteína a una concentración diez veces mayor del punto de máxima concentración de la titración (por ejemplo 32  $\mu$ M), en el tampón empleado en la última etapa de purificación; de aquí se diluyeron 2  $\mu$ l en 2  $\mu$ l del mismo tampón y así sucesivamente para obtener todo el rango de la titración diez veces más concentrado. Después se dispensó 1  $\mu$ l de cada dilución en un eppendorf que ya contenía el ADN en el siguiente tampón de reacción: Tris-HCl 50 mM pH 7.5, MgCl<sub>2</sub> 5 mM, glicerol 2%, DTT 5 mM. La concentración final de NaCl era 75 mM aproximadamente. Se incubaban en hielo 30min antes de cargar en el sistema electroforético. Igual que con Twinkle, las medidas de concentración de proteína y ADN se realizó con *Nanodrop*, y la concentración de TFAM fue corregida con el coeficiente de extinción molar estimado experimentalmente en el laboratorio ( $\epsilon=35535$ ). Las cantidades de proteína y ADN empleadas se describen para cada paso en el apartado de Resultados 8.1.1.6.

### 7.3.3 Análisis de la interacción Twinkle/ADN mediante EMSA.

Para migrar mediante geles no desnaturizantes los complejos Twinkle/ADN se emplearon polímeros diferentes, una mezcla de acrilamida 30% (w/v) y bis-acrilamida 0.8% (w/v) al 8% (Sigma-Aldrich); o, por otro lado, geles de agarosa al 1.5%. Tanto para los geles de acrilamida como de agarosa se empleó como tampón electroforético Tris-HCl 40 mM, ácido acético 20 mM y EDTA 1 mM pH 8.0 (TAE 1x) y las electroforesis se realizaron a temperatura ambiente. Se analizó la unión a ADNs de distinta longitud, de 20, 30 y 153 pares de bases (pb) para hebra doble y 29 bases para hebra sencilla. Los oligonucleótidos de 29 bases, o de 20 y 30 pb se adquirieron comercialmente (Sigma-Aldrich) y el oligonucleótido de 153 pb se produjo mediante la reacción de PCR y purificado mediante el protocolo de precipitación etanol y fenol-cloroformo (Russell 2001). Para estos estudios no se tuvo en consideración la secuencia de los oligonucleótidos porque las helicasas interaccionan con el ADN inespecíficamente con la secuencia, por tanto sólo se tuvo en consideración la longitud y ó si era cadena sencilla o cadena doble.

Para la preparación de los complejos para EMSA se realizó una dilución serial de la proteína: se diluía la proteína sola en el tampón de elución de la cromatografía de gel de exclusión por tamaño a una concentración final diez veces mayor a la que se empleó en cada punto de la titración del gel. De cada dilución se dispensaba 1  $\mu\text{l}$  en una disolución que contenía el ADN diluido en el siguiente tampón de reacción: citrato sódico 100 mM, pH 6.5  $\text{MgCl}_2$  5 mM, L-arginina 100 mM, glicerol 10%, DTT 2 mM. La concentración final de NaCl era de 100mM aproximadamente. Esta mezcla se incubó en hielo 30 min antes de cargar en el sistema electroforético. La concentración de la proteína y del ADN se midió por *Nanodrop*; para el cálculo molar de Twinkle se consideró el estado hexamérico de la proteína y como coeficiente de corrección teórico 0.88, calculado con el programa VECTOR-NTI. Las cantidades empleadas en cada ensayo de proteína y ADN se describen en el apartado de Resultados 8.2.1.4.3.

## **7.4 Calorimetría Isotérmica de Titulación (ITC) aplicada a la interacción proteína/ADN.**

### **7.4.1 Fundamentos.**

La calorimetría isotérmica de titulación es una técnica biofísica basada en la medida del calor desprendido o absorbido en una reacción química a temperatura y presión constante, permitiendo en el caso de macromoléculas biológicas la estimación directa de la entalpía asociada a dicha reacción. Esta técnica es usada frecuentemente en estudios bioquímicos para caracterizar las interacciones de macromoléculas, tanto entre ellas (proteína-proteína o proteína-ADN/ARN) como con moléculas orgánicas de pequeño tamaño. Un experimento de ITC consiste en realizar inyecciones de pequeño volumen ( $v$ ) entre 2-50  $\mu\text{l}$  de modo sucesivo de un ligando (L) en el interior de una celda aislada térmicamente que contiene la muestra diana, hasta alcanzar la saturación del sistema. Por cada inyección se produce una variación de calor ( $q_i$ ) como consecuencia de la interacción entre el ligando y la molécula; ese calor desprendido o absorbido es proporcional a la entalpía ( $\Delta H$ ) de la reacción (ecuación\_2):

$$\text{Ecuación}_2: q_i = v \times \Delta H \times \Delta L_i$$

Donde ( $q_i$ ) es la variación de calor,  $v$  es el volumen de la celda y  $\Delta L_i$  es el incremento de concentración del ligando por cada inyección  $i$ . Los picos de calor medidos consisten en el calor que debe proporcionar el equipo para conservar la temperatura constante e igual entre la celda de referencia y la celda de la muestra. Esta diferencia es menor de modo progresivo con cada inyección hasta alcanzar la saturación en la cual sólo se mide el calor de dilución del ligando en el tampón. De modo directo se estima la entalpía de la interacción por la suma de

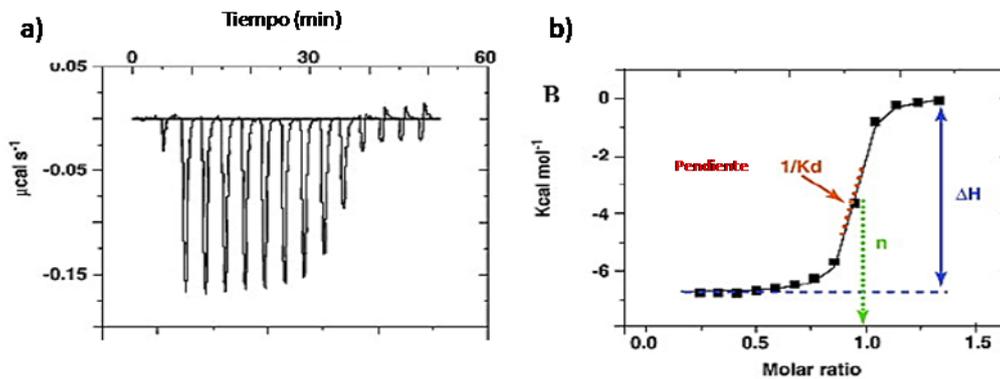
la integración de los picos de calor medidos. A partir de la pendiente de las curvas de titulación se calcula la constante de disociación ( $1/K_d$ ). La determinación del punto medio en la curva de titulación nos indica el valor de la estequiometría de la reacción ( $n$ ) (ver figura\_13).

Mediante esta técnica se puede evaluar:

- La contribución o variación entálpica ( $\Delta H$ ) y entrópica ( $\Delta S$ ).
- La constante de disociación ( $K_d$ ).
- La energía libre de Gibbs:  $\Delta G = -RT \ln K_a = \Delta H - (T\Delta S)$
- La estequiometría de unión: número de sitios de unión, y si existe cooperatividad en la unión.

**Figura\_13. Esquema de un termograma de ITC.**

**a)** Se muestra un perfil teórico de un termograma endotérmico ( $\Delta H < 0$ ); **b)** se indica la integración teórica del perfil de la izquierda, además de los parámetros termodinámicos que se calculan directamente a partir de la integración como son la entalpía ( $\Delta H$ ) (azul), la estequiometría ( $n$ ) (verde) y la constante de afinidad ( $1/K_d$ ) (rojo).



#### 7.4.2 Estudio de la interacción entre TFAM y distintas secuencias de ADN mediante ITC.

Preparación del ADN: Para el trabajo de esta tesis se hibridaron los oligonucleótidos de cadena sencilla complementarios (Tabla\_5) (Sigma-Aldrich) calentando una mezcla equimolar de cada oligonucleótido durante 15 minutos a  $99^\circ\text{C}$  en un bloque y enfriando lentamente durante toda la noche. Después de la hibridación, se dializó a TFAM y a los ADN de doble cadena usando una membrana Spectr/Por (tamaño de poro 3,500 Da) de modo independiente pero en el mismo recipiente que contenía 1 litro de tampón HEPES pH 7.5 y NaCl 150 mM, previamente desgasificado, renovado cuatro veces durante un total de 36 horas a  $4^\circ\text{C}$ . Así nos aseguramos de que el tampón sin muestra, el tampón del ADN y el de la proteína era exactamente el mismo.

**Tabla\_5. Secuencia de los Oligonucleótidos empleados en ITC.**

En la tabla se indica la secuencia de los oligonucleótidos monocatenarios empleados para la hibridación en el experimento de ITC.

| Nombre      | Secuencia                 |
|-------------|---------------------------|
| Site-X-f    | TAACAAAAAATTTCCACCAAAC    |
| Site-X-r    | GTTTGGTGGAAATTTTTTGTTA    |
| LSP-f       | TAACACTCACCCCCCACTAAC     |
| LSP-r       | GTTAGTTGGGGGGTGAGTGTTA    |
| Site-Y-f    | TAACAATTGAATGTCTGCACAG    |
| Site-Y-r    | CTGTGCAGACATTCAATTGTTA    |
| polydA-f    | AAAAAAAAAAAAAAAAAAAAAAAAA |
| polydA-r    | TTTTTTTTTTTTTTTTTTTTTTTTT |
| polydA-dC-f | AAACAAAAAAAAAACAAAAAAAAA  |
| polydA-dC-r | TTTTTTTGTTTTTTTTTTTGTTT   |

Experimento de ITC: Este trabajo se desarrolló en el “Centre Científics i Tecnològics” de la Universidad de Barcelona en colaboración con Rafel Prohens. Los experimentos de ITC se realizaron a 20 °C en el equipo “VP-ITC microcalorimeter system” (MicroCal) y los datos fueron recogidos con el programa VPViewer2000. TFAM (6-9 µM) fue termoestabilizada a 20 °C en la celda de muestra y titulada posteriormente con el ADN. La titulación se realizó con 30 inyecciones (10 µl durante 20 s) de los oligonucleótidos LSP, Site-Y, poly-dA, poly-dA-dC (Tabla\_5) en un rango de concentración de 40-50 µM para los tres primeros oligonucleótidos y 70-60 µM para el último. Para la secuencia Site-X se tituló la proteína con 15 inyecciones (20 µl durante 30 s) en el mismo rango de concentración que los anteriores (40-50 µM). La concentración de proteína y de ADN fue medida con *Nanodrop*, en el caso de la proteína se corrigió después con el coeficiente de extinción molar experimental (ver apartado 7.3.2 para la estimación de la concentración). Los datos calorimétricos fueron analizados con el programa “*MicroCal Origin*” integrando manualmente los datos crudos. Las curvas isotermas de unión fueron ajustadas mediante un algoritmo iterativo no lineal de mínimos cuadrados a los modelos “Uno y Dos Lugares de Unión”. Todas las medidas fueron replicadas tres veces y promediadas para los cálculos de los parámetros termodinámicos de unión.

## 7.5 Ensayo *in vitro* de actividad helicasa.

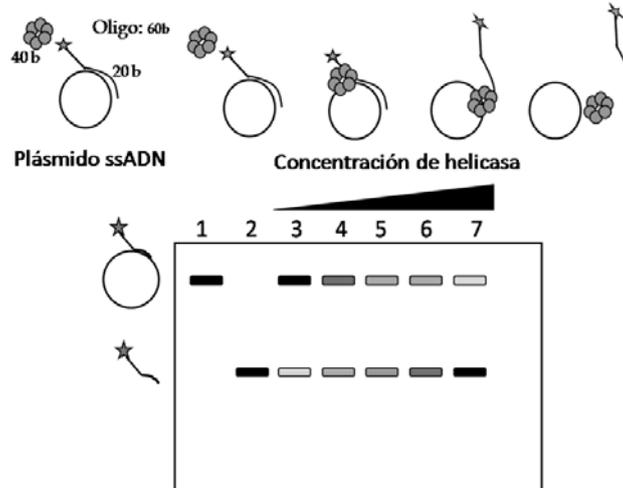
### 7.5.1 Fundamentos.

Una reacción helicasa consiste en el desenrollamiento de la doble cadena de nucleótidos de ADN, ARN o híbridos ADN/ARN, en dos cadenas sencillas, catalizada por una enzima. Esta reacción va acoplada a la hidrólisis de nucleótidos trifosfato (ATP, UTP, GTP, TTP ó CTP) por parte de la enzima helicasa. Una forma de analizar la actividad helicasa se basa en emplear un plásmido monocatenario que es complementado por un oligonucleótido que tiene un extremo

5' monocatenario libre que facilita la incorporación de la helicasa a la reacción de desenrollamiento de la doble cadena; a este sustrato se le añade cantidades crecientes de enzima y el resultado de la reacción, la liberación del oligonucleótido por la actividad separadora de las hebras, se visualiza por electroforesis (Figura\_14).

**Figura\_14. Esquema de un ensayo teórico helicasa.**

Arriba se muestra de izquierda a derecha el esquema de una reacción helicasa. Se indica el sustrato un plásmido con el oligonucleótido hibridado con el extremo libre marcado con  $^{32}\text{P}$ , y con esferas grises la enzima (en este caso Twinkle). Abajo se muestra es el esquema del resultado esperado en el gel de la reacción: carril 1) control de sustrato; 2) sustrato calentado a 95 °C; carriles 3) a 7) cada carril contiene una cantidad fija de sustrato y cantidades crecientes de enzima. Se espera que a cantidades crecientes de enzima el oligo marcado liberado aumente.



**7.5.2 Ensayo de actividad helicasa con Twinkle.**

El ensayo de esta actividad lo realizamos marcando el sustrato de la reacción, un oligonucleótido monocatenario marcado en el extremo 5' con un isótopo de fósforo 32 ( $^{32}\text{P}$ ), el cual se hibridó al plásmido M13(+) también monocatenario. El oligonucleótido era 20 pb complementario a M13(+) pero además tenía en el extremo 5' con 40 nucleótidos sin complementar quedando como cadena sencilla expuesta. Este extremo 5' libre facilita la incorporación de la helicasa a la reacción de desenrollamiento de la doble cadena. El oligonucleótido empleado para la hibridación parcial fue:

**5'ACATGATAAGATACATGGATGAGTTTGGACAAACCACAACGTAAAACGACGGCCAGTGCC-3'**, en **negrita** se indica la base con el isótopo  $^{32}\text{P}$ .

La reacción se realizó a 37 °C durante una hora, a concentraciones crecientes de la enzima Twinkle de 10, 20, 30, 60 y 100 ng frente a una concentración constante de 5  $\mu\text{M}$  de plásmido M13(+). El tampón de reacción consistía en Tris-HCl 25 mM pH 7.5,  $\text{MgCl}_2$  4.5 mM, UTP 3 mM, NaCl 27 mM, glicerol 10%, DTT 1 mM, L-arginina 0.1 M, BSA 0.1mg/ml. Como controles se emplearon: 1) como control negativo el sustrato (ADN) sin enzima (Twinkle), carril 1 figura 14; 2) el sustrato después de desnaturalizarlo térmicamente como control positivo, carril 2 figura

14. Para evitar interacciones inespecíficas se empleó el mismo oligonucleótido libre no marcado radiactivamente, denominado “oligonucleótido frío”. 3) Como control negativo adicional se comprobó la reacción sin UTP, frente a 400 pmol de sustrato marcado radiactivamente. Se realizó una titulación en un rango desde 10 a 100 ng de proteína. La reacción se detuvo por desnaturalización de la proteína añadiendo el siguiente tampón: SDS 6%, EDTA 90 mM, Tris-HCl pH 8.0, glicerol 10%, azul de xileno 0.25% y azul de bromofenol 0.25% (figura\_14). Las muestras se cargaron en un gel nativo de acrilamida al 15% empleando como tampón de electroforesis TBE 1x. El gel se corrió a 80 V durante 60 min a temperatura ambiente. Posteriormente el gel se deshidrató y se reveló empleando una lámina fotosensible para la autografía y se visualizó en un sistema de escáner fosfoiluminador.

## 7.6 Dispersión dinámica de la luz.

### 7.6.1 Fundamentos.

La técnica de dispersión dinámica de la luz (“*Dynamic light Scattering*”, DLS) es una técnica espectrométrica basada en la medida de la dispersión de un haz de luz como consecuencia de la interacción de moléculas en movimiento browniano cuando se encuentran en disolución. La intensidad de la dispersión es proporcional al coeficiente de difusión de la muestra, y la difusión será menor en partículas de mayor tamaño. La medida de la fluctuación de la intensidad con el tiempo de una muestra se compara con una base de datos que correlaciona dichos coeficientes con el tamaño de macromoléculas biológicas, y con ello se deduce el tamaño de la molécula a la vez que se comprueba que no se produzca agregación. El coeficiente de difusión se calcula a partir de la variación del coeficiente de autocorrelación en función del tiempo. Además, el cálculo del porcentaje de luz dispersada frente al radio hidrodinámico permite estimar la polidispersidad de la muestra, y por tanto permite conocer su homogeneidad (a más homogeneidad, menos polidispersión, siendo el máximo de polidispersión considerado homogéneo el de 20%).

Para realizar el experimento se disponen 50  $\mu$ l de la muestra en una cubeta de cuarzo la cual es atravesada por un haz láser, midiéndose la dispersión de la muestra con un detector ubicado a 90° respecto al haz incidente. La concentración mínima de muestra para realizar el experimento depende del tamaño de la partícula.

### 7.6.2 Estudios de DLS con Twinkle.

Se realizaron análisis de la solubilidad de los constructos Twinkle y (GB1)- $\Delta$ 358Twinkle. El constructo con el dominio helicasa (GB1)- $\Delta$ 358Twinkle fue analizado a una concentración comprendida entre 1.2 y 1.8 mg/ml en distintos tampones y pH: fosfato sódico pH 8.5, HEPES pH 7.5, Tris-HCl pH 7.5 y citrato sódico pH 6.0. Todos los tampones contenían una

concentración de NaCl 1 M y  $\beta$ ME 1mM. Por cada tampón se realizaron dos medidas, una a 4 y otra 20 °C de temperatura. El constructo Twinkle se analizó a 1.5mg/ml y con un mismo tampón: citrato sódico 50 mM pH 6.5, NaCl 1 M, L-Arg 50 mM, L-Glu 50 mM, pero con distintos aditivos: glicerol 10%,  $\beta$ ME, MgCl<sub>2</sub> o UTP. También se analizó por DLS una muestra de proteína con glicerol al 10% crioenfriada con nitrógeno líquido para estudiar si durante el proceso de congelación se producía agregación.

Con esta técnica se comprobó que el tampón de purificación empleado cumplía los requisitos necesarios para los ensayos de cristalización: Twinkle es disolución con este tampón presentaba gran homogeneidad (polidispersidad menor al 15%) y no producía agregación a altas concentraciones.

## 7.7 Barrido diferencial de fluorescencia.

### 7.7.1 Fundamentos.

La técnica de Barrido diferencial de fluorescencia (*"Differential Scanning Fluorescence"*, "DSF") se basa en la medida en las variaciones de fluorescencia de un cromóforo originada por el cambio de estado entre plegado y desplegado de las proteínas. Se realiza una desnaturalización térmica de la proteína de interés en presencia de un reactivo fluorescente por ejemplo SyPro Orange (Invitrogen) cuya emisión de fluorescencia incrementa al reaccionar con residuos hidrofóbicos expuestos al disolvente durante la desnaturalización térmica por el desplegamiento de la proteína, que genera la exposición de núcleos hidrofóbicos. A partir de la medida de la fluorescencia *versus* la temperatura se obtiene la curva de desnaturalización. La derivada de esta curva respecto de la temperatura genera una curva con un pico cuyo máximo corresponde a la temperatura de fusión de la proteína, que se define como la temperatura a la cual se produce la transición entre el estado plegado y desplegado. La comparación de la temperatura de fusión de una proteína en distintos tampones y/o ligandos permite identificar las condiciones químicas (a determinado pH, concentración de sal, efectos de ligandos, etc) en las cuales la temperatura de fusión era mayor y por tanto la proteína era más estable.

### 7.7.2 Análisis de estabilidad mediante DSF con Twinkle.

Para realizar este experimento en general se diluyeron 5  $\mu$ l de la proteína concentrada entre 2 y 5 mg/ml en cada pocillo de 200  $\mu$ l de una placa de 96 pocillos. Cada pocillo contenía una mezcla de 25  $\mu$ l de un tampón a concentración 2x (tabla\_6), 7.5  $\mu$ l de SyPro Orange (300x) (Invitrogen) y 13  $\mu$ l de agua esterilizada de pureza milliQ. Para el experimento se realizó un aumento de la temperatura a una velocidad de un 1 °C/min desde temperatura ambiente (20 °C) hasta 95 °C. La reacción se midió en un equipo de PCR en tiempo real (IQ5 Multicolor RealTimePCR Detector system – Bio-Rad) y las curvas de desnaturalización se trataron con el

programa IQ5 Bio-RAD Version 2.0 implementado en el equipo. Como control y referencia de la temperatura de fusión en el tampón de reacción se substituyó el tampón 2x por H<sub>2</sub>O, en el que tanto la proteína como el Sypro Orange se emplearon en la misma concentración.

**Tabla\_6. Lista de tampones empleados en los ensayos de DSF.**

Se enumeran los tampones empleados en los distintos ensayos de DSF y la muestra que se evaluó. Tf, temperatura de fusión.

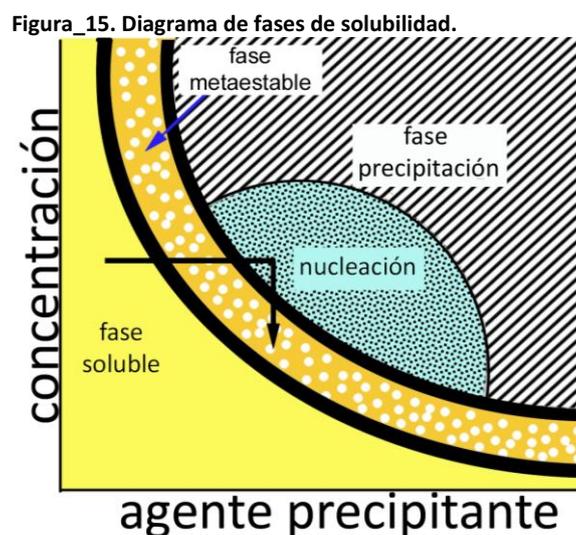
| Tampón           | pH   | Concentración NaCl | Aditivo                | Muestra            | Tf(°C) |
|------------------|------|--------------------|------------------------|--------------------|--------|
| Citrato          | 6.0  | 500                | ----                   | Twinkle            | 53.5   |
| Citrato          | 6.0  | 1000               | ----                   | Twinkle            | 55     |
| Fosfato          | 8.0  | 500                | ----                   | Twinkle            | 52.7   |
| Fosfato          | 8.0  | 1000               | ----                   | Twinkle            | 54     |
| HEPES            | 7.5  | 500                | ----                   | Twinkle            | 53     |
| HEPES            | 7.5  | 1000               | ----                   | Twinkle            | 53.75  |
| Tris-HCl         | 7.5  | 200                | ----                   | Twinkle            | 52.5   |
| Tris-HCl         | 7.5  | 200                | MgCl <sub>2</sub> 5mM  | Twinkle            | 53     |
| Tris-HCl         | 7.5  | 500                | ----                   | Twinkle            | 52.5   |
| Tris-HCl         | 7.5  | 500                | MgCl <sub>2</sub> 5mM  | Twinkle            | 52.5   |
| Tris-HCl         | 7.5  | 1000               | ----                   | Twinkle            | 53     |
| Tris-HCl         | 7.5  | 1000               | MgCl <sub>2</sub> 5mM  | Twinkle            | 54     |
| H <sub>2</sub> O | ---- | ----               | ----                   | Twinkle            | 53     |
| Citrato          | 6.5  | 1000               | ----                   | (GB1)-Δ358-Twinkle | 50     |
| Citrato          | 6.0  | 1000               | ----                   | (GB1)-Δ358-Twinkle | 47     |
| Tris-HCl         | 7.5  | 1000               | ----                   | (GB1)-Δ358-Twinkle | 47     |
| Tris-HCl         | 7.5  | 500                | ----                   | (GB1)-Δ358-Twinkle | 58     |
| Tris-HCl         | 7.5  | 1000               | MgCl <sub>2</sub> 5 mM | (GB1)-Δ358-Twinkle | 47     |
| Tris-HCl         | 8.5  | 1000               | ----                   | (GB1)-Δ358-Twinkle | 52     |
| Fosfato          | 7.0  | 1000               | ----                   | (GB1)-Δ358-Twinkle | 58     |
| Fosfato          | 7.5  | 1000               | ----                   | (GB1)-Δ358-Twinkle | 57     |
| Fosfato          | 8.0  | 1000               | ----                   | (GB1)-Δ358-Twinkle | 54     |
| Fosfato          | 8.5  | 1000               | ----                   | (GB1)-Δ358-Twinkle | 52     |
| H <sub>2</sub> O | ---- | ----               | ----                   | (GB1)-Δ358-Twinkle | 53     |

## 7.8 Cristalografía de proteínas y difracción de rayos X.

### 7.8.1 Principios y métodos de cristalización.

La cristalografía de macromoléculas es una técnica empleada para resolución de la estructura tridimensional de moléculas a resolución atómica, tanto de proteínas, ácidos nucleicos, complejos macromoleculares, etc. La cristalización de macromoléculas requiere de unas condiciones químicas donde se produzca la sobresaturación de las mismas para que precipiten de forma ordenada formando cristales. No obstante, no es predecible en qué condiciones químicas concretas se producirá la cristalización, por lo que se emplean una serie de estrategias para encontrar dichas condiciones. La primera etapa consiste en precipitar las

macromoléculas en disolución para determinar la concentración de macromolécula o de agente precipitante en la que se produce sobresaturación de la primera. Este proceso se ilustra mediante un diagrama de fases (figura\_15), en el cual se representa la fase soluble y la fase precipitada de las moléculas de interés y entre ellas la transición entre ambas fases, la zona metaestable donde se produce la sobresaturación sin precipitación en unas condiciones químicas determinadas. La sobresaturación se puede producir por múltiples factores: aumento de concentración de la proteína, presencia de agentes deshidratantes, variación de pH acercándolo al punto isoeléctrico de la muestra para que ésta pierda solubilidad, variación de temperatura para variar la difusión, etc. Incrementos de la concentración de macromolécula y/o agente precipitante en la región metaestable puede dar lugar a la nucleación cristalina que, mediante la variación sutil de las condiciones químicas o de la concentración puede que derive en cristalización. Las condiciones físico-químicas que definen la región metaestable son específicas para cada macromolécula, y para acotarlas se emplea un sistema de ensayo y error. Se prueban inicialmente una batería de condiciones de kit comerciales (por ejemplo “*crystal screen I*” de Hampton Research) donde se determina la concentración a la que la muestra tiende a precipitar (al menos en un 50% de las condiciones probadas) de forma desordenada. A partir de la concentración de la muestra donde la precipitación es frecuente, se prueban de modo aleatorio centenares de condiciones químicas, a diferentes temperaturas (4 o a 20 °C) para encontrar las condiciones de cristalización. Explorando entorno las condiciones que dan lugar a núcleos cristalinos, (inicialmente con variaciones de pH en 0.5 puntos; de agente precipitante tipo polietilenglicol en un 5%, sales en 50 mM, etc), se puede entonces mejorar el aspecto y calidad de los cristales.



Para cristalizar proteínas se emplean diferentes métodos: difusión de vapor, diálisis, difusión líquida, y “*microbatch*”.

En este trabajo se ha empleado el sistema de difusión de vapor, el cual consiste en provocar una difusión por evaporación del agua de una disolución que contiene la proteína, generándose una deshidratación lenta que da lugar a un estado de sobresaturación de la macromolécula, que precipita. La reacción se realiza dentro de un recipiente hermético, en el cual se dispone una gota de disolución (0.1-2  $\mu$ l) compuesta por proteína más una gota de disolución de cristalización (0.1-2  $\mu$ l) mezcladas entre ellas sin tener contacto con el reservorio donde se dispone un volumen 500 o 1000 veces mayor para provocar entre ambas disoluciones la difusión de vapor de agua, en exceso en la solución que contiene la proteína. Existen dos métodos de cristalización por difusión de vapor en función de la ubicación de la gota dentro del recipiente de cristalización. En un tipo la gota con proteína se dispone en la tapa del recipiente, quedando la gota suspendida frente al reservorio que la va a deshidratar; este sistema se denomina cristalización por “gota colgante”. El otro método emplea un recipiente que contiene una elevación cóncava en la cual se deposita la gota, quedando ésta rodeada por la solución de cristalización. Este método se denomina cristalización por “gota sentada”.

Comercialmente se facilitan dos tamaños de placas para las técnicas de difusión de vapor: placas de 96 pocillos para la dispensación mediante robots automatizados de gotas de un volumen de 50 a 200 nanolitros, frente a un reservorio que contiene entre 100 y 200  $\mu$ l; y placas Limbro de 24 pocillos de mayor tamaño, en las cuales se trabaja con gotas de un volumen entre 0.5-2  $\mu$ l, conteniendo el reservorio entre 0.5 y 1 ml.

### **7.8.2 Cristalización de TFAM en complejo con SITE-X.**

La cristalización del complejo TFAM/Site-X se realizó con la proteína madura TFAM (residuos 43 al 246, sin la secuencia N-terminal de translocación a mitocondria) formando un complejo con un ADN de doble cadena que contiene la secuencia de 22 pares de bases SITE-X, 5'-TAACAAAAATTTCCACCAAAC-3'; 5'-TTTGGTGGAATTTTTGTTAG-3', ambas cadenas complementarias, con los extremos cohesivos, e hibridadas previamente. La hibridación se realizó mezclando cantidades equimolares de ambos oligonucleótidos calculas con *Nanodrop* y calentando 85 °C 20 min y dejando enfriar lentamente toda la noche hasta temperatura ambiente.

Para la cristalización del complejo TFAM/Site-X, éste se preparó con un ratio de proteína:ADN de 2:1, ambas muestras previamente disueltas en tampón de purificación (HEPES pH 7.5, 750 mM NaCl, DTT 5mM). El contenido en NaCl se fue bajando hasta alcanzar 20 mM, de un modo progresivo mediante tres pasos de diálisis de concentración de NaCl 500, 100 y 20 mM. Finalmente el complejo se concentró con un sistema de filtros por centrifugación (Vivaspin-

Sartorius) con un tamaño de poro de 10 KDa hasta una concentración de 13.5 mg/ml de proteína medida mediante el método de Bradford (BioRad).

Los primeros experimentos de cristalización se realizaron en placas de 96 pocillos empleando un sistema de dispensación automático de nanogotas en la Plataforma Automatizada de Cristalografía (PAC) del IBMB empleando robots (“*Cartesian*” de Genomic Solutions y “*Phoenix*” de Rigaku). Los primeros cristales aparecidos en nanogotas se escalaron a volúmenes de 1  $\mu$ l y optimizaron en placas de gota sentada “Limbro” de 24 pocillos. Los cristales se obtuvieron en gotas que contenían 1  $\mu$ l de disolución de complejo proteína-ADN (a 13 mg/ml en 50 mM HEPES, 20 mM de NaCl y 5 mM de DTT a pH 7.5) y 1  $\mu$ l de reservorio (PEG3350 18-24%, tartrato sódico potásico 0.1-0.25 M) a 20 °C. Como crio-protector se preparó la misma disolución de cristalización de la condición a la cual cristalizó y se le añadió PEG400 logrando un 10-20% de PEG400. Los datos de difracción fueron recogidos en la línea de difracción “ID23-2” del “*European Synchrotron Radiation Facility*” (ESRF, Grenoble, Francia).

### **7.8.3 Ensayos de cristalización de Twinkle.**

Los ensayos de cristalización de los múltiples constructos de Twinkle también se realizaron empleando el método de difusión de vapor con gota sentada en placas de 96 pocillos, con nanogotas de 100 nanolitros dispensadas automáticamente por los robots mencionados anteriormente, en la PAC. Los ensayos se realizaron a dos temperaturas, 20 y 4 °C. Las condiciones en ensayos se indican en la tabla\_16 (ver en el apartado 8.2.1.5).

Además de los ensayos indicados en la tabla\_16 el constructo Twinkle también se intentó cristalizar cambiando el pH del tampón sustituyendo el citrato sódico pH 6.5 por HEPES 7.5 y Tris-HCl pH 8.0, todos ellos con 1M NaCl, L-Arg 50mM, L-Glu 50mM y/o empleando  $MgCl_2$  5 mM como aditivo en el tampón. Los kits empleados fueron “Crystal Screen I & II”, “Index”, “Natrix”, “SalTRX”, “A/S Ion Screen”, “PEG6000/LiCl”, “PEG400”, “PEG400/LiCl”, “MDP”, (Hampton Research) en el mismo rango de concentraciones de proteína que se indica en la tabla 16.

También se realizaron los ensayos de cristalización de Twinkle con el método de co-cristalización, en la cual se concentró la proteína (Vivaspin, tamaño de poro 30 KDa) hasta la concentración indicada a continuación y se mezcló con los distintos ligandos ADN y UMPNP en distintos ratios molares, también indicada a continuación. La concentración de proteína se determinó con el *Nanodrop* y teniendo en cuenta el coeficiente de extinción molar teórico del monómero. Para los cálculos molares que se emplearon para la co-cristalización se consideró siempre el peso molecular de 436 KDa correspondiente a un hexámero. Las condiciones ensayadas fueron:

- Co-cristalización Twinkle 3.2 mg/ml (7.6  $\mu$ M) con ADN de cadena sencilla de 29 bases, cadena doble de 20pb, 29pb y a distintos ratios molares de proteína:ADN : 1:1, 1:2, 1:3, 1:6, 2:1.
- Co-cristalización Twinkle-6hisC 3.7 mg/ml (8.8  $\mu$ M) con ADN de cadena sencilla de 29 bases, cadena doble de 20pb, 29pb y a distintos ratios molares de proteína:ADN : 1:1, 1:2, 1:3, 1:6, 2:1.
- Co-cristalización Twinkle 6.6 mg/ml (15.7 $\mu$ M) con UTP y UMPPNP a diferentes ratios proteína: UTP/UMPPNP de 1:3, 1:6
- Co-cristalización de Twinkle 5.0 mg/ml (11.9 $\mu$ M) con combinaciones de lo anterior: UTP y UMPPNP con las diferentes moléculas de ADN, proteína:UTP/UMPPNP:ADN a 1:3:2, 1:6:2, 1:6:3.

Para todos los demás constructos también se realizaron múltiples ensayos, los cuales se detallan en la tabla\_16.

## 7.8.4 Difracción de rayos X de cristales de macromoléculas.

### 7.8.4.1 Fundamentos.

La segunda etapa de la cristalografía de proteínas consiste en la difracción de un haz de rayos X por parte de los cristales obtenidos. Un cristal es una red ordenada, simétrica, periódica y tridimensional cuya unidad mínima que se repite en el espacio se denominada unidad asimétrica. Por operaciones de simetría cristalográfica la unidad asimétrica da lugar a la celdilla unidad, la cual también por las mismas operaciones cristalográficas genera todo el cristal. Gracias a las propiedades estructurales de orden y simetría de los cristales cuando un haz de luz, en este caso de rayos X, incide sobre ellos se produce la difracción coherente de los rayos X. El patrón de difracción es específico para cada cristal y viene definido por la ley de Bragg (ecuación\_3).

$$\text{Ecuación}_3: 2d * \text{sen}\theta = n\lambda$$

En la cual  $\lambda$  es la longitud de onda del haz incidente de rayos X, “d” es la distancia entre los planos de átomos dentro del cristal y  $\theta$  es el ángulo el rayo incidente y reflejado, el ángulo de difracción. Cada reflexión recogida por el detector está originada por todos los rayos X difractados por todos los átomos que constituyen un mismo plano del cristal. Los planos reales (o del espacio real, del cristal) por el efecto de la difracción dan lugar a puntos geométricos ordenados en tres dimensiones, que constituye la red recíproca de puntos (espacio recíproco), que se define mediante tres índices de Miller (hkl) asignados a cada punto. Esta información, en forma de reflexiones hkl de intensidad específica  $I_{hkl}$ , está recogida por el detector durante

la difracción. Según el modelo de Ewald el rayo es difractado cuando se cumple que un punto recíproco incide con una esfera virtual cuyo radio es el inverso de la longitud de onda del rayo. La difracción de cristales de macromoléculas se produce por la interacción de los fotones de los rayos X con los electrones de los átomos que componen la macromolécula. Con la medida de la difracción se puede calcular mapas de densidad electrónica mediante la función de densidad electrónica (ecuación\_4), que no es más que la transformada de Fourier de las reflexiones en espacio recíproco que da lugar a la densidad electrónica en espacio real. Las reflexiones son ondas y matemáticamente se definen como un vector con un módulo y un ángulo. Experimentalmente se miden las intensidades de las reflexiones, a partir de la cual se calcula el módulo, denominado Factor de Estructura  $F$ , mediante la relación  $I \propto F^2$ .

Ecuación\_4. Función de la densidad electrónica:

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F(h, k, l)| \exp[-2\pi i * ((hx + ky + lz) - i\Phi(hkl))]$$

Donde  $\rho$  es la densidad electrónica,  $x, y, z$  son las coordenadas cartesianas en espacio real,  $V$  es el volumen,  $F$  los factores de estructura,  $h, k, l$  son los índices de Miller, son las coordenadas en espacio recíproco,  $i$  es el número imaginario y  $\Phi$  es el ángulo de cada factor de estructura, es decir, la fase.

Para poder completar la ecuación de la transformada de Fourier, es necesario calcular el ángulo de difracción por cada reflexión, que depende de las posiciones atómicas que la difractaron, pero esta información no está contenida en el patrón de difracción (que sólo contiene las intensidades de las reflexiones). Este hecho es el llamado problema de la fase. Para resolver la fase en el caso de macromoléculas existen tres métodos clásicos:

- MIR: Reemplazo isomorfo múltiple. Este método se basa en la propiedad de que las intensidades de las reflexiones difractadas son proporcionales a la masa de la partícula difractada. Esta propiedad es empleada introduciendo, en el cristal de macromoléculas biológicas, sales de átomos pesados los cuales van a coordinarse con la molécula formando una sub-red en el cristal que dará un patrón de difracción superpuesto al de la macromolécula. Por tanto el patrón de difracción resultante (del cristal derivado con sales metales) será diferente al original (nativo). Mediante cálculos específicos esta diferencia entre cada reflexión  $hkl$  del conjunto de datos nativo y derivado permite asignar la posición  $x, y, z$  en el espacio de los átomos pesados y así calcular la fase de las reflexiones correspondientes a la difracción del átomo. A partir de estos valores iniciales se calcula las fases de las reflexiones de difracción de todo el cristal de la macromolécula.

- Dispersión anómala en una sola o en múltiples longitudes de onda (SAD y MAD, respectivamente), medidos en un solo cristal: El principio de este método se basa en la capacidad de absorción de los átomos a ciertas energías del rayo incidente, midiéndose los datos a partir de un solo cristal para minimizar el efecto anisomórfico entre cristales. Un método de introducir átomos pesados en la proteína es mediante la sustitución química de átomos en aminoácidos de las proteínas, comúnmente se realiza la sustitución de azufre de la metionina por selenio. Si se incide el cristal derivado con un haz de rayos X a una energía que corresponde a la frecuencia específica de absorción de los electrones de un átomo, (por ejemplo un átomo de selenio) entonces parte de la radiación incidente es absorbida, y por tanto el átomo pesado dispersa los rayos X con menos electrones de los que le correspondería. Según la ley de Friedel cada par de reflexiones enantioméricas y equivalentes deberían tener la misma amplitud y fase pero con signo opuesto. No obstante debido a este fenómeno de absorción las reflexiones enantioméricas tienen intensidades diferentes, por lo tanto la ley de Friedel no se cumple. La variación de la dispersión teórica ( $f_0$ ) depende del factor dispersivo ( $f'$ ) y del factor anómalo ( $f''$ ) y se cuantifican según la siguiente ecuación:

**Ecuación 5. Dispersión anómala.**

$$F = f_0 + f' + i * f''$$

En el método MAD se realizan recogidas de datos a tres longitudes de onda, una corresponde al máximo de la absorción ( $f''$ ) del átomo pesado presente en el cristal y es donde se produce el mayor efecto anómalo, y otro en el punto de inflexión ( $f'$ ), donde el término dispersivo está en su mínimo. Con esta información ya se puede discriminar entre los dos pares enantioméricos. Un conjunto de datos adicional en la longitud de onda remota a energías por encima del máximo de absorción puede añadir información anómala (aunque esto depende del átomo pesado) y a sobredeterminar la fase calculada con los datos de las otras longitudes de onda, mejorando la calidad de la fase obtenida experimentalmente. La comparación de todos estos datos permite el cálculo de la intensidad y las fases de los dispersantes anómalos. Posteriormente, a partir de estas fases iniciales, se asigna fase a todas las reflexiones y se genera así el mapa de densidad electrónica completo. Con el método SAD solo se utiliza los datos recogidos a la longitud de onda donde la absorción es máxima, y mediante los protocolos de “modificación de densidad” se calcula la fase del dispersante, y si el

enantiómero es correcto, se calculan mapas interpretables que abarcan toda la molécula (Dodson 2003).

- Remplazo molecular: Para emplear este método es necesario poseer la estructura cristalina de una molécula similar que presente un gran parecido estructural. Se basa en emplear esta estructura conocida, a partir de la cual se calculan los factores de estructura y la fase ( $F_{\text{cal}}$   $\phi_{\text{cal}}$ ) para emplearla como referencia para resolver las fases de la estructura problema, de la cual sólo se conocen los factores de estructura ( $F_{\text{obs}}$ ) y no sus fases. Se calcula para ambas moléculas la función de Patterson (ecuación 6), que matemáticamente es el cuadrado de la transformada de Fourier. En consecuencia, se elimina el parámetro de la fase, y sólo se define vectorialmente la distancia interatómica de todos los átomos que hay en la molécula, que matemáticamente es el cuadrado de los módulos de los factores de estructura, es decir dependen directamente y solamente de las intensidades medidas experimentalmente.

**Ecuación 6. Función de Patterson.**

$$P(uvw) = \frac{1}{V} \sum_{hkl}^{+\infty} |F(hkl)|^2 * \cos 2\pi [hu + kv + lw]$$

Con la síntesis de Patterson se obtiene el mapa de Patterson o de distancias interatómicas tanto para la molécula conocida como para los datos experimentales. A partir de aquí se rota en las tres direcciones x,y,z el mapa de Patterson de la proteína referencia para que coincida con el mapa de Patterson problema. Cuando se consigue un buen resultado de rotación se calcula la translación en x,y,z para encontrar la posición de la molécula respecto del origen de coordenadas. Cuando en una orientación determinada el valor del coeficiente de correlación entre ambos mapas destaca sobre los demás picos muy probablemente significará que el modelo de búsqueda inicial está en la posición y orientación correctas. Con el cálculo de las posiciones xyz encontradas se aplica la fase de la referencia ( $\phi_{\text{cal}}$ ) y los factores de estructuras experimentales ( $F_{\text{obs}}$ ) en la ecuación de Fourier, que permiten calcular un mapa de densidad electrónica con las fases del modelo de búsqueda en el cristal “problema”. Los valores de fases experimentales se mejoraran durante la etapa de refinado de la estructura cristalográfica.

Para evaluar la calidad del resultado obtenido se emplea como referencia los valores *Rwork* y *Rfree*. El primero es la medida de la discrepancia entre los valores de  $F_{\text{cal}}$  y  $F_{\text{obs}}$  (ecuación 7) entendiéndose por  $F_{\text{cal}}$  los valores de factores de estructura del modelo de la estructura final y

$F_{obs}$  los factores de estructura experimentales. Este parámetro nos dice el error existente entre el modelo de la estructura y los valores experimentales. No obstante, cálculo de  $R_{work}$  se basa en modificar las fases para minimizar el error entre los datos y la estructura, por lo que existe un sesgo del modelo. Para tener un control de que la influencia del modelo no es significativa, se emplea el cálculo del  $R_{free}$ , es el mismo cálculo pero realizado con un pequeño grupo de reflexiones (5-10% de las reflexiones totales, o hasta aproximadamente 500 reflexiones) que no se emplean en el cálculo de las fases. Los valores de  $R_{work}$  y  $R_{free}$  no deben discrepar más de un 5%.

#### Ecuación 7. $R_{work}$

$$R = \frac{\sum_{hkl} ||F_{obs}| - |F_{cal}||}{\sum_{hkl} |F_{obs}|}$$

Reemplazo Molecular fue el método que se empleó en este trabajo para la resolución estructural de la proteína TFAM en complejo con el ADN de doble cadena Site-X (ver más adelante).

#### **7.8.4.2 Procesado, escalado de los datos de difracción y resolución de la estructura cristalográfica del complejo TFAM/Site-X.**

Para la estructura del complejo TFAM/Site-X los datos de difracción fueron integrados y escalados empleando XDS (Kabsch 2010; Kabsch 2010) y SCALA (Evans 2006) perteneciente a la plataforma de programas CCP4 (Winn, Ballard et al. 2011). Para resolver el problema de la fase se empleó el método de remplazo molecular (RM) con el programa PHASER (McCoy 2007; McCoy, Grosse-Kunstleve et al. 2007), también implementado en CCP4, y PHENIX (Adams, Afonine et al. 2010). Como molécula de referencia se empleó el complejo proteína/ADN: TFAM/LSP (código PDB: 3TQ6) (Rubio-Cosials, Sidow et al. 2011). Las etapas de refinado se realizaron automáticamente con PHENIX (Adams, Afonine et al. 2010) y BUSTER (Branden and Jones 1990; Vagin and Isupov 2001; Blanc, Roversi et al. 2004), incluyendo refinado anisotrópico mediante TLS, así como simetría no cristalográfica entre complejos (4 en total) de la unidad asimétrica (u.a.) e imponiendo emparejamiento “Watson y Crick” de la doble cadena de ADN. El refinado automático de la estructura se alternó con la construcción, modelado e inspección visual del modelo realizada manualmente con COOT (Emsley and Cowtan 2004).

## **7.9 Dispersión de bajo ángulo de rayos X.**

### **7.9.1 Fundamentos.**

La dispersión de bajo ángulo de rayos X (“*Small Angle X-ray Scattering*”, SAXS) es una técnica que mide la dispersión elástica de los rayos X al incidir con macromoléculas en disolución. Los

datos de dispersión que se utilizan en un experimento de SAXS se obtienen al sustraer los datos correspondientes al tampón de dispersión de la muestra, así se obtiene la dispersión correspondiente a la macromolécula. Los datos de dispersión experimentales de SAXS corresponde a la variación de la intensidad de la dispersión en función del *momentum transfer*, que es la diferencia vectorial de la dirección de las ondas incidentes respecto a las dispersadas, aunque se representa gráficamente el logaritmo de las intensidades frente el *momentum transfer* ( $S$ ), que matemáticamente se define con la ecuación 8:

**Ecuación\_8. Momentum transfer.**

$$|S| = \frac{4\pi \sin \theta}{\lambda}$$

Siendo  $2\theta$  el ángulo de dispersión y  $\lambda$  la longitud de onda de los rayos X. El patrón de dispersión del haz de rayos X depende del tamaño y la forma de las partículas incididas, por tanto nos va a permitir obtener datos del tamaño y la forma de las macromoléculas en solución, idealmente con una resolución máxima aproximada de 10 Å.

Para obtener la curva final de un experimento de SAXS, se realizan varias series medidas a concentraciones crecientes (Tabla\_7) debido a que la región de baja resolución de las curvas de SAXS, que corresponde al inicio de la curva, se define mejor a bajas concentraciones de proteína, como consecuencias a que los fenómenos de interacciones entre partículas son mínimos. En concreto, al principio de la curva si los valores de  $\log(I)$  en una representación  $\log(I)$  vs  $S$  tienden a infinito es indicativo de agregación en la muestra. Cuando eso sucede no es posible definir el principio de la curva y por tanto se puede emplear para ningún tipo de cálculo. Al contrario sucede con la región final de curva en la cual la relación señal/ruido es mejor cuando la concentración de la proteína es mayor.

A partir de un experimento de SAXS se puede calcular:

- El parámetro de Guinier, este parámetro determina la calidad experimental de una curva de SAXS. El mínimo de valor del Guinier para aceptar una curva es cuando se logra un ajuste lineal superior al 80% con un mínimo de 30 puntos al inicio de la curva (a baja resolución). Los puntos anteriores a los empleados a este ajuste son eliminados.
- Radio de giro (Rg). El radio de giro se obtiene partir de la representación de Guinier, que consiste en la representación de  $\ln(I(S)) / S^2$ , donde  $I$  es la intensidad medida experimentalmente, y  $S$  el *momentum transfer*. En esta función lineal, la pendiente de la recta es el radio de giro, siempre y cuando el ángulo sea pequeño, con Rg menor a 1.3.

- Peso molecular ( $Pm$ ). Uno de los métodos más extendidos consiste en emplear la relación entre las intensidades ( $I(0)$ ) de la muestra en el origen (de  $S$ ) y de un estándar conocido, en este caso la proteína BSA. Se mide experimentalmente la intensidad en el origen de la BSA,  $I(0)$ , y de ahí se relaciona proporcionalmente con la  $I(0)$  experimental de la muestra problema mediante la siguiente ecuación 9:

**Ecuación\_9: Cálculo del peso molecular.**

$$Pm = \left( \frac{I(0)_{\text{experimental}}}{\text{concentración}_{\text{experimental}}} \right) * \left( \frac{Pm \text{ BSA}}{\frac{I(0)_{\text{BSA}}}{\text{concentración}_{\text{BSA}}}} \right)$$

- Función de la distribución de la distancia  $\rho(r)$ , la representación de la  $\rho(r)$  frente al radio de las partículas nos da una distribución de las distancias máximas interatómicas de todas las partículas en disolución. La  $\rho(r)$  indica la frecuencia con que se presentan las distancias interatómicas de un mismo valor.
- Distancia máxima. El valor de la distancia máxima se obtiene a partir de la representación de la  $\rho(r)$  frente al radio de giro de las partículas. El valor de distancia máxima corresponde al mayor valor de distancia intramolecular que se obtiene en esta representación.
- Determinar el estado plegado, desplegado o parcialmente desordenado de la macromolécula mediante la representación de Kratky,  $I(S)*S^2 / S$ , donde  $I$  es la intensidad medida experimentalmente y  $S$  es el *momentum transfer*. Esta ecuación proporciona información sobre el desorden estructural de las partículas, en este caso las proteínas. Cuando la proteína se encuentra en un estado plegado el perfil de la curva es del tipo campana. En cambio, cuando la proteína está totalmente desplegada el valor de  $I(S)*S^2$  crece a valores mayores de  $S$ , aunque en algunos casos alcanza un *plateau*. Si la proteína está en un estado parcialmente desordenado se observa una curva mezcla de las dos anteriores, con un perfil similar a una campana pero sin que  $I(S)*S^2$  vuelva al valor 0 frente valores de  $S$  crecientes, sino que se mantiene constante en un valor intermedio que no es máximo.
- Validar modelos tridimensionales de estructuras cristalográficas, RMN o modelos computacionales generados por predicción por homología de secuencia. A partir de cualquier modelo tridimensional atómico es posible calcular una curva teórica de SAXS tipo  $\log(I)$  vs  $S$ , y compararla con la curva experimental para así determinar si la estructura tridimensional es similar a la experimental medida por

SAXS. Este tipo de cálculos se realiza con programas como CRY SOL (Svergun D.I. 1995). El grado de divergencia entre la curva teórica del modelo y la curva experimental de SAXS se evalúa mediante el parámetro  $\chi^2$  que tiene que ser el mínimo posible y que se define mediante la siguiente ecuación:

**Ecuación\_10. Ecuación del cálculo de  $\chi^2$ .**

$$\chi^2 = \frac{1}{K-1} \sum_{j=1}^K \left[ \frac{\mu I_{teo}(s_j) - I_{exp}(s_j)}{\sigma(s_j)} \right]^2$$

Donde  $j$  es cada punto de un total de  $K$  puntos de la curva con una intensidad  $I_j$  experimental ( $I_{exp}$ ) o teórica ( $I_{teo}$ ), y  $\mu$  es un factor de escala entre ambas;  $\sigma$  es el error de la medida experimental. Cuando la diferencia entre ambas curvas es menor, el valor de  $\chi^2$  también lo es, por tanto un buen ajuste se cuantifica por el menor valor de  $\chi^2$ . Un ajuste es razonable cuando el valor de  $\chi^2$  está por debajo de 1.5, y cuanto menor sea mejor será el ajuste.

- Obtener modelos *Ab Initio* de los volúmenes de las macromoléculas. A partir de los datos monodimensionales de SAXS se puede generar un modelo tridimensional mediante diferentes aproximaciones matemáticas, la más generalizada consiste en generar de modo automático y aleatorio un conjunto de esferas que describa la forma de partícula, cuya curva teórica se compara a la experimental; mediante reajustes del modelo se optimiza el encaje de la curva experimental con la teórica.

### 7.9.2 Estudios de SAXS con Twinkle.

Con el objetivo de conocer la estructura de Twinkle en disolución se realizó un estudio mediante la técnica de SAXS. Se realizaron diferentes series de medidas a distintas concentraciones de proteína. Además de las variaciones en la concentración de proteína, se analizó la muestra con distintos tampones listados en la tabla 7. Las series Tw\_A1\_1M\_NaCl, Tw\_A2\_1M\_NaCl, tw\_C\_1M y tw\_C2\_1M corresponden a muestras contenidas en el mismo tampón de purificación, a las cuales se les añadió 10% de glicerol, necesario para su crio-conservación en nitrógeno líquido durante el transporte previo al experimento. También se analizó la muestra sin crio-congelar y por tanto estas muestras no contienen glicerol. Estas series se denominaron Tw\_1MNaCl\_no\_Gly y Tw\_1MNaCl+UMP\_no\_Gly. Además se analizó el efecto de la concentración de NaCl en el tampón y qué efectos podían presentar los ligandos UTP y UMPNNP en las series Tw\_1MNaCl+UMP\_no\_Gly\_1 y tw\_C\_1M\_NaCl UMP\_1. En la tabla\_7 se indica la concentración, la variación del tampón y/o ligandos añadidos. La diferencia entre las series Tw\_A1\_1M\_NaCl y Tw\_A2\_1M\_NaCl es el tampón empleado para la substracción, en la Tw\_A1\_1M\_NaCl se empleó el tampón recogido del concentrador tras

concentrar la muestra, y en la Tw\_A2\_1M\_NaCl el tampón es el de la última etapa de purificación pero con 10% glicerol. Las medidas de las series denominadas como tw\_C\_1M, tw\_C2\_1M, tw\_C\_0.3M\_NaCl\_1 y tw\_C\_1M\_NaCl UMP\_1 corresponden a Twinkle expresada mediante células de insecto SF9.

Finalmente, se obtuvieron curvas de SAXS de mejor calidad al fusionar distintas secciones de curvas a distintas concentración de proteínas pero en condiciones similares de tampón, de tal manera la región de baja resolución se definió con la curva de menor concentración y la parte final de la curva con aquella de mayor concentración.

**Tabla\_7. Muestras de Twinkle analizadas por SAXS.**

Se indica el nombre de la muestra medida, la concentración de Twinkle y el tampón. GF: es el tampón empleado en la etapa cromatográfica de gel por exclusión de tamaño, más 10% glicerol. Ver texto para más detalles. La concentración de proteína se estimó empleando un equipo *Nanodrop* y corregido considerando el coeficiente de extinción molar teórico.

| Nombre                 | Concentración (mg/ml) | Tampón/aditivo               |
|------------------------|-----------------------|------------------------------|
| Tw_A1_1M_NaCl_1        | 1.88                  | GF                           |
| Tw_A1_1M_NaCl_2        | 1.40                  | GF                           |
| Tw_A1_1M_NaCl_3        | 0.85                  | GF                           |
| Tw_A1_1M_NaCl_4        | 0.57                  | GF                           |
| Tw_A1_1M_NaCl_5        | 0.34                  | GF                           |
| Tw_A2_1M_NaCl_1        | 1.88                  | GF                           |
| Tw_A2_1M_NaCl_2        | 1.40                  | GF                           |
| Tw_A2_1M_NaCl_3        | 0.85                  | GF                           |
| Tw_A2_1M_NaCl_4        | 0.57                  | GF                           |
| Tw_A2_1M_NaCl_5        | 0.34                  | GF                           |
| Tw_B1_1M_NaCl_1        | 0.9                   | GF                           |
| Tw_B1_1M_NaCl_2        | 0.6                   | GF                           |
| Tw_B1_1M_NaCl_3        | 0.3                   | GF                           |
| Tw_B1_05M_NaCl_2       | 0.6                   | GF a 0.5M NaCl               |
| Tw_B1_05M_NaCl_1       | 0.9                   | GF a 0.5M NaCl               |
| Tw_B1_03M_NaCl_1       | 0.6                   | GF a 0.3M NaCl               |
| Tw_B1_03M_NaCl_2       | 0.3                   | GF a 0.3M NaCl               |
| Tw_1MNaCl_no_Gly_1     | 1.5                   | GF sin glicerol              |
| Tw_1MNaCl_no_Gly_2     | 1.1                   | GF sin glicerol              |
| Tw_1MNaCl_no_Gly_3     | 0.3                   | GF sin glicerol              |
| Tw_1MNaCl_no_Gly_4     | 0.01                  | GF sin glicerol              |
| Tw_1MNaCl+UMP_no_Gly_1 | 1.5                   | GF sin glicerol + UMPPNP 2mM |
| Tw_1MNaCl+UMP_no_Gly_2 | 1.1                   | GF sin glicerol + UMPPNP 2mM |
| Tw_1MNaCl+UMP_no_Gly_3 | 0.3                   | GF sin glicerol + UMPPNP 2mM |
| Tw_1MNaCl+UMP_no_Gly_4 | 0.01                  | GF sin glicerol + UMPPNP 2mM |
| tw_C2_1M_NaCl_1        | 1.27                  | GF                           |
| tw_C2_1M_NaCl_2        | 0.9                   | GF                           |
| tw_C2_1M_NaCl_3        | 0.6                   | GF                           |
| tw_C2_1M_NaCl_4        | 0.3                   | GF                           |
| tw_C_0.3M_NaCl_1       | 0.6                   | GF a 0.3MNaCl                |
| tw_C_1M_NaCl UMP_1     | 0.9                   | GF + UMPPNP 2mM              |
| tw_C_1M_NaCl_1         | 1.5                   | GF                           |
| tw_C_1M_NaCl_2         | 0.6                   | GF                           |
| tw_C_1M_NaCl_3         | 0.3                   | GF                           |
| tw_C_1M_NaCl_4         | 0.13                  | GF                           |

Las recogidas de datos se realizaron en la línea ID14-EH3 del sincrotrón ESRF (*“European Synchrotron Radiation Facility”* (ESRF) Grenoble, Francia). Previamente a la medida, las muestras se centrifugaron en centrífuga de mesa a 4 °C, 16000 rpm durante 20 min, y el sobrenadante filtrado con un tamaño de poro de 0.1µm (Ultrapor de Minipore) para eliminar precipitados y agregados de gran tamaño que pueden interferir con las medidas de la muestra. Todas las medidas fueron realizadas a 20 °C.

El procesado de substracción de la curva del tampón respecto a la curva de la muestra Twinkle se realizó de modo manual con el programa PRIMUS o de modo automático, empleando el mismo programa. Con el mismo programa se realizó el cálculo del parámetro Guinier con el cual se estimó la calidad de cada curva.

Para poder explicar la curva de SAXS con un modelo, ante la falta de una estructura cristalográfica se decidió generar una estructura tridimensional *in silico*. Tal y como se explica en la sección 7.10, se empleó el programa MODELLER (Eswar, Webb et al. 2006) para crear un modelo tridimensional de predicción por homología (ver apartado 7.10 para la realización del modelo). Inicialmente se generó un monómero a partir del que se crearon dos oligómeros, hexámero y heptámero, basados en la homología con la helicasa del bacteriófago T7. Posteriormente, se generó otro hexámero, a partir del mismo monómero, pero esta vez ajustado a un mapa generado por microscopía electrónica obtenido en colaboración con M. Lázaro y el Dr. M. Valle (CICbiogune, Derio (Bizkaia)).

Para lograr un valor de  $\chi^2$  aceptable durante el ajuste de la curva teórica con la experimental, fue necesario emplear el programa *“Ensemble Optimization Method”* (EOM) el cual está compuesto a su vez por dos programas independientes RANCH90 y GAJOE (Bernado, Mylonas et al. 2007). El primero permite dotar de flexibilidad tanto a toda la proteína, a una región específica o a los conectores entre dominios plegados. En las zonas de la estructura definidas como flexibles, este programa puede calcular múltiples conformaciones de la cadena polipeptídica según los parámetros que definen un enlace peptídico, y posteriormente el mismo programa selecciona aquéllas que cumplen con los criterios de los ángulos dihédricos que componen los enlaces entre carbonos- $\alpha$ . En general se calculan entre 5000 y 10000 modelos en los que la zona definida como flexible tiene conformaciones diferentes, con lo que el posicionamiento entre dominios varía respecto al de la estructura original.

Para la selección de los modelos que definen mejor a la curva experimental se empleó el otro programa incluido en EOM, GAJOE (Bernado, Mylonas et al. 2007), que incluye un algoritmo genético que permite definir no sólo uno sino un conjunto de modelos, dividido a su vez por distintos perfiles de dispersión correspondientes a distintos conformeros. Este es el caso de Twinkle, para conseguir el valor mínimo de  $\chi^2$  se definieron tres dominios plegados conectados

por dos conectores flexibles, además de unas colas N- y C-terminales también flexibles. Funcionalmente y enunciándolos del extremo N- a C-terminal, los dominios y subdominios de Twinkle se denominan “*Zinc Binding Domain*” (ZBD), “*RNA Polymerase Domain*” (RPD) y dominio helicasa, siendo éste último el responsable de la oligomerización de la proteína. El rango de residuos que define cada región se define en la siguiente tabla (tabla\_8):

**Tabla\_8. Delimitación de los residuos de las regiones flexibles de Twinkle.**

| Región          | Numeración de los residuos en el modelo | Numeración de los residuos en la proteína | Característica del dominio |
|-----------------|---|---|----------------------------|
| Cola N-terminal | 1-25                                    | 17-42                                     | Flexible                   |
| ZBD             | 26-132                                  | 43-149                                    | Globular/rígido            |
| Conector 1      | 133-137                                 | 150-154                                   | Flexible                   |
| RPD             | 138-336                                 | 155-353                                   | Globular/rígido            |
| Conector 2      | 337-341                                 | 354-358                                   | Flexible                   |
| Helicasa        | 342-614                                 | 359-631                                   | Globular/rígido            |
| Cola C-terminal | 615-667                                 | 632-684                                   | Flexible                   |

A partir de esta plantilla de clasificación de los residuos por zonas se generaron 5000 modelos diferentes mediante EOM y sus correspondientes curvas teóricas. Posteriormente, se compararon con la curva experimental y se seleccionaron un conjunto por el mínimo número de curvas teóricas con GAJOE, y por tanto modelos, que definieron entre todas una curva teórica promedio que encaja razonablemente bien con la curva experimental.

## 7.10 Construcción de proteínas por homología: MODELLER.

### 7.10.1 Fundamentos.

La predicción de estructuras tridimensionales de proteínas se basa en uno de los conceptos básicos de la biología estructural: “la estructura tridimensional de una proteína está determinada por su secuencia primaria” (Zhang 2008), y como consecuencia, dos proteínas que poseen una gran homología de secuencia podrían presentar una gran similitud estructural. Bajo este principio es posible generar de modo computacional un modelo de estructura tridimensional de una proteína a partir de su secuencia primaria siempre que se conozca previamente la estructura tridimensional de referencia de al menos una proteína con gran homología de secuencia. El grado de fidelidad del modelo está determinado por distintos factores: el porcentaje de identidad entre secuencias de aminoácidos, ya que proteínas con una identidad de secuencia del 50% o mayor permiten generar modelos de una fidelidad suficiente para comparar la interacción de ligandos; en cambio, si la identidad es menor al 30% el grado de fidelidad del modelo es sólo a nivel de plegamiento. Si además las proteínas poseen múltiples subunidades o estados oligoméricos, la fidelidad del modelo asimismo disminuye (Eswar, Webb et al. 2006).

Existen muchos servidores con distintas características para construir modelos tridimensionales a partir de la secuencia primaria. De todos ellos se eligió el programa MODELLER (Eswar, Webb et al. 2006) porque permite una gran versatilidad e interacción por parte del usuario. Para este trabajo, una de las grandes ventajas del programa radicaba en la opción de combinar diferentes estructuras cristalográficas simultáneamente para construir el modelo final. El proceso de construir un modelo empleando el programa MODELLER se divide en cuatro etapas: la primera es identificar proteínas relacionadas con la proteína de interés para las que exista un modelo tridimensional; después, se realiza el alineamiento de secuencias entre la proteína diana y las proteínas referencia; la tercera etapa consiste en construir el modelo asignando la secuencia de la proteína diana al modelo referencia; finalmente se trata de evaluar los parámetros estructurales del modelo final mediante programas de validación de estructura tales como por ejemplo VERIFY3D (Luthy, Bowie et al. 1992), PROSAIL (Sippl 1993; Wiederstein and Sippl 2007) entre otros.

### **7.10.2 Modelo tridimensional de Twinkle mediante predicción de estructura.**

Fueron empleados el servidor BLAST (Altschul, Gish et al. 1990) y SCOP (Orengo, Jones et al. 1994) (Holm and Sander 1996; Holm and Sander 1996) para encontrar aquellas proteínas de mayor homología de secuencia y, de éstas, se seleccionaron aquellas cuyas estructuras tridimensionales eran conocidas. Twinkle es una proteína multidominio, y el mejor resultado se encontró buscando independiente los dominios helicasa y primasa.

Para el dominio helicasa o C-terminal la mayor similitud se encontró con la helicasa gp4 del fago T7, para la que existen varias estructuras cristalográficas. La helicasa gp4 tiene dos dominios, unos primasa y otro helicasa, y además presenta los estados oligoméricos hexámero y heptámero, como se indicó en la Introducción. Esta proteína ha sido cristalizada por fragmentos; una de las estructuras que se compone estrictamente por el dominio C-terminal cristalizó como hexámero (código PDB 1EOJ). También se cristalizó de modo independiente el dominio N-terminal incluyendo ZBD y RPD (código pdb 1NUI). Además, hay otra estructura que forma heptámero que comprende la totalidad del dominio C-terminal y la mitad del RPD (1Q57).

Para el dominio N-terminal de Twinkle se encontró que la primasa de *E. coli* DnaG (código pdb 2AU3) también podía ser empleada como referencia. DnaG está constituida por tres subdominios: ZBD y RPD (ambos necesarios para la actividad primasa) y otro extra compuesto por tres hélices- $\alpha$  cuya función es interaccionar con la helicasa DnaB (por tanto prescindibles cuando se compara con Twinkle). La identidad de secuencia del dominio N-terminal de Twinkle frente a DnaG y frente gp4 es de 19 y 18% respectivamente. No obstante, si se compara la

estructura secundaria de estas proteínas se observa que DnaG tiene unas inserciones presentes en Twinkle (según su predicción de estructura secundaria) ausentes en gp4.

Se realizaron dos alineamientos, el primero con Twinkle, gp4 heptamérica (1Q57) y DnaG (2AU3), y el segundo con Twinkle, gp4 heptamérica (1Q57) y parte N-terminal de gp4 (1NUI). Ambos alineamientos inicialmente se realizaron teniendo en cuenta la secuencia primaria, y posteriormente se corrigió tomando como referencia los elementos de estructura secundaria predichos para Twinkle. De esta manera, se alinearon los elementos de estructura secundaria dejando “gaps” en las regiones en las cuales en las estructuras correspondían a “loops” (ver alineamiento figura\_31). Con MODELLER se pudo imponer la estructura secundaria predicha para Twinkle al modelo final. Los modelos finales fueron dos monómeros diferentes: uno basado en el dominio primasa de gp4 cristalizado independientemente (código pdb 1NUI), que incluye el subdominio ZBD y RPD, fusionado con la estructura de gp4 que contiene el subdominio RPD y el dominio helicasa (código pdb 1Q57). El otro modelo para el monómero se generó a partir de una quimera compuesta por la estructura de la proteína DnaG (código PDB 2AU3), que incluye los subdominios ZBD y RPD, y la estructura gp4 1Q57, que se compone del dominio helicasa y el subdominio RPD. En las dos combinaciones el subdominio RPD se obtuvo por la fusión de dos estructuras distintas. Para generar el heptámero se empleó el monómero creado con la proteína quimera DnaG/gp4 (2AU3/1Q57), que se “multimerizó” en base a la estructura 1Q57. Para el hexámero se empleó el mismo monómero pero como referencia de oligomerización se usó la estructura 1EOJ de gp4. Finalmente, se generó otro modelo hexamérico basado en los resultados de microscopía electrónica, como se menciona a continuación.

## **7.11 Microscopía Electrónica con Twinkle.**

### **7.11.1 Fundamentos.**

Este trabajo fue realizado por M. Lázaro y el Dr. M. Valle del CIC-bioGUNE (Derio, Bizkaia) con quien se estableció una colaboración. A continuación se describe los fundamentos de esta técnica relacionados con nuestro proyecto, en concreto el análisis por microscopía electrónica de Twinkle.

La microscopía electrónica permite adquirir imágenes de macromoléculas o macrocomplejos biológicos a resolución atómica. Los principios de la microscopía electrónica son los mismos que en la microscopía óptica: con el uso de lentes se consigue una imagen aumentada de un objeto cuando un haz de luz incide con el mismo. La principal diferencia es la sustitución de la luz visible por un haz de electrones, permitiendo alcanzar resolución atómica. Los electrones al incidir con la muestra dan lugar a dos tipos de interacciones: elásticas e inelásticas. Las

primeras son consecuencia de la interacción con el campo electrostático de la muestra y responsable del contraste de la imagen; las interacciones inelásticas son consecuencia de la interacción de los electrones del haz con los núcleos atómicos y producen daño por radiación en la muestra.

Los electrones interaccionan con aire y por ello es necesario, con los microscopios electrónicos, trabajar en vacío. Al someter las muestras al vacío perderían la capa de hidratación distorsionando gravemente la muestra. Por tanto es necesario fijar la muestra para conservar la capa de hidratación, y para ello se aplican principalmente dos técnicas: tinción negativa y crio-microscopía. La tinción negativa tiene una mayor limitación de resolución, no obstante la preparación de la muestra resulta más sencilla por lo que actualmente se emplea para estudios iniciales de la muestra. Esta tinción se basa en generar una imagen en negativo de la muestra mediante el recubrimiento de la misma con una capa fina de metales pesados (uranio, molibdeno o wolframio). El modo de conseguirlo es añadir una disolución de sales del metal pesado a la muestra previamente dispuesta en una rejilla metálica. De esta manera se deshidrata y se fija la muestra. Con los electrones se obtiene una imagen del contraste entre el átomo pesado y su ausencia, siendo esta ausencia o vacío en realidad el lugar ocupado por las partículas a analizar. La máxima resolución que se puede alcanzar es 20 Å, aproximadamente. La crio-microscopía permite obtener imágenes de alta resolución de macromoléculas. Para lograr la mejora en la resolución se trabaja a menor temperatura: 100 K (298 K para la tinción negativa). Para preservar la muestra del vacío se fija con azúcares (glucosa, sucrosa o sacarosa) o hielo vítreo, y además para evitar la formación de cristales de agua y preservar correctamente la muestra se realiza una ultra congelación en etano líquido. Para la recogida de imágenes hay que considerar que estas muestras presentan una menor protección al daño por radiación.

Cuando se recogen todas las imágenes de una partícula individual, antes de abordar la reconstrucción tridimensional deben ser procesadas bidimensionalmente para lograr imágenes con una buena señal-ruido. Las imágenes recogidas varían ya sea por contener orientaciones diferentes de la partícula, y/o conformaciones distintas de la partícula o simplemente diferencias por distorsiones provocadas por daño en la manipulación de la muestra debido a fijación, radiación, etc. Por tanto es necesario agrupar y clasificar las imágenes de cada partícula por orientaciones equivalentes de la misma. Primero se realiza una selección de las partículas en base a su morfología, en el caso que existan más de una forma estructural de la partícula. Cada tipo de morfología se alinea posteriormente mediante traslaciones y rotaciones, para lograr que todas las imágenes equivalentes bidimensionales estén en la misma orientación y origen. Cuando todas las imágenes presentan la misma orientación

respecto a un punto central de coordenadas se clasifican por distintas poblaciones o clase. Finalmente cada clase, que está compuesta por imágenes equivalentes, se promedia de modo independiente. Estas clases son las que se emplearán para la reconstrucción tridimensional.

La construcción del modelo tridimensional es un proceso iterativo en el cual hay que encontrar los ángulos reales que relacionan las imágenes bidimensionales entre sí. Los dos métodos existentes para realizar las construcciones se denominan “refinamiento angular” y “inclinación cónica al azar”. La primera de ellas es la más extendida cuando se dispone de una gran variedad de orientaciones de la partícula. En esta modalidad, se puede partir de un modelo de referencia o aplicando el método de “líneas comunes”, que se utiliza cuando la partícula presenta simetría, pudiéndose relacionar las imágenes a través de los elementos de simetría, como ejes o planos de simetría, y calcular así los ángulos que relacionan las imágenes; este último es el método que se empleó para la construcción de Twinkle. El otro método, “inclinación cónica al azar”, consiste en tomar dos micrografías de la misma área, una sin inclinar y otra inclinada (0-60°). Si se consiguen suficientes imágenes inclinadas se puede reconstruir el espacio tridimensional calculando el ángulo acimutal entre estas imágenes y el plano XY en el cual están las imágenes no inclinadas. Este otro método suele emplearse cuando las partículas presentan una orientación preferente.



**RESULTADOS**  
**Y**  
**DISCUSIÓN**



**RESULTADOS Y DISCUSIÓN  
DE TFAM EN COMPLEJO  
CON SITE-X**



## 8 RESULTADOS Y DISCUSIÓN.

### 8.1 Resultado y discusión de TFAM en complejo con Site-X.

#### 8.1.1 Resultados obtenidos con el complejo TFAM/Site-X.

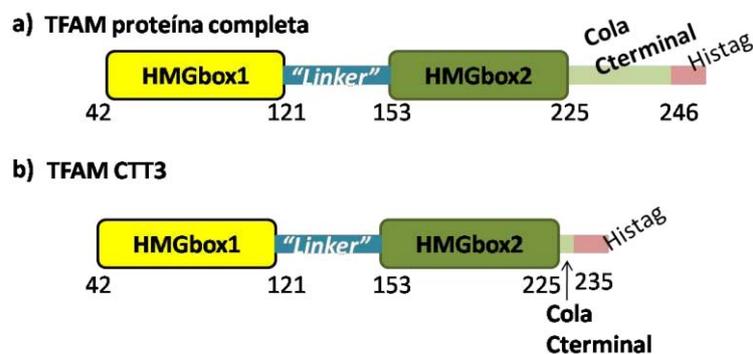
##### 8.1.1.1 Clonación, purificación y formación del complejo TFAM/Site-X.

La clonación y el protocolo de purificación están descritos con detalle en la tesis: “Análisis bioquímico y estructural del factor transcripción mitocondrial humano A, TFAM en complejo con la secuencia del promotor LSP”, porque fue parte del trabajo de tesis de la Dra. Rubio-Cosials. Por esa causa este trabajo no se detalla el protocolo, sólo se indican los clones empleados (figura\_16) y el resultado final de la purificación (figura\_17).

De los clones disponibles, para este trabajo se utilizaron dos, el primero fue TFAM (residuo 43 al 246) que comprende la proteína madura, sin la secuencia de traslocación a mitocondria, y el segundo fue TFAM-CTT3 (residuo 43 al 234) en el cual se eliminó parte de la cola C-terminal de la proteína. Ambos constructos fueron clonados con una cola de histidinas C-terminal (figura\_16).

**Figura\_16. Clones de TFAM.**

Se indica con un esquema de bloques los dominios de TFAM HMGbox1 (amarillo), el “linker” (azul), HMGbox2 (verde), la cola C-terminal (verde claro) y las seis histidinas C-terminales (rosa). En **a)** se muestra la proteína TFAM madura y en **b)** el clon con la cola C-terminal más corta, TFAM-CTT3.



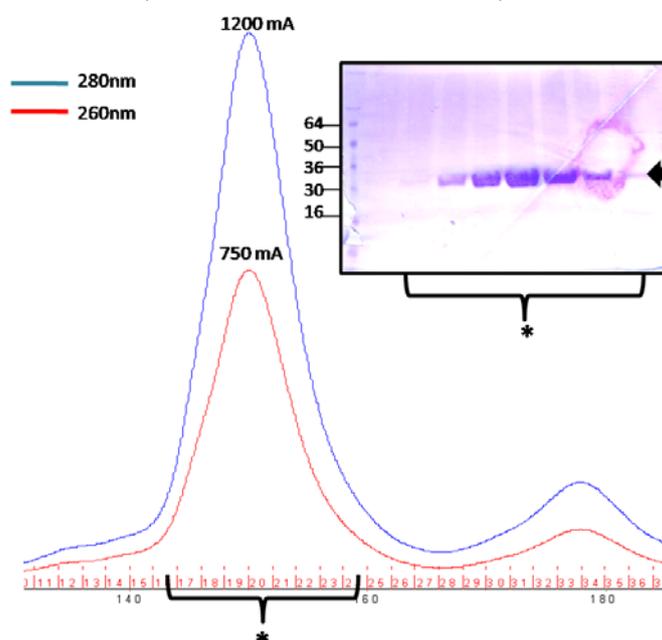
Para la purificación de TFAM y TFAM-CTT3 se siguió el protocolo publicado por Rubio-Cosials (Rubio-Cosials, Sidow et al. 2011) con el cual se obtenía proteína homogénea, sin degradación, con un pureza >95% por tinción por Comassie, según se puede apreciar en la figura 17, ideal para su posterior cristalización. El protocolo consistía en dos pasos de purificación, uno por afinidad empleando la cola de histidinas y otro por cromatografía de exclusión por tamaño. El tampón de elución de la cromatografía de gel filtración fue HEPES 50mM pH 7.5, NaCl 750 mM, DTT 5 mM.

El lugar de unión Site-X reconocido por TFAM fue descubierto por técnicas de “*footprinting*” in *organello*. La longitud del ADN contactado determinada por esta técnica era de 28 pb (Fisher,

Topper et al. 1987). Posteriormente con la información de la estructura cristalográfica de TFAM/LSP 22pb (Rubio-Cosials, Sidow et al. 2011) y TFAM/LSP 28 pb (Ngo, Kaiser et al. 2011) se determinó la longitud exacta del ADN que contactado por la proteína, de 22 pb de longitud. Con ésta información y en base al alineamiento de Site-X con LSP realizado por Fisher y coautores (Fisher, Topper et al. 1987), se definió la secuencia de Site-X empleada en la cristalización.

#### Figura\_17. Purificación TFAM.

Se muestra el cromatograma correspondiente a la cromatografía de exclusión de tamaño en la cual se indica la absorción a una longitud de onda de 280 nm (azul) y 260 nm (rojo), y su relación 260/280 es 0.625. Arriba a la derecha se muestra el gel (SDS-PAGE 15%) teñido por Comassie, con las fracciones cargadas del pico, que se indican con una llave que corresponde a la llave del cromatograma y donde se aprecia la pureza de la proteína TFAM indicada con una flecha negra. A la izquierda del gel se indican los pesos moleculares del marcador de proteína.



El protocolo de formación del complejo entre TFAM/Site-X fue el mismo que se publicó anteriormente para el complejo TFAM/LSP (Rubio-Cosials, Sidow et al. 2011), que consistió en mezclar TFAM y Site-X en una proporción molecular 2:1 proteína:ADN a una concentración de proteína entre 0.5 y 0.9mg/ml en un volumen de 3 ml, seguida de una posterior diálisis (empleando una membrana con un tamaño de poro de 3,5KDa) en la cual se reducía la concentración de NaCl en tres etapas: 2 horas en el tampón de purificación pero con 500 mM de NaCl, toda la noche con NaCl a 100 mM y finalmente 4 horas con NaCl a 20 mM. Esta reducción progresiva de NaCl aseguraba la estabilidad de TFAM durante su unión con ADN.

#### 8.1.1.2 Cristalización y resolución de la estructura TFAM/Site-X.

#### 8.1.1.3 Cristalización de TFAM/Site-X.

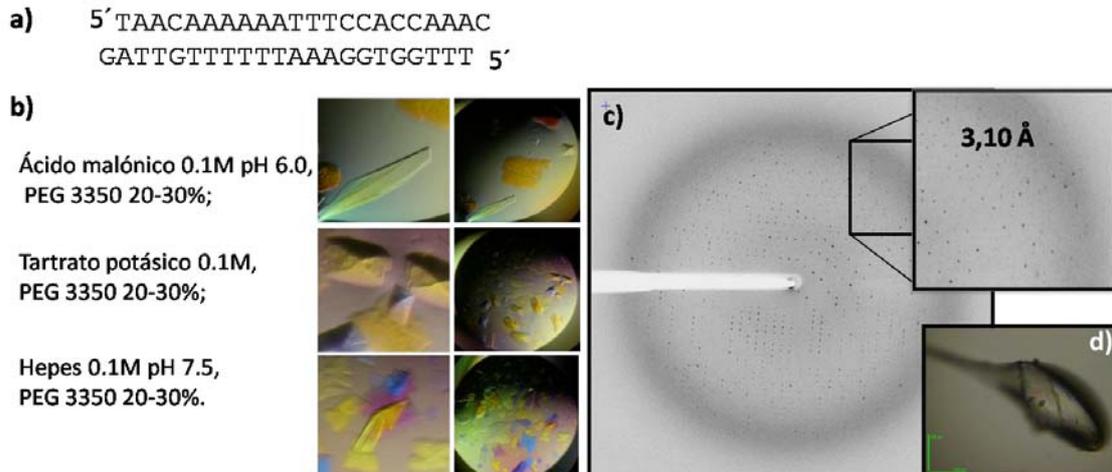
Los primeros ensayos realizados con nanogotas dieron cristales que fueron escalados a gotas de 1µl + 1µl con placas de 24 pocillos (Limbro), en gota sentada. Se obtuvieron cristales a 20 °C

con el ADN de secuencia Site-X de 22pb de longitud y extremos cohesivos de una base por cada extremo (figura\_18a) en diferentes condiciones: Ácido malónico 0.1M pH 6.0, PEG 3350 20-30%; Tartrato potásico 0.1M, PEG 3350 20-30%; HEPES 0.1M pH 7.5, PEG 3350 20-30% (figura\_18b). El complejo estaba a una concentración de proteína de 13mg/ml en el tampón de diálisis HEPES 50 mM, NaCl 20 mM, DTT 5 mM, pH 7.5.

Para reducir el daño por radiación se recogieron los datos de difracción de rayos X de cristales que habían sido crioenfriados en nitrógeno líquido. Para la criogenización de los cristales, éstos se sumergieron en una solución crio-protectora que contenía la disolución precipitante de cristalización a la que se le añadió 10-20% PEG 400.

**Figura\_18. Resultado gotas de cristalización del complejo.**

**a)** Se indica la secuencia de ADN del complejo que cristalizó, **b)** se indican las condiciones de cristalización a la izquierda y a la derecha la imagen del cristal y de la gota. **c)** En la izquierda se muestra una de las imágenes de difracción del cristal, a la derecha arriba se muestra una ampliación del patrón de difracción, cuya resolución nominal era de 3,0Å. **d)** A la derecha abajo hay una imagen del cristal en el "loop" antes de difractarlo.



#### 8.1.1.4 Procesado de datos y resolución de la estructura cristalográfica.

Los datos de difracción de los cristales fueron recogidos en la línea del sincrotrón ESRF (Grenoble) ID-23-2, una línea de microfocus en la que el rayo se focalizaba a un área menor de 10 µm x 10 µm. Los datos se recogieron a 100 K de temperatura gracias a un sistema de emisión de nitrógeno (Oxford Cryosystems 700 series) para crio-protoger los cristales del daño por radiación durante la exposición de los rayos X. Se recogieron datos con una resolución nominal de 3.0 Å (Figura\_18d). Para la difracción del cristal se empleó la técnica de recogida helicoidal (Flot, Mairs et al. 2009), que consiste en realizar un movimiento de traslación simultáneamente a la rotación del cristal, a lo largo de dos puntos definidos en el cristal. Con esta técnica se minimiza el daño por radiación, especialmente cuando se realiza la recogida en líneas en las cuales el haz del rayo es de alto flujo como es el caso de la línea microfocus ID23-2.

Los datos de difracción fueron indexados, integrados, escalados y reducidos a reflexiones únicas con XDS (Kabsch 2010; Kabsch 2010) y SCALA (Evans 2006). Los datos fueron procesados en el grupo espacial monoclinico C2, y se hallaron cuatro complejos en la unidad asimétrica (u.a.), con un valor de coeficiente de Matthews  $3.31 \text{ \AA}^3/\text{Da}$ , calculado con el conjunto de mapas de CCP4 (Winn, Ballard et al. 2011). Se procesaron también en ortorrómbico I222 por que inicialmente no se observó diferencias en la calidad del indexado entre los dos grupos espaciales, sin embargo el grupo ortorrómbico no se consiguió ninguna solución por remplazo molecular y con el grupo C2 sí se encontró solución para los cuatro complejos de la u.a.

El problema de la fase se resolvió con el método de Remplazo Molecular (RM) con el programa PHASER (McCoy, Grosse-Kunstleve et al. 2007), utilizando como estructura de referencia el complejo TFAM/LSP (Rubio-Cosials, Sidow et al. 2011). Los valores de los picos (“z-score”) encontrados para las cuatro moléculas de la unidad asimétrica se indican en la tabla\_9, tanto de las rotaciones, las traslaciones como del número de choques que presentaron las moléculas (PAK) entre ellas, tanto de la unidad asimétrica o como aquellas relacionadas por simetría.

**Tabla\_9. Valores de las soluciones de PHASER para el complejo TFAM/Site-X.**  
Se indican los valores de z-score de las rotaciones (RFZ), traslaciones (TFZ) y el número de choques (PAK) que tiene cada solución.

|            | RFZ | TFZ  | PAK |
|------------|-----|------|-----|
| Solución 1 | 3.5 | 7.7  | 1   |
| Solución 2 | 5.0 | 6.2  | 2   |
| Solución 3 | 3.7 | 6.6  | 9   |
| Solución 4 | 4.3 | 10.4 | 3   |

El criterio para considerar una solución válida en PHASER es que ésta tiene que tener un valor de TFZ mayor de 5, si es entre 5-6 es poco probable, 6-7 posible y mayor de 8 es definitivo, siempre y cuando el valor de choques entre moléculas sea menor de 10. Además, se tuvo en cuenta un análisis visual de todas las soluciones, se analizó qué empaquetamiento se generó con cada una de las moléculas de modo independiente y con las cuatro en conjunto.

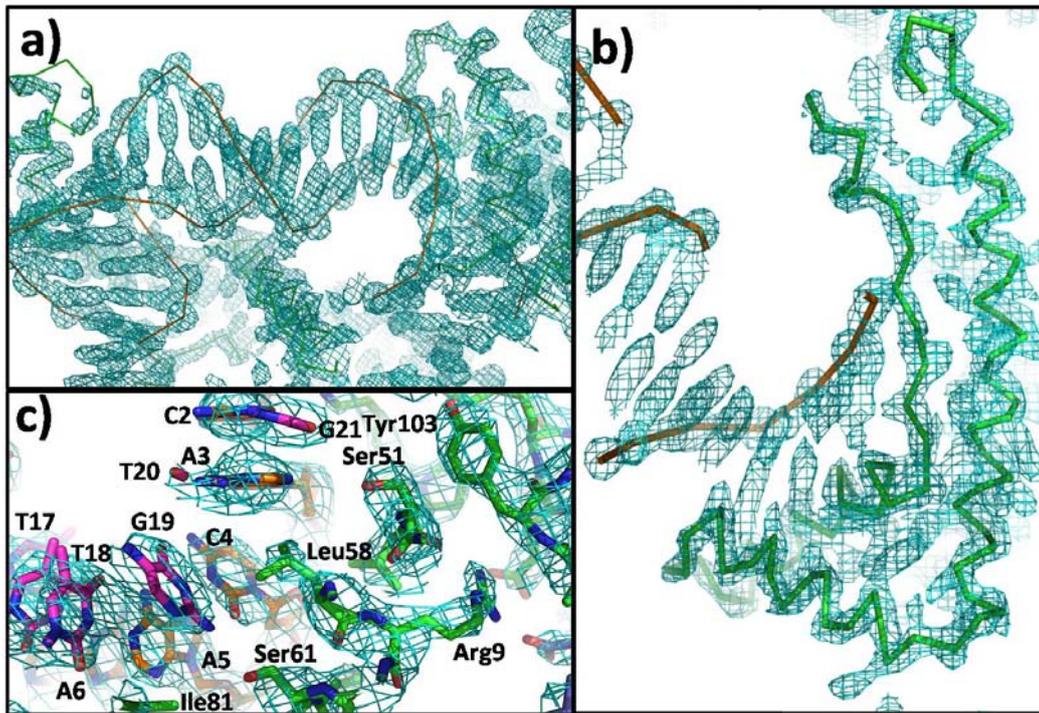
La estructura fue refinada combinando dos programas PHENIX (Adams, Afonine et al. 2010) y BUSTER (Branden and Jones 1990), en los cuales se incluyó el refinado anisotrópico de grupos mediante TLS (“*Traslation Libration Screw-motion*”), restricciones por similitud entre pares de moléculas relacionadas por simetría no cristalográfica y apareamiento de bases según “Watson y Crick” en las siguientes parejas de bases A3/T20, A7/T16, A8/T15, A9/T14, A10/T13, T11/A12, T12/A11, T13/A10, C14/G9 y C18/G5. La asignación de la secuencia del ADN se realizó teniendo en cuenta los mapas de densidad electrónica. Las etapas de refinado automático se

fueron alternando con etapas de posicionamiento de los átomos manual de modo interactivo con COOT (Emsley and Cowtan 2004) con el cual se construyó manualmente el modelo.

Finalmente la estructura se construyó con cuatro complejos de TFAM/Site-X en el interior del mapa de densidad electrónica obtenido por RM y mejorado durante el refinado (figura\_19).

**Figura\_19. Mapa de densidad electrónica.**

Mapa de densidad electrónica final para el complejo TFAM/Site-X de la complejo A/BC. En **a)** se muestra la densidad correspondiente al ADN. En **b)** se muestra la densidad correspondiente del dominio HMGbox1, se representa en su interior los C $\alpha$  en verde. En **c)** con representación tipo "stick" se muestra la inserción de la Leu58 entre G19 y T20.



En los cuatro complejos, la proteína se pudo trazar desde el residuo 43 al 234 y el ADN se trazó completamente. Cada uno de los cuatro complejos de la u.a. están constituidos por una molécula de proteína y dos hebras complementarias cohesivas de ADN, llamados proteína A/hebras B y C, D/EF, G/HI, J/KL. En los complejos D/EF y G/HI se trazó doble conformación en varias regiones, concretamente en los residuos 188 al 199, de las dos cadenas proteicas D y G y las bases 12 hasta la base 21 de la cadena H del ADN. Estas zonas de doble conformación de proteína y ADN están cercanas entre sí y se hablará de ellas tras la descripción general del complejo TFAM/Site-X.

Los valores finales de calidad del refinado  $R_{work}$  y  $R_{free}$  fueron aceptables (tabla\_10). Además la estructura cumplió los parámetros geométricos según la representación de Ramachandran. Del total de 727 residuos que forman la u.a. el 97% tienen en regiones favorables los dos ángulos  $\Psi$  (psi) y  $\phi$  (phi) que definen el enlace peptídico en regiones favorables, un 3% en regiones permitidas (22 aminoácidos) y 0% en zonas "no permitidas" (ver tabla\_10 y figura\_20).

**Tabla\_10. Datos cristalográficos de los cristales del complejo TFAM/Site-X.**

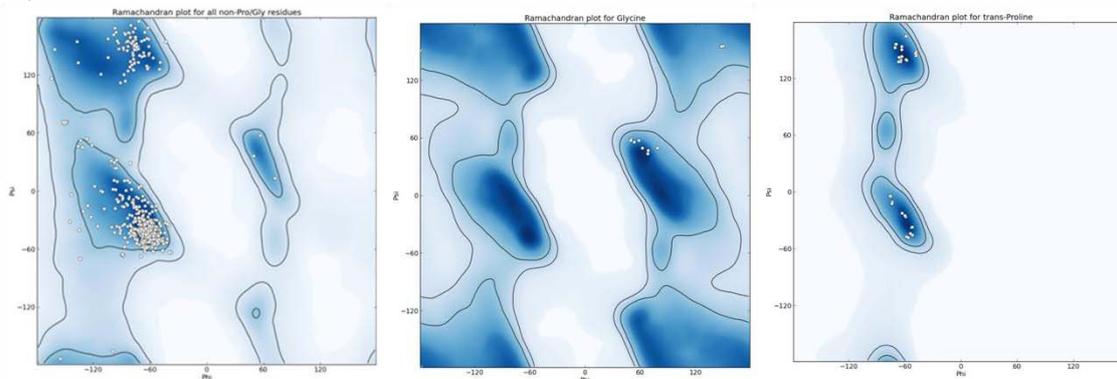
En la columna de la izquierda se indica entre paréntesis los parámetros de la última capa de resolución.

| Recogida de Datos                        |                          |
|--|--------------------------|
| Grupo espacial                           | C 2                      |
| Dimensiones de la celda:                 |                          |
| a, b, c, (Å)                             | 164.79, 145.136, 108.119 |
| α, β, γ (°)                              | 90, 130.8, 90            |
| Resolución (Å)                           | 42.21 - 3.1 (3.27 - 3.1) |
| Rmerge (%)                               | 7.6 (30.1)               |
| I/σ (I)                                  | 15.9 (4.7)               |
| Compleitud (%)                           | 99.0 (99.1)              |
| Redundancia                              | 3.8 (3.9)                |
| Disolvente (%)                           | 67                       |
| Refinamiento                             |                          |
| Resolución (Å)                           | 47.3 - 3.1 (3.27-3.1)    |
| No. Reflexiones                          | 132671 (19425)           |
| No. Reflexiones Únicas                   | 34552 (5000)             |
| Rwork/ Rfree                             | 0.1932 / 0.2316          |
| No. Átomos                               | 10107                    |
| Proteína                                 | 6510                     |
| ADN                                      | 3584                     |
| H <sub>2</sub> O                         | 13                       |
| Factor B (Å <sup>2</sup> )               | 58.53                    |
| Media de factores B                      | 65.50                    |
| Ramachandran favorables y permitidos (%) | 100                      |
| Ramachandran no permitidos (%)           | 0                        |
| desviación R.m.s:                        |                          |
| Longitud de enlace (Å)                   | 0.012                    |
| Ángulos de enlace (°)                    | 1.36                     |

$$Rwork/free = \frac{\sum_{hkl} |F_{obs}| - |F_{cal}|}{\sum_{hkl} |F_{obs}|}$$

**Figura\_20. Representación de Ramachandran.**

Se muestran tres representaciones de Ramachandran en las cuales se representan los ángulos Ψ (Psi) y φ (Phi) del enlace peptídico. A la izquierda está el diagrama de Ramachandran que contiene todos los residuos excepto las glicinas y las prolinas, debido a que estos residuos tienen sus diagramas específicos en el centro y a la derecha, respectivamente. Calculado con PHENIX (Adams, Afonine et al. 2010).



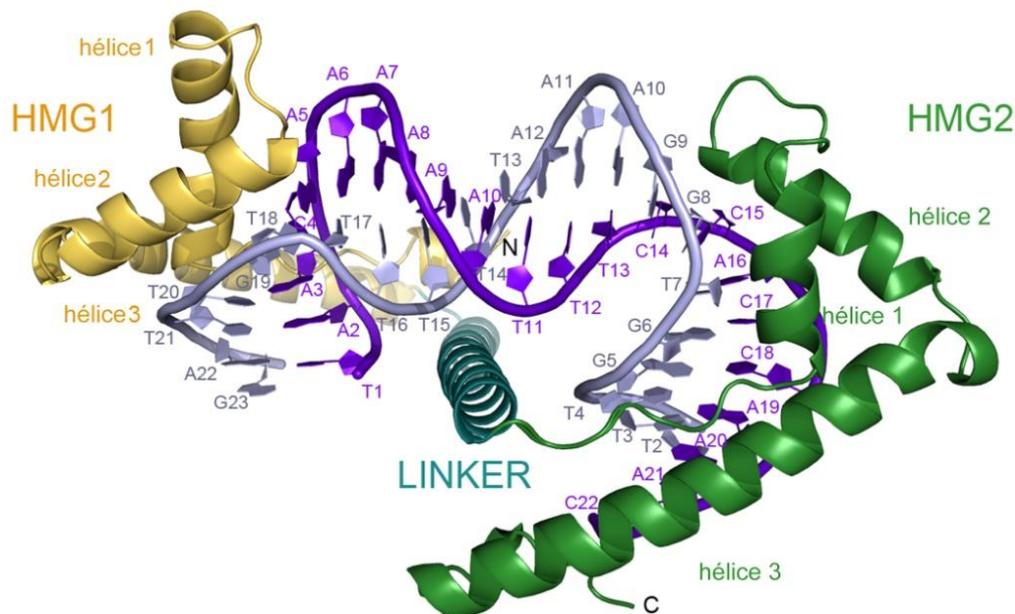
### 8.1.1.5 Estructura Tridimensional del complejo TFAM/Site-X.

#### 8.1.1.5.1 Estructura complejo TFAM/Site-X.

Tal y como se mencionó en el apartado anterior, la estructura de TFAM en complejo con la secuencia Site-X de 22 pb cristalizó con cuatro complejos en la u. a., todos ellos con la forma de un “U-turn” similar a los que anteriormente se había publicado para estructuras de esta proteína con otras secuencias de ADN: LSP22pb (Rubio-Cosials, Sidow et al. 2011), LSP28pb (Ngo, Kaiser et al. 2011), HSP1 y nsADN (no específico) (Ngo, Lovely et al. 2014). TFAM posee dos dominios globulares HMGbox1 y 2, los cuales tienen la estructura tipo L típica de este tipo de dominios, compuestos por tres hélices  $\alpha$ , dos más cortas (hélice 1 y 2) y de longitud similar, ambas formando un ángulo de aproximadamente  $80^\circ$  con la hélice 3, que es de mayor longitud y que además tiene contactos antipalarelos con el segmento extendido N-terminal del dominio (figura\_21). La parte cóncava de cada HMGbox interactúa con el surco menor del ADN, doblándolo cada una  $90^\circ$ , y por tanto se produce una curvatura total de  $180^\circ$  prácticamente en el mismo plano. Las dos HMGbox están conectadas por el “linker” que posee también una estructura de hélice  $\alpha$ . El “linker” también interactúa con el ADN, sin embargo, no provoca ninguna distorsión estructural en él, e interactúa exclusivamente con el esqueleto de fosfatos, por el lado opuesto al de las HMGbox. De un modo global, TFAM envuelve al ADN y lo dobla en forma de U (“U-turn”, figura\_21).

**Figura\_21. TFAM/Site-X.**

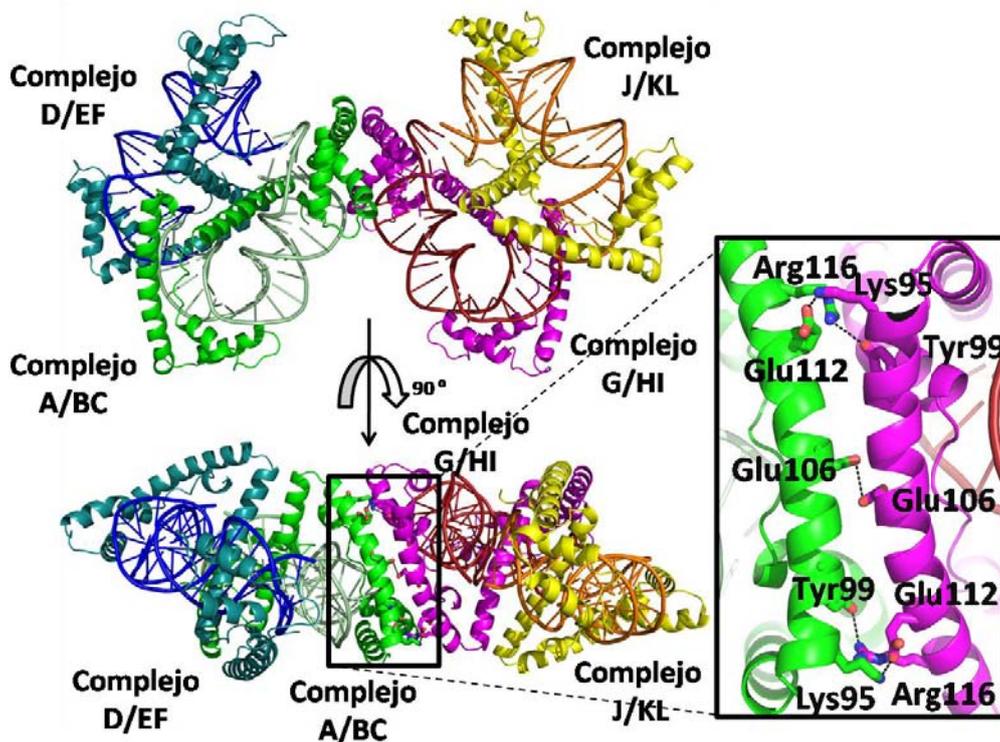
Se muestra el complejo TFAM/Site-X de las cadenas A/BC. TFAM se representa con distintos colores para cada dominio: HMG1 en amarillo, el “linker” en azul y HMG2 en verde. En violeta y azul pálido se indican respectivamente las cadenas BC del ADN Site-X. Se indica cada base con código de una letra y su posición en la secuencia de 22 pb cristalizada.



Los cuatro complejos de la u.a. interactúan entre sí por pares, formando un círculo pseudocontinuo de ADN. El par de complejos A/BC y D/EF forman un círculo y los complejos G/HI y J/KL forman otro. En un mismo círculo las proteínas interactúan por el “linker” mientras que el ADN forma pares Watson y Crick por sus extremos cohesivos, de las diferentes cadenas además de apilarse entre ellas. Entre anillos también hay interacciones, HMGbox1 de la proteína de un anillo contacta de forma antiparalela con el HMGbox1 de otro anillo. Las interacciones polares entre HMGbox1 de diferentes anillos implican a los residuos Lys95 de una molécula con Glu112 de la otra, y Tyr99 con Arg116 (figura\_22), éste último además interactúa con los residuos Glu63 y Asn64 de la hélice 1; asimismo, Glu106 interactúa con Arg59 de la hélice 1. Las no polares ocurren entre Ile98 y Ala102 de una proteína con Val109 y Ala105 de la otra; y el mismo esquema se repite a la inversa por ser la interacción antiparalela (mismos pares de residuos pero de la otra cadena).

**Figura\_22. Unidad Asimétrica de TFAM/Site-X.**

Se indican los cuatro complejos de la u.a. de TFAM/Site-X: complejo A/BC (verde), complejo D/EF, complejo G/HI (rosa), complejo J/KL (amarillo). A la derecha se indica la interacción entre las hélices 3 de los complejos A/BC y G/HI implicados en la dimerización.



Estas interacciones entre dos HMGbox1 de dos complejos TFAM/Site-X ya han sido observadas en otros complejos como y analizadas por Huu Ngo y coautores, mediante estudios por mutagénesis de cinco residuos implicados en los contactos (K95A, Y99F, E106A, E112A y R116A), determinaron que estos residuos eran esenciales para la dimerización y participaban en el empaquetamiento del ADN (Ngo, Lovely et al. 2014). También observaron mediante

ensayos de FRET que estos mutantes no influían ni en la unión al ADN ni tampoco en doblarlo y mediante ensayos de transcripción observaron que tampoco afectaban la unión a la secuencia del promotor específico para TFAM, como tampoco afectaban a la activación de la transcripción en LSP ni en HSP1, por tanto no es necesaria la dimerización para el inicio en la transcripción. Por el contrario, mediante técnicas de partícula única se demostró que estas mutaciones disruptivas del dímero afectaban la capacidad de empaquetamiento del ADN por parte de TFAM (Ngo, Lovely et al. 2014).

Los cuatro complejos son estructuralmente muy parecidos entre en sí, el valor de la desviación cuadrática media (root mean square deviation, r.m.s.d.) entre los aminoácidos 44 y 234 es tan bajo entre las cadenas polipeptídicas A y J, pero no tanto entre A y D o A y G (ver tabla\_11), es decir, el menor r.m.s.d. no es entre moléculas del mismo anillo, el complejo A/CD es más parecido al complejo J/KL y a su vez complejo D/EF al complejo G/HI (figura\_23). Este hecho se evidencia con la superposición de todos los complejos de la u.a. por HMGbox1, que muestra un gran parecido entre los complejos A/BC y J/KL pero una diferencia notable de éstos con D/EF y G/HI por un desplazamiento de la cadena polipeptídica que empieza en el “linker” y que continúa progresivamente hasta desplazar la HMGbox2 hasta 7.4 Å en el aminoácido Glu172 entre la molécula A y la molécula G (figura\_23, se indica con triángulos negros el desplazamiento).

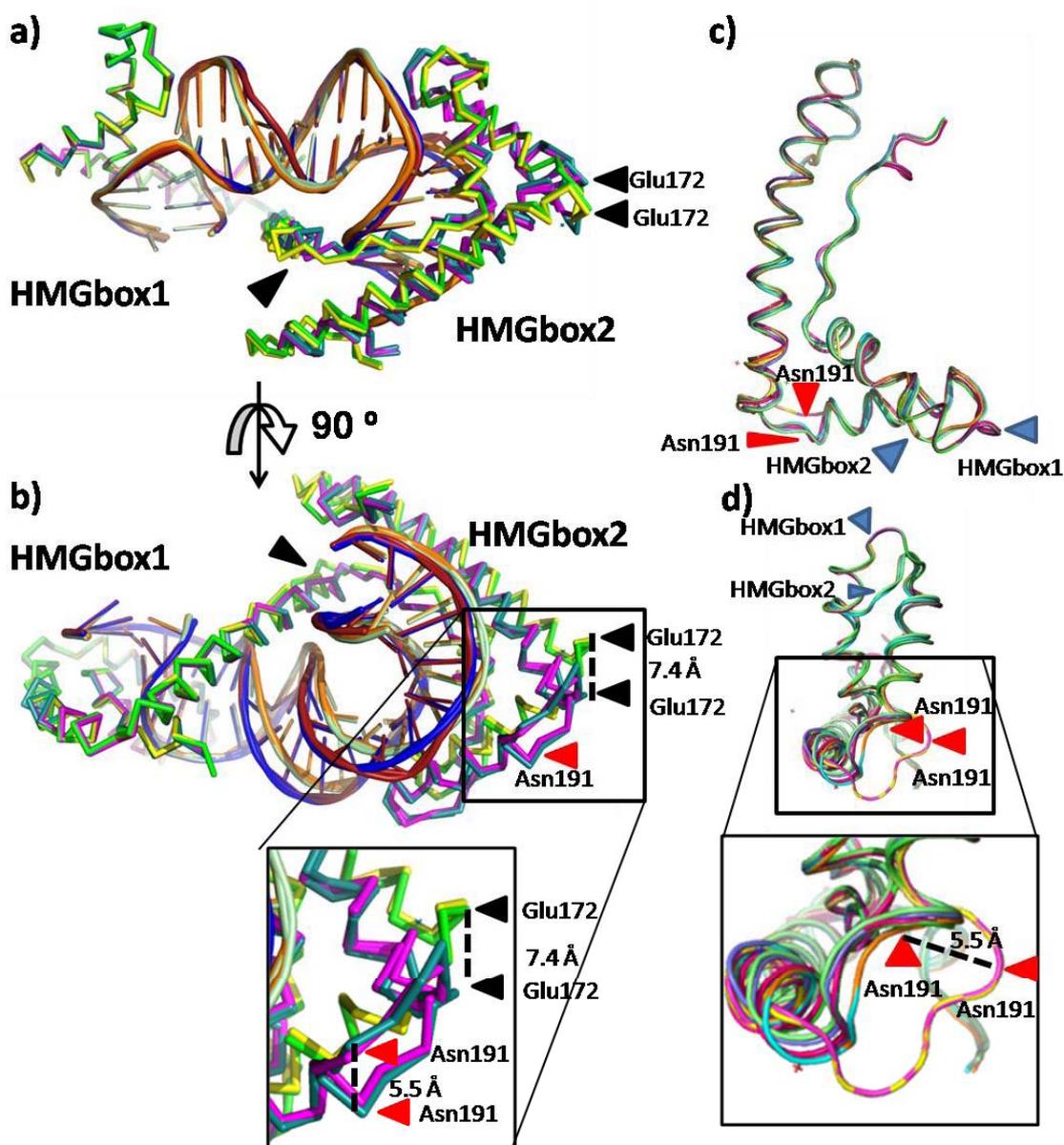
| <b>Tabla_11. Valores de los r.m.s.d de los complejos TFAM/Site-X.</b><br>Se escogió como referencia el complejo A/BC y se calculó el r.m.s.d superponiéndole a la molécula A las distintas cadenas polipeptídicas D, G y J existentes en la unidad asimétrica. |     |
|--|-----|
| Complejo D/EF  | 1.9 |
| Complejo G/HI  | 1.8 |
| Complejo J/KL  | 0.3 |

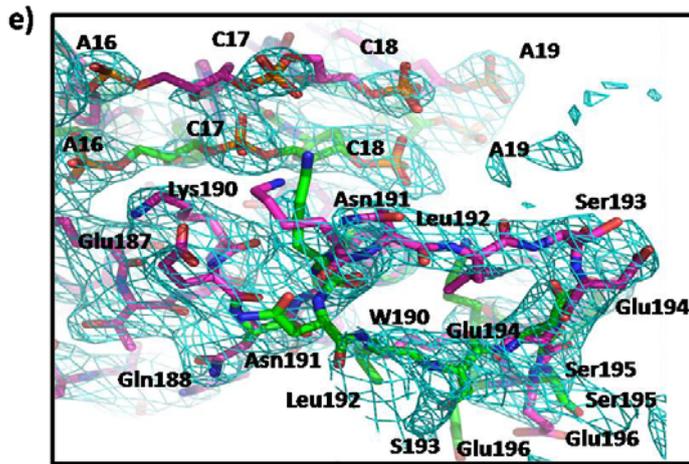
La superposición de todas las HMGbox1 con las HMGbox2 muestra que la hélice 2 de HMGbox2 tiene un giro menos que la hélice 1 de HMGbox1 (indicado con triángulo color azul en la figura\_23c y d), como ya se había observado anteriormente (Rubio-Cosials and Sola 2013). La misma superposición muestra que el “loop” que incluye los residuos desde Lys186 hasta Ile201 de HMGbox2 de los complejos D/EF y G/HI, que conectan la hélice 2 con la hélice 3 (en el “codo” de la HMGbox2), presenta dos conformaciones, una que denominamos cerrada y es la misma que la que se observa en las estructuras de los otros complejos TFAM/ADN, y otra conformación nueva denominada abierta, en la cual el “loop” que une ambas hélices se observa en una conformación más expuesta al disolvente que en la conformación cerrada (figura\_24c-d, indicado con triángulo rojo). Aparentemente, esta doble conformación esta inducida por contactos por simetría (sim) cristalográfica: ambas cadenas D y Gsim (o viceversa,

G y Dsim) contactan por este “loop” entre hélices 2 y 3, y si las dos moléculas simétricas presentan conformaciones “cerradas” entonces se producen choques estéricos. Ello sugiere que en el cristal la forma cerrada de una molécula contacta con la forma abierta de la otra molécula, es decir D-cerrada contacta a Gsim-abierta (o viceversa, G-cerrada contacta a Dsim-abierta) a lo largo de todo el cristal y es indicativo de la flexibilidad de esta zona.

**Figura\_23. Superposición complejos TFAM/Site-X y sus HMGbox.**

Se muestran la superposición de los cuatro complejos de la u.a. de TFAM/Site-X: complejo A/CB (verde), complejo J/KL (amarillo), complejo D/EF (azul) y complejo G/HI (magenta) en **a)** y rotados 90° en **b)**, con triángulos negros se marca el desplazamiento de la estructura de los complejos. Con triángulo rojo se indica la doble conformación de las cadenas D y G. En **c)** y **d)** se muestra la superposición de todas las HMGbox 1 y 2 (vista lateral y desde abajo, respectivamente). Con triángulo azul se indica el desplazamiento de Glu172 entre las HMGbox1 y 2, y con triángulo rojo el desplazamiento de Asn191 en la doble conformación de HMGbox2 en las cadenas D y G. En **e)** se muestra la doble ocupación de la cadena H del ADN y del “loop” entre los residuos 188 al 200 de la cadena G, del complejo TFAM/Site-X, en color verde la conformación abierta y en rosa la conformación cerrada.





El desplazamiento de la proteína en esa región está relacionada con un alejamiento del ADN, éste también en doble conformación en el complejo G/HI, en el cual se observan las bases desde C<sub>14</sub> hasta A<sub>20</sub> de la cadena H más expuestas al disolvente y alejadas de la proteína y de la cadena complementaria I (figura\_23e). Algo similar se observó en el complejo D/EF aunque de forma menos evidente que en la cadena H; el mapa de densidad electrónica en la cadena E era muy pobre para trazar la segunda conformación a pesar que el mapa Fo-Fc indicaba la segunda posición para los fosfatos desde el 16 al 20 de la cadena E, sugiriendo la doble conformación. Estas observaciones sugieren la importancia de los contactos de las cadenas laterales de HMGbox2 para estabilizar el ADN.

#### 8.1.1.5.2 Interacción de TFAM con el ADN Site-X.

A partir de este punto de la descripción de TFAM, dado que de los cuatro complejos en la u.a., A/BC y J/KL presentan un mapa de densidad electrónica de mejor calidad que D/EF y G/HI, los cuales poseen más desorden, se empleará el complejo A/BC como referencia para la descripción general.

La interacción entre TFAM con Site-X involucra estructuralmente tres regiones distintas: HMGbox1, HMGbox2 y el “linker” (figura\_24a). La interacción de la HMGbox1 se produce a través de las hélices 1 y 2 con el surco menor de Site-X, dando lugar a una separación de las hebras de ADN y a grandes distorsiones respecto a un ADN tipo B ideal en la inclinación (“roll”) y el giro helicoidal (“twist”) de los tres pares de bases A<sub>3</sub>-T<sub>20</sub>, C<sub>4</sub>-G<sub>19</sub> y A<sub>5</sub>-T<sub>18</sub> (numeración de las bases según el PDB, ver figura\_24). Esta fuerte distorsión se ve estabilizada por la interacción del residuo Leu58 que inserta su cadena lateral A<sub>3</sub>C<sub>4</sub>/G<sub>19</sub>T<sub>20</sub>, aunque además participan otros residuos, como Tyr57 que forma un enlace con su hidroxilo con el átomo N3 de la G<sub>19</sub>, y Ser61 que interacciona con el átomo O2 de la misma base (figura\_24). Ile81 se acerca sin acabar de intercalar las bases en el paso contiguo C<sub>4</sub>A<sub>5</sub>/T<sub>18</sub>G<sub>19</sub> en el cual también están distorsionados los parámetros de inclinación y giro helicoidal. También participan en la interacción con el ADN los

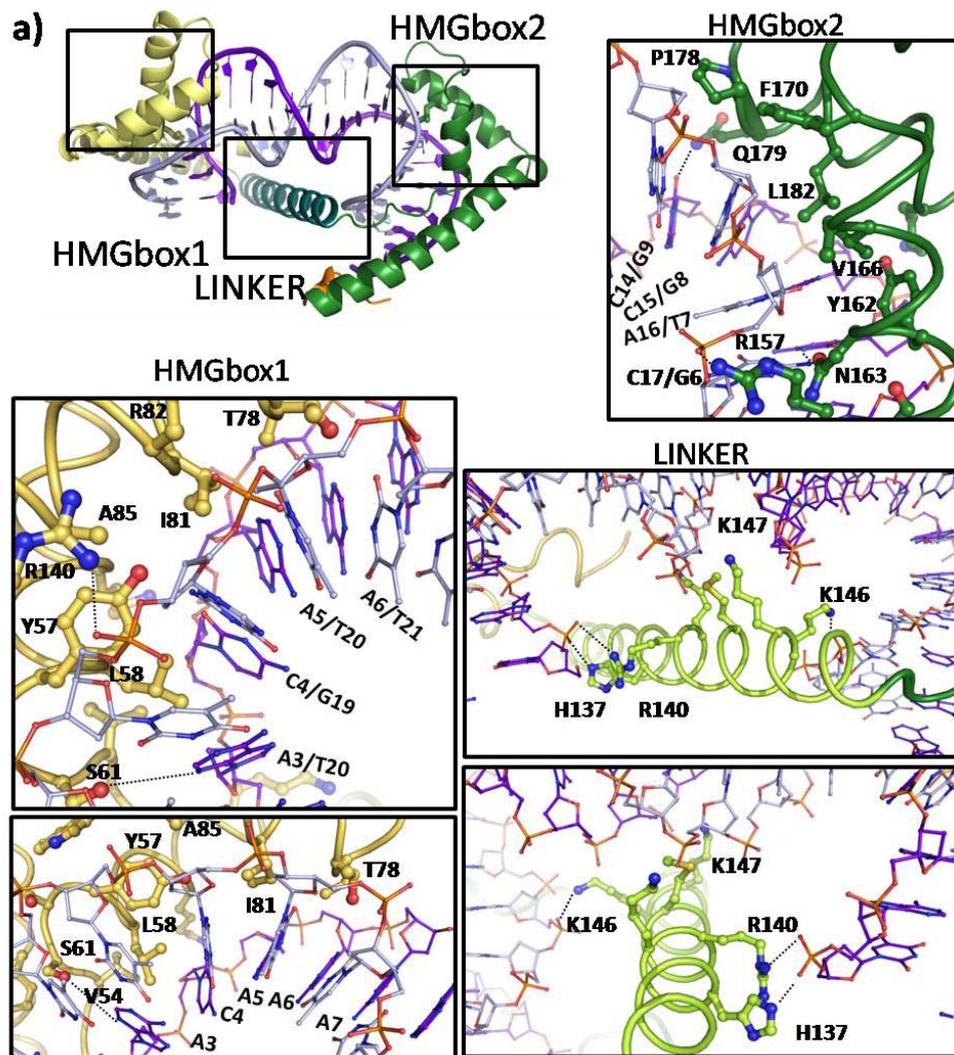
residuos Thr78, Arg82, Ala85, Trp88 y Arg89 que forman enlaces con el esqueleto de fosfatos del ADN de la hebra C desde las bases T<sub>18</sub> hasta T<sub>21</sub> (figura\_24b). La región N-terminal alargada de HMGbox1 también participa en la unión con el ADN, como el N $\zeta$  del residuo Lys52 que forma puentes de hidrógeno con el átomo N3 de la A<sub>2</sub> de la cadena B. Conjuntamente, los residuos Lys51, Pro53 y Val54 realizan contactos de van der Waals y electrostáticos con el esqueleto de fosfatos de las bases C<sub>4</sub>, A<sub>3</sub> y A<sub>2</sub>. Finalmente los residuos Gln100 y Tyr103 de la hélice 3 forman enlaces con los fosfatos de las bases T<sub>20</sub> y A<sub>22</sub> de la cadena C.

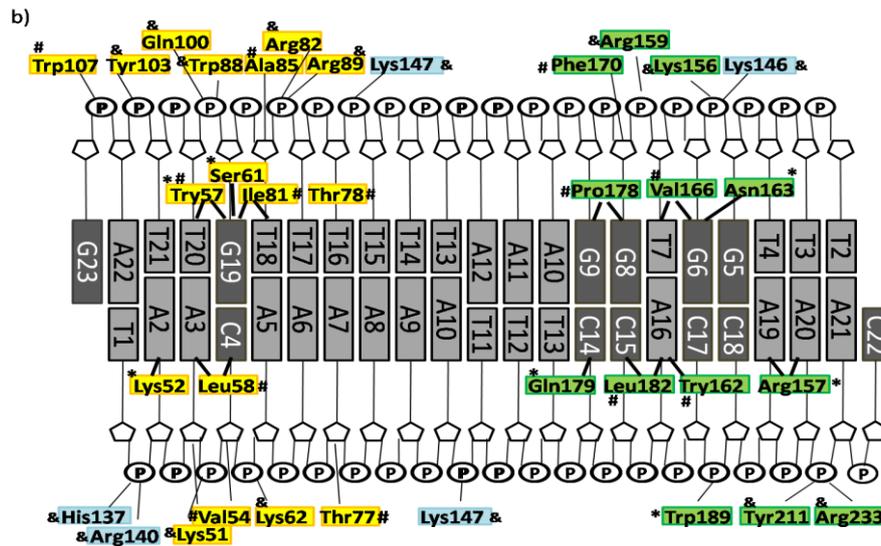
La otra región TFAM que dobla el ADN es HMGbox2, que presenta un escenario estructural semejante a HMGbox1 aunque, según el mapa de densidad electrónica, posee más desorden. La secuencia del ADN en la cual se produce la apertura y torsión de las hebras comprende las bases C<sub>14</sub>C<sub>15</sub>A<sub>16</sub> de la cadena B del DNA, complementarias a G<sub>9</sub>G<sub>8</sub>T<sub>7</sub>, respectivamente. En el paso C<sub>15</sub>A<sub>16</sub>/T<sub>7</sub>G<sub>8</sub>, en el cual la distorsión es máxima coincide con la inserción de Leu182 (figura\_24), de un modo semejante a Leu58 en HMGbox1, aunque Leu182 se encuentra en la hélice 2 y no en la hélice 1. Sin embargo, el mapa de densidad electrónica alrededor de la cadena lateral 182 sólo se ve parcialmente en el complejo J/KL, y no es visible en los demás complejos dentro de la u.a. en cuyas cadenas de proteína la mutación *in silico* por Ala generó densidad positiva en el mapa de diferencias Fo-Fc, sugiriendo la presencia de una cadena lateral más larga. Éstos datos sugieren que Leu182 presenta múltiples rotámeros y está desordenada, indicativo de cierto grado de desorden en estos complejos dentro de la u.a. En el mismo paso C<sub>15</sub>A<sub>16</sub>/T<sub>7</sub>G<sub>8</sub>, Tyr162 (hélice 1) contacta con su hidroxilo a N3 de A<sub>16</sub>, y Val166 por van der Waals a T<sub>7</sub>. En el paso anterior C<sub>14</sub>C<sub>15</sub>/G<sub>8</sub>G<sub>9</sub> se producen interacciones del N $\epsilon$ 2 la Gln179 (hélice 2) por puentes de hidrógeno con el O2 de la C<sub>14</sub>, aunque esta interacción sólo se observa en los complejos equivalentes D/EF y G/HI; además, G<sub>8</sub> y G<sub>9</sub> están separadas por contactos tipo van der Waals entre Pro178. El tercer paso de ADN que presenta grandes variaciones en la inclinación de las bases y en el giro helicoidal es posterior al sitio de inserción, A<sub>16</sub>C<sub>17</sub>/G<sub>6</sub>T<sub>7</sub>, cuyas distorsiones se estabilizan con la unión de N $\delta$ 2 de Asn163 (hélice 1) con el N3 de la G<sub>6</sub> y Val166 con la misma base mediante interacciones de van der Waals. En la zona doblada del ADN hay otras interacciones con las bases que no aparecen en todos los complejos. En los complejos A/BC y J/KL, Arg157 (del segmento alargado de N-terminal) contacta con el átomo N2 de G<sub>5</sub> y N3 de A<sub>19</sub>. En el complejo G/HI, Arg157 contacta con N3 de A<sub>19</sub> y N3 de A<sub>20</sub> y en el complejo D/EF con el átomo N3 A<sub>19</sub> y N3 de G<sub>5</sub>. Además de las interacciones con las bases se encuentran residuos que forman enlaces de van der Waals con los anillo de ribosa como Phe170 con G<sub>8</sub>; e interacciones electroestáticas con los fosfatos como Arg159, Trp189, Try211 y Arg233 con las bases de T<sub>7</sub>, C<sub>17</sub>, A<sub>20</sub> y A<sub>21</sub> respectivamente, aunque la interacción de la Arg159-T<sub>7</sub> solo está presente en los complejos A/BC y J/KL. La tercera región

de interacción se realiza mediante el “linker”, que se ubica en la parte interna del “U-turn” y evita la repulsión electrostática entre los fosfatos del esqueleto de ADN curvado. En el surco mayor distorsionado bajo HMGbox1 (figura\_24a) N $\delta$ 1 de His137 y N $\omega$ 1 de Arg140 contactan el O1 y O2 del fosfato de A<sub>2</sub>, respectivamente (cadena B), mientras en la torsión opuesta, bajo HMGbox2, N $\zeta$  de Lys146 contacta con O2P de G<sub>6</sub> (cadena C). Finalmente Lys139, Met143 y Lys147 contactan el surco menor del ADN en la zona intermedia entre las dos torsiones de 90°, Lys139 se acerca a O1 y O2 de T<sub>15</sub> de la cadena C, Met153 se dobla sobre si misma bajo el azúcar de esta misma base, y sólo Lys139 contacta con su N $\zeta$  a los O3 de ambos A<sub>10</sub> (cadena B) y T<sub>15</sub> (cadena C) simultáneamente.

**Figura\_24. Interacción TFAM con Site-X.**

**a)** se muestra el complejo TFAM/Site-X y se enmarcan las tres zonas de unión al ADN en HMGBox 1 (amarillo), “linker” (azul) y HMGbox2 (verde), cuyas ampliaciones se muestran al lado. Con código de una letra y el número se indican los residuos de TFAM que interaccionan con las bases de Site-X. **b)** Se muestra un esquema con la secuencia completa de Site-X y los residuos que interaccionan en código de tres letras con su número. En cajas amarillas se indican las que se localizan en la HMGbox1, en azul en el “linker” y verde en HMGbox2. Se marca con \* para los enlaces por puente de hidrógeno, # por van der Waals y & para las electrostáticas.





### 8.1.1.5.3 Interacción del "linker" con las HMGBbox1 y 2.

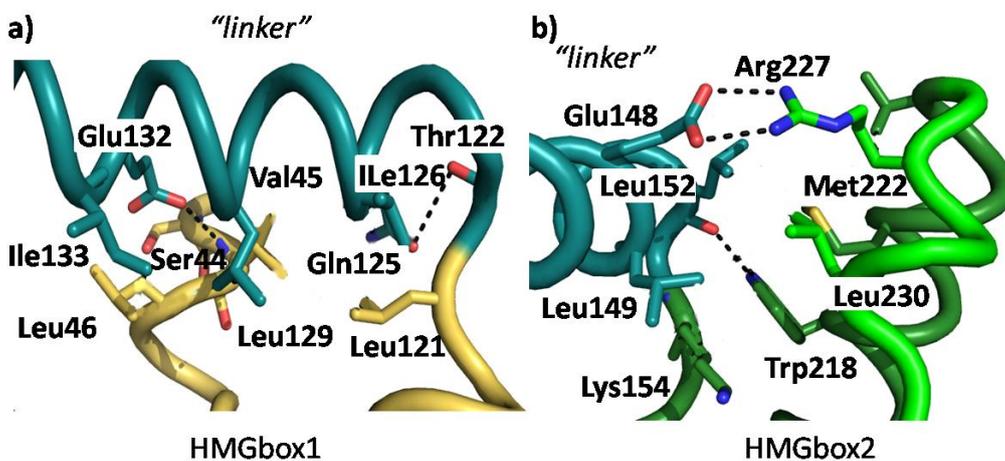
El "linker" posee como principal función evitar las repulsiones electrostáticas del esqueleto de ADN, que contacta tanto en el surco mayor doblado bajo las HMGBbox, como en el surco menor entre estos dos dominios. Anteriormente, se indicó que la posición relativa de las HMGBbox1 y 2 cambia de un complejo a otro dentro de la u.a. Para comprender mecanísticamente el movimiento relativo entre ellas, se analizaron los contactos de cada HMGBbox con el "linker", por si había un efecto puntual de bisagra. Entre HMGBbox1 y el "linker", y entre éste y HMGBbox2 (figura\_25). El "linker" (residuos 122 al 153) presenta una estructura secundaria de hélice- $\alpha$  que comienza justo después del giro al final de la hélice 3 de la HMGBbox1 (figura\_25a). En la cara interna de este giro se producen interacciones hidrofóbicas entre Leu121 ubicada en la hélice3 de HMGBbox1, y Leu129 e Ile126, situados en el "linker". Sin embargo, los principales contactos con el "linker" se producen entre los residuos del extremo N-terminal de la proteína donde comienza la región alargada, que incluyen contactos hidrofóbicos entre las cadenas laterales de Val45 y Leu46, con Ile133 e Ile126 del "linker" e interacciones de puentes de hidrógeno entre el amino de Val45 y el átomo OE1 de Glu132. Finalmente entre Thr122, en el giro entre hélice y "linker" y Gln125, en el "linker", también existen contactos polares que estabilizan la orientación del "linker" respecto HMGBbox1.

Por su parte, HMGBbox2 interactúa con el "linker" mediante residuos del final de la hélice3 y de la cola C-terminal. De entrada una colección de residuos forman un pequeño núcleo hidrofóbico: Leu149 y Leu152 (del "linker"), Met222 y Trp218 (hélice 3), y Leu230 y Arg227 (cola C-terminal) (figura\_25) entre los 3 elementos de la estructura secundaria. Este contacto hidrofóbico se ve reforzado por interacciones de puentes de hidrógeno entre el OE1 y OE2 de Glu148 (en el "linker") con Nw1 y Nw2 de Arg227 (hélice 3) y entre el carbonilo peptídico de Leu152 (en el "linker") con NE1 de Trp218 (cola C-terminal).

Las superposiciones por los residuos 123 hasta 136 de la primera mitad del “linker” de las distintas moléculas de u.a. así como las superposiciones de los distintos complejos de TFAM con las secuencias de ADN LSP, HSP1, snADN permitieron analizar la variabilidad estructural del contacto entre el “linker” y la HMGbox1. Por otra parte, la superposición de los aminoácidos desde el 140 hasta 153 sirvieron para comparar la interacción entre el “linker” y la HMGbox2 entre estructuras. El resultado de estas superposiciones fue que todos los contactos observados entre el “linker” y las dos HMGbox en el complejo TFAM/Site-X también se encuentran en los demás complejos. El valor de mayor r.m.s.d entre los aminoácidos superpuestos fue marginal, de 0.31 Å entre los complejos Site-X y nsADN, al superponer la región N-terminal del “linker”, y 0.24 entre Site-X y HSP1 al comparar la región C-terminal del “linker”. Estos datos indican que la interacción entre el “linker” y las HMGbox no es la responsable de distinta posición relativa entre ambas HMGbox en los distintos complejos. Como se indica más adelante, el origen de esa variación viene determinada por la distinta curvatura del “linker”, y por un movimiento de basculación o péndulo entre éste y HMGbox1 (ver figura 27).

**Figura\_25. Interacciones entre el “linker” y las dos HMGbox.**

En **a)** se muestra la unión entre HMGbox1 (amarillo) y el “linker” (azul). En **b)** se muestra las interacciones entre el “linker” (azul) y la HMGbox2 (verde oscuro) y la cola C-terminal (verde claro). En ambos casos los residuos que participan en la unión están indicados con código de tres letras y los enlaces con líneas negras discontinuas.



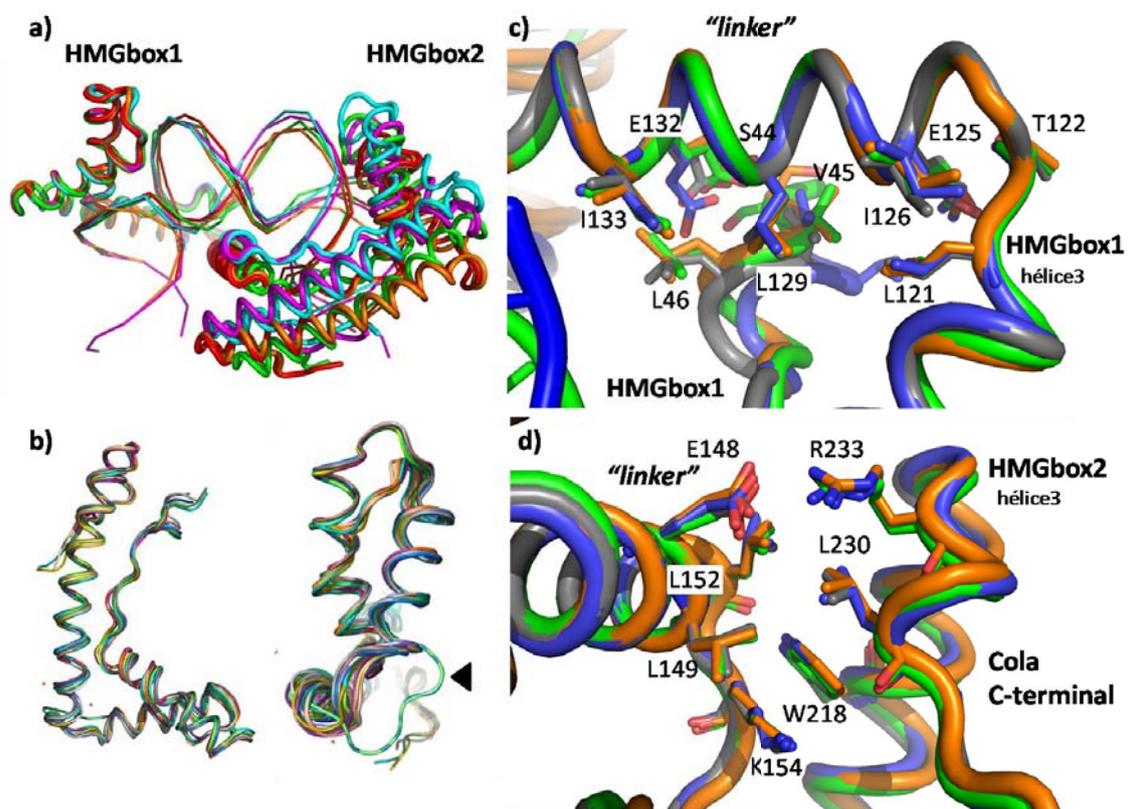
#### 8.1.1.5.4 Comparación de la estructura de TFAM/Site-X con las estructuras de otros complejos TFAM/ADN.

La superposición de las estructuras de TFAM con ADN (Site-X, LSP, HSP1 y nsADN) mostraron el mismo tipo de plegamiento del ADN que el observado para SITE-X, doblado en dos regiones un ángulo de 90° debido a la interacción con cada HMGbox adoptando una forma de “U-turn” (figura\_26). La superposición de los aa 45-234 de la cadena polipeptídica de los distintos complejos TFAM/LSP, HSP1 y nsADN con el complejo TFAM/Site-X (cadenas A/BC) dio lugar a

un valor de r.m.s.d. inferior a 2 Å (ver tabla\_12), muy similar al valor de r.m.s.d. existente entre los distintos complejos de la u.a. de TFAM/Site-X (ver tabla\_11). Tampoco se observó grandes diferencias globales de tamaño entre complejos, tal y como mostró la distancia entre los C<sup>α</sup> de Ser94 (HMGbox1) y Ser195 (HMGbox2) en los giros que unen la hélice2 y hélice3 de ambas HMGbox. En este caso, los valores oscilan entre los 80.5 Å en TFAM/Site-X y 82 Å en TFAM/nsADN. Se superpusieron las estructuras por HMGbox1 (desde residuo 45 hasta 121) o por HMGbox2 (153 hasta 225) dio lugar a un valor máximo de r.m.s.d. de 1.1 Å entre HSP1 y Site-X (ver tabla\_12). El plegamiento de las distintas HMGbox no mostró diferencias importantes entre ellas (figura\_26c y d), sólo la doble conformación en el giro entre hélice2 y 3 de HMGbox2 observada en los complejos TFAM/Site-X D/EF y G/HI la cual no se observó en los demás complejos, que sólo muestran la conformación cerrada.

**Figura\_26. Superposición de los complejos TFAM/ADN.**

Se muestra en la figura **a)** la superposición de las estructuras de los complejos TFAM/Site-X (verde), TFAM/LSP22 (cian), TFAM/LSP28 (rosa), TFAM/HSP1 (rojo) y TFAM/nsADN (naranja), superpuestas por la HMGbox1. **b)** Se muestran desde la vista lateral (izquierda) y desde abajo (derecha) la superposición de las HMGbox1 y 2 de las estructuras TFAM/Site-X (A/BC y G/HI), TFAM/LSP22 (A/BC), TFAM/LSP28 (A/BC), TFAM/HSP1 (A/CD) y TFAM/nsADN (A/CD). Con un triángulo negro se indica la conformación abierta (ver texto) del giro que conecta hélice 2 y 3 en el complejo G/HI de TFAM/Site-X. En las figuras **c)** y **d)** se representan la superposición TFAM/Site-X (verde), TFAM/LSP22 (naranja), TFAM/HSP1 (azul) y TFAM/nsADN (gris) en las regiones de contacto entre el "linker" y las HMGbox 1 y 2 respectivamente; se representan también los residuos implicados en esta interacción.



**Tabla\_12. Valores de r.m.s.d de los complejos TFAM/ADN.**

Se calcula el r.m.s.d de los distintos complejos enteros TFAM/ADN (A/BC) y los distintos dominios HMGbox1 (44-121) y 2 (153-225) para todas las comparaciones se escogió como referencia el complejo TFAM/Site-X (A/BC).

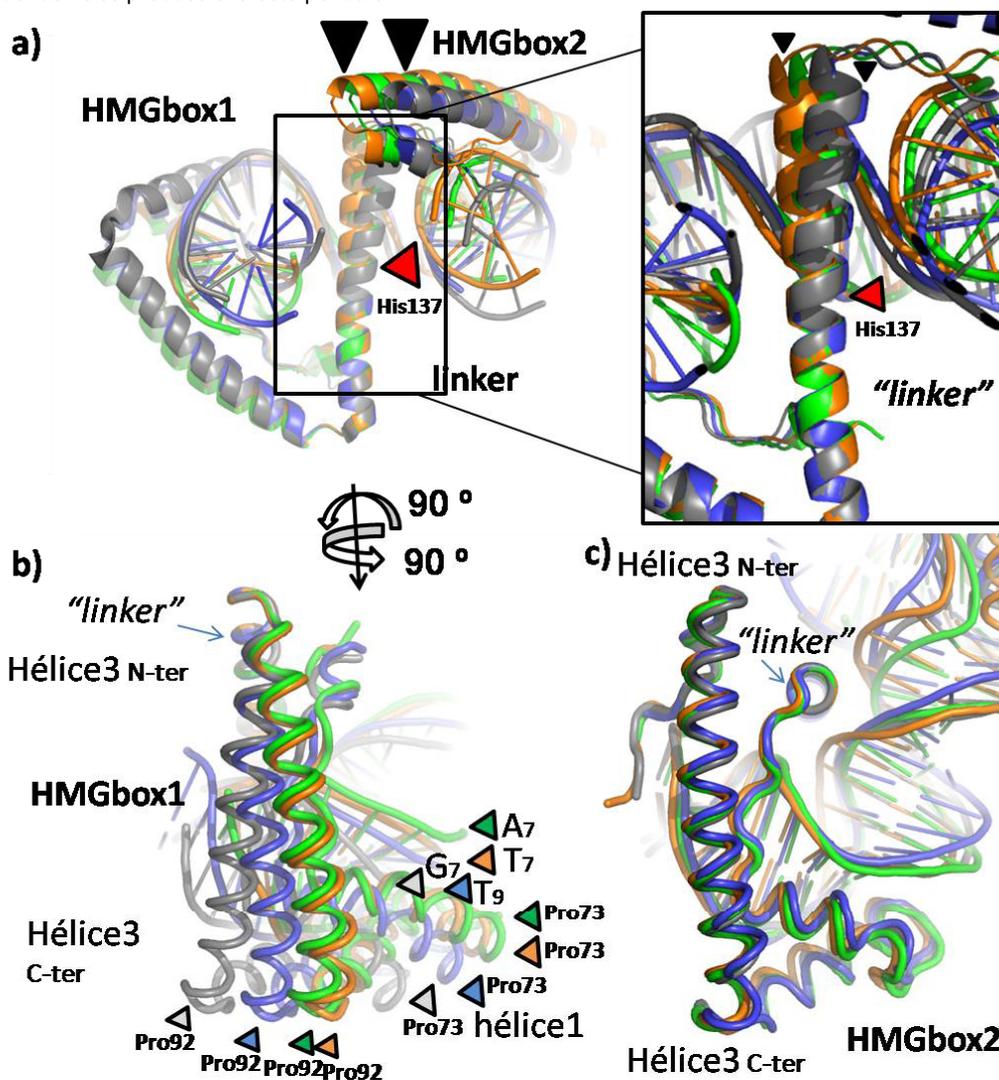
|            | Complejo (A/BC) (Å) | HMGbox1 (Å) | HMGbox2 (Å) |
|------------|---------------------|-------------|-------------|
| TFAM/LSP22 | 1.6                 | 0.7         | 0.7         |
| TFAM/HSP1  | 1.9                 | 1.1         | 0.6         |
| TFAM/nsADN | 1.8                 | 0.9         | 0.9         |

La superposición por HMGbox1 de las moléculas A de los complejos con distintos ADNs (LSP-22, X, HSP1 y nsDNA) muestra como este dominio es muy parecido en todas ellas (ver tabla 12). No obstante, la posición relativa de HMGbox2 respecto a HMGbox1 va variando de una estructura a otra, de forma muy parecida a cómo las estructuras Site-X divergen entre sí. Para identificar los puntos a partir de los cuales se produce la rotación de una HMGbox respecto a la otra, se superpusieron todas las estructuras por zonas. Una primera superposición general del “*linker*” mostró desviaciones estructurales que se hicieron más evidentes al superponer la mitad N-terminal del mismo (residuos 123 hasta 136). En esta superposición, HMGbox1 mostraba diferentes orientaciones según el complejo, en concreto una reorientación del eje de la hélice 3 respecto al “*linker*” (figura\_27b) que afectaba a la posición relativa de toda la HMGbox1 y el ADN al que contacta. Con la misma superposición por la parte N-terminal del “*linker*”, el segundo punto divergente se producía a mitad de la hélice de éste, aproximadamente en el residuo His137, que es donde empieza el contacto del “*linker*” con los distintos ADN (figura\_27a). Las diferencias en la curvatura de la hélice del “*linker*” conllevan el movimiento de HMGbox2, que se desplaza conjuntamente con su ADN.

Si se superpone la mitad C-terminal del “*linker*” se observa que HMGbox2 y “su” ADN no se reorienta entre las diferentes estructuras, como tampoco cambian las interacciones entre el “*linker*” y HMGbox2, principalmente hidrofóbicas, como se describió anteriormente. Ello sugiere que esta región C-terminal del “*linker*” y HMGbox2 (con su ADN) forman un solo bloque (figura\_27c). Por otra parte, la superposición por el extremo N-terminal del “*linker*” mostró como los contactos entre el “*linker*” y HMGbox 1 también están conservados entre estructuras (figura\_27c y d). Por tanto, los contactos entre los extremos del “*linker*” y las dos HMGbox son estables entre estructuras (ver figura\_27c y d), lo que descarta definitivamente la posibilidad que los extremos del “*linker*” funcionen como bisagras para reorientar los dominios HMGbox.

**Figura\_27. Superposición TFAM con todas las secuencias.**

Se han superpuesto las estructuras de los complejos TFAM/Site-X (verde), TFAM/LSP22 (naranja), TFAM/HSP1 (azul) y TFAM/nsADN (gris) desde el residuo 123 hasta 136. En **a)** se indica el con un triángulo negro el punto en el cual se produce el cambio de curvatura de la hélice del "linker" (His137) y con un triángulo rojo el desplazamiento que provoca en HMGbox2. A la derecha se muestra una ampliación de la misma representación. En **b)** se han rotado las estructuras mostrando el efecto "péndulo" que sufre HMGbox1 y con triángulos se indica el movimiento de la hélice3 (Pro92), la hélice1 (Pro73) y el ADN de color como en a). En **c)** se muestra las distintas HMGbox2 cuando la superposición se realiza por la región C-terminal del "linker" (desde el residuo 136 hasta 153), donde no se produce el efecto péndulo.



#### 8.1.1.5.5 Análisis de la estructura del ADN de Site-X.

##### 8.1.1.5.5.1 Distorsiones del ADN Site-X debido al contacto con TFAM.

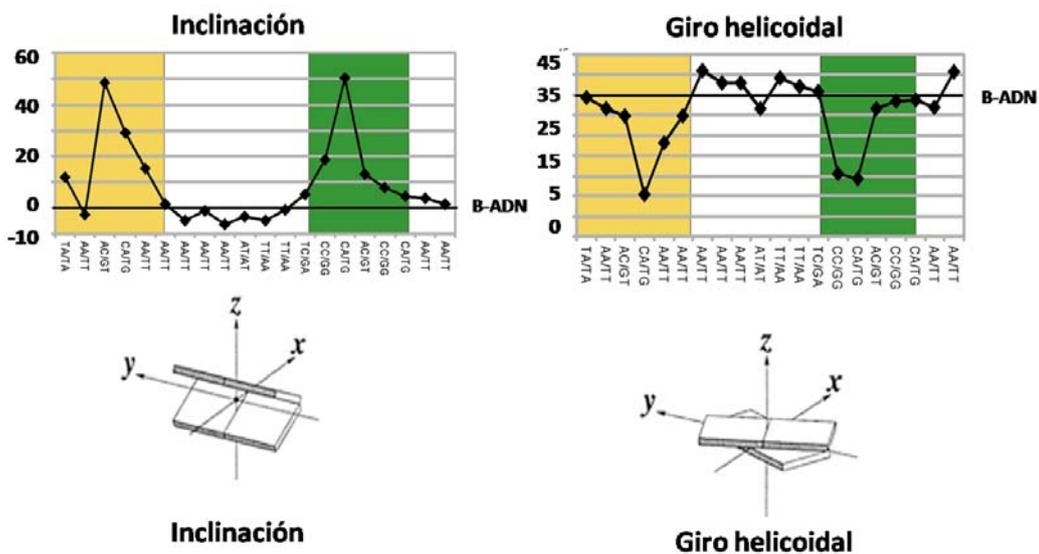
En el complejo TFAM/Site-X, el ADN presenta importantes deformaciones estructurales en las regiones donde interaccionan HMGbox 1 y 2. Ambas regiones están afectadas en dos pasos entre tres pares de bases,  $C_4A_5A_6/G_{21}T_{20}T_{19}$  y  $C_{14}C_{15}A_{16}/G_9G_8T_7$ , que muestran notables diferencias con el ADN de conformación tipo B, cuya estructura fue descrita por Watson y Crick (Watson and Crick 1953). Para evaluar las variaciones de todas las propiedades estructurales, que se describen a continuación, entre en el ADN Site-X en complejo con TFAM respecto a un

ADN tipo B ideal se empleó el programa CURVES+ ([http://gbio-pbil.ibcp.fr/cgi/Curves\\_plus/](http://gbio-pbil.ibcp.fr/cgi/Curves_plus/)) (Lavery, Moaker et al. 2009).

Los parámetros de inclinación (“roll”), giro helicoidal (“twist”) y la anchura del surco menor son los que muestran mayores divergencias. El parámetro de inclinación se define como el ángulo entre el plano que constituye un par de bases y el plano del par consecutivo, que un ADN tipo B es aproximadamente cero porque son paralelos. El parámetro de giro helicoidal define la rotación en el eje Z (en la dirección de la doble hélice de ADN) perpendicular al plano (XY) que se encuentra cada par de bases, que en el caso de ADN tipo B son 35° (figura\_28). En ambas regiones distorsionadas por HMGbox1 y 2 se alcanza una distorsión de inclinación de 50°, y una disminución en la rotación del giro de hasta 10° (de HMGbox1) y 15° (de HMGbox2). Las distorsiones de inclinación y giro helicoidal son más fuertes donde se insertan Leu58 y Leu182. Los cuatro complejos que forman la u.a. presentan aproximadamente valores muy similares tanto de inclinación, giro helicoidal, la mayor diferencia es inferior a 5°, y en el surco menor. Las dos regiones en las cuales el ADN sufre las mayores distorsiones, además de insertarse principalmente los residuos Leu58 y Leu182, también se observa la interacción con las bases de otros residuos hidrofóbicos como Ile81, Tyr57, Val166 y Tyr167. Es muy posible que estas interacciones tengan como función ayudar a estabilizar la deformación del ADN, que sufre en estas regiones fuertes alteraciones en el empaquetamiento entre bases.

**Figura\_28. Inclinación y Giro helicoidal del ADN Site-X.**

Se indican los parámetros de inclinación y giro helicoidal de la estructura TFAM/Site-X, complejo A/BC. En abscisas se indican los pasos entre pares de bases que componen la secuencia cristalizada y en ordenadas el valor del ángulo correspondiente de inclinación (izquierda) y de giro helicoidal (derecha). Con un cuadrado amarillo se indican las bases que interaccionan con HMGbox1 y en verde con HMGbox2. Con una línea recta continua se marca el valor ideal del ADN tipo B. Debajo de cada gráfica se ilustra el parámetro con un dibujo de las posiciones relativas de cada par de bases.



Precisamente los pasos en los cuales se dobla el ADN hasta 90°, y en los cuales las distorsiones de la inclinación y de la rotación son máximas, coinciden con los valores máximos de apertura del ancho del surco menor del ADN (figura\_29). En el complejo TFAM/Site-X se alcanza un valor de ancho del surco menor de hasta 23 Å en los cuatro complejos, como también se había observado en otros complejos de TFAM, como por ejemplo con LSP (figura\_29b), y a diferencia del ADN tipo B que tiene un ancho de surco menor de 12.5 Å.

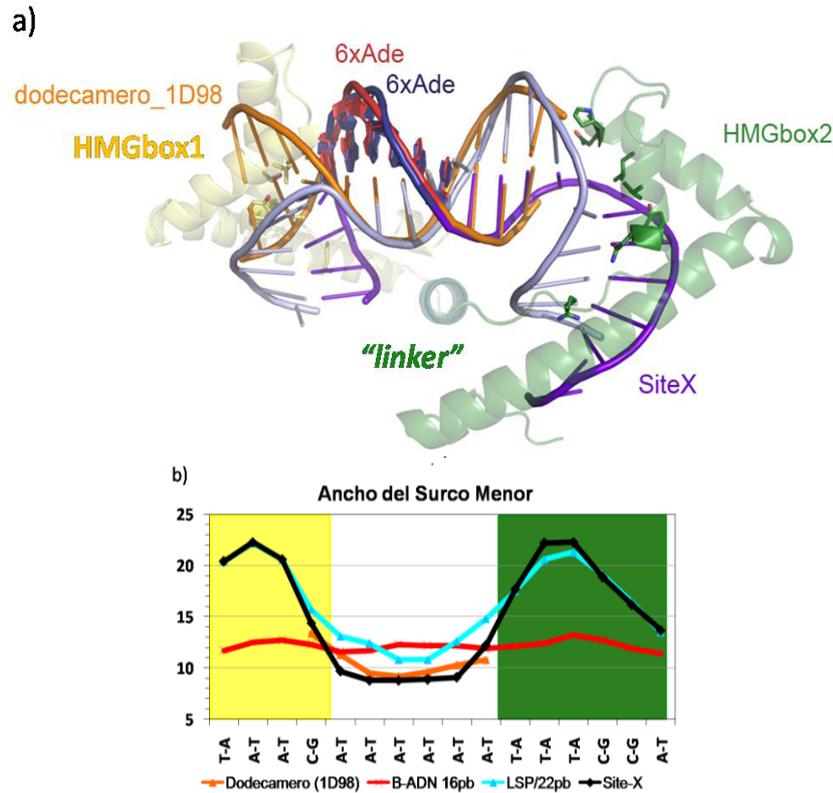
#### ***8.1.1.5.5.2 Análisis de la secuencia de poli-adeninas central de Site-X.***

La secuencia de ADN entre las regiones curvadas es tiene estructura ADN tipo B lineal, sin ningún tipo de variación importante en valores de los pares de bases en la inclinación y en el giro helicoidal. No obstante, es interesante denotar que en la región central del Site-X se encuentran una sucesión de seis adeninas consecutivas que otorgan a la secuencia unos valores estructurales singulares y diferentes al ADN tipo B, similar a como se ha observado en otros casos con anterioridad, como el de una estructura del dodecámero con secuencia CGCAAAAACGC, que contiene seis adeninas en la parte central de la secuencia (código pdb 1D98) (Nelson, Finch et al. 1987). En esta estructura dodecamérica se observó que la secuencia de poli-adeninas genera una disminución del ancho del surco menor (9.5 Å) del ADN (figura 29), provocando la compactación del apilamiento del ADN y dotándolo de mayor rigidez. Esto es consecuencia de una variación de la inclinación de la adenina hacia la timina del par de bases siguiente la cual permite que se forme un enlace de hidrógeno extra entre el átomo O4 de la adenina y el N6 de la timina (Nelson, Finch et al. 1987).

La superposición de Site-X con el dodecámero por la región rica en adeninas (desde A<sub>4</sub> a C<sub>12</sub> del dodecámero y desde A<sub>5</sub> a T<sub>13</sub> en Site-X) (figura\_29a) da lugar a un valor de r.m.s.d. de 1.3 Å. El ancho del surco menor es muy parecido entre ambos ADN ricos en adeninas (figura\_29b), tienen un ancho medio de 9.5 Å (calculado con CURVES+ (Lavery, Moaker et al. 2009; Lavery, Zakrzewska et al. 2009). De hecho todos los complejos el ADN unido con TFAM tiene un surco menor más estrecho que un ADN tipo B, como por ejemplo CTACAATGTTGCAA (código pdb 3BSE) (figura\_29b) pero en cambio la región central de Site-X se estrecha en un tramo más de pares de bases (figura\_29b). Estos resultados sugieren que esta región del ADN conserva las propiedades estructurales previas a la interacción y que TFAM no genera ninguna variación o distorsión de la misma.

**Figura\_29. Comparación Site-X con Dodecámero.**

**a)** se muestra la superposición de los ADN de la estructura 1D98 (naranja) con Site-X (color violeta cadena B y azul pálido cadena C) en complejo con TFAM (semi-transparente), en la cual se observa el gran parecido estructural entre ambas en la región poli-adeninas. **b)** se representa el valor de la anchura del surco menor de las estructuras de Site-X (negro), del dodecámero (1D98) (naranja), LSP 22 pb (cian) y B-ADN de 16pb (rojo). En ordenadas se indica el ancho del surco y en abscisas la secuencia de Site-X. Con el fondo amarillo se indica la región del ADN que une a HMGbox1 y con un fondo verde la región unida por HMGbox2.



### 8.1.1.5.6 Comparación las interacciones con el ADN entre distintos complejos de TFAM.

Los principales residuos implicados en unir, insertar y doblar al ADN no varían en las distintas estructuras de TFAM en complejo con los diferentes ADN. Se analizaron los valores de inclinación y de giro helicoidal de las bases en las zonas donde se produce el doblamiento del ADN y se observó que todos los complejos poseen valores muy parecidos, poseen un valor máximo de inclinación de aproximadamente  $50^\circ$ , y además se produce en todos exactamente donde se insertan las Leu58 y Leu182, como ya habían indicado trabajos anteriores de Rubio-Cosials, y coautores, y Ngo y coautores (Rubio-Cosials, Sidow et al. 2011; Ngo, Lovely et al. 2014). Sí se observan reorientaciones de rotámeros asociadas a desplazamientos de la molécula de ADN (especialmente nsADN, por ejemplo) debido a que la distancia entre algunos residuos y el ADN sea diferente entre TFAM/Site-X y los distintos complejos. En este sentido en SITE-X se encuentran algunas interacciones de HMGbox1 con el ADN que no están presentes en TFAM/LSP22: Lys51 interactúa con el fosfato de A<sub>3</sub> (de la cadena B), Lys62 N $\zeta$  con el de C<sub>4</sub> (de la cadena B), Gln100 con el de T<sub>20</sub>, Tyr103 con el de A<sub>22</sub>, y Trp107 con el de G<sub>23</sub>. El residuo

Val54 contacta los anillos de ribosa de las bases A<sub>3</sub> y C<sub>4</sub> (cadena B). Los contactos con las bases que no están presentes en TFAM/HSP1 y sí en TFAM/Site-X son Val166 por van der Waals con T<sub>7</sub> (cadena C) y Arg157 con N3 de A<sub>19</sub> (cadena B); además de los contactos de Arg82 con el fosfato de T<sub>18</sub>, Gln100 con el de T<sub>20</sub> y Arg159 con T<sub>7</sub>; la Lys52 produce un cambio de rotámero y contacta con un fosfato en HSP1 y no con la base como en Site-X. Es de destacar que la hebra B de TFAM/snADN se aleja de la HMGbox1 y se pierden los siguientes contactos con los fosfatos de Lys51 con los de A<sub>3</sub>, Lys62 con los de C<sub>4</sub>, Tyr103 con los de G<sub>22</sub>, Trp107 con los de G<sub>23</sub> y Val54 con la ribosa de C<sub>4</sub> que si están presentes en TFAM/Site-X.

Las diferencias más notables entre TFAM/Site-X y el resto de estructuras ocurren en el contacto del “*linker*” con el surco menor del ADN. Las diferencias se dan por cambios de la orientación del eje longitudinal del “*linker*” respecto de las dos hebras de la molécula de ADN, que de hecho va variando de una estructura a otra. Un factor que parece condicionar la posición de la hélice del “*linker*” respecto del surco menor es la anchura de éste, ya que permite (o no) a las cadenas laterales entrar en él. Tal como se ha mencionado en el apartado 8.1.1.5.5., el surco es especialmente estrecho en TFAM/Site-X. La distancia entre los fosfatos de las bases A<sub>10</sub> (cadena B) y T<sub>17</sub>(C) es de 9.03 Å; T<sub>11</sub>(B) -T<sub>16</sub>(C), 8.4 Å; T<sub>12</sub>(B)-T<sub>15</sub>(C), 9.1 Å, pero también lo es en la estructura TFAM/nsADN (8.7, 9.2 y 11.6 Å en los fosfatos equivalentes) y un poco menos estrecho en TFAM/HSP1 (10.3, 10.4, 12.6 Å). En cambio, en LSP 22pb el surco es más ancho (10.4, 11.2, 13.5 Å), lo que permite un acercamiento de las cadenas principales y laterales hacia el interior del surco (Rubio-Cosials, Sidow et al. 2011). Por ejemplo, en TFAM/Site-X la distancia entre Cα de Lys147 con los fosfatos de T<sub>11</sub> (B) y T<sub>16</sub> (C) es de 6.1 y 8.3Å, respectivamente, más largas que en LSP, donde las distancias con los fósforos equivalentes son 5.2 y 7.1 Å, debido al acercamiento del “*linker*” al surco menor en esta última estructura. A pesar de las diferencias entre estructuras, en todas ellas la cadena lateral del residuo Lys147 se encuentra siempre extendida y orientada al surco menor del ADN y se ubica entre las dos cargas de fosfato de hebras diferentes.

#### **8.1.1.6 Análisis complejo TFAM/DNA mediante EMSA.**

La interacción entre Site-X y TFAM se analizó mediante electroforesis no desnaturizante (EMSA), lo que permitió establecer las condiciones de ratio proteína:ADN en el complejo y de esta manera abordar los ensayos de cristalización, así como comparar la interacción de TFAM con las distintas secuencias que se emplearon para los estudios calorimétricos.

Para realizar los ensayos de ITC se precisó una secuencia control en la cual el ADN no pudiera ser doblado por TFAM (poli-dA) y otra que debería ser doblada fácilmente (poli-dAC), tal y como se explica con más detalle en el siguiente apartado 8.1.1.7, así que como análisis previo

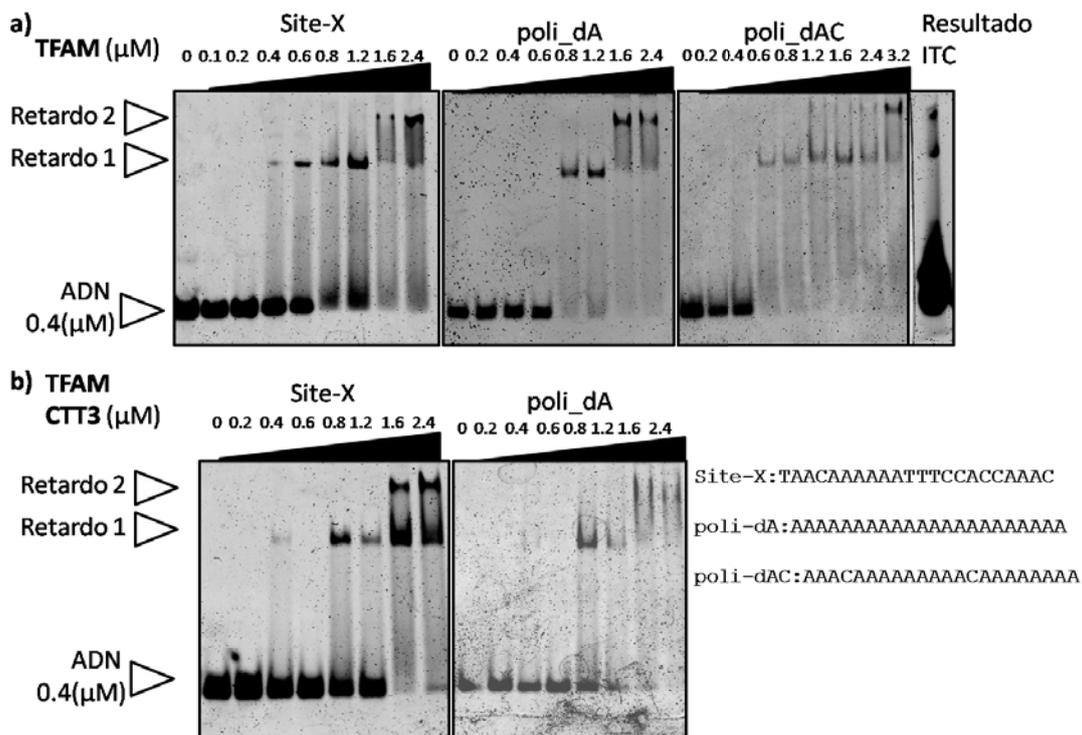
a los ensayos calorimétricos se quiso comprobar que se producía unión entre TFAM y los oligonucleótidos poli-dA y poli-dAC mediante EMSA, además de comparar su unión con la de Site-X.

Con las tres secuencias (site-X, poli-dA y poli-dAC; figura\_30) el protocolo que se empleó en los ensayos EMSA consistió en aumentar la concentración de TFAM de 0.2 a 3.2  $\mu\text{M}$  respecto a una cantidad fija de ADN (0.4  $\mu\text{M}$ ) Site-X, poli-dA y poli-dAC (figura\_30a). Se observó que cuando el ratio proteína:ADN era 2:1 aparecía una primera banda del ADN retardado. Posteriormente, el aumento la concentración de proteína hasta un ratio 4:1 dio lugar a un segundo retardo superior, tal y como anteriormente se había visto con la secuencia LSP empleando el método de diálisis para la generación de los complejos (ver apartado Materiales, Métodos y Fundamentos en la Tesis Rubio-Cosials, A.).

Tampoco se apreciaron diferencias al repetir el mismo protocolo del EMSA con el constructo de TFAM-CTT3 al que se le sustrajo la cola C-terminal (desde el residuo 235 al final) con las secuencias Site-X y poli-dA (0.4  $\mu\text{M}$ ) (figura\_30b). El resultado fue la aparición de dos bandas retardadas, la primera con un ratio 2:1 proteína:ADN y la segunda con 4:1, para ambas secuencias.

**Figura\_30. Geles EMSA de TFAM.**

Se muestran los geles EMSA de los retardos de TFAM, en **a)** y del constructo TFAM-CTT3 (sin los 7 últimos residuos), en **b)**, con las secuencias Site-X y poli-dA para ambas y poli-dAC solo con TFAM. Con las flechas de la izquierda se indican las bandas de ADN en las cuales corresponde al ADN libre o en complejo con TFAM (retardos 1 y 2). Además se muestra en **a)** la migración en gel y las bandas retardadas del resultado de la titulación por ITC del experimento de TFAM con ADN (ver más adelante). Se indican en la parte inferior izquierda las secuencias 5' de los oligonucleótidos empleados en los ensayos EMSA.



Por tanto, en conclusión, por electroforesis se observó el mismo efecto con ratio proteína:ADN similares entre todas las secuencias (figura\_31a), siendo este resultado coherente con el valor de estequiometría que se obtuvo por calorimetría (como se verá en el apartado siguiente).

#### **8.1.1.7 Calorimetría de Titulación Isoterma de la interacción de TFAM con el ADN.**

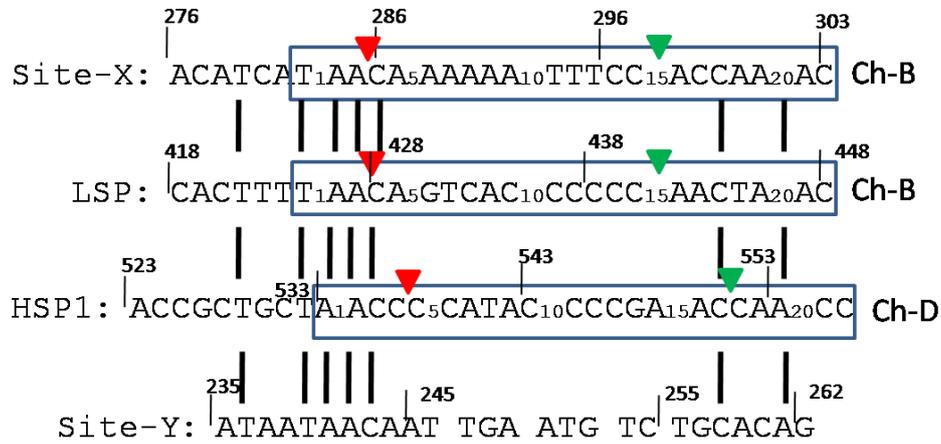
Se realizó la caracterización de termodinámica del complejo TFAM/Site-X mediante calorimetría de titulación isoterma (ITC), obteniendo los valores de entalpía, entropía, la constante de afinidad ( $K_a$ ) y la estequiometría del complejo. Como control y referencia se realizó también la caracterización TFAM/LSP, analizada anteriormente por Malarkey y coautores (Malarkey, Bestwick et al. 2011), cuyo análisis mostraron que los procesos de formación de los complejos fueron endotérmicos. El estudio calorimétrico del complejo TFAM/Site-X también mostró un proceso endotérmico (figura\_32), lo que significa que fue entálpicamente desfavorable pero compensado por una variación positiva de la entropía, dando lugar a una diferencia negativa de energía libre de Gibbs ( $\Delta G$ ), siendo por tanto un proceso espontáneo (figura 32) (ver tabla\_13). Estudios previos con otras proteínas con dominios HMGbox (hSRY, mSox, mLEF1, NHP6A, HMGB-box 2 y HMG-D100 (Dragan, Read et al. 2004) que forman complejo con secuencias de ADN específicas y no-específicas también mostraron en todos los casos un proceso endotérmico (Dragan, Read et al. 2004; Malarkey and Churchill 2012). Dichos estudios permitieron relacionar que las proteínas de unión al surco menor que doblan el ADN forman el complejo proteína/DNA mediante un proceso efectivamente endotérmico (Jen-Jacobson, Engler et al. 2000).

Es notorio que en las secuencias específicas de unión de algunas proteínas con dominios HMGbox como son SRY, Sox-5 y LEF-1 se produce una inserción en la secuencia (T/A) $\downarrow$ CAA (Dragan, Read et al. 2004), que coincide con el patrón observado para el primer sitio insertado por Leu58 en los complejos TFAM/Site-X y TFAM/LSP (figura\_31).

Para comprobar la relevancia del patrón de la secuencia A $\downarrow$ CAA en la dobladura del ADN por TFAM se diseñó como control negativo un oligonucleótido de doble cadena con una secuencia de 22 adeninas como control negativo de unión sin doblar el ADN, debido a que una secuencia rica de adeninas consecutivas (poli-dA) produce una gran rigidez al oligonucleótido (Nelson, Finch et al. 1987) (McConnell and Beveridge 2001) y por tanto debería ser más difícil de doblar por parte de TFAM. Además, en los complejos TFAM/Site-X y TFAM/LSP la máxima curvatura del ADN corresponde con la región en la cual Leu58 (HMGbox1) y Leu182 (HMGbox2) insertan el paso AA $\downarrow$ CA y CC $\downarrow$ AC respectivamente, a una distancia de 10pb. Así que como control positivo en los ensayos calorimétricos, se diseñó un oligonucleótido de poli-adeninas en el cual sí estaba el motivo de inserción CA-10pb-AC (poli-dAC) (figura\_31).

**Figura\_31. Alineamiento secuencias de ADN que une TFAM.**

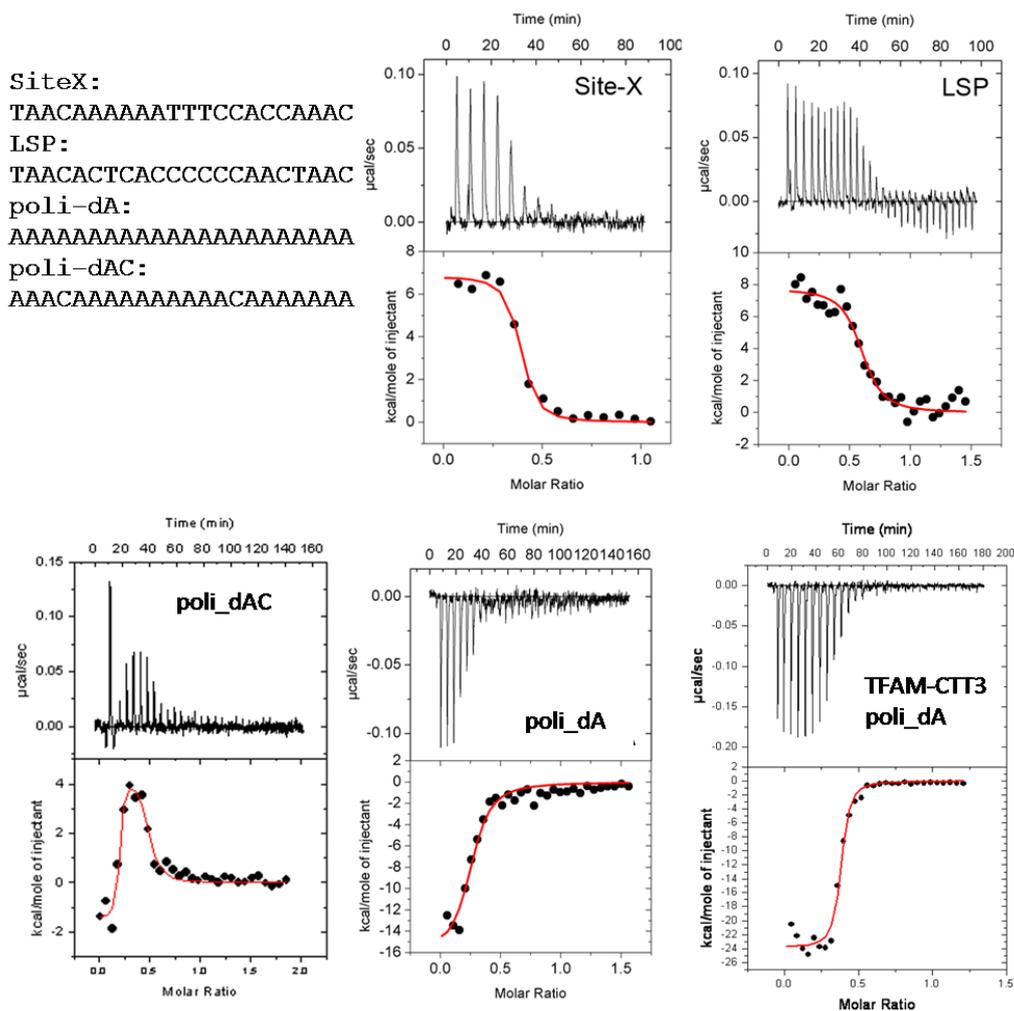
En la figura se muestra el alineamiento de las secuencias de unión de TFAM que Fisher identificó por "footprinting" (Fisher, Topper et al. 1987). En la parte superior de cada secuencia se indica la numeración de la secuencia del genoma. Con un rectángulo se distingue el fragmento cristalizado en cada complejo y con un subíndice se indica la numeración en el fichero PDB y a la derecha el nombre de la cadena correspondiente. Con un triángulo rojo y uno verde se indican las posiciones de semi-inserción de Leu58 y Leu182 respectivamente. Figura adaptada de (Fisher, Topper et al. 1987).



Como se mencionó anteriormente, se estudió la formación del complejo TFAM/poli-dA como control negativo. El resultado fue que TFAM formó un complejo con poli-dA termodinámicamente diferente, un proceso exotérmico, es decir, entálpicamente favorable, lo que sugiere que no se produce ningún tipo de dobladura en el ADN por el surco menor. Dado que este tipo de secuencias son rígidas y en este caso no presenta ningún motivo de reconocimiento para TFAM tipo CA-10pb-AC, pensamos que esa interacción no ocurría por las HMG box sino que podría estar mediada por la cola C-terminal de un modo similar a como ocurre con las proteína HMGA1-AT-hook (Dragan, Liggins et al. 2003). Las proteínas HMGA1-AT-hook interaccionan con el ADN por el surco menor con una cola C-terminal sin estructura secundaria, tal y como se ha descrito que es la cola C-terminal de TFAM. Se realizó el ensayo de ITC con el oligonucleótido poli-dA con el constructo TFAM-CTT3 en el cual se había eliminado la cola C-terminal (tabla\_13). El resultado fue muy parecido al de la proteína con cola, durante la formación del complejo TFAM-CTT3/poli-dA se producía un proceso exotérmico con las magnitudes de entalpía y entropía muy similares a TFAM/poli-dA. Este resultado fue coherente con los datos observados por EMSA en los cuales no se advirtió ninguna diferencia entre el retardo del ADN entre TFAM y TFAM-CTT3 con poli-dA (figura\_30).

**Figura\_32. Resultados ITC.**

Se indican las secuencias empleadas en el experimento de ITC en la parte superior izquierda. Se muestran un resultado y la integración de los datos del total de las tres medidas realizadas para cada uno de los complejos, indicando la secuencia a la que corresponde: Site-X, LSP, poli-dA, poli-dAC para complejo con TFAM y TFAM-CTT3/poli-dA para el constructo sin cola C-terminal.



El resultado de la interacción de TFAM con poli-dAC fue totalmente diferente a los anteriores, ya que se dieron dos procesos termodinámicos. Al inicio de la titulación se detectó un proceso exotérmico (llamado poli-dAC1) parecido al de poli-dA. No obstante, al aumentar la cantidad de ADN en la celda de medida se detectó un proceso endotérmico (poli-dAC2), consistente con que TFAM doblaría el ADN por el surco menor, como sucede en el caso de las secuencias Site-X, LSP, y las observadas para las demás HMGbox.

Se realizó un ajuste de los datos con el algoritmo de “un sitio de unión” para todos los datos excepto para el complejo TFAM-poli-dAC que se empleó el algoritmo “dos sitios secuenciales de unión”.

En la tabla 13 se puede observar la gran variabilidad del valor de la estequiometría, sin embargo se obtiene un valor medio de  $n=0.3$  para los ADN Site-X, LSP, poli\_dAC2 y poli-

dA/TFAM-CTT3, que significa que tres proteínas se unen con una molécula de ADN. Para los ADN poli-dA y poli-dAC1 el valor de n disminuye a 0.22, lo que significarían cuatro proteínas unidas a una molécula de ADN. Después de la titulación, la muestra TFAM/Site-X se analizó por EMSA (figura 30) y se observó que existen dos bandas de retardo, una de ellas migra igual que el complejo proteína:ADN 2:1, y la segunda con el ratio 4:1. Esta observación sugiere que la variabilidad en las medias que oscilan de 0.25 a 0.45 sería debida al equilibrio entre ambos estados estequiométricos.

Tabla\_13. Resultados ITC.

| Secuencias        | Ka (M <sup>-1</sup> ) | n    | ΔH (kcal.mol <sup>-1</sup> ) | ΔS (kcal.mol <sup>-1</sup> .K <sup>-1</sup> ) | ΔG (kcal.mol <sup>-1</sup> ) |
|-------------------|-----------------------|------|------------------------------|---|------------------------------|
| Site-X            | 4.25E6                | 0.32 | 7721                         | 56.7  | -8892                        |
| Site-X            | 3.10E7                | 0.36 | 6851                         | 57.6  | -10026                       |
| Site-X            | 3.36E7                | 0.41 | 8566                         | 63.6  | -10069                       |
| LSP               | 1.30E7                | 0.59 | 7737                         | 58.9  | -9521                        |
| LSP               | 1.32E7                | 0.22 | 10430                        | 68.2  | -9553                        |
| LSP               | 1.09E7                | 0.39 | 6939                         | 55.9  | -9440                        |
| poli_dA           | 2.43E6                | 0.36 | -16110                       | -24   | -9078                        |
| poli_dA           | 5.20E6                | 0.22 | -19690                       | -36   | -9142                        |
| poli_dA           | 9.12E6                | 0.18 | -14040                       | -16   | -9352                        |
| poli_dAC1         | 1.14E9                | 0.18 | -1498                        | 136   | -22063                       |
| poli_dAC1         | 6.84E9                | 0.20 | -1220                        | 140.8   | -23958                       |
| poli_dAC1         | 7.56E7                | 0.33 | -605                         | 134   | -19192                       |
| poli_dAC2         | 1.04E7                | 0.30 | 4405                         | 247   | -17093                       |
| poli_dAC2         | 1.64E7                | 0.31 | 3296                         | 244   | -17575                       |
| poli_dAC2         | 2.95E6                | 0.37 | 5967                         | 250   | -15760                       |
| TFAM-CTT3/poli_dA | 5.79E7                | 0.37 | -23850                       | -45.8   | -22934                       |
| TFAM-CTT3/poli_dA | 3.37E7                | 0.42 | -22500                       | -42.3   | -21654                       |
| TFAM-CTT3/poli_dA | 1.66E7                | 0.28 | -27150                       | -59.6   | -25958                       |

\*todos los resultados corresponden a ensayos de TFAM con distintos ADN excepto aquellos que se indican como TFAM-CTT3/ADN.

La caracterización termodinámica también permitió constatar que no hay diferencia importantes de afinidad de TFAM con las distintas secuencias excepto para el ADN poli-dAC donde se apreció un aumento de dos órdenes de magnitud de la K<sub>a</sub> del proceso exotérmico, respecto a todas las demás secuencias.

#### **8.1.1.8 Análisis computacional de las propiedades físicas de la secuencia Site-X.**

Para intentar comprender mejor la formación del complejo entre TFAM/Site-X se realizó un estudio teórico de su rigidez (o flexibilidad) y de la estructura en disolución previamente a su contacto con la proteína por métodos computacionales y de dinámica molecular en colaboración con la Dra. F. Battistini del grupo del Prof. M. Orozco del Institut de Recerca Biomèdica (Barcelona).

La predicción de la flexibilidad de una secuencia de ADN se realizó mediante una dinámica molecular, de 100 ns a partir de la cual se calcularon los valores medios de la inclinación, giro helicoidal y estiramiento. A partir de estos parámetros, que desde un punto de vista mecánico tienen un comportamiento armónico, se estimó el mínimo de energía de la secuencia (Lavery, Zakrzewska et al. 2009). Como medio de disolución en el cálculo se consideró agua (Perez, Luque et al. 2007).

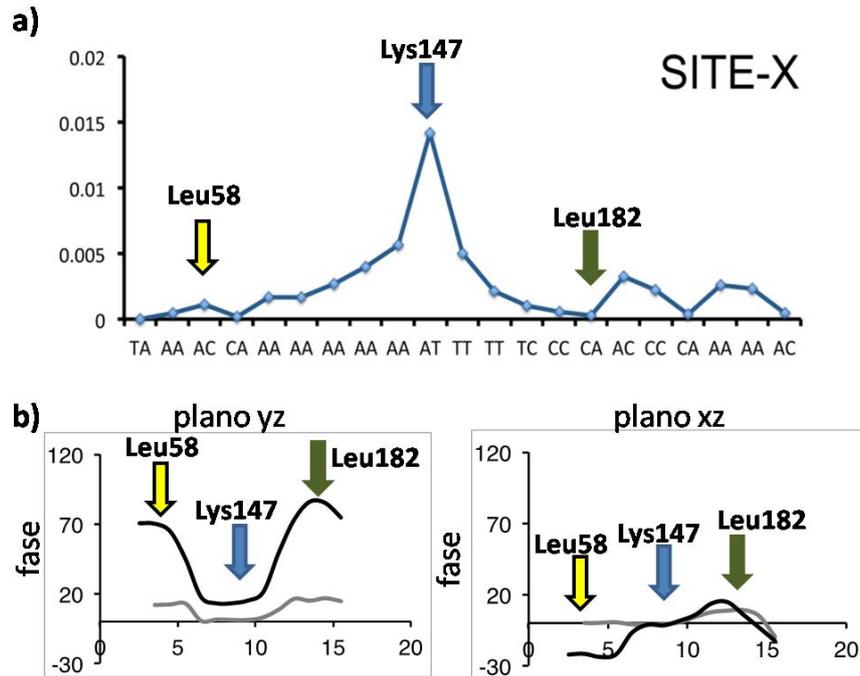
El cálculo del resultado de la rigidez de la secuencia Site-X mostró un perfil en el cual se produce un máximo asociado a la secuencia rica en adeninas (figura\_33), justo en el paso donde la cadena lateral del residuo Lys147 se extiende hacia el surco menor quedando el grupo amino entre los fosfatos de las timinas T<sub>11</sub> y T<sub>16</sub> de la cadena B y C, respectivamente (ver sección 8.1.1.5.2.). Por el contrario, los pasos insertados por Leu58 y Leu182 muestran una rigidez menor, especialmente es mínima para ésta última, es decir, son mucho más flexibles.

El cálculo de la linealidad de la molécula de ADN se calculó a partir de los valores de inclinación y giro helicoidal calculados anteriormente, utilizando los valores de la media de una trayectoria de dinámica molecular de 80ns, en el equilibrio. El producto del valor de la inclinación por el giro helicoidal proporciona un parámetro denominado fase que determina la linealidad del ADN en tres dimensiones, y se representa con dos gráficas bidimensionales, en cada una se representa el valor de la fase en un plano distinto xz e yz (Battistini, Hunter et al. 2010). Mediante los parámetros derivados por dinámica molecular mencionados para cada media de cuatro pares de bases consecutivos se pudo predecir los parámetros de inclinación y de giro helicoidal que poseen la secuencia Site-X en disolución sin unir a TFAM y se compararon con los valores de Site-X en la estructura cristalográfica en complejo con TFAM (figura\_33b).

Se observó que el giro helicoidal no presentaba casi variaciones en las estructuras cristalográficas, mientras que la inclinación entre pares de bases sí mostró una gran variabilidad a lo largo de la secuencia, siendo este el parámetro predominante que define la curvatura general del ADN en los cristales. En la figura\_33b se muestra el gran doblamiento que se produce en los primeros cinco pasos donde la inclinación alcanza casi los 100° en el plano yz. En el plano xz se encuentra una pequeña pero relevante inclinación en los cinco pasos del extremo, desde T<sub>1</sub>AAC<sub>4</sub> hasta A<sub>5</sub>AAA<sub>8</sub> al principio del oligonucleótido, y desde C<sub>15</sub>ACC<sub>18</sub> hasta A<sub>19</sub>AAC<sub>22</sub>. La predicción de la inclinación a lo largo de Site-X (sin unir) mostró un hecho interesante: los valores predichos para cada paso no son tan marcados pero muestran el mismo signo que para la estructura cristalográfica, especialmente donde se insertan Leu58 y lo que sugiere que la secuencia Site-X libre ya está doblada antes de ser contactada, y con la interacción se agudiza el grado de torsión.

**Figura\_33. Propiedades mecánicas de Site-X.**

En **a)** se indica el perfil de rigidez de cada paso de dos pares de bases de la secuencia Site-X. En **b)** se indica en gris la predicción de la inclinación de las bases del ADN en función del plano zy a la izquierda y el plano xz a la derecha. Con las flechas de colores se indica el paso en el cual los residuos que insertan, indicando el nombre y número.



## 8.1.2 Discusión de los resultados del estudio de TFAM en complejo con Site-X.

### 8.1.2.1 Estructura TFAM/Site-X.

La estructura cristalográfica del complejo TFAM/Site-X muestra como el ADN con la secuencia Site-X es doblado hasta  $180^\circ$  debido a la unión de los dos dominios HMGBbox al surco menor insertando dos leucinas, una por dominio HMGBbox, provocando que se doble  $90^\circ$  por cada unión. De modo global, la estructura presenta una estructura de “U-turn”, típica de todas las estructuras cristalográficas de TFAM en complejo con distintas secuencias de ADN. Las diferencias de los valores de r.m.s.d. de los cuatro complejos de la u.a. son muy pequeñas, y son generadas por variaciones en la posición relativa entre las HMGBbox1 y 2 en cada complejo.

Los contactos entre residuos de la hélice 3 del dominio HMGBbox1 que están implicados en la dimerización en los complejos TFAM/HSP1 y TFAM/nsADN (Ngo, Lovely et al. 2014) también fueron observados entre distintas moléculas de la u.a. del complejo TFAM/Site-X. Además, estos contactos se han descubierto en otros complejos como en las estructuras TFAM/LSP22 (Rubio-Cosials, Sidow et al. 2011) TFAM/LSP28 (Ngo, Kaiser et al. 2011) entre proteínas de la unidad asimétrica y proteínas que forman el empaquetamiento cristalino.

En la estructura del complejo TFAM/Site-X se observa que la interacción entre dos núcleos hidrofóbicos centrados en los residuos Leu58 y Leu182 son los responsables de la gran distorsión que sufre el ADN en sus parámetros de inclinación y giro helicoidal, además de la unión entre ambas macromoléculas. Este patrón es el mismo que se observó anteriormente en los otros complejos TFAM/LSP (Rubio-Cosials, Sidow et al. 2011), TFAM/HSP1 y TFAM/nsADN (Ngo, Lovely et al. 2014).

Los contactos que se producen entre los residuos del *"linker"* y las dos HMGbox con el ADN en TFAM/Site-X son prácticamente los mismos en todos los complejos de TFAM con otras secuencias, lo que sugiere que esas interacciones tienen una gran importancia en la formación y/o estabilidad del complejo proteína/ADN.

El *"linker"* presenta una gran flexibilidad y capacidad de adaptación para interactuar con distintas moléculas de ADN, y también le permite adaptarse a las pequeñas rotaciones de la HMGbox1 respecto a la HMGbox2. Las divergencias entre los complejos se originan en dos puntos localizados: el primero está definido a partir de Cys48 y Lys118 en HMGbox1, a partir de estos residuos se produce una basculación de las HMGbox1 y *"su"* ADN respecto a la posición del *"linker"* en los distintos complejos; el segundo se origina por variaciones en la curvatura de la hélice del *"linker"* en la región donde interactúa con el ADN, a partir del residuo His137. Los movimientos de la mitad C-terminal del *"linker"* desplazan conjuntamente a HMGbox2 y el ADN que interactúa con ella.

Los residuos del *"linker"* que contactan al ADN son muy similares en los distintos complejos de TFAM con el ADN, sin embargo los contactos se realizan con fosfatos no equivalentes en las distintas estructuras, gracias a cambios de rotámetros de las cadenas laterales de los susodichos residuos. Una de las funciones de los residuos básicos del *"linker"* es evitar las interacciones repulsivas como consecuencia del acercamiento de los fosfatos en las zonas dobladas. Hay dos residuos que sí presentan la misma orientación en casi todas las estructuras: Met137 y Lys147, cuyas cadenas laterales se sitúan entre los fosfatos del surco menor en la región opuesta y central entre las regiones dobladas del ADN. Estas interacciones podrían facilitar el posicionamiento del *"linker"* respecto al ADN.

La región central del complejo TFAM/Site-X, que posee una región rica en adeninas, no presentó diferencias estructurales con la región de seis adeninas de un ADN dodecámero. El surco menor presenta el mismo valor de anchura, y la superposición de ambas estructuras en esta región presenta un valor de r.m.s.d muy bajo (1.3 Å). Este hecho sugiere que la región central de Site-X conserva la misma estructura antes y después de unirse TFAM, a pesar de la torsión del ADN.

La comparación de los dos dominios de HMGbox1 y 2 entre ellos de las estructuras de todos los complejos TFAM/ADN muestran diferencias mínimas en el plegamiento, indicativo que de modo independiente al ADN que se unan conservan un mismo plegamiento. Sin embargo, la HMGbox2 de los complejos TFAM/Site-X (D/EI) y (G/HI), presenta en el giro entre las hélices 2 y 3 dos conformaciones diferentes, la primera es casi idéntica al resto de estructuras, mientras que en la segunda conformación se observa un desplazamiento del giro acompañado de una doble conformación del ADN, que también se desplaza. Esta doble conformación que se produce conjuntamente entre la proteína y el ADN evidencia la dinámica y plasticidad de la unión en esa región concreta, y la necesidad de HMGbox2 para estabilizar la forma doblada del ADN dando lugar a la conformación preferente que se encuentra en todos los complejos. Otro dato que refuerza la idea que la unión de HMGbox2 con el ADN más dinámica o desordenada que HMGbox1 es el hecho que la densidad electrónica de las cadenas laterales de algunos residuos de HMGbox2 implicados en la unión con el ADN es más pobre que los residuos equivalentes de HMGbox1.

Los residuos de HMGbox1 y 2 que doblan Site-X se encuentran también en los distintos complejos con diferentes secuencias de ADN realizando los mismos contactos, además las dos zonas de inserción en todos los complejos distan entre sí 10 pb. Este hecho es indicativo de la importancia de estas interacciones para la formación del complejo. Las principales diferencias que se han encontrado entre los complejos están localizados en algunos rotámeros de residuos que interaccionan con el esqueleto de fosfatos del ADN, posiblemente estas divergencias vienen determinadas por las variaciones estructurales intrínsecas a cada secuencia de ADN Site-X, LSP, HSP1 y nsADN. En la estructura de TFAM/Site-X en la región de unión de HMGbox2 se observa un mapa de densidad pobre en las cadenas laterales de Leu182 en los complejos X/YZ de la u.a. Este fenómeno es indicativo del desorden (probablemente en varios rotámeros de la leucina) en esta zona de la estructura y refleja la dinámica de la unión, hecho que se refuerza con la doble conformación de la cadena principal y del ADN de los complejos TFAM/Site-X D/EI y G/HI.

El posible mecanismo de unión entre TFAM y Site-X podría darse por la interacción inicial de HMGbox1 que impone una fuerte torsión del ADN y según la resistencia mecánica de cada oligonucleótido determina la posición relativa de HMGbox1, que contacta al "linker" de manera rígida, el cual gracias a su por otra parte flexibilidad estructural se adapta a la estructura del ADN apantallando los contactos repulsivos entre los fosfatos del esqueleto de ADN y simultáneamente condicionando (gracias a la de los contactos entre ambos) la posición de unión de HMGbox2 en el ADN. El contacto de HMGbox2 con el ADN no parece tan estable como el de HMGbox1, tal y como sugiere el desorden observado en TFAM/Site-X G/HI.

### **8.1.2.2 Propiedades mecánicas del ADN Site-X y su reconocimiento por TFAM.**

El análisis por dinámica molecular del ADN Site-X mostró que la región central de la secuencia presenta una gran rigidez, el paso con mayor rigidez muestra el contacto con Lys147 del “linker”. La zona en la cual se inserta Leu58 presenta una rigidez moderada, y según estudios de dinámica molecular, presenta un doblamiento previo a la unión de TFAM pero de menor magnitud al que se observa en la estructura cristalográfica. A una distancia de 10pb, se encuentra una región donde la rigidez es mínima y previamente doblado, exactamente en el paso de ADN donde inserta Leu182.

Por tanto, Site-X presenta unas propiedades físicas intrínsecas favorables para que TFAM pueda formar el complejo y doblarlo en forma de “U-turn”, tal y como que se encontró en la estructura cristalina. Las HMGbox reconocerían un ADN con una estructura pre-doblada que se distorsionaría hacia la forma doblada final de 180° y que se estabilizaría por la inserción de los residuos Leu58 y Leu182 en dos pasos del ADN desfavoreciendo el apilamiento entre pares de bases.

### **8.1.2.3 Análisis termodinámico de la formación del complejo TFAM/Site-X.**

La formación del complejo de TFAM con Site-X, al igual que ya había sido descrito para las secuencias LSP, HSP1 y un ADN no específico (Malarkey, Bestwick et al. 2011) presenta un proceso termodinámico endotérmico en el cual el aumento entrópico dirige la formación del complejo. Este comportamiento no es exclusivo de TFAM, ha sido descrito anteriormente en proteínas HMGbox (Dragan, Read et al. 2004) y otras proteínas de unión al surco menor que las cuales doblan al ADN (Jen-Jacobson, Engler et al. 2000).

Las interacciones que se observan en la estructura cristalográfica entre TFAM y Site-X en las regiones dobladas presentan una combinación de contactos hidrofílicos, principalmente con el esqueleto de ADN, e hidrofóbicos insertándose entre las bases del ADN. Según los trabajos en otras HMGbox (SRY, Sox-5 y LEF-1) (Dragan, Liggins et al. 2003; Dragan, Read et al. 2004; Dragan and Privalov 2008) la inserción de residuos en el surco menor provocan la liberación de moléculas de agua enclaustradas previamente en el surco menor y como consecuencia generan una gran ganancia entrópica en el sistema compensando el gasto energético que conllevaría doblar al ADN. Las interacciones hidrofóbicas descritas en las estructuras cristalográficas de TFAM con distintas secuencias de ADN son cuantitativamente similares entre ellas, así como también lo son los valores termodinámicos de la diferencia de energía libre de Gibbs y por tanto la constante de afinidad; estos datos sugieren una relación entre los enlaces observados en la estructura y la estabilidad termodinámica de los mismos.

En la estructura cristalográfica de TFAM/Site-X, los complejos D/EI y G/HI muestran una doble conformación en el giro entre la hélice 2 y 3 de HMGbox2. El desplazamiento de la proteína en esa región está relacionado con un alejamiento del ADN, incrementándose la apertura del surco menor. Estas dos conformaciones son indicativas de la flexibilidad de la HMGbox2 en su unión al ADN. Se han descrito para otras proteínas HMGbox una flexibilidad intrínseca que produce un cambio conformacional adoptando un estado más rígido como consecuencia de su unión al ADN (Privalov, Jelesarov et al. 1999; Dragan, Liggins et al. 2003; Dragan, Read et al. 2004). Es posible que las variaciones estructurales observadas en la HMGbox2 de TFAM sean originadas por un proceso de plegamiento como el observado en otras HMGbox.

Para estudiar los posibles mecanismos de reconocimiento de TFAM al ADN se realizaron la caracterización termodinámica de los complejos TFAM/poli-dA y TFAM/poli-dAC. TFAM/poli-dA contiene 22 adeninas por tanto tiene una elevada rigidez y un estrecho surco menor y, en consecuencia, TFAM no debería ser capaz de doblarlo. Como control al mismo ADN se sustituyeron dos adeninas por dos citosinas a una distancia de 10pb, coincidiendo con el modelo de plegamiento de la estructura TFAM/Site-X y TFAM/LSP, en los cuales se insertan los residuos implicados en doblar el ADN.

Para TFAM/poli-dA se observó una inversión en la naturaleza termodinámica del proceso de formación del complejo ya que fue exotérmico, por tanto la estabilidad del complejo vendría determinada por una entalpía negativa favorable, lo que sería indicativo de que, a pesar de que se produce interacción, y además con una afinidad muy parecida, no se dobla el ADN. Sin embargo, cuando la secuencia posee AACA y, a diez pares de bases, la repetición invertida ACAA, se producen dos procesos simultáneos, uno exotérmico TFAM/poli-dAC1 y otro endotérmico TFAM/poli-dAC2. Esto sugiere que hay un proceso de interacción como se produjo entre TFAM/poli-dA y otro de la misma naturaleza termodinámica que TFAM/Site-X, en el contacto tipo poli-dAC2 el ADN se dobla pero en poli-dAC1 no.

El ensayo de ITC del constructo TFAM-CTT3, en el cual se eliminó la cola C-terminal para formar complejo con el ADN poli-dA, dio el mismo resultado que para la proteína entera TFAM. Este resultado descarta la posibilidad que la interacción entre TFAM y poli-dA fuera consecuencia de unión por la cola C-terminal. En cambio, los datos termodinámicos de TFAM y poli-dA sugieren una interacción en la cual serían mayoritarias las interacciones electrostáticas, relacionadas con una entalpía favorable, entre el esqueleto del ADN y los residuos básicos de TFAM, aunque no se pudo discriminar exactamente qué dominio o región de la proteína interacciona. La posibilidad de la interacción entre el "linker" y el ADN de un modo similar al observado en la estructura cristalográfica podría ser una posibilidad.

Por ITC se ha encontrado un valor estequiométrico  $n=0.3$ , lo que significaría la unión de tres proteínas por una molécula de ADN. Cuando las muestras tras la titulación de calorimetría isotérmica se analizaron por EMSA se observó que existen dos que migraron igual que los dos retardos correspondientes a un ratio proteína:ADN de 2:1 y 4:1. Ello sugiere que el valor 3:1 observado por ITC corresponde a un valor medio de dos estados que se encuentran en disolución, 2:1 y 4:1. El primer retardo, en el cual el ratio es 2:1, sugiere que corresponde al dímero que observó con HSP1 mediante FRET (Ngo, Lovely et al. 2014).

**RESULTADOS Y DISCUSIÓN  
DEL ESTUDIO ESTRUCTURAL  
DE  
TWINKLE**



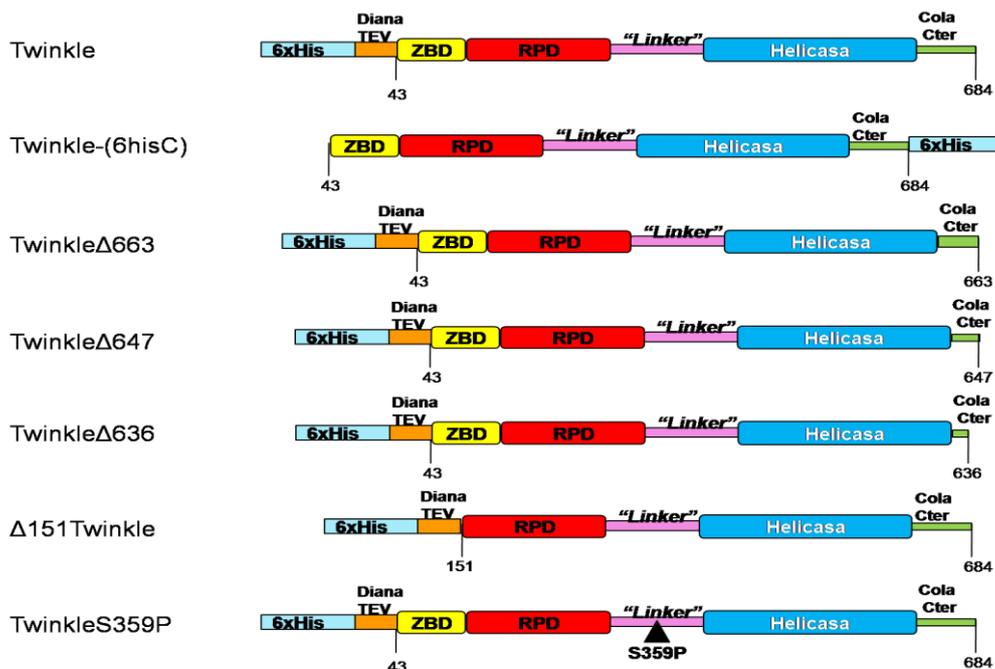


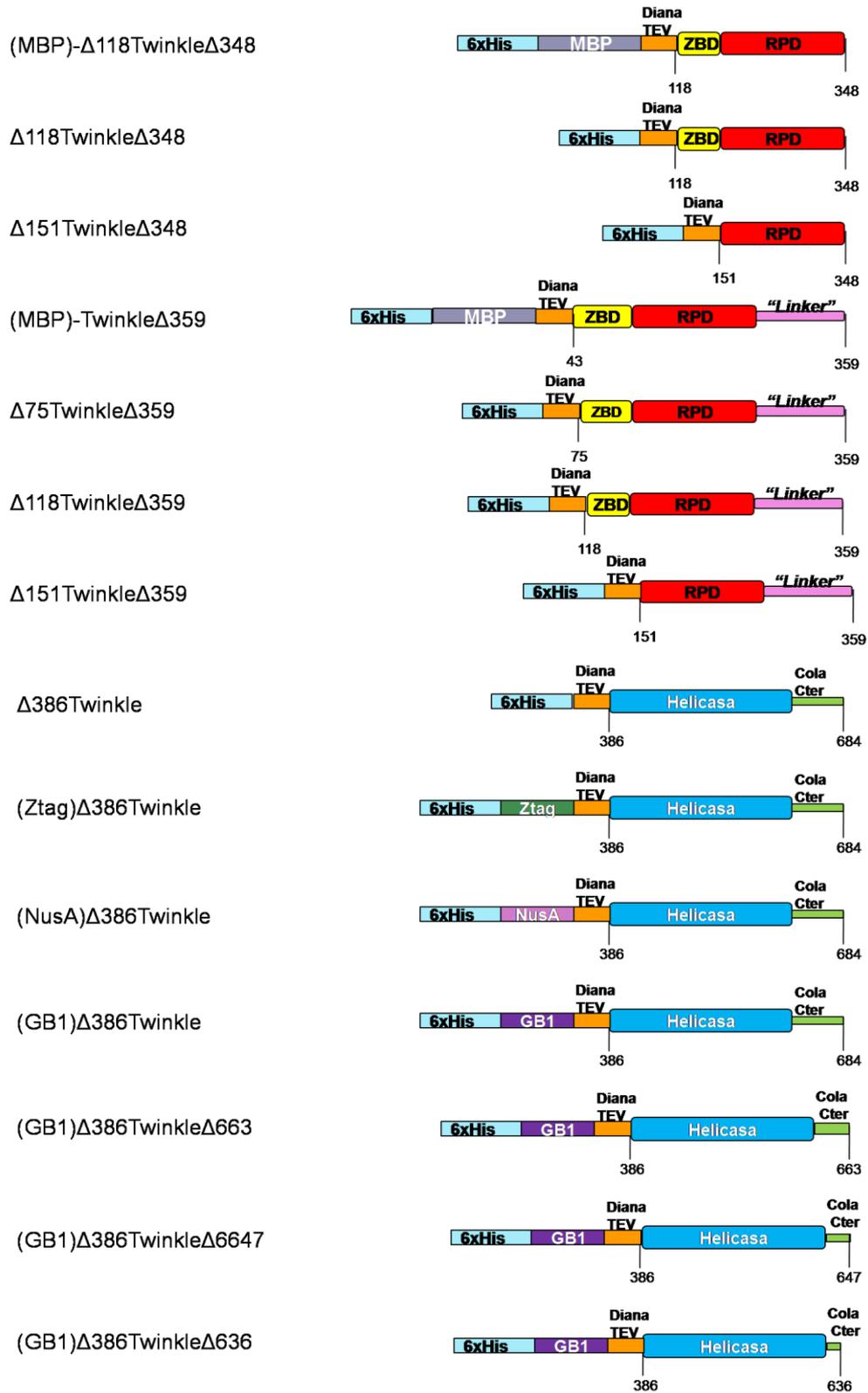
constructos con la intención que fueran dominios globulares estables y solubles, para lo que no se debía interrumpir ningún elemento de estructura secundaria, así como tampoco presentaran extremos desordenados que dificultaran la cristalización. Inicialmente se definieron 10 fragmentos Twinkle potencialmente estables: Twinkle (del residuo 43 al 684),  $\Delta 151$ Twinkle (151 al 684),  $\Delta 118$ Twinkle $\Delta 348$  (118 al 348),  $\Delta 151$ Twinkle $\Delta 348$  (151 al 348), Twinkle $\Delta 359$  (43 al 359),  $\Delta 75$ Twinkle $\Delta 359$  (75 al 359),  $\Delta 118$ Twinkle $\Delta 359$  (118 al 359),  $\Delta 151$ Twinkle $\Delta 359$  (151 al 359),  $\Delta 386$ Twinkle (386 al 684),  $\Delta 358$ Twinkle (348 al 684) (figura\_35a). Todos los clones que incluían el segmento N-terminal intacto empiezan en el residuo 43 de la secuencia ya que se eliminó la secuencia de traslocación de la mitocondria (figura\_35b). Estos fragmentos se fusionaron con diferentes proteínas que pudieran ayudar a la expresión, solubilidad y/o purificación, tales como MBP, GB1, TRX, NusA y Ztag (ver apartado 7.1 para su descripción). Todos los fragmentos, con o sin proteína de fusión, llevaban cola de histidinas en N-terminal y para el fragmento Twinkle se probó además con la cola de histidinas en C-terminal. De todo ello se obtuvieron 18 constructos diferentes (figura\_35a).

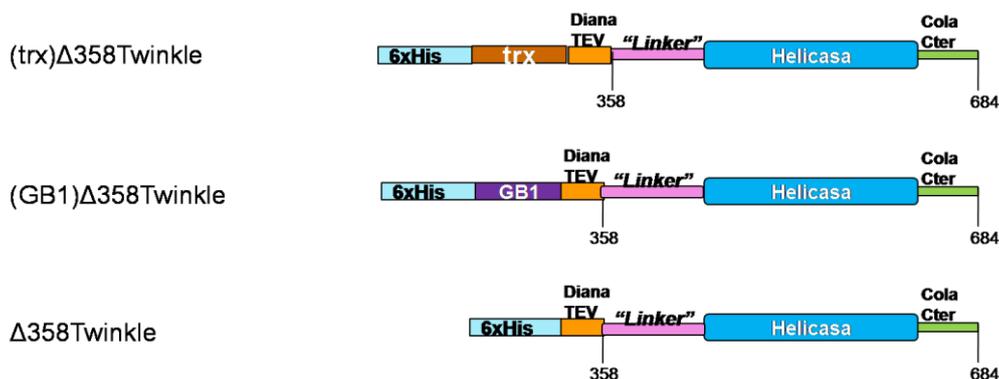
**Figura\_35. Constructos de Twinkle.**

En **a)** se indica el nombre de cada constructo y con un esquema de cajas de distinto color se indican los dominios ZBD (amarillo), RPD (rojo), "linker" (rosa), helicasa (azul), cola C-terminal (verde). Se indica el primer y el último residuo de la secuencia de Twinkle que define a cada constructo. Además se indica si tiene está fusionado a otra proteína MBP (morado), GB1 (violeta), TRX (marrón), NusA (rosa), Ztag (verde oscuro), colas de histidina (6xHis) (azul celeste) y el sitio de corte de TEV (naranja) y las mutaciones puntuales (triángulo negro). En **b)** se detalla la secuencia inicial del constructo Twinkle, además de indicar el peso molecular. En **c)** se muestra el alineamiento de la cola C-terminal de distintas secuencias de Twinkle. La conservación de los residuos se marca con distintos colores, con un cuadrado rojo se indica identidad y con una letra roja los residuos similares.

**a)**







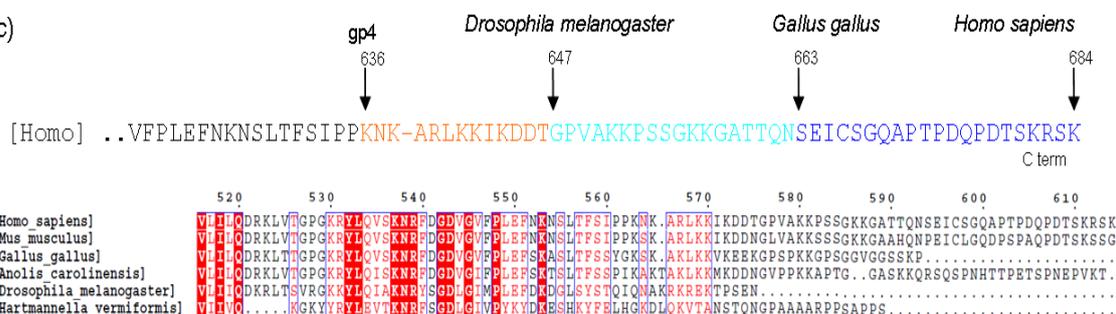
b)

CONSTRUCTOTWINKLE



Peso molecular 72KDa, Desde el residuo 43 al 684

c)



Entre la secuencia de la proteína de fusión y los constructos de Twinkle se incluyó la secuencia codificante para el sitio de corte de la proteasa TEV para que durante la purificación se pudiese escindir la proteína de fusión (figura\_35b).

Después de haber establecido las condiciones de expresión, solubilidad, purificación y abordar los ensayos cristalográficos con los constructos anteriores (ver apartados siguientes), se diseñaron clones adicionales donde la cola C-terminal se acortó en más o menos longitud con el objetivo de mejorar los resultados en la cristalización, ya que los datos de SAXS indicaron que esta región es flexible y desordenada (apartado 8.2.1.8), lo que podría impedir la cristalización. En base al alineamiento de las secuencias de Twinkle de distintos organismos *Drosophila melanogaster*, *Mus musculus*, *Anolis carolinensis*, *Hartmannella vermiformis*, *Gallus gallus* y *Homo sapiens* se observó que la cola C-terminal presenta distintas longitudes entre los distintos organismos (figura\_35c), por lo que se diseñaron los clones TwinkleΔ663, TwinkleΔ647, TwinkleΔ636, (GB1)-Δ386TwinkleΔ663, (GB1)-Δ386TwinkleΔ647, (GB1)-Δ386TwinkleΔ636, TwinkleS359P (figura\_35a).

### **8.2.1.2 Expresión de Twinkle recombinante en *E. coli* y ensayos de solubilidad.**

Los ensayos de expresión y solubilidad se realizaron siguiendo el protocolo descrito anteriormente en el apartado de “Materiales, Métodos y Fundamentos: 7.1.2 y 7.1.3”. Inicialmente se evaluó la expresión de los 18 primeros clones Twinkle, Twinkle-(6hisC),  $\Delta 151$ Twinkle,  $\Delta 118$ Twinkle $\Delta 348$ , (MBP)- $\Delta 118$ Twinkle $\Delta 348$ ,  $\Delta 151$ Twinkle $\Delta 348$ , Twinkle $\Delta 359$ , (MBP)-Twinkle $\Delta 359$ ,  $\Delta 75$ Twinkle $\Delta 359$ ,  $\Delta 118$ Twinkle $\Delta 35$ ,  $\Delta 151$ Twinkle $\Delta 359$ ,  $\Delta 386$ Twinkle, (Ztag)- $\Delta 386$ Twinkle, (NusA)- $\Delta 386$ Twinkle, (GB1)- $\Delta 386$ Twinkle,  $\Delta 358$ Twinkle, (GB1)- $\Delta 358$ Twinkle y (TRX)- $\Delta 358$ Twinkle (figura\_35) con las siguientes cepas de *E. coli*: BL21 (DE3), C43(DE3), GroEL/GroES (DE3), Rosetta (DE3), Origami2 (DE3), BL21pLys (DE3), Rosetta2pLys (DE3), BL21\*(DE3).

Las cepas que expresaron mayor número de clones fueron BL21 (DE3), BL21\*(DE3) y Origami2, aunque los resultados con Origami2 mostraron una banda de sobreexpresión de menor intensidad en comparación con las otras dos cepas. Combinando todos los resultados positivos de las cepas anteriores se encontró expresión en todos los constructos excepto para Twinkle. Las cepas Rosetta, BL21pLys y Rosetta2pLys, ofrecieron resultados positivos con muy pocos clones (ver tabla 14) y ninguno fue el clon denominado Twinkle. El resultado de la expresión se detalla en la tabla\_14. Un ejemplo del análisis de la expresión de uno de los clones, en este caso (GB1)- $\Delta 358$ Twinkle, en distintas cepas, se muestra en la figura\_36a.

Aquellos constructos que dieron una expresión positiva se emplearon para transformar, crecer e inducir las cepas que los expresaban, y posteriormente se dividieron los cultivos en ocho alícuotas y cada alícuota fue lisada en un tampón distinto (ver Materiales y métodos) (tabla\_4). El producto de lisis fue centrifugado y el sobrenadante analizado por geles SDS-PAGE (protocolo en el apartado 7.1.3). A pesar de encontrar una gran cantidad de resultados positivos en la expresión de los constructos con sólo la cola de histidinas, los análisis de solubilidad mostraron que ninguno era soluble en ningún tampón, ni en ninguna de las cepas empleadas (ver tabla 16). Sólo se observó solubilidad en algunos constructos expresados con alguna proteína de fusión como GB1, TRX o MBP. Aquellos constructos en los cuales se detectó solubilidad mediante SDS-PAGE (tabla 16 “solubilidad/escala en gel”), se crecieron en cultivos de mayor volumen (0.5L) y el sobrenadante obtenido tras la lisis, una vez filtrado, se cargó en una columna de afinidad. Las fracciones eluidas se analizaron por SDS-PAGE y las fracciones más puras se cargaron en una columna de exclusión de tamaño para verificar la agregación de la muestra (en tabla\_15 “solubilidad gran escala”).

Los constructos clonados con proteínas de fusión como MBP, GB1 y TRX dieron mejores resultados preliminares, sin embargo los constructos del dominio N-terminal clonados con MBP presentaban una gran degradación, quedando soluble sólo la proteína MBP (tabla\_15).

**Tabla 14. Resultado expresión de los constructos de Twinkle en distintas cepas bacterianas.**

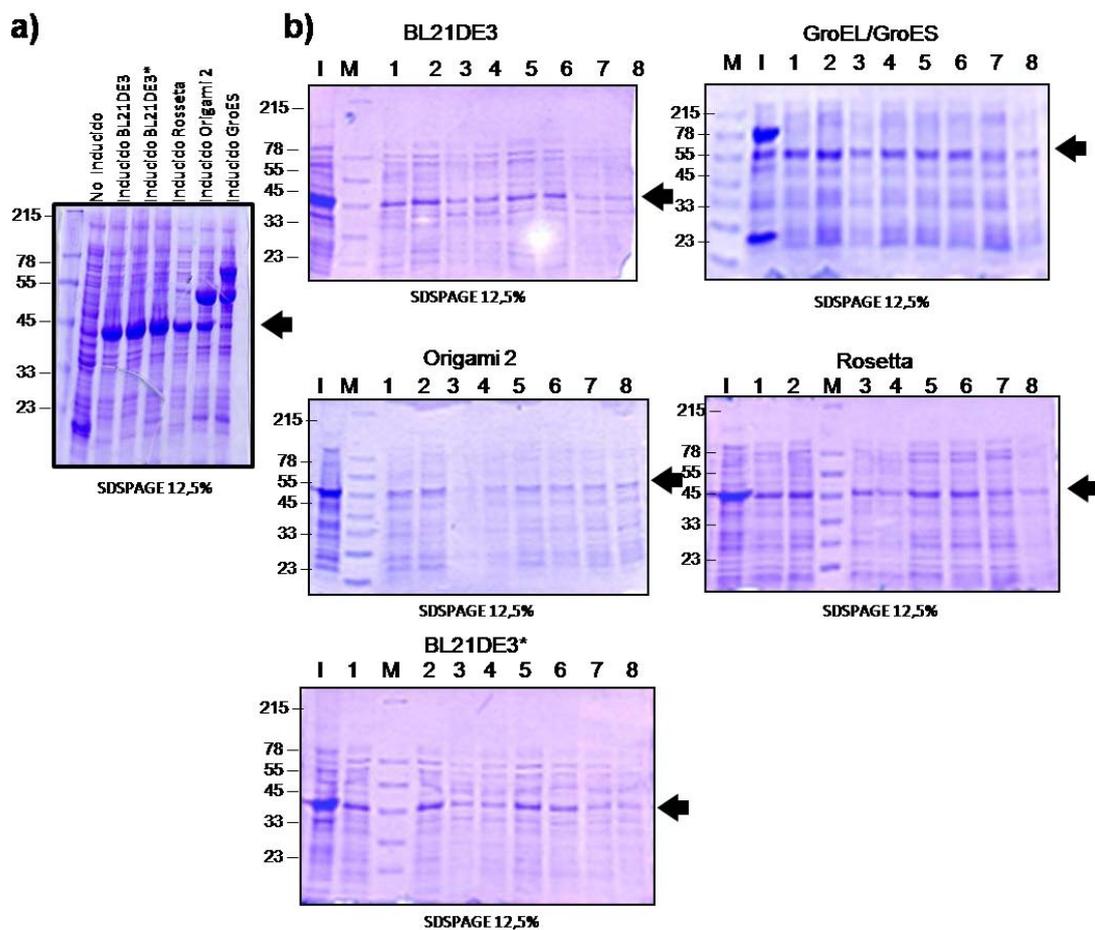
Con “✓” se indica que se observó sobreexpresión, y con una “X” que no se observó sobreexpresión. Con – se indica que no hay datos.

| Cepa                  | Constructo |     |             |         |           |          |                  |          |             |
|-----------------------|------------|-----|-------------|---------|-----------|----------|------------------|----------|-------------|
|                       | BL21 (DE3) | C43 | GroES/GroEL | Rosetta | Origami 2 | BL21pLys | Rosettagami2pLys | BL21DE3* | Takara KJE8 |
| Twinkle               | X          | X   | X           | X       | X         | X        | X                | X        | ✓           |
| Twinkle-(6hisC)       | -          | -   | -           | -       | -         | -        | -                | -        | ✓           |
| TwinkleΔ663           | -          | -   | -           | -       | -         | -        | -                | -        | X           |
| TwinkleΔ647           | -          | -   | -           | -       | -         | -        | -                | -        | X           |
| TwinkleΔ636           | -          | -   | -           | -       | -         | -        | -                | -        | X           |
| Δ151-Twinkle          | X          | X   | X           | X       | X         | X        | X                | ✓        | ✓           |
| TwinkleS359P          | -          | -   | -           | -       | -         | -        | -                | -        | ✓           |
| (MBP)-Δ118TwinkleΔ348 | ✓          | X   | X           | ✓       | ✓         | X        | X                | ✓        | -           |
| Δ118-Twinkle-Δ348     | ✓          | X   | X           | X       | ✓         | X        | X                | ✓        | ✓           |
| Δ151-Twinkle-Δ348     | ✓          | X   | X           | X       | ✓         | X        | X                | ✓        | ✓           |
| (MBP)-TwinkleΔ359     | ✓          | ✓   | ✓           | X       | ✓         | X        | X                | X        | ✓           |
| Δ75-Twinkle-Δ359      | ✓          | X   | X           | X       | ✓         | X        | X                | ✓        | X           |
| Δ118-Twinkle-Δ359     | X          | X   | ✓           | X       | ✓         | X        | X                | X        | X           |
| Δ151-Twinkle-Δ359     | ✓          | X   | X           | X       | X         | X        | X                | ✓        | ✓           |
| Δ386-Twinkle          | ✓          | X   | x           | ✓       | ✓         | ✓        | ✓                | ✓        | ✓           |
| (Ztag)-Δ386Twinkle    | ✓          | ✓   | X           | X       | X         | -        | -                | -        | -           |
| (NusA)-Δ386Twinkle    | ✓          | ✓   | X           | ✓       | X         | -        | -                | -        | -           |
| (GB1)-Δ386Twinkle     | ✓          | ✓   | X           | ✓       | ✓         | ✓        | ✓                | ✓        | ✓           |
| (GB1)-Δ386TwinkleΔ663 | -          | -   | -           | -       | -         | -        | -                | -        | ✓           |
| (GB1)-Δ386TwinkleΔ647 | -          | -   | -           | -       | -         | -        | -                | -        | ✓           |
| (GB1)-Δ386TwinkleΔ636 | -          | -   | -           | -       | -         | -        | -                | -        | ✓           |
| Δ358Twinkle           | ✓          | X   | X           | ✓       | ✓         | ✓        | ✓                | ✓        | -           |
| (GB1)-Δ358Twinkle     | ✓          | X   | ✓           | ✓       | ✓         | X        | X                | ✓        | ✓           |
| (TRX)-Δ358Twinkle     | ✓          | X   | ✓           | ✓       | ✓         | ✓        | ✓                | ✓        | ✓           |

Las fusiones con GB1 y TRX eran solubles en diversos tampones mientras se conservaba la proteína de fusión (tabla\_15). De entre todos los constructos analizados, los constructos del dominio helicasa ( $\Delta 358$ Twinkle y  $\Delta 386$ Twinkle) clonados con GB1 o TRX mostraron mejores resultados; eran solubles con diferentes cepas, de todas ellas la mejor cepa fue Rosetta y en múltiples tampones, como Tris-HCl 100mM pH 7.5, NaCl 1M, BME 5mM; Bicina 100mM pH 9.5, NaCl 1M, BME 5mM; y citrato sódico 100mM pH 6.5, NaCl 700mM, BME 5mM, los cuales corresponden a los carriles 2, 5 y 6 en la figura\_36b, respectivamente. Además, posteriormente se trabajó más profundamente con ambos, pero principalmente con el constructo GB1- $\Delta 358$ Twinkle (figura\_36).

**Figura\_36. Resultado de expresión y solubilidad de (GB1)- $\Delta 358$ Twinkle analizado por geles SDS-PAGE.**

En **a)** se muestra el resultado de la expresión positiva del constructo (GB1)- $\Delta 358$ Twinkle. El primer carril contiene la muestra antes de la inducción y en los carriles adyacentes se indica la sobreexpresión en diferentes cepas. La banda de expresión se indica con una flecha negra. Las bandas adicionales de expresión en los carriles de inducido Origami2 y GroES corresponden a chaperonas. En **b)** se enseñan los geles de SDS-PAGE correspondientes a las cepas en las cuales hubo sobreexpresión de (GB1)- $\Delta 358$ Twinkle; en cada gel se cargó la muestra inducida, el marcador de peso molecular de proteína y el sobrenadante del producto de lisis de distintos tampones; la numeración es la misma que en la tabla\_4. Con la flecha negra se indica la banda que corresponde a (GB1)- $\Delta 358$ Twinkle.



En una segunda fase se repitieron las pruebas de solubilidad utilizando la cepa Takara-KJE8 y un nuevo tampón optimizado con 50 mM arginina y 50 mM glutámico (ver apartado 8.2.1.3.1) con (GB1)- $\Delta$ 358Twinkle y con todos los constructos clonados. Este segundo intento sirvió para encontrar expresión y solubilidad positiva para otros constructos en los cuales no se había observado anteriormente, como la proteína entera Twinkle con la cola 6xHis en posición N-terminal, entre otros (tabla\_14 y 15).

**Tabla\_15. Resultado de la expresión, solubilidad y purificación.**

En la tabla se indica el nombre del constructo y si se pudo expresar o no. La solubilidad se analizó en dos condiciones: a baja concentración "solubilidad /escala gel" (ver apartado 7.1.3); y a concentraciones superiores a 1 mg/ml "solubilidad/gran escala".

| Constructo                              | Expresión | Solubilidad /escala gel | Solubilidad /gran escala |
|---|-----------|-------------------------|--------------------------|
| Twinkle                                 | ✓         | ✓                       | ✓                        |
| Twinkle-(6hisC)                         | ✓         | ✓                       | ✓                        |
| Twinkle $\Delta$ 663                    | ✗         |                         |                          |
| Twinkle $\Delta$ 647                    | ✗         |                         |                          |
| Twinkle $\Delta$ 636                    | ✗         |                         |                          |
| $\Delta$ 151Twinkle                     | ✓         | ✓                       | ✓                        |
| TwinkleS359P                            | ✓         | ✓                       | ✓                        |
| (MBP)- $\Delta$ 118Twinkle $\Delta$ 348 | ✓         | ✓                       | ✗                        |
| $\Delta$ 118Twinkle $\Delta$ 348        | ✓         | ✗                       |                          |
| $\Delta$ 151Twinkle $\Delta$ 348        | ✓         | ✗                       |                          |
| (MBP)-Twinkle $\Delta$ 359              | ✓         | ✓                       | ✗                        |
| $\Delta$ 75Twinkle $\Delta$ 359         | ✓         | ✗                       |                          |
| $\Delta$ 118Twinkle $\Delta$ 359        | ✓         | ✗                       |                          |
| $\Delta$ 151Twinkle $\Delta$ 359        | ✓         | ✗                       |                          |
| $\Delta$ 386Twinkle                     | ✓         | ✗                       |                          |
| (Ztag)- $\Delta$ 386Twinkle             | ✓         | ✗                       |                          |
| (NusA)- $\Delta$ 386Twinkle             | ✓         | ✗                       |                          |
| (GB1)- $\Delta$ 386Twinkle              | ✓         | ✓                       | ✓                        |
| (GB1)- $\Delta$ 386Twinkle $\Delta$ 663 | ✓         | ✓                       | ✓                        |
| (GB1)- $\Delta$ 386Twinkle $\Delta$ 647 | ✓         | ✓                       | ✓                        |
| (GB1)- $\Delta$ 386Twinkle $\Delta$ 636 | ✓         | ✓                       | ✓                        |
| $\Delta$ 358Twinkle                     | ✓         | ✗                       |                          |
| (GB1)- $\Delta$ 358Twinkle              | ✓         | ✓                       | ✓                        |
| (TRX)- $\Delta$ 358Twinkle              | ✓         | ✓                       | ✓                        |

### **8.2.1.3 Purificación y estudios biofísicos del dominio helicasa de Twinkle: (GB1)- $\Delta$ 358Twinkle.**

#### **8.2.1.3.1 Purificación de (GB1)- $\Delta$ 358Twinkle.**

El protocolo inicial de purificación de (GB1)- $\Delta$ 358Twinkle permitió obtener varios mg de proteína al expresar el constructo con la cepa Rosetta y purificarlo con el tampón Tris-HCl 25 mM pH 7.5, NaCl 1 M, BME 5 mM. La purificación se realizó por cromatografía de afinidad mediante incubación y elución por gravedad (figura\_37a). Por SDS-PAGE se observó la banda con el peso molecular que corresponde al constructo (figura\_37a, indicado con flecha negra) y se confirmó por “*fingerprint*” (análisis por espectrometría de masas de la banda del gel). No obstante, tal y como se observa en el gel, se producía una fuerte degradación, la banda inferior en el gel corresponde al peso molecular de la proteína de fusión GB1 (figura\_37a indicada con una flecha roja). La fusión era soluble sólo a concentraciones inferiores a 1 mg/ml, lo cual era indicativo de su inestabilidad.

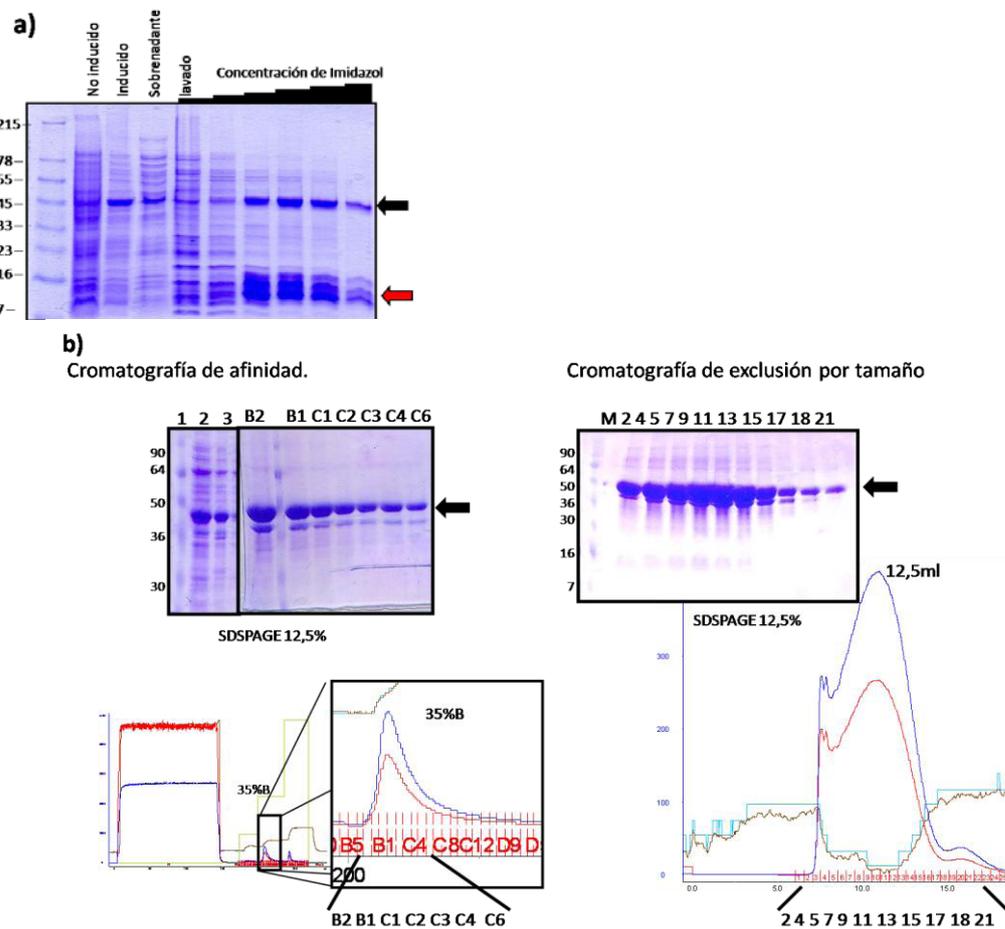
Este avance inicial permitió mejorar las condiciones del tampón de purificación al añadir los aditivos L-arginina y L-glutámico al tampón que, con el cambio de cepa de expresión a la cepa takara-KJE18, mejoró la producción de proteína soluble y estable hasta concentraciones 7-8 mg/ml. En producciones posteriores del constructo (GB1)- $\Delta$ 358Twinkle se emplearon sistemas automáticos de purificación de proteínas. En la fase de purificación por afinidad empleando colas de histidinas se observó la elución a 35% de tampón B (175mM de imidazol) (figura\_37b). La etapa siguiente por cromatografía de exclusión se observó que la muestra no tenía contaminación con ADN, como demuestra el ratio 260/280 de 0.65 (ver Materiales y Métodos). Esta misma cromatografía mostró como eluían en el volumen de exclusión los agregados (los dos picos al principio de la elución en la figura\_37b) de la proteína soluble de la fracción no agregada. La presencia de agregados suele indicar cierta inestabilidad del constructo. Finalmente, el pico de elución era extraordinariamente ancho, siendo indicativo de una gran heterogeneidad en la muestra. A pesar de la gran cantidad de proteína obtenida, el análisis de estas fracciones por SDS-PAGE y tinción por Comassie mostró una pureza aceptable >90% pero seguía sufriendo una degradación importante (figura\_37b). No obstante, la calidad de la muestra fue suficiente para realizar ensayos biofísicos como DSF y DLS (ver apartado 8.2.1.3.2), ensayos de cristalización (ver apartado 8.2.1.5) y un análisis por microscopía electrónica por el método de tinción negativa (ver apartado 8.2.1.3.3).

Este protocolo de purificación en el cual la proteína fue expresada en la cepa KJE-18 y solubilizada en presencia de arginina y glutámico, ambos a 50 mM, también se empleó para los

constructos de la helicasa (TRX)- $\Delta$ 358Twinkle, (GB1)- $\Delta$ 386Twinkle, (GB1)- $\Delta$ 386Twinkle $\Delta$ 636, (GB1)- $\Delta$ 386Twinkle $\Delta$ 647 y (GB1)- $\Delta$ 386Twinkle $\Delta$ 663, obteniendo idénticos resultados para todos los constructos: proteínas solubles que se pudieron concentrar hasta valores superiores a 5 mg/ml, sin contaminación de nucleótidos, y picos muy anchos en la cromatografía de exclusión por tamaño.

**Figura\_37. Purificación de (GB1)- $\Delta$ 358Twinkle.**

En **a)** se muestra el resultado por SDS-PAGE (12.5%) de la purificación por cromatografía de afinidad por gravedad. En los carriles de izquierda a derecha del gel se cargó: marcador de peso molecular, cultivo no inducido, inducido, el sobrenadante después de la lisis, la muestra después del lavado de la resina, la elución a 75, 100, 150, 200, 300, 500 mM de imidazol. Las flechas negra y roja indican la banda del constructo (GB1)- $\Delta$ 358Twinkle y GB1 degradado, respectivamente. En **b)** a la izquierda abajo se indica la cromatografía por afinidad (acoplada a un sistema AKTA) y se indica el pico B que eluyó a 35% tampón B. En el gel se indican los carriles 1-marcaador de peso molecular, 2-inducido, 3-sobrenadante, los carriles siguientes se cargaron las fracciones de la purificación B2, B1, C1, C2, C3, C4 y C6. A la derecha abajo se indica la cromatografía de exclusión por tamaño, en el gel (parte superior derecha) se indica M-marcaador de peso molecular, fracción2, fracción4, fracción5, fracción7, fracción9, fracción11, fracción13, fracción15, fracción17, fracción18 y fracción21. En ambas cromatografías la línea azul corresponde a la absorción a la longitud de onda 280 nm y la roja a 260 nm.



Dado que todos estos constructos tenían una fusión, se realizaron ensayos de digestión con TEV para separar la proteína de fusión GB1 del constructo  $\Delta$ 358Twinkle, a distintos ratios TEV: GB1- $\Delta$ 358Twinkle 1:500, 1:100, 1:50 y 1:25, a dos temperaturas: 4 y 20 °C, a dos concentraciones de NaCl 0.5 y 1 M y con diferentes tiempos de reacción 1, 2, 4 horas y toda la

noche. Todos los ensayos mostraron los siguientes resultados: TEV solo presentaba actividad a 20 °C a 0.5 M NaCl con un ratio 1:50 y 1:25. En el resto de condiciones la digestión de TEV fue marginal o no hubo digestión. El constructo (GB1)- $\Delta$ 358Twinkle agregaba en estas condiciones de 20 °C a 0.5 M NaCl. No se encontraron condiciones donde la actividad de TEV y la estabilidad de (GB1)- $\Delta$ 358Twinkle fuesen compatibles.

#### **8.2.1.3.2 Barrido diferencial de fluorescencia (DSF) y dispersión dinámica de la luz (DLS) del constructo (GB1)- $\Delta$ 358Twinkle.**

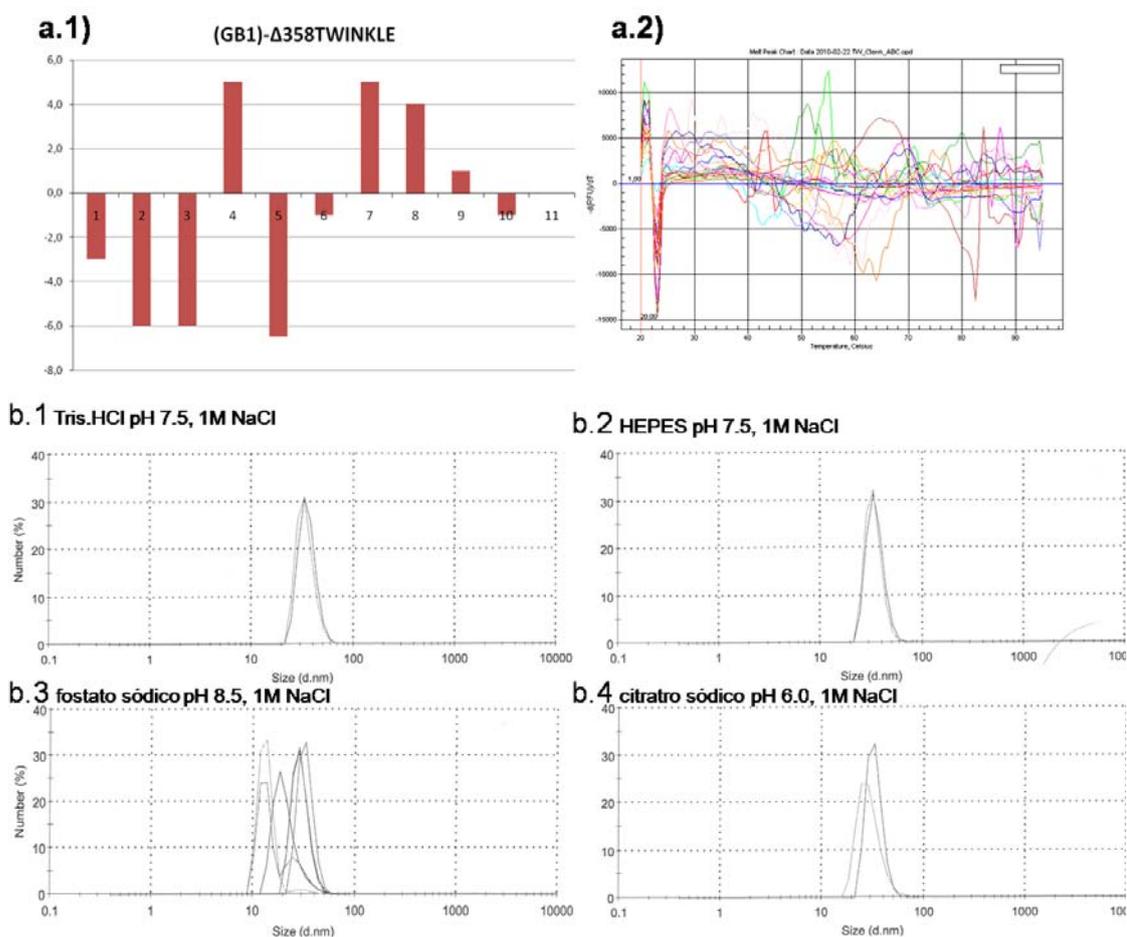
Con el objetivo de conocer el tampón en el que (GB1)- $\Delta$ 358Twinkle era más estable se realizaron ensayos por DSF. Para ello, (GB1)- $\Delta$ 358Twinkle diluida en el tampón de purificación Tris-HCl 25 mM pH 7.5, NaCl 1 M, BME 5 mM, L-Arg 50 mM, L-Glu 50 mM se concentró la muestra hasta 4.2 mg/ml (más detalles en el apartado 7.7). En las curvas de desnaturalización (figura\_38a2) se observaron múltiples picos, muy anchos y difícilmente interpretables. Cada pico corresponde a la emisión de Sypro Orange por contacto a un núcleo hidrofóbico, que podría ser de los distintos dominios en las distintas subunidades y simultáneo a la transición de un estado oligomérico superior (hexámero o heptámero) a otro inferior (dímeros, monómeros), produciendo la exposición consecutiva y/o simultánea de núcleos hidrofóbicos previos a la desnaturalización completa de la proteína. La existencia de múltiples picos en DSF podría estar relacionada con la heterogeneidad observada por la cromatografía de exclusión por tamaño mencionada en el apartado anterior. Respecto al análisis de la estabilidad a distintos tampones se observó que el constructo (GB1)- $\Delta$ 358Twinkle presentaba una mayor estabilidad en los tampones a pH 7.5 y 8.0 con alta concentración de NaCl (1 M) (figura\_38-a1). Tras los ensayos de estabilidad por DSF se analizó la solubilidad y la heterogeneidad de la muestra por DLS. La muestra se concentró entre 1.2 y 1.8 mg/ml y se analizó la solubilidad con tampones a distintos pH: fosfato sódico pH 8.5, HEPES pH 7.5, Tris-HCl pH 7.5, y citrato sódico pH 6.0, todos ellos con una concentración de NaCl 1 M y  $\beta$ ME 1mM. Por cada tampón se realizó una réplica a dos temperaturas a 4 y 20 °C (figura\_38b).

Los mejores resultados obtenidos fueron a pH 7.5, tanto con Tris-HCl como con HEPES, ya que ambos a las dos temperaturas presentaron una polidispersidad menor al 15% y no se observó agregación (figura\_38 b.1 y 2 respectivamente). Las medidas realizadas con fosfato sódico a pH 8.5 mostraron una gran heterogeneidad (>30%) a ambas temperaturas (figura\_38b.3). En cambio con citrato sódico a pH 6.0 se consiguieron buenos resultados a ambas temperaturas pero especialmente a 4 °C, al mostrar una polidispersidad menor al 15% y no agregación (figura\_38b.3 y 4 respectivamente).

**Figura\_38. Termograma de DSF de (GB1)- $\Delta$ 358Twinkle y ensayos por DLS.**

En la parte superior se detalla en la tabla la lista de tampones empleados en DSF. La numeración de la tabla es la misma empleada en la figura a.1), en la cual se muestra con la gráfica de barras la diferencia en la temperatura de fusión entre cada tampón y el control H<sub>2</sub>O (el eje de ordenadas) y en abscisas cada tampón empleado. En la gráfica a.2) se muestra las curvas de desnaturalización de cada tampón, en la cual se observa la diferencia de fluorescencia (en ordenadas) durante el incremento de temperatura (en abscisas). En las figuras de los experimentos de DLS b.1,b.2,b.3 y b.4 se muestra el número de eventos (%) frente al tamaño (nm) empleando los tampones Tris-HCl pH 7.5, 1M NaCl; HEPES pH 7.5, 1M NaCl; fosfato sódico pH 8.5, 1M NaCl, citrato sódico pH 6.0, 1M NaCl, respectivamente.

| Número | Tampón  | Número | Tampón                        |
|--------|---|--------|-------------------------------|
| 1      | Citrato sódico pH6.5 NaCl 1 M                 | 6      | Tris-HClpH8.5 NaCl 1 M        |
| 2      | Citrato sódico pH6.0 NaCl 1 M                 | 7      | Fosfato sódico pH7.0 NaCl 1 M |
| 3      | Tris-HClpH7.5 NaCl 1 M                        | 8      | Fosfato sódico pH7.5 NaCl 1 M |
| 4      | Tris-HClpH7.5 NaCl 0.5 M                      | 9      | Fosfato sódico pH8.0 NaCl 1 M |
| 5      | Tris-HClpH7.5 NaCl 1 M MgCl <sub>2</sub> 5 mM | 10     | Fosfato sódico pH8.5 NaCl 1 M |
|        |   | 11     | Control: H <sub>2</sub> O.    |



### 8.2.1.3.3 Microscopía electrónica del constructo (GB1)- $\Delta$ 358TWINKLE.

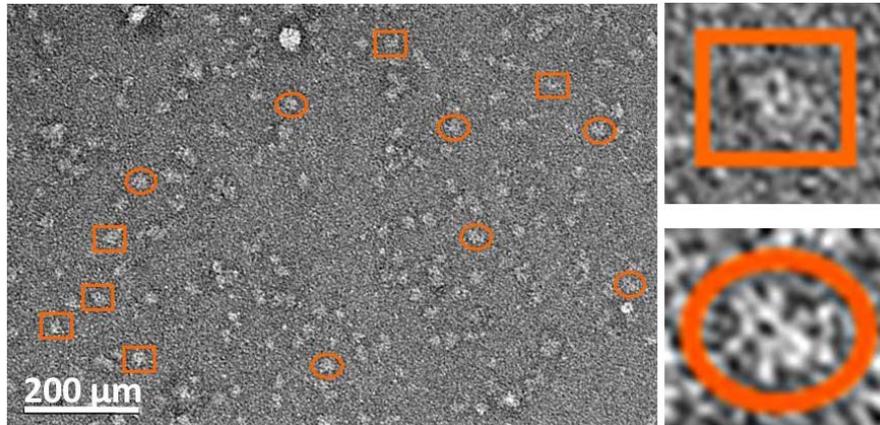
Para analizar la integridad estructural de la muestra se decidió hacer un análisis por tinción negativa de la muestra por microscopía electrónica. El trabajo que se describe a continuación fue realizado por M. Lazado y el Dr. M. Valle.

Las imágenes permitieron comprobar que el dominio helicasa con "linker" (GB1)- $\Delta$ 358TWINKLE presentaba una estructura de anillo (figura\_39 indicado con círculos), con una simetría de orden 6. También se observó la heterogeneidad de la muestra tal y como se observa en la imagen de la figura\_39, en la cual existen múltiples tipos de partículas diferentes e incluso

agregados (figura\_39), tal y como anteriormente ya se había observado con los estudios por cromatografía por exclusión de tamaño. Con estos resultados se decidió no acometer ensayos más sofisticados por EM como la crio-microscopía.

**Figura\_39. Microscopía electrónica de (GB1)- $\Delta$ 358Twinkle.**

Están marcadas las vistas laterales con algo de inclinación con una elipse naranja (tienen algo de simetría radial). Algunas vistas laterales (perpendiculares al eje de simetría radial) están dentro de rectángulos naranjas. Éstas tienen una simetría de orden 2. Ambas son muy típicas de este tipo de hexámeros.



**8.2.1.4 Purificación y estudios biofísicos del constructo Twinkle.**

**8.2.1.4.1 Purificación de Twinkle.**

Inicialmente el protocolo de purificación empleado con Twinkle se basó en el protocolo mejorado para el constructo del anillo C-terminal (GB1)- $\Delta$ 358-Twinkle. Posteriormente a las primeras pruebas de purificación se realizaron los ensayos biofísicos por DSF y DLS (apartado 8.2.1.4.2 más adelante) que permitieron una posterior mejora del protocolo de purificación de Twinkle entera, el cual se describe a continuación.

En la primera etapa de purificación de Twinkle por cromatografía de afinidad se empleó el tampón optimizado para (GB1)- $\Delta$ 358-Twinkle (descrito en el apartado 7.2.3) consiguiéndose una pureza superior al 95% según lo analizado por SDS-PAGE y tinción por Comassie. Las fracciones de mayor pureza se juntaron y concentraron hasta valores superiores a 4 mg/ml. Con la subsiguiente cromatografía de exclusión por tamaño se consiguió una muestra con una pureza >98% y sin degradación. El ratio 260/280 indicó que la muestra no estaba contaminada con ningún nucleótido, ya que presentó un valor de 0.65 (figura\_40). Sin embargo el constructo Twinkle mostró un pico de elución relativamente ancho y era precedido de una pendiente en la cual la señal de la muestra incrementaba paulatinamente, hasta resolverse el pico en cuestión. Las muestras analizadas por SDS-PAGE indicaban un grado de pureza y homogeneidad en todas las fracciones, por lo que la anchura de la elución sugería un elevado grado de heterogeneidad de Twinkle. El pico principal correspondía con peso molecular aparente de 550KDa, que correspondería aparentemente a ocho monómeros. La elución de partículas de Twinkle más grande antes del pico principal sugería un elevado grado de

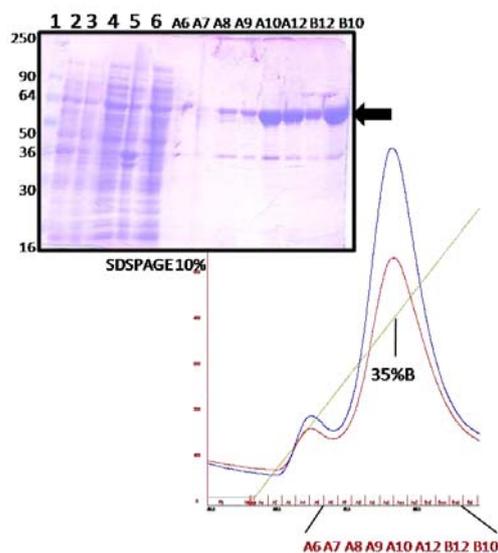
heterogeneidad estructural y sugería un equilibrio entre distintos estados oligoméricos, posiblemente entre heptámeros y hexámeros como se había publicado para la proteína homóloga gp4 (Ahnert, Picha et al. 2000; Ziebarth, Gonzalez-Soltero et al. 2010), o incluso partículas más grandes o más alargadas.

Este protocolo de purificación también fue empleado para los constructos Twinkle-(6hisC), Twinkle-S359P y  $\Delta$ 151-Twinkle, obteniendo los mismos resultados, en todos los casos la pureza por SDS-PAGE fue similar y el perfil de la cromatografía de exclusión por tamaño también fue muy parecido, mostrando una pendiente creciente con anterioridad al pico de elución de la muestra, aunque menos pronunciado en el constructo  $\Delta$ 151-Twinkle. El pico principal eluyó sistemáticamente en el mismo volumen 12.5ml. Solo cabe destacar que el constructo  $\Delta$ 151-Twinkle agregaba aproximadamente un 40% cuando se concentraba de la proteína obtenida en la primera etapa de purificación para cargarla a la cromatografía de exclusión de tamaño.

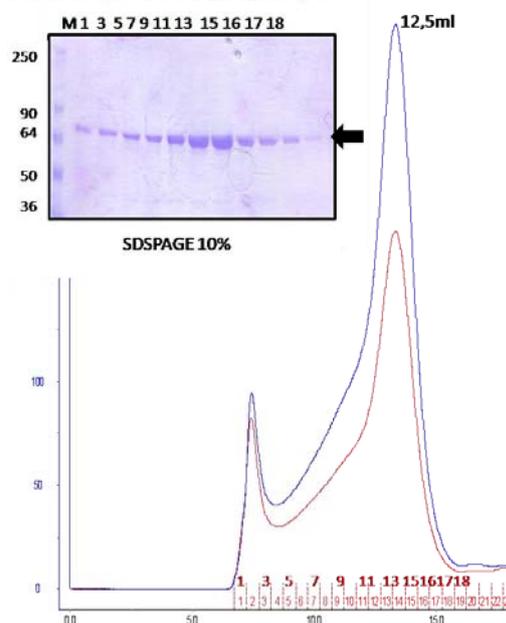
**Figura\_40. Purificación de Twinkle.**

En la izquierda de la figura se indica el cromatograma de la primera etapa de purificación por afinidad (acoplada a un sistema ÄKTA). En el gel (SDS-PAGE 10%) se indican los carriles 1, marcador de peso molecular; 2, no inducido; 3, inducido; 4, sobrenadante de lisis; 5, pellet de lisis; 6, lavado y las siguientes fracciones de elución en la carril A6; A7; A8; A9; A10; A12; B12; B11. A la derecha se indica la cromatografía de exclusión por tamaño empleando una columna Superosa 6 10/300 (GE-Heathcare). En el gel se indica M, marcador de peso molecular; fracción1; fracción3; fracción5; fracción7; fracción9; fracción11; fracción13; fracción15; fracción16; fracción17 y fracción18. En ambas cromatografías la línea azul corresponde a la absorción a 280 nm de longitud de onda y la roja a 260 nm.

**Cromatografía de afinidad.**



**Cromatografía de exclusión por tamaño**



#### 8.2.1.4.2 Barrido diferencial de fluorescencia (DSF) y dispersión dinámica de la luz (DLS) del constructo Twinkle.

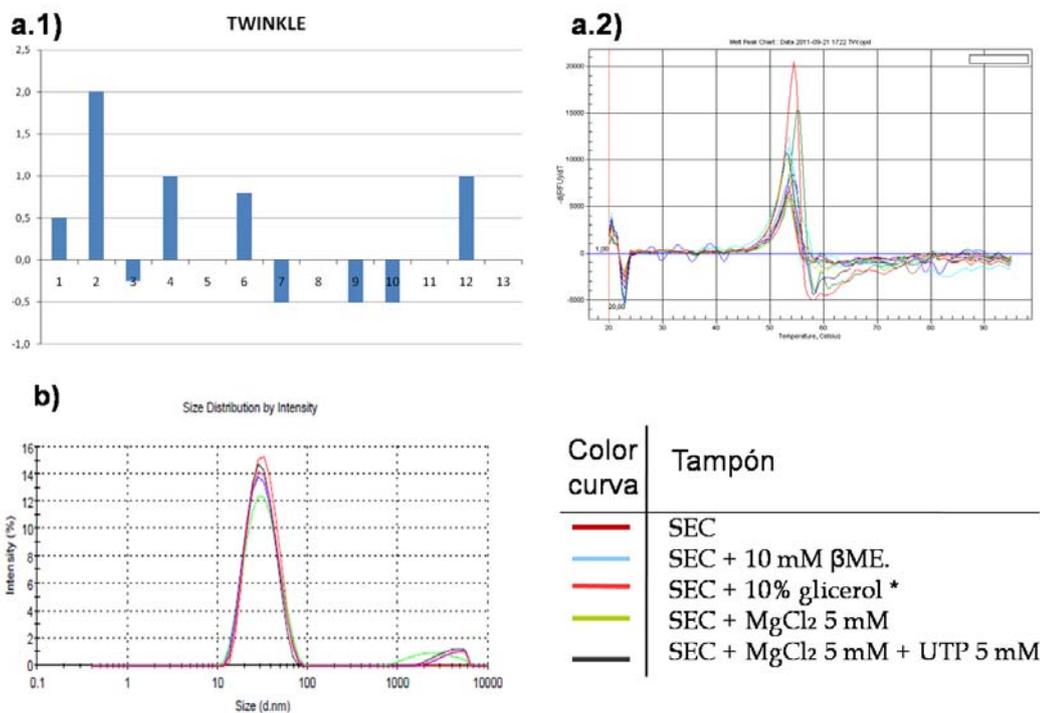
El análisis por DSF permitió optimizar las condiciones de purificación de Twinkle comparando la estabilidad térmica de Twinkle en distintos tampones. Se realizó un barrido de tampones con distintos pH (6.5, 7.5 y 8.0) y se observó que Twinkle era más estable a pH 6.5 (ver figura\_41a1) ya que produjo un aumento de la temperatura de desnaturalización de 2 °C, a

diferencia respecto del constructo (GB1)- $\Delta$ 358Twinkle, donde la mejora fue a pH 7.5. Para además analizar el efecto que podría tener la concentración del NaCl se realizaron dos medidas a distinta concentración de NaCl, 0.5 y 1 M, por cada pH 6.5, 7.5 y 8.0. El resultado fue que a cualquier de los tres pH medidos, la muestra a 1 M NaCl era más estable que a 0.5M. También se analizó el efecto añadir MgCl<sub>2</sub> 5 mM en la estabilidad de la muestra y se observó que, en función de la concentración de NaCl, el MgCl<sub>2</sub> tenía distintos efectos en la estabilidad de la muestra, por ejemplo a 200 mM NaCl estabilizaba ligeramente a la proteína pero a 500 mM NaCl no se observó ningún efecto y a 1 M mejoró la estabilidad ligeramente (Figura\_41a.1).

**Figura\_41. Resultado y termograma de DSF del constructo Twinkle.**

Arriba se enumeran las condiciones químicas empleadas en el experimento, el número indica la posición que le corresponde en la gráfica **a.1)**, en la cual se representa la diferencia en la temperatura de fusión medida en cada uno de los tampones y la medida de la temperatura de fusión al emplear una disolución con H<sub>2</sub>O. Los valores positivos indican el incremento de la temperatura de fusión. En la gráfica **a.2)** se representan las curvas de desnaturalización del experimento de DSF. Se puede observar que sólo hay un pico por tampón correspondiente a la transición plegado-desplegado de la proteína. En **b)** se indican las curvas de DLS con distintos tampones, que se listan a la derecha. Se ha denominado SEC al tampón de elución de la cromatografía de exclusión de tamaño.

| Número | Tampón                          | Número | Tampón   |
|--------|---------------------------------|--------|--|
| 1      | Citrato sódico pH6.5 NaCl 0.5 M | 7      | Tris-HCl pH7.5 NaCl 0.2 M                        |
| 2      | Citrato sódico pH6.5 NaCl 1 M   | 8      | Tris-HCl pH7.5 NaCl 0.2 M MgCl <sub>2</sub> 5 mM |
| 3      | Fosfato sódico pH8.0 NaCl 0.5 M | 9      | Tris-HCl pH7.5 NaCl 0.5 M                        |
| 4      | Fosfato sódico pH8.0 NaCl 1 M   | 10     | Tris-HCl pH7.5 NaCl 0.5 M MgCl <sub>2</sub> 5 mM |
| 5      | HEPES pH7.5 NaCl 0.5 M          | 11     | Tris-HCl pH7.5 NaCl 1 M                          |
| 6      | HEPES pH7.5 NaCl 1 M            | 12     | Tris-HCl pH7.5 NaCl 1 M MgCl <sub>2</sub> 5 mM   |
|        |                                 | 13     | Control: H <sub>2</sub> O.                       |



Las curvas de desnaturalización de Twinkle por DSF (figura\_41a.2) mostraron un solo pico correspondiente a una única temperatura de fusión por cada condición, indicando que sólo se

produjo una transición entre los estados plegado y desplegado, a diferencia del constructo (GB1)- $\Delta$ 358-Twinkle (figura\_41a.2) que mostró múltiples transiciones, como se indicó anteriormente. Este dato podría ser indicativo de una mayor homogeneidad de Twinkle respecto al constructo del dominio helicasa (GB1)- $\Delta$ 358-Twinkle.

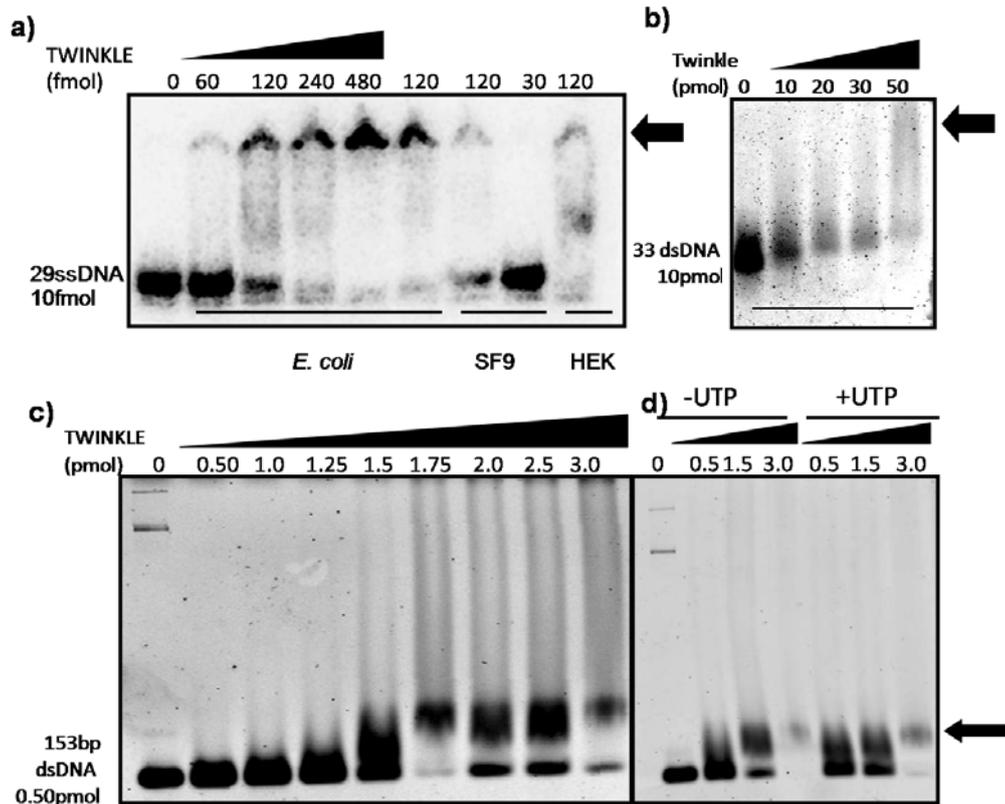
Posteriormente al análisis por DSF, Twinkle se analizó por DLS para evaluar su posible agregación y su homogeneidad en presencia de distintos aditivos como  $\beta$ ME,  $MgCl_2$  y UTP, a una concentración de 1.5mg/ml. No se observó ningún efecto en la solubilidad ni en la polidispersidad de Twinkle por DLS al añadir ninguno de estos aditivos. También se comprobó que la muestra se mantenía soluble tras crio-congelarla con nitrógeno líquido y descongelarla, para lo que se añadió glicerol con una concentración final de 10% para preservar la muestra (figura\_41b).

#### **8.2.1.4.3 Análisis del complejo Twinkle-ADN por técnicas electroforéticas: EMSA.**

Twinkle es una enzima cuya función es desenrollar el ADN bicatenario y por tanto tiene la capacidad de interactuar con el ADN. Para investigar esta interacción proteína/ADN se realizaron una serie de ensayos por EMSA. Primero, se quiso comprobar que Twinkle recombinante obtenida en distintos sistemas de expresión como *E.coli*, células de insecto SF9 y células HEK no mostraba diferencias en su capacidad de unión al ADN. Para esta comparación se utilizó un ADN monocatenario de (ssADN) 29 bases de longitud. En principio, al ser Twinkle una helicasa que une el ADN de forma no específica, la naturaleza de la secuencia de ADN no debería ser determinante para este tipo de análisis. Para una cantidad constante de ADN de 10 fmoles se realizó una titración de Twinkle (*E.coli*) desde 60 a 120 fmol (es decir, a ratios de proteína:ADN 6:1, 12:1, 24:1, 48:1, considerando para el cálculo molar el peso molecular del hexámero) y se comparó con las mismas cantidades de Twinkle con varios ratios proteína:ADN 12:1, 3:1 (para SF9) y 12:1 (para HEK). Las condiciones de reacción fueron las mismas en todos los casos (ver apartado 7.3.1). Curiosamente, las cantidades de hexámero de Twinkle necesarias para retardar el ADN eran anómalamente alta en todos la proteína expresada con los tres sistemas distintos, mínimo de un ratio 12:1 proteína:ADN (figura\_42a). Aunque no se pudieron comparar cuantitativamente si se observó que Twinkle interactuaba con el ssADN independientemente de los tres sistemas de expresión ensayados (figura\_42a). Seguidamente se analizó la capacidad de Twinkle de unir un ADN bicatenario de 33 pb. Para la muestra bicatenaria se observó que la banda de ADN libre ya desaparecía a un ratio 5:1 (proteína:ADN) (figura\_42b). Por el contrario, la muestra de ssADN fue completamente retardada sólo cuando se alcanzó un ratio 24:1. Se concluyó que la afinidad de Twinkle por un ADN bicatenario era mayor que por el monocatenario.

**Figura\_42. Ensayos de interacción Twinkle y el ADN mediante EMSA.**

En **a)** se muestra un gel EMSA utilizando 10 fmol ssADN de 29 bases al que se añadieron cantidades crecientes de enzima en cada carril, de izquierda a derecha Twinkle expresada en *E.coli* 60, 120, 240 y 480 fmoles; Twinkle expresada en células de insecto SF9 (SF9) a dos cantidades 30 y 120 fmoles; y finalmente se añadieron en el último carril 120 fmoles de Twinkle expresada en células de mamífero HEK (HEK). En **b)** se muestra el ensayo EMSA con el dsADN de 33 pb (10 pmol), la titración se hizo con cantidades crecientes de Twinkle de 10 a 50 pmoles de izquierda a derecha. En **c)** se muestra la titración de dsADN-153pb (0.5 pmol) de Twinkle desde a 0.5 a 3.0 pmol. En **d)** se muestran dos titraciones frente al oligonucleótido dsADN-153 pb (0.5 pmol) con Twinkle a concentraciones de 0.5, 1.5 y 3.0 pmol sin UTP (-UTP) y con 5 mM UTP (+UTP).



Posteriormente, se analizó la influencia de la longitud del ADN y su interacción con Twinkle. Para ello se realizó un EMSA con un ADN doble cadena (dsADN) de 153 pb. Al contrario que para los dos casos anteriores ssADN 29 bases y dsADN de 33 pb, en este caso se observó que con un ratio 4:1 (proteína:ADN) ya aparecía una banda de retardo bien definida a la vez que casi todo el ADN estaba retardado. Por tanto, la conclusión fue que la afinidad de Twinkle por ADN cortos es menor que para ADN largos (figura\_42c).

El UTP es el cofactor que requiere Twinkle para realizar la reacción helicasa, por lo tanto se quiso conocer si podría influir en la unión al ADN, o sólo era necesario para la actividad enzimática. Se realizaron dos titraciones con la misma muestra de proteína y el mismo ADN (dsADN 153 pb) donde la única diferencia fue la presencia de UTP en una de las titraciones. Mediante la técnica de EMSA no se observó ninguna diferencia en las dos titraciones (figura\_42-d).

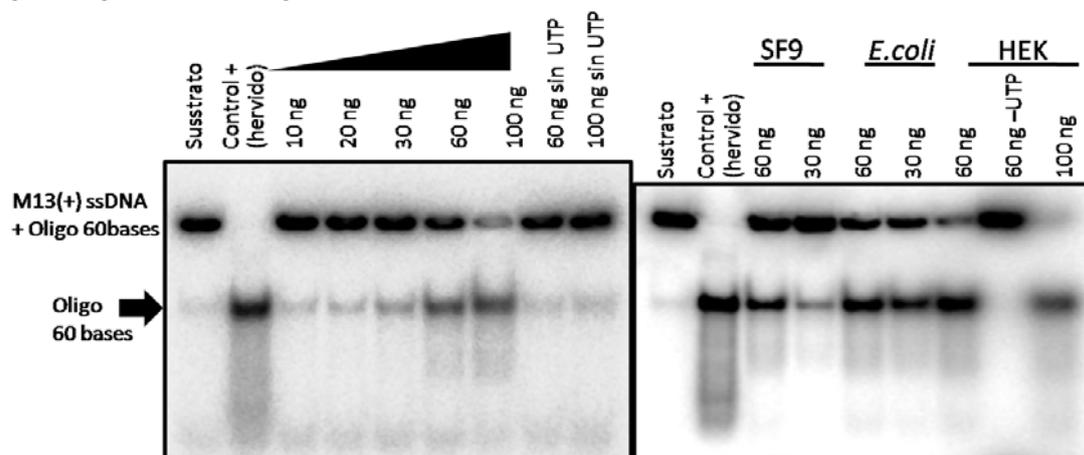
#### 8.2.1.4.4 Análisis de la actividad de Twinkle: ensayo helicasa.

Twinkle es una helicasa cuya función es separar las cadenas del ADN durante la replicación del genoma mitocondrial. En general, para evaluar *in vitro* la capacidad de las helicastas de separar dos hebras de ADN se emplea un sustrato con estructura de horca. En nuestro caso se generó dicha horca con el plásmido M13(+), que es monocatenario, al cual se le hibridó un extremo de un oligonucleótido de 60 bases marcado radiactivamente en el extremo 5'. Solamente las últimas 20 bases del oligonucleótido complementaban la secuencia de M13(+), quedando 40 bases del extremo 5' monocatenario libre, sin complementar (ver más detalles en el apartado 7.5).

Para comprobar si Twinkle tenía actividad helicasa se realizó un ensayo helicasa con cantidades crecientes de enzima (Twinkle) (de 10 a 100 ng) frente a un sustrato a concentración constante. El resultado fue que a cantidades crecientes de Twinkle aumentó proporcionalmente el producto de la reacción, el oligonucleótido de 60 bases libre. Ello evidenció que Twinkle producida en *E.coli* tenía la capacidad de desenrollar el ADN de doble cadena. Para la reacción es indispensable emplear algún NTP como fuente de energía, y por tanto como control negativo se realizó la reacción sin UTP. El resultado fue que no hubo actividad al eliminar el UTP de la reacción (figura\_43-a).

**Figura\_43. Ensayo helicasa de la actividad de Twinkle.**

A la izquierda se muestra el resultado de la reacción de desenrollamiento del dsADN en gel de acrilamida, en el cual se ha cargado, de izquierda a derecha, Sustrato sin enzima; un control con el sustrato hervido (en el cual el oligonucleótido se separó del plásmido M13(+)); los siguientes 5 carriles indican la cantidad de enzima empleada: 10, 20, 30, 60 y 100 ng; los dos últimos carriles son el control negativo en el cual no se añade UTP, con 60 y 100 ng de proteína. A la derecha se comparan la actividad de la enzima en función del sistema de expresión: células de insecto (SF9) (carriles 3-4), *E.coli* (carriles 5-6) y células de mamíferos (HEK) (carriles 7-8-9). Se han cargado en el gel, de izquierda a derecha, en sustrato sin enzima; sustrato hervido; y en el resto de carriles se indican las distintas cantidades de enzima empleadas, carril 3, 60 ng; 4, 30 ng; 5, 60 ng; 6, 30; 7,60 ng; 8, 60 ng Sin UTP; 9, 100 ng.



Posteriormente se comparó la actividad helicasa de proteínas de Twinkle obtenidas mediante sistemas de expresión distintos, concretamente *E.coli*, células de insecto (SF9) y células de mamífero (células HEK), tal y como ya se analizó anteriormente con los ensayos EMSA mencionados. Se realizó el ensayo a dos concentraciones de proteína por sistema de expresión: 30 y 60 ng para Twinkle de *E.coli* y de células SF9, y 60 y 100ng para Twinkle de células HEK. La intensidad de banda de oligonucleótido libre fue similar cuando la cantidad de Twinkle era 60 ng en los tres casos. También la intensidad de banda del oligonucleótido libre aumentaba de acuerdo con el aumento de la enzima (Figura 43b).

#### **8.2.1.5 Cristalización de Twinkle y distintos constructos.**

Twinkle, como se mencionó anteriormente en la introducción, es una proteína multimérica y compuesta por varios dominios y subdominios. Con este escenario, desde el principio del proyecto se planeó como estrategia el diseño de múltiples clones definiendo distintos fragmentos de la proteína que permitiera encontrar algunos susceptibles de ser cristalizados. Para ello se intentó definir dominios globulares sin regiones flexibles ni desordenadas, tal y como se menciona en el apartado 8.2.1.1. Con esta finalidad se clonaron 24 constructos distintos, de los cuales 12 eran solubles y se purificaron 10, lo cuales se emplearon en los ensayos de cristalización. En la tabla 16 indican los constructos, las concentraciones de proteína en las que se hicieron los ensayos de cristalización, y las condiciones de cristalización con sus nombres comerciales (aunque se preparasen a mano por parte del Departamento de Biología Estructural del IBMB-CSIC y la PAC) que se ensayaron para cada constructo (tabla\_16). También se intentó cristalizar Twinkle formando complejos con ADN. Inicialmente se intentó realizar el complejo mediante diálisis. Mezclando distintas proporciones equimolares de ADN y Twinkle (por ejemplo 3mg/ml). Se había observado que Twinkle era estable sólo a concentraciones muy altas de NaCl (entre 0.75 y 1.0 M), que se atribuyó a que los iones del NaCl compensaban las cargas de la proteína necesaria para su interacción con él. Por tanto, para intercambiar los iones del NaCl por la interacción con el ADN evitando exponer a Twinkle a un medio desestabilizador sin suficiente NaCl, se incorporó el ADN directamente en las a la proteína purificada a 1M NaCl y mediante diálisis se fue reduciendo la concentración de NaCl hasta alcanzar 50 mM. Sin embargo y a pesar de las precauciones se observó una fuerte precipitación en todos los ensayos.

**Tabla\_16. Resultado cristalización de Twinkle y distintos constructos.**

| Constructo            | Concentración (mg/ml) | Condiciones de cristalización  |
|-----------------------|-----------------------|--|
| Twinkle               | 3.75, 2.8 mg/ml       | "Crystal Screen I & II", "Index", "Natrix", "SalTRX", "A/S Ion Screen", "Quick phosphate", "Ammonium sulphate", "PEG6000/LiCl", "PEG400", "PEG400/LiCl", "MDP", "PEG-Ion", "NaCl", "complex screen", " (Hampton Research); "Wizard I & II" (Rigaku-EmeralBiosystem); "Protein-DNA Screen" (KeraFast); "Pi-minimal screen"(Jena Bioscience); "PACT premier HT-96" (Molecular Dimensions). |
| Twinkle-(6hisC)       | 7.2, 5.0 mg/ml        | "Protein-DNA Screen" (KeraFast); "Pi-minimal screen"(Jena Bioscience); "Crystal Screen I & II", "Index", "Natrix", "SalTRX", "A/S Ion Screen", "Quick phosphate", "PEG6000/LiCl", "PEG400", "PEG400/LiCl" (Hampton Research).  |
| Δ151Twinkle           | 3.9 mg/ml             | "Crystal Screen I & II", "Index",  |
| TwinkleS359P          | 4.5 mg/ml             | "Crystal Screen I & II", "Index", "Protein-DNA Screen" (KeraFast) "PEG6000/LiCl", "PEG400", "PEG400/LiCl", "MDP"   |
| (GB1)-Δ358Twinkle     | 12.5mg/ml             | "Crystal Screen I & II", "Index", "Natrix", "SalTRX", "A/S Ion Screen", "Quick phosphate", "Ammonium sulphate", "PEG6000/LiCl", "PEG400", "PEG400/LiCl", "MDP", "PEG-Ion", "NaCl", "complex screen", " (Hampton Research); "Wizard I & II" (Rigaku-EmeralBiosystem); "Protein-DNA Screen" (KeraFast); "Pi-minimal screen"(Jena Bioscience); "PACT premier HT-96" (Molecular Dimensions). |
| (TRX)-Δ358Twinkle     | 8.8mg/ml              | "Protein-DNA Screen" (KeraFast); "Pi-minimal screen"(Jena Bioscience); "Crystal Screen I & II", "Index", "Natrix", "SalTRX", "A/S Ion Screen", "Quick phosphate", "PEG6000/LiCl", "PEG400", "PEG400/LiCl" (Hampton Research).  |
| (GB1)-Δ386Twinkle     | 9.5 mg/ml             | "Protein-DNA Screen" (KeraFast); "Pi-minimal screen"(Jena Bioscience); "Crystal Screen I & II", "Index", "Natrix", "SalTRX", "A/S Ion Screen", "Quick phosphate", "PEG6000/LiCl", "PEG400", "PEG400/LiCl" (Hampton Research).  |
| (GB1)-Δ386TwinkleΔ663 | 6.4mg/ml              | "Crystal Screen I & II", "Index"   |
| (GB1)-Δ386TwinkleΔ647 | 3.3mg/ml              | "Crystal Screen I & II", "Index"   |
| (GB1)-Δ386TwinkleΔ636 | 4.4mg/ml              | "Crystal Screen I & II", "Index"   |

Una estrategia alternativa fue la co-cristalización: se dispensaba Twinkle en su tampón con alta concentración de NaCl (1M NaCl) en las gotas de cristalización, las cuales contenían cantidades casi-equimolares de ADN (proteína:ADN, 1:1,2). Además de la co-cristalización con el ADN también se intentó co-cristalizar con ligandos como UTP y UMPPNP y la combinación de ambos. Las pruebas que se hicieron fueron

- Co-cristalizacion Twinkle 3.2 mg/ml (7.6 μM) con ADN de cadena sencilla de 29 bases, cadena doble de 20pb, 29pb y a distintos ratios molares de proteína:ADN : 1:1, 1:2, 1:3, 1:6, 2:1.
- Co-cristalizacion Twinkle-6hisC 3.7 mg/ml (8.8 μM) con ADN de cadena sencilla de 29 bases, cadena doble de 20pb, 29pb y a distintos ratios molares de proteína:ADN, 1:1, 1:2, 1:3, 1:6, 2:1.
- Co-cristalizacion Twinkle 6.6 mg/ml (15.7μM) con UTP y UMPPNP a diferentes ratios proteína: UTP/UMPPNP de 1:3, 1:6

- Co-cristalización de Twinkle 5.0 mg/ml (11.9 $\mu$ M) con combinaciones de UTP y UMPPNP con las diferentes moléculas de ADN, proteína:UTP/UMPPNP:ADN 1:3:2, 1:6:2, 1:6:3.

Aunque se alcanzaron concentraciones en las cuales se producía precipitación en al menos dos tercios de las condiciones ensayadas por placa de cristalización, lo que era indicativo de que la concentración de proteína a la cual se realizaba el experimento era suficiente para precipitar (y con ello cristalizar), no se logró ningún agregado ni precipitado cristalino tanto para la proteína sola como la co-cristalización con los ligandos de ADN y/o UTP.

Los resultados de SAXS y EM, que se describen más adelante, sugieren que la principal razón por la cual se entiende este resultado cristalográfico fue la flexibilidad que poseían los constructos empleados.

#### **8.2.1.6 Generación de un modelo de Twinkle in silico con MODELLER.**

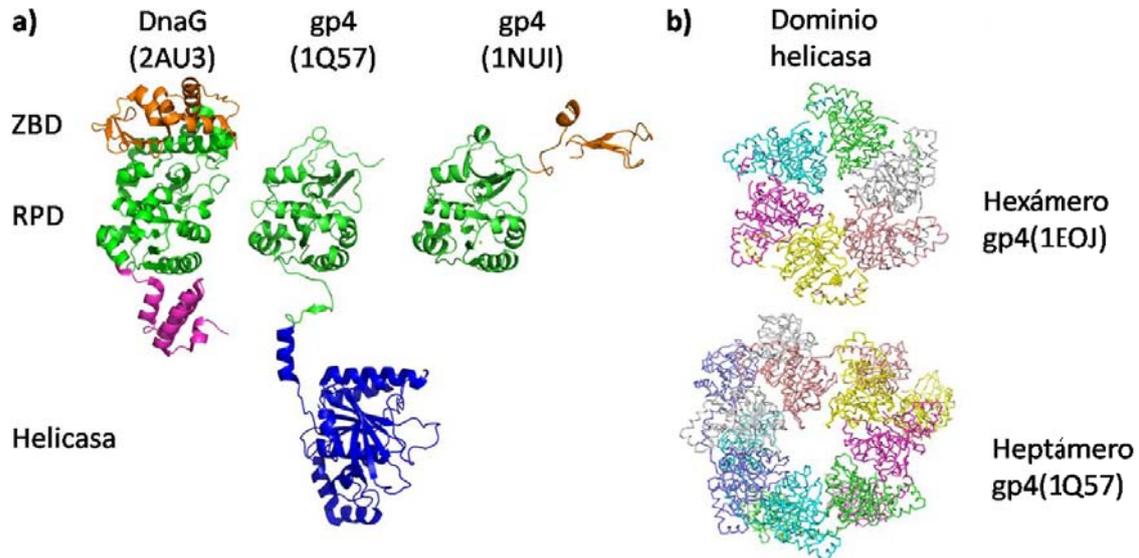
Para poder interpretar los resultados de SAXS y de microscopía electrónica era necesario un modelo estructural tridimensional de Twinkle. Al no haber disponibilidad de una estructura a resolución atómica, se generó un modelo por predicción a partir de la estructura de proteínas homólogas. Para realizar este modelo se empleó el programa MODELLER (Eswar, Webb et al. 2006), tal y como se describió en el apartado 7.10.

Se escogieron las estructuras cristalográficas cuyas secuencias primarias y estructuras secundarias presentaban mayor parecido con la secuencia y la predicción de estructura de Twinkle. En base a este criterio se identificaron cuatro estructuras depositadas en el “*Protein data bank*” (PDB), de las cuales tres correspondían a distintos fragmentos de la helicasa/primasa gp4 del fago T7. La primera de estas estructuras consiste en el hexámero formado por el dominio helicasa y el “*linker*” (código del “*protein data bank*”, PDB, 1EOJ). El segundo fragmento cristalizado de gp4 es un heptámero formado por el dominio helicasa y una parte del dominio primasa, concretamente el subdominio RPD (PDB 1Q57). La tercera estructura consiste en un dímero formado por el dominio N-terminal primasa de gp4 que incluye los dos subdominios ZBD y RPD (PDB 1NUI). Además de las estructuras de gp4, también se identificó la estructura de la primasa DnaG (PDB 2AU3), que no forma multímeros si no es en contacto con la helicasa DnaB (Ver figura\_44).

Después de seleccionar las estructuras que se emplearon como plantillas para la construcción de los modelos de Twinkle, la etapa más crítica fue el alineamiento de secuencia primaria, ya que esta etapa determina la estructura tridimensional final del modelo.

**Figura\_44. Estructuras empleadas como plantilla para MODELLER.**

En **a)** se muestran las estructuras empleadas como plantillas para generar el modelo del monómero de Twinkle. De izquierda a derecha, DnaG (2AU3), dominio helicasa y RPD de gp4 (1Q57) y dominio primasa con los subdominios ZBD y RPD de gp4 (1NUI). Se indican los distintos dominios ZBD (naranja), RPD (verde), helicasa (azul), y el dominio de interacción de DnaG con DnaB (rosa). En **b)** se indican las dos estructuras de gp4 que se utilizaron para generar el hexámero (1EOJ) y heptámero (1Q57).

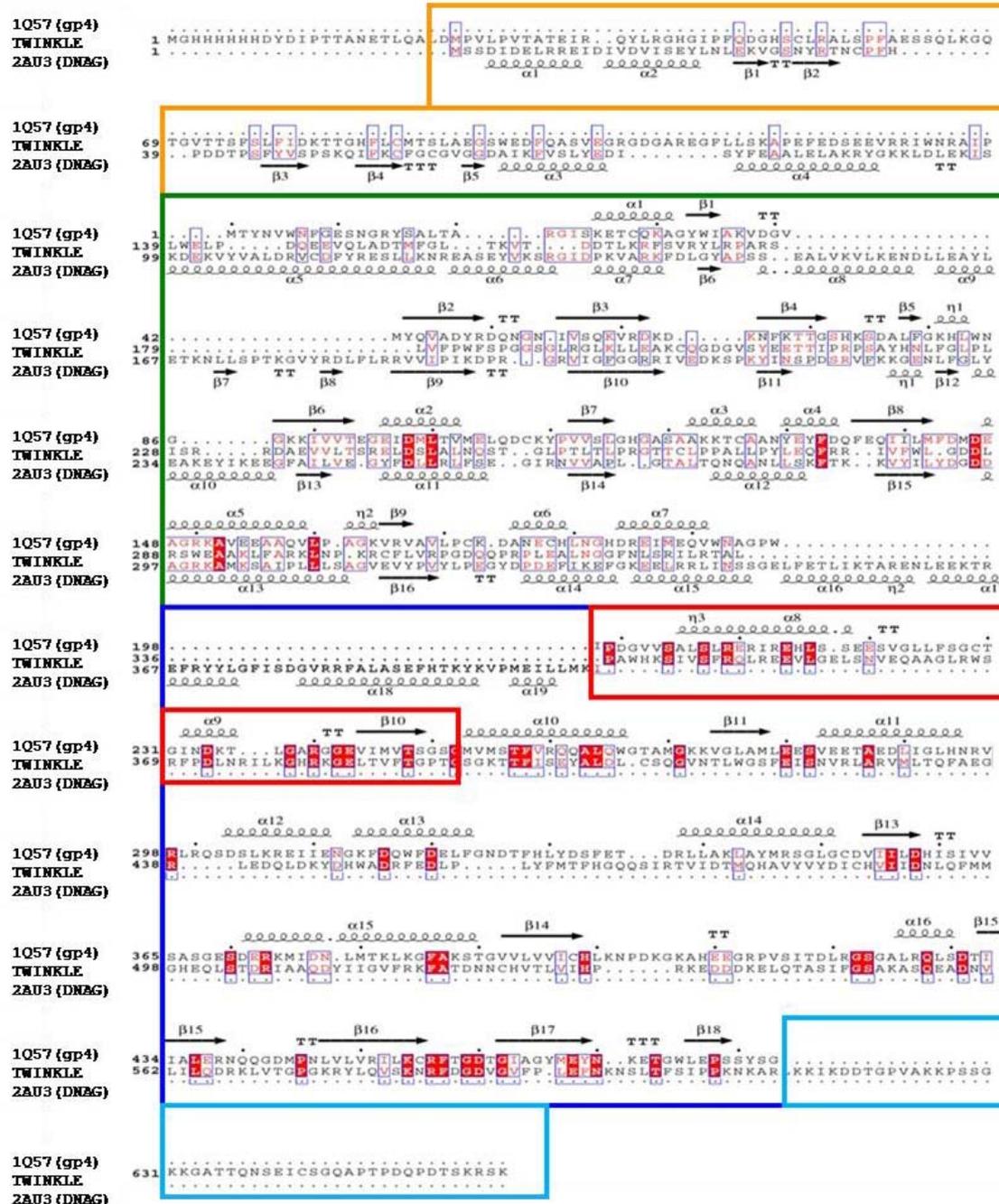


Inicialmente se realizó el alineamiento de la secuencia de Twinkle con la secuencia de todas las estructuras mencionadas, pero al final se descartó la estructura 1NUI de gp4, principalmente porque la secuencia del dominio primasa de gp4 presenta una delección en el dominio N-terminal si se la compara con Twinkle y DnaG, además de que parte del dominio ZBD no está trazado. Por el contrario, la secuencia del dominio primasa de DnaG y Twinkle son igual de extensas. La secuencia de 1EOJ no se consideró en el alineamiento porque es la misma que la secuencia de 1Q57, por tanto como primer paso se alinearon las secuencias de las estructuras 2AU3 (DnaG), 1Q57 (gp4) con la secuencia primaria de Twinkle (figura\_45).

El alineamiento se generó en dos etapas. En una primera etapa se alineó la secuencia primaria de las tres proteínas con el programa CLUSTALW (Thompson, Gibson et al. 2002). Este alineamiento se confirmó para DnaG y gp4 mediante la superposición de cada elemento de estructura secundaria de las estructuras respectivas de 2AU3 y 1Q57, para validar qué aminoácidos de una estructura y otra eran equivalentes estructuralmente. Gracias a este alineamiento estructural, se corrigió el alineamiento de secuencia primaria inicial entre DnaG y gp4. La segunda etapa consistió en modificar el alineamiento de la secuencia de Twinkle con las secuencias de las otras dos proteínas teniendo en cuenta la predicción de estructura secundaria de la primera, es decir haciendo que coincidieran los elementos de estructura secundaria predichos para Twinkle con los de las estructuras de DnaG y gp4 superpuestas (2AU3, 1Q57) (ver figura\_45).

Figura\_45. Alineamiento para la elaboración del modelo atómico 3D.

Se indica el alineamiento que se ha empleado para generar el modelo de Twinkle, que incluye las secuencias de la estructura de DnaG (2AU3), de gp4 (1Q57) y del constructo de Twinkle empleado tanto para los experimentos de SAXS como para EM. Se marca los distintos dominios ZDB, RPD, "linker", helicasa y cola C-terminal con rectángulos en color en naranja, verde, rojo, azul y cian respectivamente. Se indica con una espiral la estructura de hélice y con una flecha las hebras  $\beta$  de las estructuras 2AU3 y 1Q57.



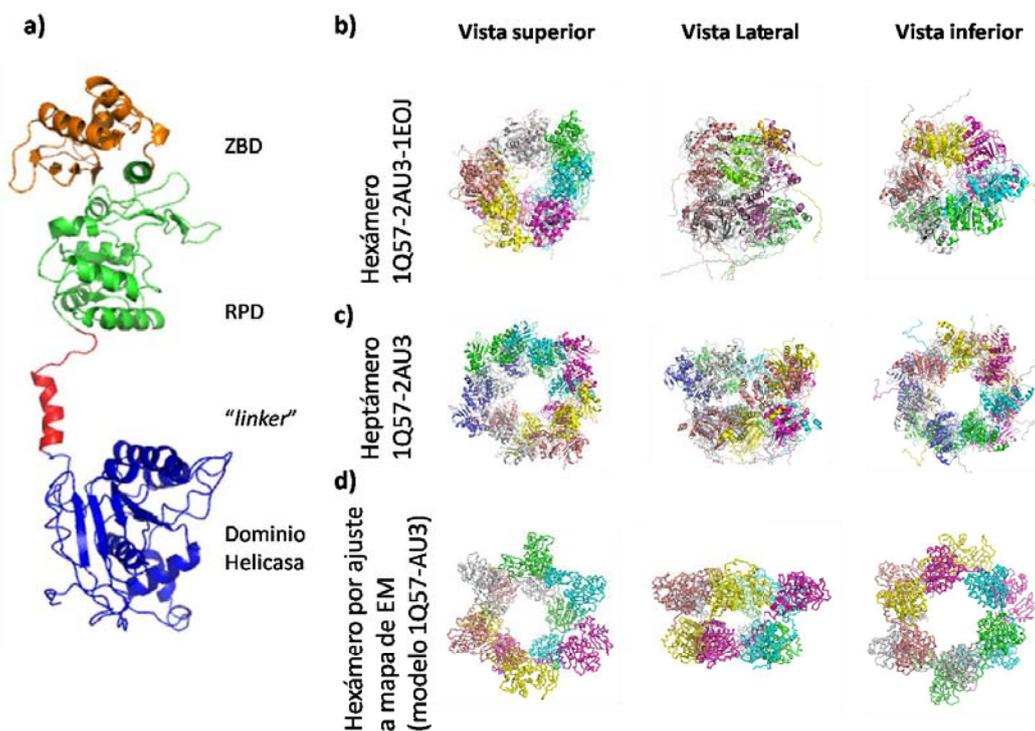
El resultado de similitud e identidad entre los distintos subdominios fueron 17.2 y 13.4% para el ZBD (desde el residuo 43 al 138), 39.8 y 3.4% para el RPD (desde el residuo 139 al 337), y 23.0 y 16.1 % para el dominio helicasa (desde el residuo 338 al 637), en este último se incluyó la secuencia del "linker" pero no la cola C-terminal. Hay que indicar que para el cálculo de similitud e identidad del dominio ZBD sólo se tuvieron en consideración las secuencias de

DnaG y Twinkle, porque no se empleó la secuencia de gp4 para modelar este fragmento en Twinkle. De igual manera, para el fragmento del “*linker*” y del dominio helicasa sólo se consideraron las secuencias de gp4 (PDB:1Q57) y Twinkle (figura 45), el valor final resultante de similitud e identidad en el alineamiento entre las secuencias en los distintos subdominios fue diferente al inicial obtenido por CLUSTALW porque tuvo se le dio más importancia al alineamiento estructural que no sólo al de secuencia primaria.

En la siguiente tercera etapa se definieron con MODELLER todos los elementos de estructura secundaria predichos para Twinkle en base al análisis estructural con las proteínas similares, para que el programa los tuviera en cuenta para la elaboración del modelo. Para la zona N-terminal de Twinkle (del residuo 43 al 361) se emplearon como referencia las estructuras 2AU3 de DnaG (del residuo 1 al 345) y 1Q57 (del residuo 64 al 258), mientras que para la zona “*linker*” y C-terminal sólo se empleó 1Q57 de gp4 (desde el residuo 259 al 549). De esta manera el modelo estructural para el monómero de Twinkle se construyó como una “quimera estructural” entre DnaG y gp4 (ver figura\_46). Con este monómero se construyó el hexámero que apareció en el mapa de densidad electrónica obtenido por crio-microscopía electrónica (ver más adelante, apartado 8.2.1.7).

**Figura\_46. Modelos generados con MODELLER.**

En **a)** se indica la estructura del monómero de Twinkle basado en las estructuras de las helicasas de gp4 y DnaG, indicándose los dominios que lo componen: ZBD (naranja), RPD (verde), “*linker*” (rojo) y helicasa (azul). Además se muestran las vistas superiores, laterales e inferiores del hexámero (1EOJ-2AU3-1EOJ), heptámero (1Q57-2AU3) y hexámero obtenido por EM en **b), c) y d)** respectivamente.



Para realizar los estudios de SAXS, se construyeron dos modelos adicionales que seguían la estructura cuaternaria de hexámero y de heptámero de gp4 (PDBs 1EOJ y 1Q57, de gp4, respectivamente). Para la entrada a MODELLER se “multimerizó” el monómero tridimensional creado para EM, descrito en el párrafo anterior, siete veces para generar el heptámero y seis para el hexámero. En concreto, para generar el heptámero se tomó como referencia el anillo heptamérico de 1Q57 de gp4: se copió siete veces el alineamiento anterior, se septuplicó el PDB 2AU3 para generar siete cadenas de la parte N-terminal, que se “fusionarían” con la parte N- y C-terminal de las siete cadenas que componen el PDB 1Q57 de gp4. De modo similar se generó el hexámero: se sextuplicó el PDB 2AU3 de DnaG para crear los dominios N-terminal, que se fusionaron con seis cadenas de 1Q57 con la secuencia del hexámero 1EOJ (idéntica a 1Q57 pero de menor longitud en la parte N-terminal) se generaron seis cadenas en base a 1EOJ con MODELLER, tomando como referencia el anillo hexamérico de 1EOJ (figura\_46-b).

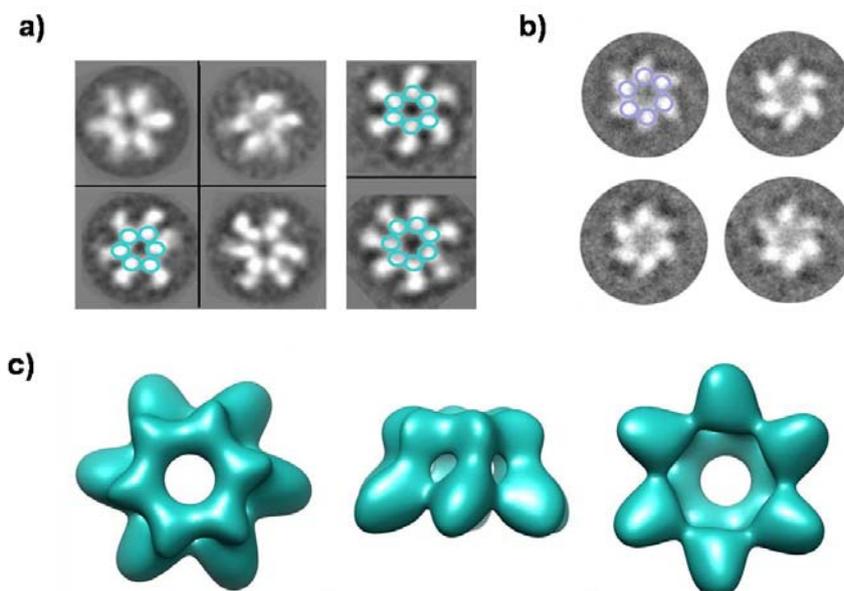
#### **8.2.1.7 Microscopía electrónica.**

Este trabajo fue realizado por M. Lázaro y Dr. M. Valle del CIC-bioGUNE (Derio, Bizkaia), con quien se estableció una colaboración. El análisis por microscopía electrónica del constructo parcial (GB1)- $\Delta$ 358Twinkle se abordó pero no prosperó por el alto grado de heterogeneidad de la muestra. Por el contrario, sí fue posible este tipo de análisis estructural con el constructo Twinkle ya que se purificó con la pureza y homogeneidad necesaria para estudios posteriores por crio-microscopía. Los resultados obtenidos por tinción negativa mostraron que la muestra contenía una mezcla de especies con simetría seis y simetría siete, tal y como se pudo interpretar en la orientación frontal de las partículas (véase las imágenes en la figura\_47a), lo cual fue indicativo de la presencia de una mezcla de heptámeros y hexámeros, pero en un porcentaje variable según la muestra. Con estos datos por tinción negativa se seleccionaron las imágenes con partículas de orden 6 (figura\_47a), con las que se pudo construir un mapa de una resolución aproximada de 30 Å (figura\_47c).

Otro resultado interesante con las micrografías de tinción negativa fue observar la gran flexibilidad de la molécula, en las imágenes se observó claramente como había un dominio que se distribuía al azar respecto al anillo central, a pesar de conservar el estado oligomérico de seis subunidades (figura\_47a). El alto grado de flexibilidad de este dominio (que se asignó como N-terminal) respecto del anillo central (asignado como dominio C-terminal y parte del “linker”), impidió abordar directamente los estudios por crio-microscopía.

**Figura\_47. Imágenes de microscopía electrónica y mapa a baja resolución.**

a) Se muestran las vistas frontales de algunas imágenes de tinción negativa tanto con simetría 7 como simetría 6. En b) se muestran las imágenes de crio-microscopía. Con círculos se indica la región no flexible en las dos imágenes de a) y b). Abajo en c) se muestra el mapa obtenido con tinción negativa desde la vista, de izquierda a derecha, inferior, lateral y superior.

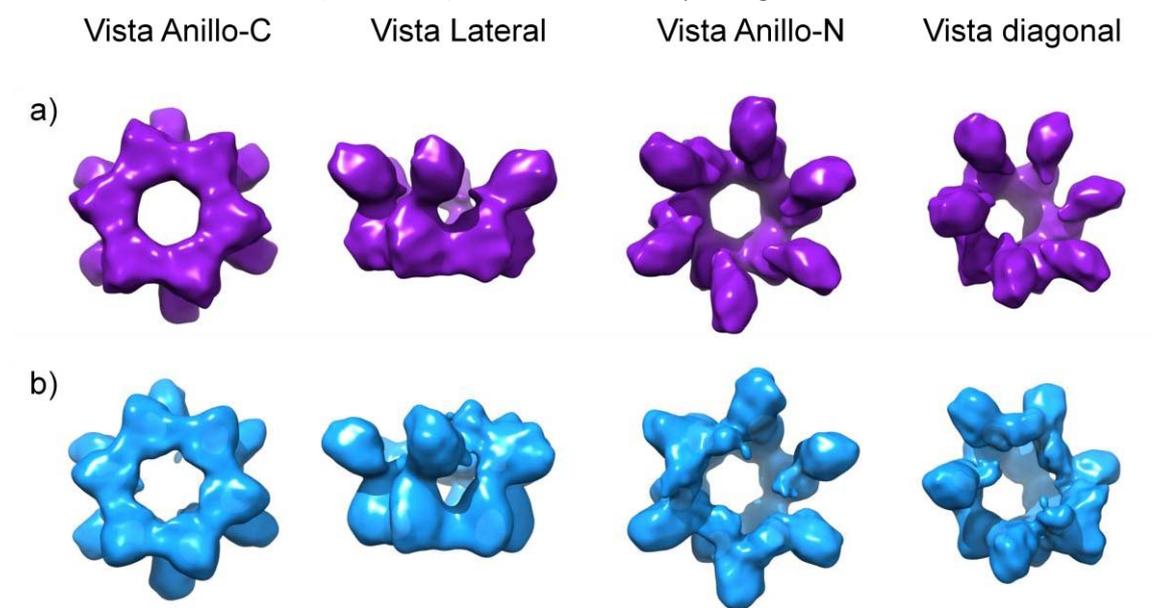


Para reducir la flexibilidad y por tanto tener una muestra homogénea conformacionalmente se realizó un tratamiento de químico de entrecruzamiento irreversible con glutaraldehído (0.075%), para tratar de fijar distintas conformaciones. Para ello, se fijó la proteína a la vez que se separaron las distintas especies conformacionales con un gradiente de glicerol con glutaraldehído por ultracentrifugación (Kastner, Fischer et al. 2008). La fijación de la muestra lo realizó S. Cainz en el laboratorio del Dr. J. Spelbrink de la Universidad de Radboud (Nijmegen-Holanda). El análisis de las imágenes tomadas por microscopía electrónica de las distintas fracciones del gradiente de glicerol mostró que se había conseguido aislar la especie hexamérica en un estado más homogéneo que sin emplear glutaraldehído (figura\_48b). Estas fracciones más homogéneas se crioenfriaron para su análisis por crio-microscopía electrónica, de las que se tomaron 12790 imágenes y con las cuales se elaboró un mapa de densidad electrónica de mayor resolución (13.6 Å) que el anterior generado con imágenes de tinción negativa (30 Å aproximadamente). En el procesado de las imágenes se generaron dos mapas diferentes. Inicialmente, para generar un primer mapa se aplicó simetría de orden 6 a toda la partícula (mapa C6-C6). En este mapa se observó una densidad continua en forma de anillo en la zona Anillo-C, que se asignó al dominio helicasa C-terminal debido al parecido con la estructura de gp4 (esta asunción se comprobó posteriormente con el ajuste del modelo generado por MODELLER, ver más adelante). En la región Anillo-N (atribuida a los 6 dominios N-terminales) se observó la densidad de seis dominios independientes que no contactaban entre sí (figura\_48a y b). Posteriormente, debido a la heterogeneidad observada con

anterioridad en la orientación del dominio N-terminal respecto al C-terminal en las distintas subunidades, se calculó un nuevo mapa donde la parte del Anillo-N no seguía la simetría de orden 6, que se sustituyó por una de orden 1 (C1-C6). En este segundo mapa se encontraron variaciones en la densidad entre los seis dominios (figura\_48a y b), en dos de ellos la densidad electrónica era de menor tamaño y en los otros cuatro se observó una densidad continua entre dominios, indicativo de la presencia de contactos entre los dominios N-terminal.

**Figura\_48. Mapa de crio-microscopía electrónica.**

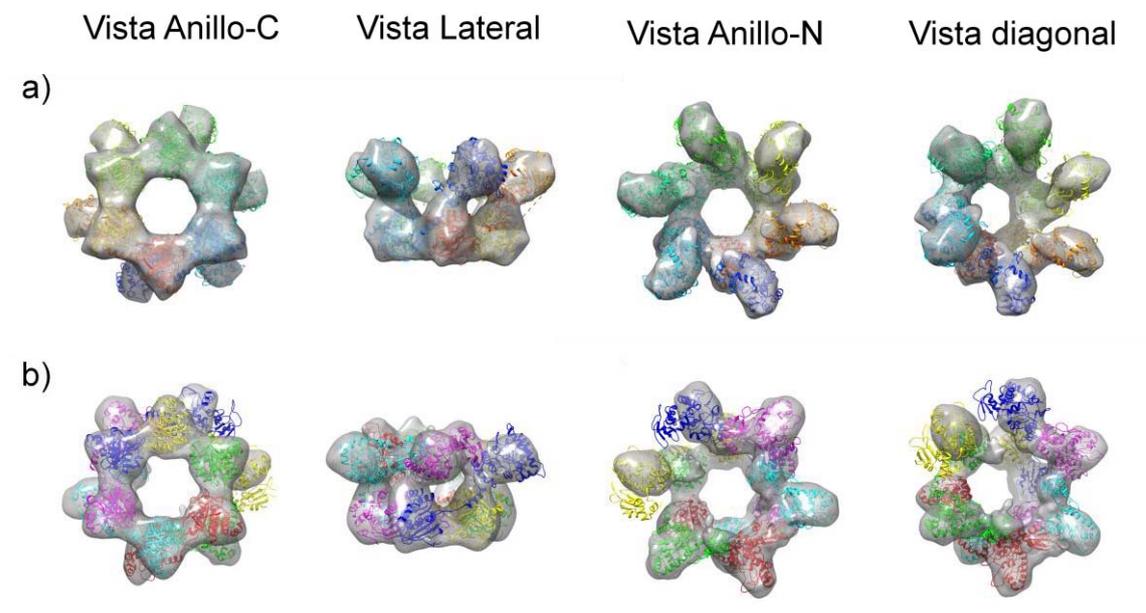
En **a)** se muestra el mapa con simetría C6-C6 (color violeta) De izquierda a derecha, visto desde el anillo-C, vista lateral, visto desde el anillo-N y en diagonal. En **b)** se muestra el mapa con simetría C1-C6 (color azul). De izquierda a derecha visto desde el anillo-C, vista lateral, visto desde el anillo-N y en diagonal.



Con estos mapas se procedió al ajuste del modelo que se había generado con MODELLER, empleado un ajuste flexible por dinámica molecular, tanto en el mapa con simetría C6-C6 como con C1-C6 (figura\_49a y b). El ajuste del modelo a la región Anillo-C-terminal no mostró diferencias entre los dos mapas, pero el ajuste de la región Anillo-N no era igual en C6-C6 que en C1-C6. En el mapa con simetría C6-C6 los dominios N-terminales tenían todos la misma orientación, con los subdominios ZBD orientados hacia la cara externa de la partícula (figura\_49a). En cambio, para el mapa C1-C6, el subdominio ZBD variaba su orientación respecto al subdominio RPD entre monómeros dentro del hexámero. Tal y como se ha mencionado, en cuatro dominios N-terminales se observaron contactos entre el ZBD de un monómero con el RPD del monómero siguiente. Los otros subdominios ZBD que no contactaban al RPD no ajustaron completamente dentro de la densidad debido a una contracción de la misma, seguramente por el desorden generado por el movimiento del subdominio ZBD. Estos datos indicaron claramente la elevada flexibilidad del "linker" que conecta el ZBD con el RPD (figura\_49a y b).

**Figura\_49. Ajuste de modelos 3D a los mapas de crio-EM.**

Se muestra el ajuste del seis moléculas generadas por MODELLER a los dos mapas de densidad electrónica obtenidos por crio-EM, en **a)** cuando se aplica una simetría C6-C6 y en **b)** al aplicar C1-C6.



Los contactos entre subdominios de distintos monómeros se denominan contactos en *trans*, y se han observado en otras helicinas/primasas como gp4 y en el heterocomplejo DnaB/DnaG. Estos contactos se han interpretado como un mecanismo de cebamiento del ADN. En estas proteínas estos movimientos realizarían la función de unir y fijar al ADN. En estas proteínas se ha observado que el ZBD une al ADN para fijarlo al subdominio RPD del otro monómero, el cual catalizaría la reacción de polimerización de los nucleótidos que compondrían el cebador unido a la cadena sencilla del ADN (Bailey, Eliason et al. 2007; Geertsema, Kulczyk et al. 2014). No obstante, al no tener Twinkle actividad primasa, la función de estos dominios en esta proteína es totalmente desconocida.

Con el resultado del ajuste de seis monómeros en el mapa de densidad electrónica C6-C6 se obtuvo un modelo de hexámero de Twinkle el cual se empleó para las curvas de ajuste de SAXS, como se mencionó anteriormente.

### **8.2.1.8 Estructura de Twinkle en solución por Dispersión de rayos X de ángulo bajo (SAXS).**

#### **8.2.1.8.1 Medida experimental de Twinkle mediante SAXS.**

Gracias a la técnica de SAXS se realizaron varios estudios de Twinkle en disolución, con cuatro objetivos principales. El primero de ellos fue estudiar el equilibrio entre estados oligoméricos de hexámero y heptámero que se producen en función de la concentración de NaCl según lo observado por otros autores empleando la microscopía electrónica (Ziebarth, Gonzalez-Soltero

et al.) El segundo objetivo fue analizar si la unión del ligando UMPPNP a Twinkle generaba alguna variación estructural en la proteína, bien en su flexibilidad o si estabilizaba algún estado oligomérico particular, ya que SAXS es sensible a este tipo de cambios. El motivo de analizar este ligando fue porque es el análogo no hidrolizable del UTP, utilizado por Twinkle como reactivo energético para catalizar el desenrollamiento del ADN. El tercer objetivo fue comprobar si la muestra expresada en células de insecto (Twinkle-SF9) o con *E. coli* (Twinkle) presentaba diferencias en la estructura cuaternaria o en la flexibilidad. Para ello se analizó Twinkle-SF9 en las mismas condiciones con las que se había analizado con anterioridad Twinkle por SAXS (ver tabla\_17). Finalmente, el último pero más importante era intentar obtener un modelo estructural de Twinkle en disolución, empleando como partida los modelos generados por MODELLER y ajustados a las curvas experimentales de SAXS.

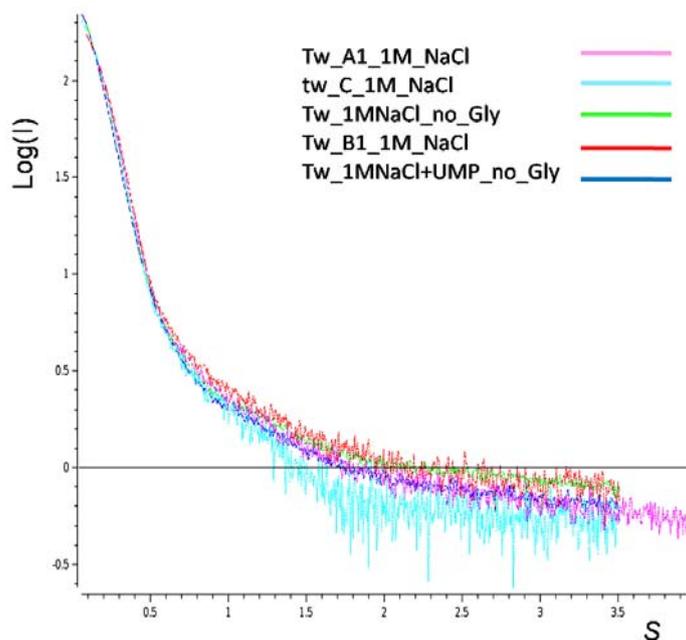
Para todo este trabajo se realizaron los ensayos mediante SAXS con Twinkle en las condiciones químicas:

- Tampón SEC: citrato sódico 50 mM pH 6.5, NaCl 1 M, glicerol 10%, L-Arg 50 mM, L-Glu 50 mM, BME 2mM.
- Tampón SEC-0.5M: citrato sódico 50 mM pH 6.5, NaCl 0.5M, glicerol 10%, L-Arg 50 mM, L-Glu 50 mM, BME 2 mM.
- Tampón SEC-0.3M: citrato sódico 50 mM pH 6.5, NaCl 0.3 M, glicerol 10%, Arg 50 mM, Glu 50 mM, BME 2mM.
- Tampón SEC-Gli: citrato sódico 50mM pH 6.5, NaCl 0.5M, L-Arg 50mM, L-Glu50 mM, BME 2mM.
- Tampón SEC-Gli+UMP: citrato sódico 50 mM pH 6.5, NaCl 0.5 M, L-Arg 50 mM, L-Glu 50 mM, BME 2 mM, UMPPNP 5 mM.

Para cada condición experimental se realizaron medidas con cinco concentraciones distintas de proteína, en un rango entre 0.3 de mínimo a un máximo de 1.8 mg/ml, con valores intermedios que se variaron 0.2-0.3 mg/ml, aunque las concentraciones no fueron idénticas en todas las series (ver tabla\_7 de Materiales y Métodos). De todas las medidas experimentales se obtuvieron finalmente nuevas curvas experimentales, que se detallan en la tabla 17. De las nueve curvas, sólo cinco pudieron ser utilizadas para los estudios citados anteriormente por que algunas de ellas no tenían calidad suficiente para poder ser fiables. Sólo se emplearon aquellas con un valor de calidad del Guinier mayor al 80% (tabla 17), las cuales se representan en la figura\_50.

**Figura\_50. Curvas de SAXS de Twinkle.**

Se muestran las curvas de SAXS de Twinkle que presentaron un valor de Guinier superior al 80%. Las curvas que se muestran abajo son Tw\_A1\_1M\_NaCl (en rosa), tw\_1Mb3\_002\_PB (azul celeste), Tw\_1MNaCl\_no\_Gly (verde), Tw\_B1\_1M\_NaCl (rojo), Tw\_b2SECGliUMP (azul oscuro).



El análisis de los datos de las curvas de SAXS (tabla 17) indica que las medidas que se realizaron de Twinkle en tampones con una concentración de NaCl inferior a 1M (curva Tw\_B1\_05M\_NaCl, Tw\_B1\_03M\_NaCl y tw\_C\_1M\_NaCl UMP) presentaban agregación y por tanto el valor de Guinier era muy bajo (<80%) signo de mala calidad de los datos, por lo tanto se descartaron estas curvas del estudio. Este resultado no fue inesperado porque todos los ensayos biofísicos realizados con anterioridad mostraron la inestabilidad de Twinkle a concentraciones de NaCl menores a 1 M. Finalmente sólo se pudieron emplear las medidas de las muestras con una concentración de 1M NaCl.

Respecto a la comparación entre las curvas de Twinkle sin ligando, y con UMPPNP (curvas Tw\_A1\_1M\_NaCl y Tw\_B1\_1M\_NaCl sin UMPPNP y Tw\_1MNaCl+UMP\_no\_Gly con UMPPNP) no se observó diferencias entre las mismas, aunque la interacción de UMPPNP con Twinkle podría verse afectada por las altas concentraciones de NaCl con las que se realizaron las medidas.

Como tercer objetivo, la comparación de las muestras expresadas con células de insecto (Twinkle-SF9) o con *E. coli* (Twinkle) no mostraron diferencias destacables, ni en el perfil de las curvas, ni la distancia de máxima de la partícula (Dmax), ni en el radio de giro (Rg). Tal y como se describe en el siguiente apartado, se pudo generar modelos de estructuras de Twinkle para ambas curvas, pero tampoco se encontraron diferencias destacables por SAXS. Estos datos son

indicativos que la estructura en disolución de Twinkle no se afectada por el uso de distintos sistemas de expresión recombinante (ver tabla\_18).

De todas las pruebas, se obtuvieron tres curvas con un valor de ajuste de Guinier mayor al 99% (ver tabla 17), de las cuales se detalla a continuación la curva Tw\_B1\_1M\_NaCl (representada en la figura\_51) porque con ella se consiguieron los mejores resultados de ajuste de los modelos (generados por MODELLER) como se indica en el siguiente apartado. En concreto, esta curva tuvo un valor de ajuste de Guinier de 99,2% (ver Tabla\_17), indicativo de su buena calidad.

**Tabla\_17. Experimentos de SAXS con Twinkle.**

Abajo se representan las curvas experimentales de las curvas de SAXS (*log(I) vs S*).

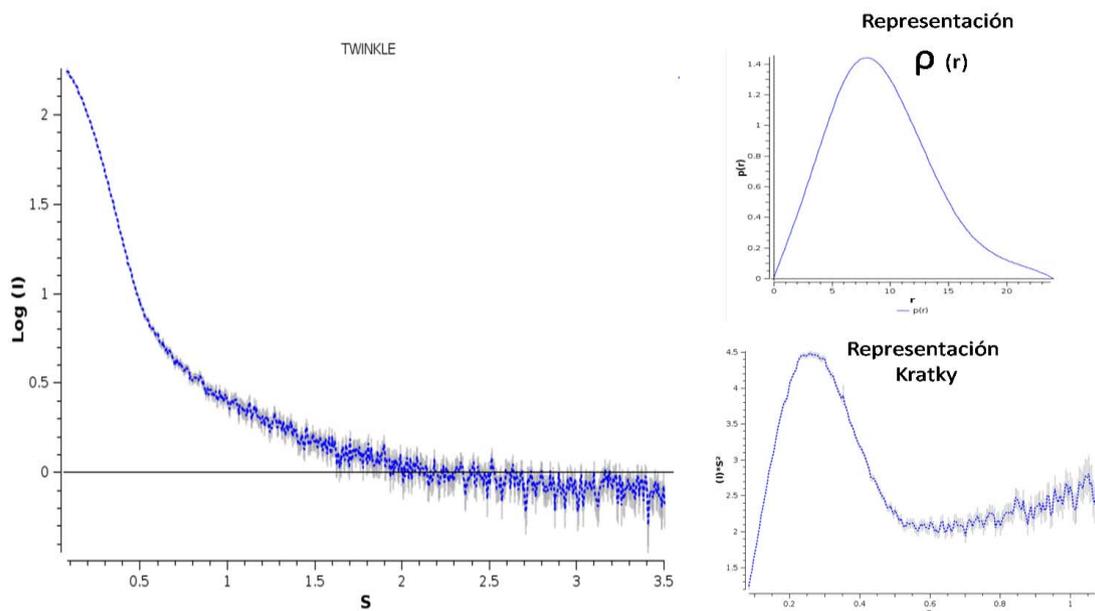
| Muestra     | Nombre de la curva   | Rango concentración (mg/ml) | Condiciones del tampón  | Calidad del Guinier |
|-------------|----------------------|-----------------------------|-------------------------|---------------------|
| Twinkle     | Tw_A2_1M_NaCl        | 1.88 – 0.34                 | SEC                     | 0.751               |
| Twinkle     | Tw_A1_1M_NaCl        | 1.88 – 0.34                 | SEC                     | 0.992               |
| Twinkle     | Tw_B1_1M_NaCl        | 0.9 – 0.33                  | SEC                     | 0.987               |
| Twinkle     | Tw_B1_05M_NaCl       | 0.9 – 0.33                  | SEC-0.5M                | 0.655               |
| Twinkle     | Tw_B1_03M_NaCl       | 0.9 – 0.33                  | SEC-0.3M                | 0.00                |
| Twinkle     | Tw_1MNaCl_no_Gly     | 1.5 – 0.33                  | SEC-Sin-glicerol        | 0.845               |
| Twinkle     | Tw_1MNaCl+UMP_no_Gly | 1.5 – 0.33                  | SEC-Sin-glicerol+UMPPNP | 0.812               |
| Twinkle-SF9 | tw_C_1M_NaCl         | 1.27– 0.33                  | SEC                     | 0.993               |
| Twinkle-SF9 | tw_C_1M_NaCl UMP     | 0.9                         | SEC+UMPPNP              | 0.493               |

La distancia máxima la Twinkle se calculó a partir de la representación de la distribución de la  $\rho(r)$  versus el radio y el resultado fue de 240.2 Å. En esta representación además se observó que la parte final de la curva muestra una interrupción en la pendiente mostrando una cola al final del perfil, que puede ser debido a una partícula más alargada o a la presencia de flexibilidad en la partícula (ver figura\_51). La representación de Kratky mostró un perfil correspondiente a una proteína globular pero con cierto grado de flexibilidad, tal y como también se observó en la curva "*log(I) vs S*" la cual mostraba el perfil típico de una proteína flexible (figura\_51).

El peso molecular de Twinkle se calculó empleando el BSA (peso molecular de 66KDa) como un estándar de referencia. La intensidad  $I(0)$  de Twinkle fue 221 a una concentración de 1.5mg/ml, y la intensidad de BSA fue 77 a una concentración de 3.6mg/ml, por tanto el resultado final del peso molecular de Twinkle fue 454 KDa (para detalles del cálculo ver apartado 7.9 De Materiales y métodos).

**Figura\_51. Curva SAXS Twinkle.**

A la izquierda se muestra la curva experimental de SAXS de la curva Tw\_1M\_B1. A la derecha arriba se muestra la representación de la distribución  $\rho(r)$  versus el radio de giro, que permite estimar la distancia máxima. A la derecha abajo la representación de Kratky, consistente con una proteína globular compacta.



#### 8.2.1.8.2 Ajuste de los modelos de Twinkle a las curvas experimentales de SAXS.

La segunda parte del estudio de Twinkle por SAXS consistió en emplear los modelos generados por predicción de estructura con MODELLER, tanto del hexámero (2AU3-1Q57-1EOJ) como del heptámero (2AU3-1Q57) para seleccionar aquellos que mejor describían el estado de Twinkle en disolución. Los ajustes iniciales de la curva teórica con la curva experimental de estos modelos como si de una proteína globular y rígida se tratase, con la curva experimental, dieron unos valores realmente malos, el valor de  $\chi^2$  fue 18 y 11, hexámero y heptámero respectivamente, claro signo que esos modelos no representaban en absoluto a la partícula en disolución (figura\_52).

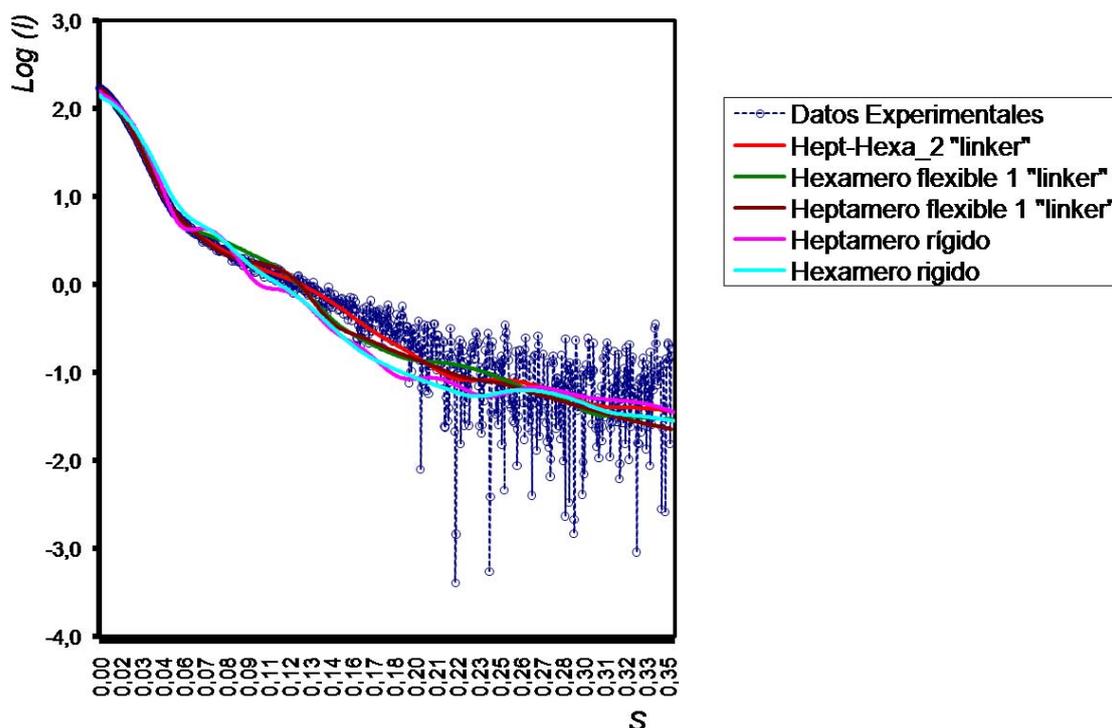
Como la representación Kratky y la curva de  $\text{Log}(I)$  vs  $S$  ya indicaban cierta flexibilidad y asociado a los malos resultados del ajuste de los modelos anteriores se decidió repetir el ajuste dotando de flexibilidad al modelo.

En base a los conocimientos de gp4, se le dio flexibilidad al "linker" que conecta al dominio primasa con el dominio helicasa. Para ello hubo que emplear la última versión de EOM (Bernado, Mylonas et al. 2007) la cual permitió asignar una región flexible en la proteína y generar 5000 modelos en los cuales la región flexible se disponía en el espacio aleatoriamente, y por tanto un dominio respecto al otro se desplazaba al azar en el espacio pudiendo dejar uno fijo como referencia todos los monómeros. Finalmente se generaron 5000 modelos por cada estado oligomérico, en los cuales el dominio helicasa formaba un anillo sin flexibilidad e idéntico en todos los modelos y el dominio primasa tenía también una estructura rígida y

globular pero se distribuía aleatoriamente alrededor del anillo que formaba el dominio helicasa gracias al “linker” flexible. De estos dos grupos de 5000 modelos EOM aplicó un algoritmo genético con el que seleccionó un conjunto 20 por cada estado oligomérico de aquéllos cuyas curvas teóricas promediadas describen mejor la curva experimental. A pesar de que el resultado mejoró notablemente, ya que el valor de  $\chi^2$  descendió de 18 a 4.1 para el hexámero (curva verde en la figura\_52) y de 11 a 2.5 para heptámero (curva marrón en la misma figura), el ajuste continuó siendo un valor insuficientemente bueno. Por otro lado, con los modelos del hexámero generados con MODELLER basados en la estructura de gp4 (1EOJ) no se logró obtener buenos valores de ajuste, ya que el valor de  $\chi^2$  no se lograba bajar de 2.5. Por este motivo, se decidió continuar con el modelo de hexámero que se obtuvo a partir de los mapas de crio-microscopía electrónica, el cual mostró un valor de  $\chi^2$  más bajo ( $\chi^2=1.97$ ).

**Figura\_52. Ajuste de los modelos de Twinkle a las curvas de SAXS.**

a) En azul oscuro y círculos huecos se representa la curva experimental Tw\_B1\_1M.dat y con línea continua se representa las distintas curvas teóricas de ajuste calculadas a partir de los modelos con la molécula rígida hexámero (azul celeste) y heptámero (rosa), con un “linker” flexible entre los dominios RPD y helicasa con el hexámero (verde) y el heptámero (marrón), y en rojo la curva de la combinación de heptámero y hexámeros (de microscopía electrónica) con dos “linkers” flexibles entre el ZBD y el RPD, y entre el RPD y el dominio helicasa.



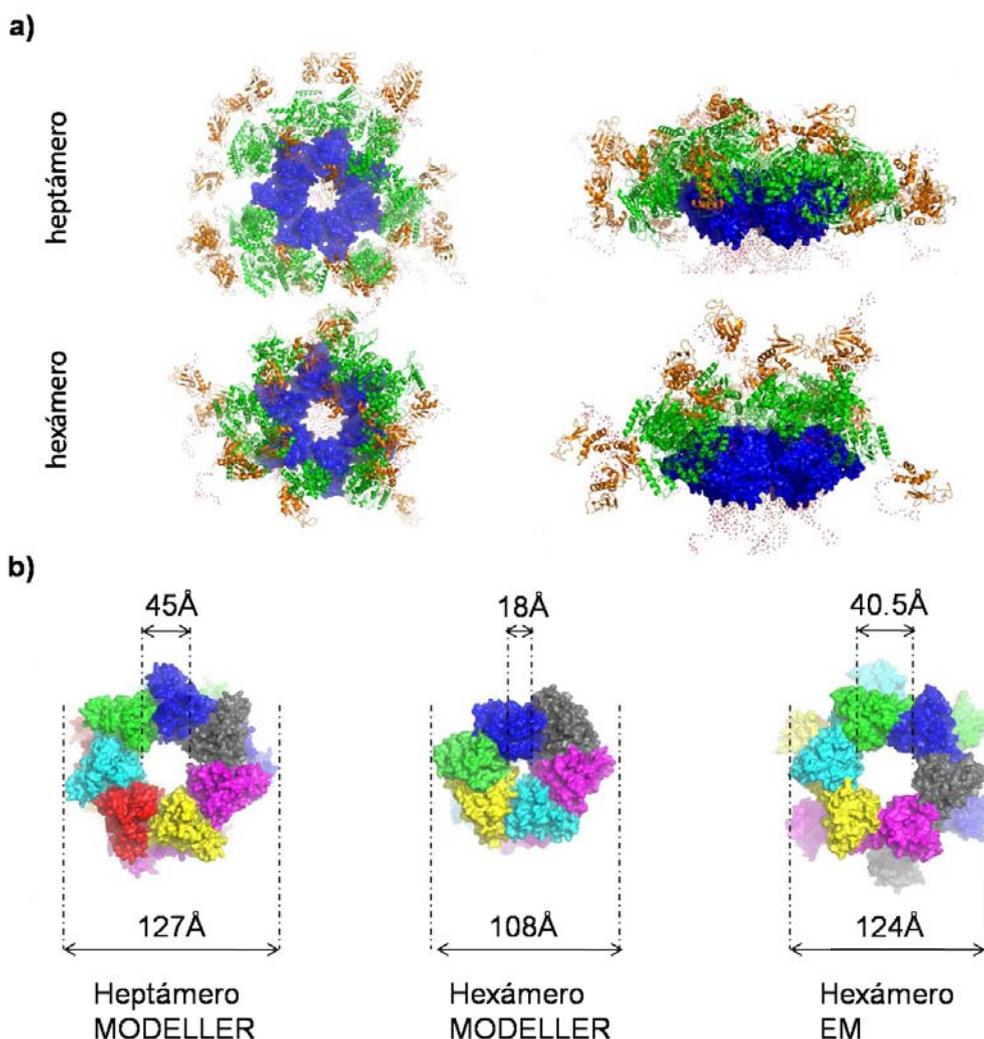
Aplicando el mismo protocolo descrito anteriormente con el programa EOM se logró mejorar el ajuste de las curvas con los modelos cuando se definió una segunda región flexible adicional, el “loop” que conecta el subdominio ZBD con el subdominio RPD, pero entonces se emplearon el hexámero obtenido por crio-EM y el heptámero basado en gp4 (figura\_53a). Con los dos “linkers” flexibles, se mejoraron los valores de  $\chi^2$  a 0.81 para el hexámero y 0.78 para el heptámero. El mejor resultado fue la combinación de los modelos de hexámeros y heptámeros

ya que dio un valor de  $\chi^2$  de 0.789 (curva roja), con la curva Tw\_B1\_1M.dat (ver figura\_52). Por tanto, con estos resultados ya se identificaron la existencia de estas dos formas oligoméricas en solución.

Es interesante destacar que las dimensiones del modelo hexamérico de Twinkle generado a partir del mapa de microscopía electrónica presenta unas dimensiones en el anillo helicasa (diámetro canal 40.5 Å y diámetro externo 122 Å) más parecidas al heptámero de Twinkle generado mediante MODELLER basado en la helicasa gp4 (diámetro canal 45 Å y diámetro externo 127 Å) que el hexámero obtenido en base a gp4 (diámetro canal 18 Å y diámetro externo 108 Å) (ver figura\_53b).

**Figura\_53. Resultado del ajuste a la curva de SAXS y modelos EOM.**

En **a)** se muestran las vistas frontales y laterales de los modelos que incluyen 20 estructuras del heptámero y del hexámero cuya curva de SAXS teórica promediada describen la curva experimental de SAXS. En la figura se ha representado la superficie del dominio helicasa (en azul), y las estructuras secundarias del RPD (en verde) y del ZBD (en naranja), y con un punto los carbonos  $\alpha$  del enlace peptídico de las regiones flexibles (rojo). En **b)** Se muestran las estructuras de Twinkle empleadas para SAXS orientadas desde el anillo que forman el dominio helicasa de izquierda a derecha el heptámero y hexámero generados mediante MODELLER basadas en gp4 y el hexámero generado por ajuste al mapa de EM.



Siguiendo el protocolo anterior, se calcularon los valores de ajuste de los dos grupos de 5000 modelos generados anteriormente para la curva Tw\_B1\_1M.dat a las otras cuatro medidas de Twinkle que ofrecieron un mejor valor de Guinier. Se realizó tanto el ajuste de los hexámeros y heptámero de modo independiente y su combinación. En dos de los cinco ensayos realizados se observó una mezcla de dos especies oligoméricas, hexámeros y heptámeros, en diferente proporción. En estas curvas los cálculos de ajuste de los distintos estados oligoméricos de modo independiente no presentó una diferencia importante respecto a la combinación de ambos. Sin embargo, en los otros tres ensayos sólo se encontró la proteína formando heptámeros cuando se calculaba el ajuste con la combinación de ambos. En estos cálculos sí se observó que el ajuste sólo del hexámero de crio-EM daba un valor más alto de  $\chi^2$  que la combinación entre hexámero y heptámero, aunque el ajuste empleando sólo el heptámero no mejoró el valor de  $\chi^2$  con respecto a la combinación de hexámero y heptámero (ver Tabla\_18). Estos datos evidencian que se puede conocer la proporción de cada modelo teórico al ajustarlo a la curva experimental, y por tanto saber si existen sólo una o varias especies en la disolución. Por otra parte, aunque las cinco curvas presentaban unos buenos valores de Guinier, sólo se consiguió un buen ajuste de la curva teórica calculada con los modelos para los tres experimentos en los que la muestra contenía glicerol. Las dos curvas de la muestras sin glicerol dieron un valor de  $\chi^2$  mayores a 1.5, indicativo que EOM no era capaz de genera modelos para todas las especies que definen esa curva, lo que sugiere que existían otras conformaciones en disolución (ver tabla\_18). Por tanto, los mejores valores de ajuste fueron obtenidos a partir de la combinación de dos poblaciones de 5000 modelos cada una de hexámero y heptámeros. Una a partir del modelo del heptámero (2AU3-1Q57) y la otra a partir del modelo conseguido por el ajuste al mapa de crio-microscopía electrónica (ver tabla\_18).

**Tabla\_18. Resultados de SAXS con Twinkle.**

SEC es el tampón de elución de la cromatografía de exclusión por tamaño.

| Curva                | Proteína     | Condiciones de tampón | $\chi^2$ | Ratio hexa/hepta | Rg    | Dmax   |
|----------------------|--------------|-----------------------|----------|------------------|-------|--------|
| Tw_A1_1M_NaCl        | Twinkle      | SEC                   | 1.339    | 7/13             | 72.04 | 270.22 |
| Tw_B1_1M             | Twinkle      | SEC                   | 0.798    | 13/7             | 69.82 | 260.61 |
| Tw_1MNaCl+UMP_no_Gly | Twinkle      | SEC-glicerol+UMPPNP   | 4.869    | 0/20             | 78.38 | 288.61 |
| Tw_1MNaCl_no_Gly     | Twinkle      | SEC-glicerol          | 4.964    | 0/20             | 78.68 | 288.76 |
| tw_C_1M_NaCl         | Twinkle(SF9) | SEC                   | 1.285    | 0/20             | 76.52 | 280.74 |

En conclusión, el análisis de los modelos que ajustaron a la curva de SAXS permitió asignar los dominios globulares y las regiones flexibles de la proteína en solución, tanto para el hexámero

como el heptámero. De esta manera, se observó que en disolución el ZBD está unido al RPD por un “*linker*” muy flexible, y se distribuye de modo totalmente aleatorio alrededor del RPD. A su vez, el RPD se une al dominio helicasa por un “*linker*” flexible pero de menor longitud, por lo que la distribución respecto al dominio helicasa se limita en el espacio. También se observó que los anillos que constituyen el dominio helicasa del heptámero y el hexámero no son muy diferentes entre sí por el tamaño, ya que el diámetro del anillo del heptámero mide 127 Å y el hexámero mide 124 Å. En los dos casos podrían acomodar un ADN de doble cadena en el canal interior, el cual tiene un diámetro de 40.5 y 45.0 Å, para el hexámero y el heptámero respectivamente (figura\_53b).

### 8.2.2 Discusión de los resultados de Twinkle.

La expresión recombinante en *E. coli* de Twinkle entera es soluble y activa. Aunque no se conoce la función del dominio N-terminal de esta proteína se ha observado que la pérdida del dominio ZBD y/o del RPD afecta drásticamente a la solubilidad de la proteína. Los constructos de Twinkle en los cuales se han dividido los dos dominios N-terminal y C-terminal y expresado de modo independiente no son solubles cuando se concentra por encima de 1 mg/ml sin una proteína de fusión. Incluso la producción del constructo  $\Delta 151$ Twinkle al cual sólo se le eliminó el subdominio ZBD, aunque fue soluble a concentraciones superiores a 1mg/ml presentó una gran agregación. Estos datos sugieren que los contactos entre los subdominios RPD y ZBD, y RPD con el dominio helicasa estabilizan al complejo porque al eliminar cualquiera de los dos subdominios la solubilidad de proteína se ve afectada. Datos bioquímicos de otros grupos ya indicaron que la escisión de parte del domino N-terminal da lugar a un descenso de la actividad de la enzima (Korhonen, Pham et al. 2004), lo que podría ser debido a su menor solubilidad, posiblemente generando inestabilidad.

El dominio helicasa con el “*linker*” ( $\Delta 358$ Twinkle) conservó el estado oligomérico de hexámero, según los datos de cromatografía de exclusión en gel y microscopía electrónica. El constructo helicasa sin el “*linker*” ( $\Delta 386$ Twinkle) purificado por cromatografía de exclusión por tamaño eluyó en el mismo volumen que el constructo con “*linker*”, lo que sugiere que también conserva la misma oligomerización que el constructo  $\Delta 358$ Twinkle. Sin embargo, ambos constructos presentaron una gran heterogeneidad estructural, lo que sugiere que el subdominio RPD ayuda a reducir la heterogeneidad, posiblemente mediante su interacción con el dominio helicasa. Para este dominio, los análisis por DSF mostraron múltiples picos en el termograma los cuales podrían ser ocasionados por cambios en la estructura cuaternaria de este constructo de hexámero a otro oligómero menor: trímero, dímero o monómero. La proteína entera se comportó diferente, la transición entre plegada y desplegada se produjo en

una sola etapa, y la pérdida de la estructura cuaternaria ocurrió como consecuencia del proceso de desplegamiento en esta misma etapa.

Recientemente, se ha observado como la estructura de anillo del hexámero de DnaB, la cual es una helicasa de la SF4 como Twinkle, cambia en función de si no está interaccionando con ninguna proteína, si forma un hetero-complejo con DnaC, que la responsable de cargar a la helicasa en el ADN, y si cuando DnaC se libera e interacciona con DnaG (Arias-Palomo, O'Shea et al. 2013). Es posible que el dominio N-terminal ayude a fijar en Twinkle las posiciones de las distintas subunidades en el hexámero y heptámero en Twinkle aunque el dominio responsable de la oligomerización se el C-terminal, de modo análogo a heterocomplejo DnaC/DnaB y DnaG/DnaB.

Las imágenes de microscopía electrónica demostraron que el dominio helicasa es capaz de adoptar una estructura cuaternaria de simetría seis. Esa simetría también se observó en la proteína entera, además de una segunda simetría de orden siete. Se elaboraron tres mapas de densidad electrónica con la estructura de hexámero, el primero por tinción negativa a una resolución de 30 Å y dos por crio-microscopía tras el tratamiento con glutaraldehído alcanzando una resolución de 13.5 Å. Los mapas generados a partir de las imágenes generadas gracias a la crio-microscopía permitieron ajustar seis modelos del monómero elaborado por predicción estructural y reconstruir dos modelos con estructura de hexámeros. La diferencia entre ambos fue la simetría en el dominio N-terminal C1 para uno y C6 para el segundo, mostrando la flexibilidad del N-terminal, observada en las imágenes sin tratamiento de glutaraldehído, y más concretamente en el subdominio ZBD. Con este modelo se ha comprobado la similitud estructural entre Twinkle y gp4. En ambas proteínas, el dominio responsable de la oligomerización es el dominio helicasa, por contactos entre la superficie en la cual acomoda al NTP y además por las interacciones a través del "linker" de un monómero con el dominio helicasa del monómero adyacente, ubicando el dominio RPD junto el monómero vecino, provocando algunos contactos. Estos contactos podrían permitir una coordinación entre los movimientos relativos de los dominios helicasa y el subdominio RPD durante la translocación del ADN en Twinkle, de modo similar a como se ha descrito para gp4 (Lee and Richardson 2011; Geertsema, Kulczyk et al. 2014) .

También se observó en el mapa de crio-EM los contactos entre el ZBD y el RPD tipo *trans*, y los movimientos que a pesar de la fijación con glutaraldehído siguió teniendo el subdominio ZBD, posicionado diferente según el monómero, indicativo de la flexibilidad del "linker" que une los subdominios RPD y ZBD. En gp4 y DnaG los movimientos de ZBD respecto al RPD van relacionados con la actividad primasa.

El estudio de SAXS permitió confirmar, al igual que se observó por EM, que Twinkle en disolución tiene dos estructuras cuaternarias distintas de hexámero y heptámero donde el dominio helicasa es el responsable de la oligomerización, interaccionando entre sí para formar una estructura con forma de anillo. En gp4 el cambio de oligomerización va relacionado con la unión al ssADN (Crampton, Ohi et al. 2006). En Twinkle se podría hipotetizar un mecanismo similar de unión al ADN en el cual se produce una transición de heptámero a hexámero.

En modelos generados para SAXS, el tamaño del anillo es similar entre el heptámero generado a partir de la estructura cristalográfica de gp4 (código PDB, 1Q57) y el hexámero generado por microscopía electrónica, tienen un tamaño de diámetro de 130-121 Å, respectivamente y 40 Å en el canal interno. El tamaño anillo del heptámero Twinkle es similar al que muestra la estructura cristalográfica de gp4 (código pdb:1Q57), sin embargo no sucede igual entre los hexámeros de ambas proteínas ya que difieren notablemente en su tamaño, la estructura de gp4 (1EOJ) tiene un canal del anillo de 15 Å de diámetro (con capacidad para una sola hebra de ADN) mientras que el del hexámero de Twinkle mide 40 Å (capaz de alojar un ADN de doble cadena). Esta mayor anchura del anillo interno del hexámero de Twinkle podría justificar en cierta medida la mayor afinidad que presenta la enzima por el dsADN respecto al ssADN tal y como se ha observado en los ensayos EMSA en esta tesis y Sen y co-autores (Sen, Nandakumar et al. 2012).

En disolución los dos subdominios ZBD y RPD que componen el dominio primasa están conectados entre sí por un "linker" desordenado y a su vez el dominio RPD está unido al dominio helicasa por otro "linker" también parcialmente desestructurado. Estos dos "linker" desordenados provocan que Twinkle tenga una gran flexibilidad en su dominio N-terminal, respecto al anillo rígido helicasa y en los dos subdominios. Esta flexibilidad también se encontró en las distintas estructuras cristalográfica de la helicasa gp4 (1NUI y 1Q57) de fagoT7. Podría relacionarse esta flexibilidad con la función de estas enzimas que realizan simultáneamente la reacción de traslocación y cebamiento, además de interaccionar con otras proteínas como la polimerasa para coordinarse con la polimerización del ADN. No obstante, no se ha demostrado pero Twinkle una actividad primasa, a la vez, la delección de este dominio desestabiliza el hexámero C-terminal, lo que implica una comunicación entre dominios, probablemente necesaria para la actividad.

Además, mediante SAXS se ha podido observar la flexibilidad y desordenada de la cola C-terminal en Twinkle, aunque estudios sobre su posible interacción con la polimerasa podría desempeñar una función análoga a la cola C-terminal de gp4, la cual también se encuentra desordenada y está implicada en la interacción con gp5, teniendo un papel fundamental durante la replicación (Lee and Richardson 2011).

Las dos técnicas anteriores, microscopía electrónica y SAXS, mostraron una gran congruencia entre ellas, a pesar de estas basadas en principios metodológicos dispares. Ambas técnicas mostraron el equilibrio heptámero/hexámero, la flexibilidad de los subdominios ZDB y RPD, y los modelos estructurales tridimensionales generados por predicción se ajustaban con un elevado grado de precisión a los resultados tanto de las curvas de SAXS como a los mapas de crio-microscopía electrónica. Por otra parte, la fallida cristalización de Twinkle debió ser originada por la gran flexibilidad del sistema. La obtención de un constructo sin colas desordenadas y en el cual los subdominios ZBD y RPD se encuentren fijados en una posición estable mediante algún heterocomplejo con ADN, será determinante para poder lograr su cristalización.



# **CONCLUSIONES**



## 9 CONCLUSIONES.

### 9.1 Conclusiones del proyecto TFAM en complejo con Site-X.

- TFAM forma un complejo con forma de “U-turn” con la secuencia Site-X en el cual el ADN es doblado en dos puntos 90 °.
- La unión de TFAM con Site-X se realiza de un modo coordinado por tres regiones diferentes: HMGbox1, HMGbox2 y “linker”.
- Los residuos implicados en doblar el ADN en HMGbox2 presentan más desorden. Además, y la HMGbox2 posee una región con distintos trazados de la cadena polipeptídica, lo que sugiere mayor flexibilidad de este dominio respecto a HMGbox1.
- El reconocimiento de unión entre TFAM y Site-X puede venir determinada por un mecanismo de lectura indirecta asociados a las propiedades físicas intrínsecas a la secuencia: una región central muy rígida reconocida por el “linker”, flanqueada una distancia definida de dos regiones pre-dobladas del ADN, una un poco rígida de unión con HMGbox1, una flexible de unión a HMGbox2.
- La formación del complejo TFAM/Site-X es termodinámicamente endotérmico, y tiene una afinidad similar al complejo TFAM/LSP. Además presenta la misma naturaleza termodinámica que otras HMGbox como Syr, Ler, Sox y otros factores de transcripción que unen el ADN por el surco menor y lo doblan.

### 9.2 Conclusiones Twinkle.

- Se ha realizado la expresión recombinante de modo soluble en escala de mg en *E. coli* múltiples constructos de la helicasa humana mitocondrial Twinkle, definiendo diferentes dominios: ZBD, RPD y helicasa. Los constructos obtenidos son: Twinkle (desde el residuo 43-684) con la cola de histidina en N y C-terminal, el dominio helicasa con la cola C-terminal de distinta longitud: 386/684, 386/636, 386/647, 386/663, y el dominio helicasa con el “linker” que conecta al dominio RPD: 358/684 con dos proteínas de fusión distintas: GB1 y TRX.
- Se han purificado con una pureza mayor al 90% por Comassie de distintos constructos de Twinkle: Twinkle, Twinkle-(6hisC),  $\Delta$ 151Twinkle, TwinkleS359P, (GB1)- $\Delta$ 358Twinkle, (TRX)- $\Delta$ 358Twinkle, (GB1)- $\Delta$ 386Twinkle, (GB1)- $\Delta$ 386Twinkle $\Delta$ 663, (GB1)- $\Delta$ 386Twinkle $\Delta$ 647, (GB1)- $\Delta$ 386Twinkle $\Delta$ 636.
- La proteína recombinante Twinkle expresada en *E. coli* presenta actividad helicasa.

- El dominio N-terminal, tanto el dominio RPD como el ZBD influyen a la estabilidad y homogeneidad de Twinkle, la cual es más insoluble a medida que se recorta el dominio primasa.
- Los constructos de Twinkle con el dominio N-terminal tienen mayor homogeneidad que los constructos sin él, según se observó por cromatografía de exclusión en tamaño y DSF.
- Los experimentos de SAXS y EM han permitido conocer la estructura de Twinkle a baja resolución en disolución:
  - En disolución Twinkle tiene dos estados de oligomerización: hexámero y heptámero.
  - El tamaño del hexámero y del heptámero es parecido: 124-127 Å de diámetro respectivamente, y 40.5 y 45 Å en el canal interno.
  - La disposición relativa de los dominios helicasa de Twinkle en estado heptamérico es similar a gp4, pero no en el caso en estado de hexámero.
  - Twinkle se compone por tres dominios globulares: ZDB, RPD y helicasa, siendo este último dominio el responsable de la oligomerización.
  - Los tres dominios ZBD, RPD y helicasa están conectados entre sí por dos “*linker*” flexibles, fenómeno que otorga una gran flexibilidad a la totalidad de la proteína.
  - Los datos de entrecruzamiento químico existe una interacción tipo *trans* entre los subdominios ZBD y el RPD.
  - Twinkle tiene una cola C-terminal desordenada y totalmente desplegada.
- Twinkle presenta demasiada flexibilidad para ser cristalizada directamente, sería necesario realizar un complejo que estabilice las posiciones del dominio primasa.

# **BIBLIOGRAFÍA**



## 10 BIBLIOGRAFÍA.

- Adams, P. D., P. V. Afonine, et al. (2010). "The Phenix software for automated determination of macromolecular structures." Methods **55**(1): 94-106.
- Ahnert, P., K. M. Picha, et al. (2000). "A ring-opening mechanism for DNA binding in the central channel of the T7 helicase-primase protein." EMBO J **19**(13): 3418-27.
- Akabayov, B., S. R. Akabayov, et al. (2010). "Conformational dynamics of bacteriophage T7 DNA polymerase and its processivity factor, Escherichia coli thioredoxin." Proc Natl Acad Sci U S A **107**(34): 15033-8.
- Alam, T. I., T. Kanki, et al. (2003). "Human mitochondrial DNA is packaged with TFAM." Nucleic Acids Res **31**(6): 1640-5.
- Altschul, S. F., W. Gish, et al. (1990). "Basic local alignment search tool." J Mol Biol **215**(3): 403-10.
- Anderson, S., A. T. Bankier, et al. (1981). "Sequence and organization of the human mitochondrial genome." Nature **290**(5806): 457-65.
- Arenas, J., E. Briem, et al. (2003). "The V368i mutation in Twinkle does not segregate with AdPEO." Ann Neurol **53**(2): 278.
- Arias-Palomo, E., V. L. O'Shea, et al. (2013). "The bacterial DnaC helicase loader is a DnaB ring breaker." Cell **153**(2): 438-48.
- Attardi, G. and J. Montoya (1983). "Analysis of human mitochondrial RNA." Methods Enzymol **97**: 435-69.
- Bailey, S., W. K. Eliason, et al. (2007). "Structure of hexameric DnaB helicase and its complex with a domain of DnaG primase." Science **318**(5849): 459-63.
- Barbour, J. A. and N. Turner (2014). "Mitochondrial Stress Signaling Promotes Cellular Adaptations." Int J Cell Biol **2014**: 156020.
- Barcena, M., C. S. Martin, et al. (1998). "Polymorphic quaternary organization of the Bacillus subtilis bacteriophage SPP1 replicative helicase (G40 P)." J Mol Biol **283**(4): 809-19.
- Battistini, F., C. A. Hunter, et al. (2010). "Structural mechanics of DNA wrapping in the nucleosome." J Mol Biol **396**(2): 264-79.
- 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. and A. Agresti (2005). "HMG proteins: dynamic players in gene regulation and differentiation." Curr Opin Genet Dev **15**(5): 496-506.
- 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.
- Bogenhagen, D. F. (2012). "Mitochondrial DNA nucleoid structure." Biochim Biophys Acta **1819**(9-10): 914-20.
- 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.
- Branden, C. and A. Jones (1990). "BUSTER." Nature **343** (687-689).
- Brendza, K. M., W. Cheng, et al. (2005). "Autoinhibition of Escherichia coli Rep monomer helicase activity by its 2B subdomain." Proc Natl Acad Sci U S A **102**(29): 10076-81.
- Brown, T. A. and D. A. Clayton (2006). "Genesis and wanderings: origins and migrations in asymmetrically replicating mitochondrial DNA." Cell Cycle **5**(9): 917-21.
- Brown, T. A., A. N. Tkachuk, et al. (2011). "Superresolution fluorescence imaging of mitochondrial nucleoids reveals their spatial range, limits, and membrane interaction." Mol Cell Biol **31**(24): 4994-5010.
- Bustin, M. (2001). "Revised nomenclature for high mobility group (HMG) chromosomal proteins." Trends Biochem Sci **26**(3): 152-3.

## Bibliografía

- Campbell, C. T., J. E. Kolesar, et al. (2012). "Mitochondrial transcription factor A regulates mitochondrial transcription initiation, DNA packaging, and genome copy number." Biochim Biophys Acta **1819**(9-10): 921-9.
- Clayton, D. A. (2003). "Mitochondrial DNA replication: what we know." IUBMB Life **55**(4-5): 213-7.
- Clayton, R. A., O. White, et al. (1997). "The first genome from the third domain of life." Nature **387**(6632): 459-62.
- Crampton, D. J., M. Ohi, et al. (2006). "Oligomeric states of bacteriophage T7 gene 4 primase/helicase." J Mol Biol **360**(3): 667-77.
- 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.
- Deschauer, M., R. Kiefer, et al. (2003). "A novel Twinkle gene mutation in autosomal dominant progressive external ophthalmoplegia." Neuromuscul Disord **13**(7-8): 568-72.
- Dodson, E. (2003). "Is it jolly SAD?" Acta Crystallogr D Biol Crystallogr **59**(Pt 11): 1958-65.
- Double, S. and T. Ellenberger (1998). "The mechanism of action of T7 DNA polymerase." Curr Opin Struct Biol **8**(6): 704-12.
- Dragan, A. I., J. Klass, et al. (2003). "DNA binding of a non-sequence-specific HMG-D protein is entropy driven with a substantial non-electrostatic contribution." J Mol Biol **331**(4): 795-813.
- Dragan, A. I., J. R. Liggins, et al. (2003). "The energetics of specific binding of AT-hooks from HMGA1 to target DNA." J Mol Biol **327**(2): 393-411.
- Dragan, A. I. and P. L. Privalov (2008). "Use of fluorescence resonance energy transfer (FRET) in studying protein-induced DNA bending." Methods Enzymol **450**: 185-99.
- 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.
- 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.
- Enemark, E. J. and L. Joshua-Tor (2006). "Mechanism of DNA translocation in a replicative hexameric helicase." Nature **442**(7100): 270-5.
- Enemark, E. J. and L. Joshua-Tor (2008). "On helicases and other motor proteins." Curr Opin Struct Biol **18**(2): 243-57.
- Erzberger, J. P. and J. M. Berger (2006). "Evolutionary relationships and structural mechanisms of AAA+ proteins." Annu Rev Biophys Biomol Struct **35**: 93-114.
- Erzberger, J. P., M. L. Mott, et al. (2006). "Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling." Nat Struct Mol Biol **13**(8): 676-83.
- Eswar, N., B. Webb, et al. (2006). "Comparative protein structure modeling using Modeller." Curr Protoc Bioinformatics **Chapter 5**: Unit 5 6.
- Evans, P. (2006). "Scaling and assessment of data quality." Acta Crystallogr D Biol Crystallogr **62**(Pt 1): 72-82.
- Falkenberg, M., N. G. Larsson, et al. (2007). "DNA replication and transcription in mammalian mitochondria." Annu Rev Biochem **76**: 679-99.
- Farge, G., T. Holmlund, et al. (2008). "The N-terminal domain of TWINKLE contributes to single-stranded DNA binding and DNA helicase activities." Nucleic Acids Res **36**(2): 393-403.
- Fernandez-Silva, P., J. A. Enriquez, et al. (2003). "Replication and transcription of mammalian mitochondrial DNA." Exp Physiol **88**(1): 41-56.
- Fish, J., N. Raule, et al. (2004). "Discovery of a major D-loop replication origin reveals two modes of human mtDNA synthesis." Science **306**(5704): 2098-101.
- 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., 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.
- Fletcher, R. J., B. E. Bishop, et al. (2003). "The structure and function of MCM from archaeal *M. Thermoautotrophicum*." *Nat Struct Biol* **10**(3): 160-7.
- Flot, D., T. Mairs, et al. (2009). "The ID23-2 structural biology microfocuss beamline at the ESRF." *J Synchrotron Radiat* **17**(1): 107-18.
- Fouts, E. T., X. Yu, et al. (1999). "Biochemical and electron microscopic image analysis of the hexameric E1 helicase." *J Biol Chem* **274**(7): 4447-58.
- Fuste, J. M., S. Wanrooij, et al. (2010). "Mitochondrial RNA polymerase is needed for activation of the origin of light-strand DNA replication." *Mol Cell* **37**(1): 67-78.
- 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.
- Gaspari, M., N. G. Larsson, et al. (2004). "The transcription machinery in mammalian mitochondria." *Biochim Biophys Acta* **1659**(2-3): 148-52.
- Geertsema, H. J., A. W. Kulczyk, et al. (2014). "Single-molecule studies of polymerase dynamics and stoichiometry at the bacteriophage T7 replication machinery." *Proc Natl Acad Sci U S A*.
- Gerlitz, G. and M. Bustin (2009). "Nucleosome binding proteins potentiate ATM activation and DNA damage response by modifying chromatin." *Cell Cycle* **8**(11): 1641.
- 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.
- Goodwin, G. H., C. Sanders, et al. (1973). "A new group of chromatin-associated proteins with a high content of acidic and basic amino acids." *Eur J Biochem* **38**(1): 14-9.
- Gray, M. W. (1999). "Evolution of organellar genomes." *Curr Opin Genet Dev* **9**(6): 678-87.
- Ha, T., I. Rasnik, et al. (2002). "Initiation and re-initiation of DNA unwinding by the Escherichia coli Rep helicase." *Nature* **419**(6907): 638-41.
- Hamdan, S. M., B. Marintcheva, et al. (2005). "A unique loop in T7 DNA polymerase mediates the binding of helicase-primase, DNA binding protein, and processivity factor." *Proc Natl Acad Sci U S A* **102**(14): 5096-101.
- He, J., H. M. Cooper, et al. (2012). "Mitochondrial nucleoid interacting proteins support mitochondrial protein synthesis." *Nucleic Acids Res* **40**(13): 6109-21.
- 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.
- Hickman, A. B., Z. N. Perez, et al. (2005). "Molecular architecture of a eukaryotic DNA transposase." *Nat Struct Mol Biol* **12**(8): 715-21.
- Hingorani, M. M. and S. S. Patel (1996). "Cooperative interactions of nucleotide ligands are linked to oligomerization and DNA binding in bacteriophage T7 gene 4 helicases." *Biochemistry* **35**(7): 2218-28.
- Holm, L. and C. Sander (1996). "Alignment of three-dimensional protein structures: network server for database searching." *Methods Enzymol* **266**: 653-62.
- Holm, L. and C. Sander (1996). "The FSSP database: fold classification based on structure-structure alignment of proteins." *Nucleic Acids Res* **24**(1): 206-9.
- Holt, I. (2000). "Age is not a barrier to individuality." *Br J Community Nurs* **5**(2): 56.
- Holt, I. J., J. He, et al. (2007). "Mammalian mitochondrial nucleoids: organizing an independently minded genome." *Mitochondrion* **7**(5): 311-21.
- Holt, I. J. and H. T. Jacobs (2003). "Response: The mitochondrial DNA replication bubble has not burst." *Trends Biochem Sci* **28**(7): 355-6.

## Bibliografía

- Holt, I. J. and A. Reyes (2012). "Human mitochondrial DNA replication." Cold Spring Harb Perspect Biol **4**(12).
- Hood, D. A. (2001). "Invited Review: contractile activity-induced mitochondrial biogenesis in skeletal muscle." J Appl Physiol (1985) **90**(3): 1137-57.
- Iborra, F. J., H. Kimura, et al. (2004). "The functional organization of mitochondrial genomes in human cells." BMC Biol **2**: 9.
- Ilyina, T. V. and E. V. Koonin (1992). "Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaeobacteria." Nucleic Acids Res **20**(13): 3279-85.
- Jakobs, S. (2006). "High resolution imaging of live mitochondria." Biochim Biophys Acta **1763**(5-6): 561-75.
- Jemt, E., G. Farge, et al. (2011). "The mitochondrial DNA helicase TWINKLE can assemble on a closed circular template and support initiation of DNA synthesis." Nucleic Acids Res **39**(21): 9238-49.
- Jen-Jacobson, L., L. E. Engler, et al. (2000). "Structural and thermodynamic strategies for site-specific DNA binding proteins." Structure **8**(10): 1015-23.
- Jones, D. T. (1999). "Protein secondary structure prediction based on position-specific scoring matrices." J Mol Biol **292**(2): 195-202.
- Kabsch, W. (2010). "Integration, scaling, space-group assignment and post-refinement." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 133-44.
- Kabsch, W. (2010). "Xds." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 125-32.
- Kastner, B., N. Fischer, et al. (2008). "GraFix: sample preparation for single-particle electron cryomicroscopy." Nat Methods **5**(1): 53-5.
- Kato, M., T. Ito, et al. (2003). "Modular architecture of the bacteriophage T7 primase couples RNA primer synthesis to DNA synthesis." Mol Cell **11**(5): 1349-60.
- 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.
- Kawashima, S., S. Ohta, et al. (1994). "Widespread tissue distribution of multiple mitochondrial DNA deletions in familial mitochondrial myopathy." Muscle Nerve **17**(7): 741-6.
- Koh, J., I. Shkel, et al. (2011). "Nonspecific DNA binding and bending by HUalpha: interfaces of the three binding modes characterized by salt-dependent thermodynamics." J Mol Biol **410**(2): 241-67.
- Koonin, E. V. (1993). "A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication." Nucleic Acids Res **21**(11): 2541-7.
- Korhonen, J. A., M. Gaspari, et al. (2003). "TWINKLE Has 5' -> 3' DNA helicase activity and is specifically stimulated by mitochondrial single-stranded DNA-binding protein." J Biol Chem **278**(49): 48627-32.
- Korhonen, J. A., V. Pande, et al. (2008). "Structure-function defects of the TWINKLE linker region in progressive external ophthalmoplegia." J Mol Biol **377**(3): 691-705.
- Korhonen, J. A., X. H. Pham, et al. (2004). "Reconstitution of a minimal mtDNA replisome in vitro." EMBO J **23**(12): 2423-9.
- Korolev, S., J. Hsieh, et al. (1997). "Major domain swiveling revealed by the crystal structures of complexes of E. coli Rep helicase bound to single-stranded DNA and ADP." Cell **90**(4): 635-47.
- 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.
- Kukat, C., C. A. Wurm, et al. (2011). "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.

- Lavery, R., R. Moaker, et al. (2009). "Conformation analysis fo nucleic acids revisited:Curves+." Nucleic Acids Res **37**: 5917-5929.
- Lavery, R., K. Zakrzewska, et al. (2009). "A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA." Nucleic Acids Res **38**(1): 299-313.
- Lecrenier, N., P. Van Der Bruggen, et al. (1997). "Mitochondrial DNA polymerases from yeast to man: a new family of polymerases." Gene **185**(1): 147-52.
- Lee, J., P. D. Chastain, 2nd, et al. (1998). "Coordinated leading and lagging strand DNA synthesis on a minicircular template." Mol Cell **1**(7): 1001-10.
- Lee, S. J. and C. C. Richardson (2002). "Interaction of adjacent primase domains within the hexameric gene 4 helicase-primase of bacteriophage T7." Proc Natl Acad Sci U S A **99**(20): 12703-8.
- Lee, S. J. and C. C. Richardson (2011). "Choreography of bacteriophage T7 DNA replication." Curr Opin Chem Biol **15**(5): 580-6.
- Lee, S. J., B. Zhu, et al. (2010). "Mechanism of sequence-specific template binding by the DNA primase of bacteriophage T7." Nucleic Acids Res **38**(13): 4372-83.
- Leipe, D. D., Y. I. Wolf, et al. (2002). "Classification and evolution of P-loop GTPases and related ATPases." J Mol Biol **317**(1): 41-72.
- Lewis, S., W. Hutchison, et al. (2002). "Clinical and molecular features of adPEO due to mutations in the Twinkle gene." J Neurol Sci **201**(1-2): 39-44.
- Linder, P. and P. Lasko (2006). "Bent out of shape: RNA unwinding by the DEAD-box helicase Vasa." Cell **125**(2): 219-21.
- Liu, Z., Y. Ding, et al. (2008). "A novel Twinkle (PEO1) gene mutation in a Chinese family with adPEO." Mol Vis **14**: 1995-2001.
- Lohman, T. M. and K. P. Bjornson (1996). "Mechanisms of helicase-catalyzed DNA unwinding." Annu Rev Biochem **65**: 169-214.
- Lonnqvist, T., A. Paetau, et al. (2009). "Recessive twinkle mutations cause severe epileptic encephalopathy." Brain **132**(Pt 6): 1553-62.
- Lowell, B. B. and B. M. Spiegelman (2000). "Towards a molecular understanding of adaptive thermogenesis." Nature **404**(6778): 652-60.
- Luthy, R., J. U. Bowie, et al. (1992). "Assessment of protein models with three-dimensional profiles." Nature **356**(6364): 83-5.
- Machwe, A., L. Xiao, et al. (2005). "RecQ family members combine strand pairing and unwinding activities to catalyze strand exchange." J Biol Chem **280**(24): 23397-407.
- Malarkey, C. S., M. Bestwick, et al. (2011). "Transcriptional activation by mitochondrial transcription factor A involves preferential distortion of promoter DNA." Nucleic Acids Res **40**(2): 614-24.
- Malarkey, C. S. and M. E. Churchill (2012). "The high mobility group box: the ultimate utility player of a cell." Trends Biochem Sci **37**(12): 553-62.
- Masse, J. E., B. Wong, et al. (2002). "The *S. cerevisiae* architectural HMGB protein NHP6A complexed with DNA: DNA and protein conformational changes upon binding." J Mol Biol **323**(2): 263-84.
- Matsushima, S., T. Ide, et al. (2006). "Overexpression of mitochondrial peroxiredoxin-3 prevents left ventricular remodeling and failure after myocardial infarction in mice." Circulation **113**(14): 1779-86.
- Matsushima, Y., C. L. Farr, et al. (2008). "Physiological and biochemical defects in carboxyl-terminal mutants of mitochondrial DNA helicase." J Biol Chem **283**(35): 23964-71.
- Matsushima, Y. and L. S. Kaguni (2007). "Differential phenotypes of active site and human autosomal dominant progressive external ophthalmoplegia mutations in *Drosophila* mitochondrial DNA helicase expressed in Schneider cells." J Biol Chem **282**(13): 9436-44.

## Bibliografía

- Matsushima, Y. and L. S. Kaguni (2009). "Functional importance of the conserved N-terminal domain of the mitochondrial replicative DNA helicase." *Biochim Biophys Acta* **1787**(5): 290-5.
- McConnell, K. J. and D. L. Beveridge (2001). "Molecular dynamics simulations of B'-DNA: sequence effects on A-tract-induced bending and flexibility." *J Mol Biol* **314**(1): 23-40.
- McCoy, A. J. (2007). "Solving structures of protein complexes by molecular replacement with Phaser." *Acta Crystallogr D Biol Crystallogr* **63**(Pt 1): 32-41.
- McCoy, A. J., R. W. Grosse-Kunstleve, et al. (2007). "Phaser crystallographic software." *J Appl Crystallogr* **40**(Pt 4): 658-674.
- 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.
- Milenkovic, D., S. Matic, et al. (2013). "TWINKLE is an essential mitochondrial helicase required for synthesis of nascent D-loop strands and complete mtDNA replication." *Hum Mol Genet* **22**(10): 1983-93.
- 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.
- Montoya, J., G. L. Gaines, et al. (1983). "The pattern of transcription of the human mitochondrial rRNA genes reveals two overlapping transcription units." *Cell* **34**(1): 151-9.
- Montoya, J., D. Ojala, et al. (1981). "Distinctive features of the 5'-terminal sequences of the human mitochondrial mRNAs." *Nature* **290**(5806): 465-70.
- Musicco, C., M. Roberti, et al. (2000). "Regulation of the expression of the sea urchin mitochondrial D-loop binding protein during early development." *Biochem Biophys Res Commun* **277**(2): 299-304.
- Nelson, H. C., J. T. Finch, et al. (1987). "The structure of an oligo(dA).oligo(dT) tract and its biological implications." *Nature* **330**(6145): 221-6.
- Ngo, H. B., J. T. Kaiser, et al. (2011). "The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA." *Nat Struct Mol Biol* **18**(11): 1290-6.
- Ngo, H. B., G. A. Lovely, et al. (2014). "Distinct structural features of TFAM drive mitochondrial DNA packaging versus transcriptional activation." *Nat Commun* **5**: 3077.
- Nikali, K., A. Suomalainen, et al. (2005). "Infantile onset spinocerebellar ataxia is caused by recessive mutations in mitochondrial proteins Twinkle and Twinky." *Hum Mol Genet* **14**(20): 2981-90.
- 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.
- Ojala, D., C. Merkel, et al. (1980). "The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA." *Cell* **22**(2 Pt 2): 393-403.
- Orengo, C. A., D. T. Jones, et al. (1994). "Protein superfamilies and domain superfolds." *Nature* **372**(6507): 631-4.
- Pandey, M., S. Syed, et al. (2009). "Coordinating DNA replication by means of priming loop and differential synthesis rate." *Nature* **462**(7275): 940-3.
- 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.
- Park, K., Z. Debyser, et al. (1998). "Formation of a DNA loop at the replication fork generated by bacteriophage T7 replication proteins." *J Biol Chem* **273**(9): 5260-70.
- Patel, S. S. and M. M. Hingorani (1995). "Nucleotide binding studies of bacteriophage T7 DNA helicase-primase protein." *Biophys J* **68**(4 Suppl): 186S-189S; discussion 189S-190S.
- Perez, A., F. J. Luque, et al. (2007). "Dynamics of B-DNA on the microsecond time scale." *J Am Chem Soc* **129**(47): 14739-45.

- 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.
- Pohjoismaki, J. L., S. Wanrooij, et al. (2006). "Alterations to the expression level of mitochondrial transcription factor A, TFAM, modify the mode of mitochondrial DNA replication in cultured human cells." *Nucleic Acids Res* **34**(20): 5815-28.
- Privalov, P. L., A. I. Dragan, et al. (2011). "Interpreting protein/DNA interactions: distinguishing specific from non-specific and electrostatic from non-electrostatic components." *Nucleic Acids Res* **39**(7): 2483-91.
- Privalov, P. L., A. I. Dragan, et al. (2007). "What drives proteins into the major or minor grooves of DNA?" *J Mol Biol* **365**(1): 1-9.
- Privalov, P. L., I. Jelesarov, et al. (1999). "The energetics of HMG box interactions with DNA: thermodynamics of the DNA binding of the HMG box from mouse sox-5." *J Mol Biol* **294**(4): 997-1013.
- Reyes, A., L. Kazak, et al. (2013). "Mitochondrial DNA replication proceeds via a 'bootlace' mechanism involving the incorporation of processed transcripts." *Nucleic Acids Res* **41**(11): 5837-50.
- Ringel, R., M. Sologub, et al. (2011). "Structure of human mitochondrial RNA polymerase." *Nature* **478**(7368): 269-73.
- Rubio-Cosials, A., J. F. Sidow, et al. (2011). "Human mitochondrial transcription factor A induces a U-turn structure in the light strand promoter." *Nat Struct Mol Biol* **18**(11): 1281-9.
- Rubio-Cosials, A. and M. Sola (2013). "U-turn DNA bending by human mitochondrial transcription factor A." *Curr Opin Struct Biol* **23**(1): 116-24.
- Russell, J. S. a. D. W. (2001). "Commonly Used Techniques in Molecular Cloning," "Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA Volume 3(Appendix 8).
- Sanchez-Martinez, A., M. Calleja, et al. (2012). "Modeling pathogenic mutations of human twinkle in Drosophila suggests an apoptosis role in response to mitochondrial defects." *PLoS One* **7**(8): e43954.
- Sarzi, E., S. Goffart, et al. (2007). "Twinkle helicase (PEO1) gene mutation causes mitochondrial DNA depletion." *Ann Neurol* **62**(6): 579-87.
- Schenkel, L. C. and M. Bakovic (2014). "Formation and Regulation of Mitochondrial Membranes." *Int J Cell Biol* **2014**: 709828.
- Sen, D., D. Nandakumar, et al. (2012). "Human mitochondrial DNA helicase TWINKLE is both an unwinding and annealing helicase." *J Biol Chem* **287**(18): 14545-56.
- Shadel, G. S. and D. A. Clayton (1997). "Mitochondrial DNA maintenance in vertebrates." *Annu Rev Biochem* **66**: 409-35.
- Shi, Y., A. Dierckx, et al. (2012). "Mammalian transcription factor A is a core component of the mitochondrial transcription machinery." *Proc Natl Acad Sci U S A* **109**(41): 16510-5.
- Shutt, T. E. and M. W. Gray (2006). "Twinkle, the mitochondrial replicative DNA helicase, is widespread in the eukaryotic radiation and may also be the mitochondrial DNA primase in most eukaryotes." *J Mol Evol* **62**(5): 588-99.
- Singleton, M. R., M. S. Dillingham, et al. (2007). "Structure and mechanism of helicases and nucleic acid translocases." *Annu Rev Biochem* **76**: 23-50.
- Singleton, M. R., M. R. Sawaya, et al. (2000). "Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides." *Cell* **101**(6): 589-600.
- Sippl, M. J. (1993). "Recognition of errors in three-dimensional structures of proteins." *Proteins* **17**(4): 355-62.
- Skordalakes, E. and J. M. Berger (2003). "Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading." *Cell* **114**(1): 135-46.
- 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.

## Bibliografía

- Spelbrink, J. N., F. Y. Li, et al. (2001). "Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria." *Nat Genet* **28**(3): 223-31.
- Stros, M., D. Launholt, et al. (2007). "The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins." *Cell Mol Life Sci* **64**(19-20): 2590-606.
- Svergun D.I., B. C. a. K. M. H. J. (1995). "CRY SOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates " *J. Appl. Cryst.* **28**, : 768-773.
- Taanman, J. W. (1999). "The mitochondrial genome: structure, transcription, translation and replication." *Biochim Biophys Acta* **1410**(2): 103-23.
- Tabor, S., H. E. Huber, et al. (1987). "Escherichia coli thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7." *J Biol Chem* **262**(33): 16212-23.
- Thompson, J. D., T. J. Gibson, et al. (2002). "Multiple sequence alignment using ClustalW and ClustalX." *Curr Protoc Bioinformatics* **Chapter 2**: Unit 2 3.
- Thorington, G. and L. Margulis (1981). "Hydra viridis: transfer of metabolites between Hydra and symbiotic algae." *Biol Bull* **160**(1): 175-88.
- 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.
- Toth, E. A., Y. Li, et al. (2003). "The crystal structure of the bifunctional primase-helicase of bacteriophage T7." *Mol Cell* **12**(5): 1113-23.
- Tyynismaa, H., H. Sembongi, et al. (2004). "Twinkle helicase is essential for mtDNA maintenance and regulates mtDNA copy number." *Hum Mol Genet* **13**(24): 3219-27.
- 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.
- Vandenberghe, W., K. Van Laere, et al. (2009). "Neurodegenerative Parkinsonism and progressive external ophthalmoplegia with a Twinkle mutation." *Mov Disord* **24**(2): 308-9.
- Wanrooij, P. H., J. P. Uhler, et al. (2012). "A hybrid G-quadruplex structure formed between RNA and DNA explains the extraordinary stability of the mitochondrial R-loop." *Nucleic Acids Res* **40**(20): 10334-44.
- Wanrooij, P. H., J. P. Uhler, et al. (2010). "G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation." *Proc Natl Acad Sci U S A* **107**(37): 16072-7.
- Wanrooij, S., J. M. Fuste, et al. (2008). "Human mitochondrial RNA polymerase primes lagging-strand DNA synthesis in vitro." *Proc Natl Acad Sci U S A* **105**(32): 11122-7.
- Wanrooij, S., S. Goffart, et al. (2007). "Expression of catalytic mutants of the mtDNA helicase Twinkle and polymerase POLG causes distinct replication stalling phenotypes." *Nucleic Acids Res* **35**(10): 3238-51.
- Washington, M. T., A. H. Rosenberg, et al. (1996). "Biochemical analysis of mutant T7 primase/helicase proteins defective in DNA binding, nucleotide hydrolysis, and the coupling of hydrolysis with DNA unwinding." *J Biol Chem* **271**(43): 26825-34.
- Watson, J. D. and F. H. Crick (1953). "The structure of DNA." *Cold Spring Harb Symp Quant Biol* **18**: 123-31.
- Wiederstein, M. and M. J. Sippl (2007). "ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins." *Nucleic Acids Res* **35**(Web Server issue): W407-10.
- Wigley, D. B. (2000). "DNA helicases: one small step for PcrA, one giant leap for RecBC?" *Curr Biol* **10**(12): R444-6.

- Williams, D. C., Jr., M. Cai, et al. (2004). "Molecular basis for synergistic transcriptional activation by Oct1 and Sox2 revealed from the solution structure of the 42-kDa Oct1.Sox2.Hoxb1-DNA ternary transcription factor complex." *J Biol Chem* **279**(2): 1449-57.
- Winn, M. D., C. C. Ballard, et al. (2011). "Overview of the CCP4 suite and current developments." *Acta Crystallogr D Biol Crystallogr* **67**(Pt 4): 235-42.
- Yakubovskaya, E., K. E. Guja, et al. (2014). "Organization of the human mitochondrial transcription initiation complex." *Nucleic Acids Res.*
- Yasukawa, T., M. Y. Yang, et al. (2005). "A bidirectional origin of replication maps to the major noncoding region of human mitochondrial DNA." *Mol Cell* **18**(6): 651-62.
- Ye, J., A. R. Osborne, et al. (2004). "RecA-like motor ATPases--lessons from structures." *Biochim Biophys Acta* **1659**(1): 1-18.
- 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.
- Yu, X., M. M. Hingorani, et al. (1996). "DNA is bound within the central hole to one or two of the six subunits of the T7 DNA helicase." *Nat Struct Biol* **3**(9): 740-3.
- Zeviani, M., S. Servidei, et al. (1989). "An autosomal dominant disorder with multiple deletions of mitochondrial DNA starting at the D-loop region." *Nature* **339**(6222): 309-11.
- Zhang, H., S. J. Lee, et al. (2011). "Helicase-DNA polymerase interaction is critical to initiate leading-strand DNA synthesis." *Proc Natl Acad Sci U S A* **108**(23): 9372-7.
- Zhang, Y. (2008). "Progress and challenges in protein structure prediction." *Curr Opin Struct Biol* **18**(3): 342-8.
- Ziebarth, T. D., C. L. Farr, et al. (2007). "Modular architecture of the hexameric human mitochondrial DNA helicase." *J Mol Biol* **367**(5): 1382-91.
- Ziebarth, T. D., R. Gonzalez-Soltero, et al. "Dynamic effects of cofactors and DNA on the oligomeric state of human mitochondrial DNA helicase." *J Biol Chem* **285**(19): 14639-47.
- Ziebarth, T. D., R. Gonzalez-Soltero, et al. (2010). "Dynamic effects of cofactors and DNA on the oligomeric state of human mitochondrial DNA helicase." *J Biol Chem* **285**(19): 14639-47.



## 11 ANEXO

Acrónimos y anglicismos:

- ADN: Acrónimo de ácido desoxiribonucleico.
- ARN: Acrónimo de ácido ribonucleico.
- BME: Acrónimo de  $\beta$ mercaptetanol.
- CSB: Acrónimo de *Conserved Sequence Blocks*, Bloques de secuencia conservada.
- Crio-EM: Acrónimo del inglés de *cryo-electrón microscopy*.
- *Crosslinkin*: Entrecruzamiento.
- *Cryo-electrón microscopy*: Técnica denominada crio-microscopía electrónica.
- dsADN: acrónimo del inglés *double strand DNA*, Ácido desoxiribonucleico de doble cadena.
- DLS: acrónimo del inglés *Dynamic light scattering*, (dispersion dinámica de la luz).
- DSF: acrónimo del inglés *Differential scan fluoresnce*.
- DTT: dithiothreitoll
- *Electrón microscopy*: Técnica denominada microscopía electrónica.
- EM: Acrónimo del inglés de *electrón microscopy*.
- EMSA: Acrónimo del inglés de *electrophoretic mobility shift assay*, ensayo de cambio en la migración electroforética.
- *Fingerprint*: Huella dactilar.
- *Footprinting*: Huella plantar.
- *Gaps*: huecos
- ITPG: acrónimo de isopropil  $\beta$ -D-1tiogalactopiranosido.
- *Linker*: Conector.
- *Loop*: Lazo.
- *Microbatch*: micro grupo.
- *Mismatch*: discordancia.
- *Polylinker*: Conector múltiple.
- *Primer*: cebador.
- *propoller twist*: giro de una base de ADN respecto aquélla que estable los enlaces de hidrógeno tipo Watson y Crick.
- *Protein data bank*: Base de datos de proteínas de denominada Banco de datos de proteínas.
- PDB: Acrónimo del inglés de *protein data bank*.
- *Pellet*: sedimento.
- *RNA polymerase domain*: Dominio de la ARN polimerasa.
- *Roll*: Inclinación.
- RPD: Acrónimo del inglés de *RNA polymerase domain*,
- SAXS: Acrónimo del inglés de de *Small angle X-ray Scattering*.
- *Small angle X-ray Scattering*: técnica denominada “dispersión de rayosX a bajo ángulo”.
- SDS-PAGE: Acrónimo de del inglés *sodium dodecyl sulfate polyacrylamide gel electrophoresis*, electroforesis en gel de poliacrilamida con dodecilsulfato sódico.

- *Step*: Paso.
- ssADN: acrónimo del inglés *single strand DNA*, Ácido desoxiribonucleico de cadena sencilla.
- *Turn*: giro.
- *Twist*: Giro helicoidal.
- *Vortex* (agitador): agitador o mezclador de vórtice.
- ZBD: Acrónimo del inglés de *Zinc Bind Domain*, Dominio de unión al Zinc.