

Universidad Complutense de Madrid

Facultad de Ciencias Biológicas

Departamento de Bioquímica y Biología Molecular I



Biología estructural de la reparación de  
roturas de doble cadena en el ADN

TESIS DOCTORAL

Angel Rivera Calzada

Madrid, 2008





Universidad Complutense de Madrid

Facultad de Ciencias Biológicas

Departamento de Bioquímica y Biología Molecular I



Memoria presentada para optar  
al grado de Doctor por

Angel Rivera Calzada

Universidad Complutense de Madrid

Julio de 2008

Esta Tesis ha sido realizada en el **Centro de Investigaciones Biológicas**  
(CIB-CSIC) bajo la dirección del **Dr. Óscar Antonio Llorca Blanco**.





*Para Lucía, Eduardo, Alberto y Laura*



## Agradecimientos

Comencé esta tesis el 3 de junio del 2002. Han sido seis años en los que he tenido la suerte de hacer lo que me gusta rodeado de gente excepcional. En todo este tiempo han sido muchas las personas que han invertido tiempo y esfuerzo en este proyecto y por ello quiero hacer patente mi agradecimiento.

Oskar nada de esto habría sido posible sin ti, sin tu esfuerzo, dedicación, criterio, motivación, ilusión, paciencia infinita y un largo etcétera. En estos años te he visto pasar de “post-doc” a Investigador Científico y no he dejado nunca de aprender de ti, tanto en lo profesional como en lo personal. Oskar te llevas mi más sincero y profundo agradecimiento.

Este trabajo tampoco habría sido posible sin la colaboración de Laurence H. Pearl. Gracias por proporcionarnos los distintos especímenes analizados, especialmente a Laura Spagnolo y a Neil Rzechorzek que han sido los encargados de lidiar con su purificación. Además también me gustaría agradecer a Pablo Chacón la ayuda prestada durante los experimentos de análisis vibracional de DNA-PKcs y ATM. Gracias a José Jesús Fernández por solucionarnos todas las dudas durante la corrección de CTF. Me gustaría expresar mi agradecimiento al servicio de microscopía electrónica del CNB por habernos facilitado en todo momento el uso y disfrute del FEG (gracias Rocío y Cris).

No puedo olvidarme de los organismos e instituciones que me han financiado durante este período de tiempo. Mi más interesado agradecimiento para el Ministerio de Ciencia y Tecnología y para el programa I3P del CSIC.

Desde un punto de vista más personal tengo también muchas cosas que agradecer.

Jaska de ti no me iba a olvidar, ¿dónde estaría yo ahora si no te hubieras acordado de mí?. Gracias Jaskita por estar siempre ahí.

¿Y vosotros compañeros?, ¡no sabéis lo importante que sois para mí!. Hemos pasado muchas horas juntos y en todo este tiempo siempre he estado encantado de teneros cerca. Gracias por saber escuchar, por saber discutir, por saber aportar. Gracias a Magi, a Eva, a Bego, a Rafa y a Gabriel. Gracias sobre todo a ti Ernesto, por ser un compañero excepcional y por muchas más cosas.

Gracias también al servicio de Microscopía Electrónica del CIB, a Lola Guirao y Carmen Terrón, por conseguir que las horas de microscopio se hagan mucho más llevaderas.

En este tiempo he coincidido con compañeros de muy diversos grupos. Muchos de ellos ya no están en el CIB y otros se acaban de incorporar, y a todos ellos me gustaría expresar mi agradecimiento por hacer que mi paso por este centro haya merecido la pena.

Luque e Isra, gracias por seguirme hasta el infinito y más allá (entiéndase por infinito “la valla”). En definitiva gracias por darlo todo cuando hay que darlo.

Gracias también a los compañeros de los laboratorios de Pablo Chacón, Germán Rivas, Jesús Jiménez Barbero, Francisco Javier Cañada, Antonio Romero, Guillermo Jiménez, Fernando Díaz, Luis Rivas, Rafa Giraldo, Miguel Ángel Vidal, Vicente Larraga, José Manuel Andreu, Rosa Lozano y Teresa Suárez. Concretamente gracias a Nuria, Ángeles, Elena, Carmen, Pilar, José Ángel, Chus, Bea, Ariadna, Mercedes, Patrick, Carmencita, Sonia, Terepower, Mon, Nacho, Víctor, Inés, Mónica, Carlos, Arturito, Iñaki, Mariqui, Félix, Mateo, Blas, Gonn, Juan, Jose, Marian, Rubén, Isabel, J. Gayarre, Fran, el Niño, Joao, Marta, ...

Gracias también al B-14 porque fue allí donde empezó todo. Jaimito no te me enfades, ya sabes que te adoro. Gracias a los compañeros de los laboratorios de J. L. Carrascosa, José María Valpuesta, Cristina, J.R y José María Carazo. Siempre da gusto visitaros. Gracias a Sjors y a Javi Chichón.

Gracias a los Biolocos por hacer inolvidables los años de facultad. Gracias Raúl, Gabi, Marta, Tomás, Lucía, Isabel, Cristina, Yolanda y Ángela.

Gracias a los Bioquímicos por aceptar mis rarezas y por no dejarme nunca en paz. Gracias Fer, Rafa, Chus, Montse, Gloria, Ana, Juanfran, Laura, Raquel, Mamen, Eva, Pilar y Víctor.

Y porque hay vida más allá de la ciencia gracias también a David y Fernando, a Rodrigo (la arqueología no era lo mío golfo), a Patricia (¡al final lo conseguí!, ¡ahora te toca a ti!), a Juanma por hacerme correr más que a su abuela, a Wyatt (my best Chinese friend), a Iñigo (¡gracias jartolari!), a Idoia, Bevan, Jorge, Paco, Moncho, Andrés, José María, ... Gracias al balonmano (Gustavo) y al Valdeluz (Nélida, María José y Padre Soto). Gracias a la Antilla por existir.

Sobre todo y más que a nadie gracias a vosotros familia, a mis padres Eduardo y Lucía y a mi hermano Alberto. Sois lo más valioso que tengo. Todo os lo debo a vosotros, a vuestro ejemplo. Vosotros sabéis quién soy y siempre estaréis a mi lado. Tenéis toda la culpa de esto, ¿cómo se os ocurrió regalarme el Micronova?.

Por último me gustaría darte las **Gracias** a ti Laura. Gracias por creer en mi, por demostrármelo todo y gracias, en definitiva, por lo más importante, por hacerme FELIZ.

Bueno señor@s, hasta aquí llegamos. Relájense y disfruten.

Va por ustedes.

ARC (Angelito/illo)

Madrid 12+1 Junio 2008



## *Summary*

Double strand DNA breaks (DSBs) are one of the most challenging threats for cell viability. The main mechanism to repair DSBs in higher eukaryotes is the process known as non-homologous end-joining (NHEJ) which consists in the ligation of the generated ends after the break using no template. The components that participate in this repair pathway play a fundamental role in maintaining genomic stability. Furthermore they participate in programmed DNA breaks mediating biological processes such as V(D)J recombination. During NHEJ the Ku protein binds to the generated ends and recruits the catalytic subunit of the DNA Dependent Protein Kinase (DNA-PKcs) to the DSB. The DNA-PKcs:Ku complex formed on a DNA end acts then as a scaffold for the additional enzymatic activities required during the repair process. Besides, the interaction of two DNA-PKcs:Ku:ADN complexes can create a synaptic complex that holds both DNA ends together avoiding their diffusion. Some DSBs could require the processing of the generated ends prior to the final ligation of both ends. The nuclease Artemis has been implicated in this processing in association with DNA-PKcs and, additionally, the DNA-PKcs:Artemis complex has been proposed to act as the enzyme that opens the hairpin present in the coding ends during the V(D)J recombination process.

The focus of my Ph.D. project consists in the structural determination of molecular complexes implicated in the repair of DSB by NHEJ. To carry out this analysis I have used structural techniques based on the combination of electron microscopy and single particle analysis. More specifically I have solved the structure of full-length Ku alone and after binding to DNA. These structures have provided a structural model that explains the conformational changes that take place in Ku upon DNA end recognition and their functional implications. Also, I have determined the structure of DNA-PKcs that, combined with further analysis, has provided a structural model that explains the overall topology of this kinase. The conformational changes that take place in DNA-PKcs upon recognition of DNA have also been characterised. Upon DNA binding significant movements in key domains of DNA-PKcs seem to act as transducer of DNA recognition to the catalytic domain. Additionally I have solved the structure of the DNA-PKcs:Ku:DNA complex providing important clues in the structural basis of the NHEJ reaction. When analysing the DNA-PKcs:Ku:DNA complex the presence of a dimeric complex constituted by two DNA-PKcs:Ku:DNA complexes facing each other was observed. The structure of this dimeric complex is fully compatible with the

requirements described for the synaptic complex that participates in the NHEJ process. Finally I have also determined the biochemical requirements to form a complex between DNA-PKcs and Artemis proteins in our experimental conditions. This analysis has provided the initial characterization required for future structural studies of the DNA-PKcs:Artemis complex.

Overall, the results presented in this Thesis have contributed to our understanding of the structural basis of the NHEJ DNA repair reaction. The work provides the structural foundations of some of the biological functions and regulatory mechanisms of DNA-PKcs and Ku.

# Índice General

1. Introducción	1
<b>1.1 Mantenimiento de la estabilidad genómica</b>	<b>4</b>
<b>1.2 Sistemas de reparación de daño en el ADN</b>	<b>6</b>
<b>1.3 Sistemas de reparación de DSB basados en recombinación homóloga</b>	<b>8</b>
<b>1.4 Reparación de DSB por unión de extremos no homólogos</b>	<b>9</b>
1.4.1 Ku	12
1.4.2 DNA-PKcs	16
1.4.3 Artemis	21
1.4.4 XRCC4:ADN Ligasa IV	23
1.4.5 XLF o Cernunnos	25
1.4.6 Componentes adicionales del proceso de reparación por NHEJ	26
1.4.7 Efectos de las alteraciones en los componentes del proceso de NHEJ	27
1.4.8 Conclusión	30
<b>1.5 Microscopía electrónica y procesamiento digital de partículas individuales</b>	<b>31</b>
1.5.1 Introducción al microscopio electrónico y generación de contraste	31
1.5.2 Formación de la imagen en el microscopio electrónico	32
1.5.3 Corrección de CTF	32
1.5.4 Técnicas de preparación del espécimen	34
<u>Tinción negativa</u>	<u>34</u>
<u>Congelación y criomicroscopía electrónica</u>	<u>35</u>
1.5.5 Toma de micrografías	36
1.5.6 Digitalización	37
1.5.7 Ángulos de Euler	38
1.5.8 Estrategias generales de procesamiento de partículas individuales	41
<u>Inclinación cónica al azar o RCT (del inglés “Random Conical Tilt”)</u>	<u>41</u>
<u>Método de reconstitución angular</u>	<u>42</u>

<u>Métodos de alineamiento multirreferencia respecto a imágenes de proyección de una estructura conocida</u>	<u>42</u>
<u>Elipsoides artificiales</u>	<u>43</u>
<u>Ruido</u>	<u>44</u>
1.5.9 Estrategia de procesamiento del paquete informático EMAN	44
1.5.10 Cálculo de la resolución	46
<b>2. Objetivos</b>	<b>47</b>
<b>3. Materiales y Métodos</b>	<b>51</b>
<b>3.1 Obtención de moléculas de ADN de doble cadena</b>	<b>53</b>
<b>3.2 Purificación de proteínas humanas</b>	<b>55</b>
3.2.1 Purificación del heterodímero Ku	55
3.2.2 Purificación de la proteína DNA-PKcs	56
3.2.3 Purificación del complejo DNA-PKcs:Ku:ADN	56
3.2.4 Purificación de la proteína Artemis	57
3.2.5 Purificación de la proteína XRCC4	58
<b>3.3 Electroforesis en condiciones desnaturizantes</b>	<b>58</b>
<b>3.4 Electroforesis en condiciones nativas</b>	<b>59</b>
<b>3.5 Tinción de geles nativos con bromuro de etidio</b>	<b>60</b>
<b>3.6 Tinción reversa de geles nativos</b>	<b>60</b>
<b>3.7 Ensayos de “Western-blot”</b>	<b>61</b>
<b>3.8 Tinción con plata</b>	<b>62</b>
<b>3.9 Ensayos bioquímicos</b>	<b>63</b>
3.9.1 Interacción del heterodímero Ku con ADN	63
3.9.2 Ensayo de la actividad quinasa de la proteína DNA-PKcs	63
3.9.3 Formación de complejos anticuerpo:DNA-PKcs	64
3.9.4 Interacción de la proteína DNA-PKcs con ADN	64
3.9.5 Marcaje con oro de complejos proteína:ADN	65
3.9.6 Ensayo de defosforilación de Artemis	65
3.9.7 Análisis de la muestra de Artemis mediante cromatografía de exclusión molecular	66
3.9.8 Interacción de Artemis con ADN	66
3.9.9 Interacción entre DNA-PKcs y Artemis	67
3.9.10 Interacción entre DNA-PKcs y Artemis mediante tinción reversa de geles nativos	67

<b>3.10 Preparación de muestras mediante tinción negativa para su observación en el microscopio electrónico</b>	<b>68</b>
<u>Heterodímero Ku y complejo Ku:ADN</u>	<u>68</u>
<u>Complejo Ku:ADN analizado mediante RCT</u>	<u>69</u>
<u>DNA-PKcs</u>	<u>69</u>
<u>Anticuerpo:DNA-PKcs</u>	<u>69</u>
<u>DNA-PKcs:ADN</u>	<u>69</u>
<u>DNA-PKcs:ADN:estreptavidina:oro</u>	<u>69</u>
<u>DNA-PKcs:Ku:ADN</u>	<u>69</u>
<u>DNA-PKcs:Ku:ADN:estreptavidina:oro</u>	<u>70</u>
<u>DNA-PKcs:Ku:ADN autofosforilado</u>	<u>70</u>
<u>Artemis</u>	<u>70</u>
<b>3.11 Preparación de muestras mediante criomicroscopía electrónica para su observación en el microscopio electrónico</b>	<b>71</b>
<b>3.12 Digitalización de micrografías electrónicas</b>	<b>72</b>
<b>3.13 Procesamiento digital de imágenes</b>	<b>73</b>
3.13.1 Selección de partículas individuales	74
3.13.2 Corrección de CTF	74
3.13.3 Preparación de las imágenes	74
3.13.4 Análisis bidimensional de partículas individuales	75
<u>Análisis de la simetría rotacional</u>	<u>75</u>
<u>Alineamiento, clasificación y promediado bidimensional libres de patrón</u>	<u>76</u>
• <i>Máxima verosimilitud</i>	77
• <i>Métodos de alineamiento y clasificación libres de patrón disponibles en el paquete informático EMAN</i>	78
3.13.5 Reconstrucción tridimensional de partículas individuales	79
<u>Líneas comunes</u>	<u>79</u>
<u>Elipsoides artificiales</u>	<u>80</u>
<u>Ruido</u>	<u>80</u>
<i>Análisis del complejo Ku:ADN mediante el método de inclinación cónica al azar o RCT</i>	81
<i>Estrategia de procesamiento del complejo DNA-PKcs:ADN</i>	81
3.13.6 Cálculo de la resolución del volumen final	82
3.13.7 Procesamiento tridimensional multirreferencia	82
<b>3.14 Visualización</b>	<b>83</b>

<b>3.15 Análisis de la varianza</b>	<b>85</b>
<b>3.16 Ajuste computacional de estructuras atómicas en volúmenes de microscopía electrónica</b>	<b>88</b>
3.16.1 Ajuste de estructuras atómicas en los volúmenes obtenidos para el heterodímero Ku y el complejo Ku:ADN	88
3.16.2 Ajuste de estructuras atómicas en el volumen obtenido para la proteína DNA-PKcs mediante criomicroscopía electrónica	89
3.16.3 Ajuste de una molécula de ADN en la estructura obtenida mediante tinción negativa del complejo DNA-PKcs:ADN	90
3.16.4 Ajuste de la estructura atómica de Ku en los volúmenes obtenidos para el complejo DNA-PKcs:Ku:ADN	91
3.16.5 Ajuste del volumen de tinción negativa del complejo DNA-PKcs:Ku:ADN en la estructura del complejo dimérico de DNA-PKcs:Ku:ADN	91
<b>3.17 Predicción de movimientos macromoleculares utilizando análisis vibracional</b>	<b>92</b>
<b>3.18 Predicción de la estructura terciaria de Artemis</b>	<b>93</b>
<b>4. Resultados</b>	<b>95</b>
<b>4.1 Análisis estructural del heterodímero Ku</b>	<b>97</b>
4.1.1 Análisis de la interacción de la proteína Ku con ADN	97
4.1.2 Reconstrucción tridimensional del heterodímero Ku	99
4.1.3 Reconstrucción tridimensional del complejo Ku:ADN	101
4.1.4 Experimentos control que confirman la consistencia de las estructuras obtenidas para el heterodímero Ku y el complejo Ku:ADN	103
4.1.5 Experimentos de ajuste en el volumen obtenido para el heterodímero Ku	106
4.1.6 Experimentos de ajuste en el volumen obtenido para el complejo Ku:ADN	107
4.1.7 Localización del dominio SAP de Ku70 y del dominio CTD de Ku80	109
<i>Localización del dominio CTD de Ku80 en el volumen resuelto para el heterodímero Ku</i>	109
<i>Localización del dominio SAP de Ku70 y del CTD de Ku80 en el volumen resuelto para el complejo Ku:ADN</i>	110
4.1.8 Conclusión	112

<b>4.2 Análisis estructural de la proteína DNA-PKcs mediante tinción negativa</b>	<b>112</b>
4.2.1 Purificación de la proteína DNA-PKcs	112
4.2.2 Reconstrucción tridimensional de la proteína DNA-PKcs	113
4.2.3 Conclusión	116
<b>4.3 Análisis estructural de la proteína DNA-PKcs mediante criomicroscopía electrónica</b>	<b>116</b>
4.3.1 Reconstrucción tridimensional de la proteína DNA-PKcs obtenida mediante criomicroscopía electrónica	116
4.3.2 Localización de los diferentes dominios de la proteína DNA-PKcs en la estructura obtenida mediante criomicroscopía electrónica	120
A) <u>Localización de la región C-terminal conservada</u>	<u>121</u>
B) <u>Asignación y modelado de los dominios que componen la región cabeza</u>	<u>124</u>
<i>Dominio catalítico</i>	124
<i>Dominio FATC</i>	126
<i>Dominio FAT</i>	127
C) <u>Asignación y modelado de los dominios que componen las regiones brazo y palma</u>	<u>128</u>
<i>Región Brazo</i>	129
<i>Región Palma</i>	129
4.3.3 Conclusión	131
<b>4.4 Análisis estructural del complejo DNA-PKcs:ADN mediante tinción negativa</b>	<b>132</b>
4.4.1 Caracterización bioquímica de la interacción entre la proteína DNA-PKcs y el ADN	132
4.4.2 Reconstrucción tridimensional del complejo DNA-PKcs:ADN	134
4.4.3 Localización del ADN en los complejos DNA-PKcs:ADN	138
4.4.4 Cambios conformacionales en la proteína DNA-PKcs al unirse a ADN	140
4.4.5 Análisis vibracional	141
4.4.6 Conclusión	142
<b>4.5 Análisis estructural del complejo DNA-PKcs:Ku:ADN mediante tinción negativa y criomicroscopía electrónica</b>	<b>142</b>
4.5.1 Purificación y caracterización bioquímica del complejo DNA-PKcs:Ku:ADN	143

4.5.2 Reconstrucción tridimensional del complejo DNA-PKcs:Ku:ADN obtenida por tinción negativa	144
4.5.3 Localización del heterodímero Ku en la estructura resuelta mediante tinción negativa para el complejo DNA-PKcs:Ku:ADN	149
4.5.4 Observación al microscopio electrónico del complejo DNA-PKcs:Ku:ADN autofosforilado	150
4.5.5 Reconstrucción tridimensional del complejo DNA-PKcs:Ku:ADN obtenida mediante criomicroscopía electrónica	152
4.5.6 Localización de la molécula de ADN en el complejo DNA-PKcs:Ku:ADN	154
4.5.7 Reconstrucción tridimensional del complejo dimérico de DNA-PKcs:Ku:ADN obtenida a partir de especímenes teñidos negativamente	156
4.5.8 Orientación de los complejos DNA-PKcs:Ku:ADN en la estructura del complejo dimérico	161
4.5.9 Localización del ADN en el complejo dimérico	161
4.5.10 Conclusión	162
<b>4.6 Caracterización bioquímica de la interacción entre las proteínas Artemis y DNA-PKcs</b>	<b>163</b>
4.6.1 Caracterización bioquímica de la proteína Artemis	164
4.6.2 Análisis bidimensional de imágenes de microscopía electrónica correspondientes a la proteína Artemis	166
4.6.3 Generación de un modelo tridimensional para la proteína Artemis	168
4.6.4 Caracterización bioquímica de la interacción entre las proteínas Artemis y DNA-PKcs	168
4.6.5 Conclusión	172
<b>5. Discusión</b>	<b>173</b>
<b>5.1 Modelo topológico de DNA-PKcs</b>	<b>175</b>
5.1.1 Comparación entre el modelo estructural propuesto y la información existente para DNA-PKcs obtenida mediante criomicroscopía electrónica	177
5.1.2 Datos adicionales refuerzan el modelo topológico propuesto para DNA-PKcs	181



<b>5.2 El reconocimiento de extremos de ADN por DNA-PKcs conlleva la activación de la quinasa mediante cambios conformacionales</b>	<b>183</b>
<b>5.3 Existe un modelo estructural común para los miembros de la familia de las PIKKs</b>	<b>187</b>
<b>5.4 Modelo de las etapas iniciales del proceso de NHEJ</b>	<b>192</b>
5.4.1 La unión del heterodímero Ku a extremos de ADN conlleva cambios conformacionales en los dominios C-terminales de Ku70 y Ku80	192
5.4.2 El reclutamiento de DNA-PKcs por el complejo Ku:ADN conlleva cambios conformacionales en el dominio CTD de Ku80	194
5.4.3 La formación del complejo holoenzima DNA-PK conlleva cambios conformacionales en DNA-PKcs	196
5.4.4 Las características del complejo dimérico de DNA-PKcs:Ku:ADN son compatibles con las del complejo sináptico que interviene en el proceso de NHEJ	198
<b>5.5 Reparación de DSB en el contexto de la cromatina</b>	<b>202</b>
<b>5.6 Otras proteínas adicionales son requeridas para completar la reparación del ADN</b>	<b>204</b>
<b>5.7 Conclusión</b>	<b>206</b>
<b>6. Conclusiones</b>	<b>209</b>
<b>7. Bibliografía</b>	<b>213</b>
<b>8. Publicaciones</b>	<b>233</b>
♦ “Structural model of full-length human Ku70-Ku80 heterodimer and its recognition of DNA and DNA-PKcs”	235
♦ “Visualization of DNA-induced conformational changes in the DNA repair kinase DNA-PKcs”	254
♦ “Three-Dimensional Structure and Regulation of the DNA-Dependent Protein Kinase Catalytic Subunit (DNA-PKcs)”	262
♦ “Three-Dimensional Structure of the Human DNA-PKcs/Ku70/Ku80 Complex Assembled on DNA and Its Implications for DNA DSB repair”	276



# Índice de Figuras y Tablas

## 1. Introducción

<b>Figura 1.1.1.</b> Ejemplo de alteraciones genéticas en células tumorales.	5
<b>Figura 1.4.1.</b> Representación esquemática del proceso de NHEJ.	10
<b>Figura 1.4.2.</b> Características estructurales del heterodímero Ku.	14
<b>Figura 1.4.3.</b> Descripción estructural de DNA-PKcs.	18
<b>Figura 1.4.4.</b> Características estructurales de la familia de metalo- $\beta$ -lactamasas $\beta$ -CASP y esquema de la proteína Artemis.	22
<b>Figura 1.4.5.</b> Descripción estructural de los componentes que forman el complejo XRCC4:ADN Ligasa IV.	24
<b>Figura 1.4.6.</b> Descripción estructural de la proteína XLF/Cernunnos.	26
<b>Figura 1.4.7.</b> Esquema del proceso de recombinación V(D)J.	28
<b>Figura 1.5.1.</b> Técnicas de preparación de muestras para su análisis posterior mediante microscopía electrónica y procesamiento digital de imágenes individuales.	36
<b>Figura 1.5.2.</b> Ángulos de Euler y problema de determinación de la “mano” en volúmenes resueltos por microscopía electrónica y procesamiento digital de imágenes individuales.	39
<b>Figura 1.5.3.</b> Representación de los ángulos de Euler del conjunto de partículas individuales utilizadas en la reconstrucción tridimensional.	40
<b>Figura 1.5.4.</b> Esquema general del proceso de refinamiento angular característico del paquete informático EMAN.	45
<b>Figura 1.5.5.</b> Estimación de la resolución en estructuras resueltas por microscopía electrónica y procesamiento digital de imágenes individuales.	46

## 2. Objetivos

## 3. Materiales y Métodos

<b>Tabla 3.1.</b> Moléculas de ADN de doble cadena utilizadas en el ensamblaje de complejos proteína:ADN.	53
<b>Figura 3.1.1.</b> Diseño esquemático de las moléculas de ADN utilizadas en el ensamblaje del complejo DNA-PKcs:Ku:ADN.	55
<b>Tabla 3.2.</b> Anticuerpos utilizados en los diferentes ensayos de “Western-blot”.	62

**Tabla 3.3.** Resumen de los principales parámetros asociados a los diferentes procesamientos de partículas individuales. 83

**Tabla 3.4.** Datos de las diferentes estructuras depositadas en la “Macromolecular Structure Database”. 85

## 4. Resultados

**Figura 4.1.1.** Análisis de la interacción del heterodímero Ku con ADN. 98

**Figura 4.1.2.** Reconstrucción tridimensional del heterodímero Ku. 99

**Figura 4.1.3.** Reconstrucción tridimensional del complejo Ku:ADN. 102

**Figura 4.1.4.** Reconstrucciones tridimensionales del complejo Ku:ADN obtenidas mediante RCT. 104

**Figura 4.1.5.** Análisis de la varianza en las estructuras resueltas para el heterodímero Ku y el complejo Ku:ADN. 105

**Figura 4.1.6.** Experimentos de ajuste en el volumen obtenido para el heterodímero Ku. 107

**Figura 4.1.7.** Experimentos de ajuste en el volumen obtenido para el complejo Ku:ADN. 108

**Figura 4.1.8.** Localización del dominio CTD de Ku80 en el volumen obtenido para el heterodímero Ku. 110

**Figura 4.1.9.** Localización del dominio SAP de Ku70 y del CTD de Ku80 en el volumen obtenido para el complejo Ku:ADN. 111

**Figura 4.2.1.** Purificación y análisis mediante microscopía electrónica de especímenes teñidos negativamente de la proteína DNA-PKcs. 114

**Figura 4.2.2.** Reconstrucción tridimensional de la proteína DNA-PKcs. 115

**Figura 4.3.1.** Reconstrucción tridimensional de la proteína DNA-PKcs mediante criomicroscopía electrónica. 117

**Figura 4.3.2.** Estructura resuelta mediante criomicroscopía electrónica para la proteína DNA-PKcs. 119

**Figura 4.3.3.** Comparación de estructuras resueltas mediante microscopía electrónica de las PIKKs ATM y DNA-PKcs. 121

**Figura 4.3.4.** Localización de la región C-terminal de la proteína DNA-PKcs mediante marcaje con anticuerpos. 123

**Figura 4.3.5.** Localización del dominio catalítico en la región cabeza. 125

**Figura 4.3.6.** Localización del dominio FATC en la región cabeza. 127

**Figura 4.3.7.** Localización del dominio FAT en la región cabeza. 128

<b>Figura 4.3.8.</b> Localización de los dominios que componen las regiones brazo y palma.	130
<b>Figura 4.3.9.</b> Organización tridimensional de la proteína DNA-PKcs.	131
<b>Figura 4.4.1.</b> Análisis bioquímico de la interacción entre la proteína DNA-PKcs y el ADN.	133
<b>Figura 4.4.2.</b> Análisis mediante microscopía electrónica de especímenes teñidos negativamente de la muestra de DNA-PKcs incubada con ADN.	135
<b>Figura 4.4.3.</b> Representación de los ángulos de Euler y cálculo de la resolución para las estructuras resueltas correspondientes a DNA-PKcs y al complejo DNA-PKcs:ADN.	137
<b>Figura 4.4.4.</b> Localización del ADN presente en el complejo DNA-PKcs:ADN.	139
<b>Figura 4.4.5.</b> Cambios conformacionales en la proteína DNA-PKcs provocados por la unión al ADN.	141
<b>Figura 4.5.1.</b> Purificación del complejo DNA-PKcs:Ku:ADN.	145
<b>Figura 4.5.2.</b> Análisis del complejo DNA-PKcs:Ku:ADN mediante microscopía electrónica de especímenes teñidos negativamente.	146
<b>Figura 4.5.3.</b> Reconstrucción tridimensional del complejo DNA-PKcs:Ku:ADN.	148
<b>Figura 4.5.4.</b> Localización del heterodímero Ku en la estructura resuelta para el complejo DNA-PKcs:Ku:ADN.	149
<b>Figura 4.5.5.</b> Ajuste de la estructura atómica correspondiente al núcleo de Ku en la estructura resuelta para el complejo DNA-PKcs:Ku:ADN.	151
<b>Figura 4.5.6.</b> Reconstrucción tridimensional del complejo DNA-PKcs:Ku:ADN mediante criomicroscopía electrónica.	153
<b>Figura 4.5.7.</b> Estructura resuelta mediante criomicroscopía electrónica para el complejo DNA-PKcs:Ku:ADN, ajuste de la estructura atómica correspondiente al núcleo de Ku y marcaje del ADN presente en el complejo.	155
<b>Figura 4.5.8.</b> Análisis estructural del complejo dimérico de DNA-PKcs:Ku:ADN a partir de especímenes teñidos negativamente.	156
<b>Figura 4.5.9.</b> Reconstrucción tridimensional del complejo dimérico de DNA-PKcs:Ku:ADN.	159
<b>Figura 4.5.10.</b> Evidencias adicionales que confirman la presencia del complejo dimérico en la muestra del ensamblado DNA-PKcs:Ku:ADN purificado.	160

<b>Figura 4.5.11.</b> Determinación de la orientación del ensamblado DNA-PKcs:Ku:ADN en la estructura resuelta para el complejo dimérico y localización de las moléculas de ADN.	162
<b>Figura 4.6.1.</b> Análisis bioquímico de la proteína Artemis purificada.	165
<b>Figura 4.6.2.</b> Análisis por microscopía electrónica y predicción de un modelo atómico para la proteína Artemis.	167
<b>Figura 4.6.3.</b> Análisis bioquímico de la interacción entre las proteínas Artemis y DNA-PKcs.	170
<b>5. Discusión</b>	
<b>Figura 5.1.1.</b> Modelo de organización tridimensional para DNA-PKcs.	175
<b>Figura 5.1.2.</b> Estructura de DNA-PKcs a 7Å de resolución obtenida utilizando criomicroscopía electrónica.	179
<b>Figura 5.1.3.</b> Experimentos adicionales refuerzan el modelo topológico propuesto para DNA-PKcs.	182
<b>Figura 5.2.1.</b> Modelo de regulación de la actividad quinasa de DNA-PKcs dependiente de la unión a ligando.	186
<b>Figura 5.3.1.</b> Arquitectura común de las PIKKs DNA-PKcs, ATM y TOR.	188
<b>Figura 5.3.2.</b> Organización tridimensional de la subunidad catalítica de la PI3K $\alpha$ .	191
<b>Figura 5.4.1.</b> Modelo esquemático de los cambios conformacionales que tienen lugar en el heterodímero Ku completo.	193
<b>Figura 5.4.2.</b> Comparación de la estructura de Ku en los volúmenes obtenidos para los complejos DNA-PKcs:Ku:ADN y Ku:ADN.	195
<b>Figura 5.4.3.</b> Cambios conformacionales en DNA-PKcs asociados a la formación del ensamblado DNA-PKcs:Ku:ADN.	197
<b>Figura 5.4.4.</b> Localización de las moléculas de ADN presentes en el complejo dimérico de DNA-PKcs:Ku:ADN.	200
<b>6. Conclusiones</b>	
<b>7. Bibliografía</b>	
<b>8. Publicaciones</b>	

## Abreviaturas y símbolos

**Å:** ångström

**Ab:** anticuerpo (del inglés “**Antibody**”)

**ABL1:** del inglés “v-abl Abelson murine leukaemia viral oncogene homologue 1”

**ARM:** armadillo

**ATM:** del inglés “**A**taxia **t**elangiectasia **m**utated”

**ATP:** Adenosina trifosfato (del inglés “**A**denosine **t**riphosphate”)

**ATR:** del inglés “**A**TM and **R**ad3 related”

**β-CASP:** del inglés “Metallo-**β**-lactamase-associated **C**PSF **A**rtemis **S**NM1/**P**SO2”

**BCR:** del inglés “**B**reakpoint **c**luster **r**egion”

**BER:** del inglés “**B**ase **e**xcision **r**epair”

**Biot:** biotinilado o iminobiotinilado

**Bis-Tris:** del inglés “**Bis**(2-hydroxyethyl)iminotris(hydroxymethyl)methane”

**BOG:** β-octil-glucósido

**BRCT:** del inglés “**B**reast **C**ancer gene 1 **c**arboxy **t**erminal”

**BSOFT:** del inglés “**B**ernard’s **s**oftware”

**°C:** grado centígrado

**C-terminal:** carboxilo terminal

**CCD:** del inglés “**C**harged **c**oupled **d**evice”

**cDNA:** del inglés “**C**omplementary **D**NA”

**Ci:** curio

**cm:** centímetro

**colores:** del inglés “**C**orrelation based **l**ow **r**esolution docking”

**CPM:** cuentas por minuto

**CPSF-73:** del inglés “73 kDa subunit of the **c**leavage and **p**olyadenylation **s**pecificity factor”

**Crio:** criomicroscopía electrónica

**CRX:** cristalografía de rayos X

**CSIC:** Consejo Superior de Investigaciones Científicas

**CTD:** del inglés “**C**-terminal **d**omain”

**CTF:** del inglés “**C**ontrast **t**ransfer **f**unction”

**2D:** bidimensional

**DNA-PK:** del inglés “**D**N A-dependent **p**rotein **k**inase”

**DNA-PKcs:** del inglés “**D**N A-dependent **p**rotein **k**inase **c**atalytic **s**ubunit”

**dP:** defosforilado

**DSB:** del inglés “**D**ouble-**s**trand **D**N**A** **b**reaks”

**DTE:** ditioeritrol

**DTT:** ditiotreitrol

**E:** elipsoides

**e<sup>-</sup>:** electrones

**ECL:** del inglés “**E**nhanced **c**hemiluminescence”

**EDTA:** del inglés “**E**thylendiaminetetraaceticacid”

**EGTA:** del inglés “**E**thyleneglycoltetraacetic acid”

**EMAN:** del inglés “**E**lectron **m**icrograph **a**nalysis”

**EMSA:** del inglés “**E**lectrophoretic **m**obility **s**hift **a**ssay”

**et al.:** “y otros”, del latín “**e**t **a**lii”

**FAT:** del inglés “**F**RAP, **A**TM, and **T**RAAP”

**FATC:** del inglés “**F**RAP, **A**TM, **T**RAAP and **C**-terminal”

**FEG:** del inglés “**F**ield **e**mission **g**un”

**FISH:** del inglés “**F**luorescent in **s**itu **h**ybridization”

**fmol:** femtomol

**FPLC:** del inglés “**F**ast **p**erformance **l**iquid **c**hromatography”

**FSC:** del inglés “**F**ourier **s**hell **c**orrelation”

**HEAT:** del inglés “**H**untingtin, **e**longation factor 3, **A** subunit of PP2A, **T**OR1”

**HeLa:** **H**enrietta **L**acks

**HEPES:** del inglés “**N**-2-**H**ydroxyethylpiperazine-**N'**-2-**e**thanesulfonic acid”

**HPLC:** del inglés “**H**igh **p**erformance **l**iquid **c**hromatography”

**HRP:** del inglés “**H**orseradish **p**eroxidase”

**kDa:** kilodalton

**KerDenSOM:** del inglés “**K**ernel **d**ensity **s**elf **o**rganizing **m**ap”

**KIP:** del inglés “**K**inase interacting **p**roteína”

**kV:** kilovoltio

**L:** técnica de líneas comunes

**lazo D:** del inglés “**D**isplacement-loop”

**Lif1p:** del inglés “**L**igase interacting **f**actor **1**”

**LigD:** ADN **L**igasa **D**

**LRR:** del inglés “**L**eucine **r**ich **r**egion”

**µg:** microgramo

**µl:** microlitro

**M:** molar

**m:** metro

**mAu:** del inglés “**m**ili**A**bsorbance **u**nits”



**mCi:** milicurio  
**mg:** miligramo  
**min:** minutos  
**MI:** ml\_align2d  
**ml:** mililitro  
**ml\_align2d:** del inglés “**Maximum-likelihood alignment in 2D**”  
**mM:** milimolar  
**mmol:** milimol  
**MMR:** del inglés “**Mismatch repair**”  
**mQ:** mili-Q  
**mTOR:** del inglés “**Mammalian target of rapamycin**”  
**N-terminal:** amino terminal  
**NER:** del inglés “**Nucleotide excision repair**”  
**ng:** nanogramo  
**NHEJ:** del inglés “**Nonhomologous end joining**”  
**nm:** nanómetro  
**nt:** nucleótido  
**oligo:** oligo-desoxirribonucleótido  
**OH:** hidróxido  
**P:** fosfato  
**p/v :** peso/volumen  
**Pa:** pascal  
**PAGE :** del inglés “**Polyacrylamide-gel electrophoresis**”  
**pb:** pares de bases  
**PBS:** del inglés “**Phosphate buffered saline**”  
**PDB:** del inglés “**Protein data bank**”  
**PFT:** del inglés “**Protein farnesy transferase**”  
**PhD:** del latín “**Philosophiæ Doctor**”  
**Phyre:** del inglés “**Protein homology/analogy recognition engine**”  
**PI3K:** del inglés “**Phosphatidylinositol-3-kinase**”  
**PI3Ks:** del inglés “**Phosphatidylinositol-3-kinases**”  
**PIKKs:** del inglés “**Phosphatidylinositol-3-kinase related protein kinases**”  
**pmol:** picomol  
**PMSF:** del inglés “**Phenylmethylsulfonyl fluoride**”  
**PNK:** del inglés “**Polynucleotide kinase**”  
**pol:** polimerasa  
**PolDom:** del inglés “**Polimerase domain**”

**PP2A:** del inglés “**P**rotein **p**hosphatase **2A**”

**PSA:** persulfato amónico

**PVDF:** del inglés “**P**olyvinylidene **d**ifluoride”

**R:** ruido

**RAG:** del inglés “**R**ecombinase-activating **g**enes”

**RCT:** del inglés “**R**andom **C**onical **T**ilt”

**Ref:** refine2d\_py

**RMN:** resonancia **m**agnética **n**uclear

**rpm:** revoluciones **p**or **m**inuto

**RSS:** del inglés “**R**ecombination-**s**ignal **s**equence”

**RS-SCID:** del inglés “**R**adiosensitive-**SCID**”

**RSSs:** del inglés “**R**ecombination-**s**ignal **s**equences”

**SAGA:** del inglés “**S**pt-**A**da-**G**cn5-**A**cetyltransferase”

**SAP:** del inglés “**S**AF-**A**/B, **A**cinus and **P**IAS”

**SCID:** del inglés “**S**evere **c**ombined **i**mmune **d**eficiency”

**SDS:** del inglés “**S**odium **d**odecylsulphate”

**SMG1:** del inglés “**S**uppressor with **m**orphological **e**ffect on **g**enitalia **1**”

**SPIDER:** del inglés “**S**ystem for **p**rocessing **i**mage **d**ata from **e**lectron **m**icroscopy and related fields”

**SREBP-2:** del inglés “**S**terol **r**egulatory **e**lement-**b**inding **p**rotein **2**”

**SSBR:** del inglés “**S**ingle **s**trand **b**reak **r**epair”

**St:** startnrclasses

**SVD:** del inglés “**S**ingular **v**alue **d**ecomposition”

**TEMED:** N,N,N',N'-tetrametileten-diamina

**TEV:** del inglés “**T**obacco **e**tch **v**irus”

**TOR:** del inglés “**T**arget **o**f **r**apamycin”

**Tn:** tinción negativa

**Tra1:** del inglés “**T**ranscription-**a**ssociated **p**rotein **1**”

**Tris:** del inglés “**T**ris(hydroxymethyl)aminomethane”

**TRRAP:** del inglés “**T**ransformation/transcription domain associated **p**rotein”

**Tween20:** polisorbato 20

**UV:** ultravioleta

**V:** voltio

**WRN:** Werner

**XLF:** del inglés “**X**RCC4-like **f**actor”

**XMIPP:** del inglés “**X**-Window-based **m**icroscopy **i**mage **p**rocessing **p**ackage”

**XRCC4:** del inglés “**X**-ray **c**ross-**c**omplementation **g**roup **4**”



# 1. Introducción



## Biología estructural de la reparación de roturas de doble cadena en el ADN

Los procesos biológicos característicos de los seres vivos son realizados por ensamblados macromoleculares dinámicos de complejidad variable (Gavin et al., 2002; Ho et al., 2002). Los componentes que constituyen estos ensamblados interactúan entre ellos específicamente permitiendo en último término la supervivencia de los organismos. Por ello, para entender el mecanismo por el que se llevan a cabo estos procesos, es necesario caracterizar estructuralmente los componentes implicados así como las interacciones que establecen.

Esta tesis se ha centrado en el análisis estructural del proceso de reparación de roturas de doble cadena en el ADN por unión de extremos no homólogos o NHEJ (del inglés “**Non**homologous end joining”) (Lieber, 2008). Este proceso, esencial para el mantenimiento de la estabilidad génica en organismos superiores, es llevado a cabo por un grupo definido de enzimas algunas de las cuales han sido identificadas en los últimos años (Moshous et al., 2001; Ahnesorg et al., 2006; Buck et al., 2006). Además este sistema de reparación tiene importancia terapéutica ya que los distintos componentes implicados pueden utilizarse como dianas en los tratamientos antitumorales (Pastwa y Malinowski, 2007). Durante el desarrollo de esta tesis se ha realizado la caracterización estructural de proteínas clave del proceso de NHEJ así como de los complejos fisiológicos que forman. Para ello se ha empleado la técnica estructural basada en la combinación de microscopía electrónica de transmisión y procesamiento digital de partículas individuales. Esta técnica, complementaria a otras técnicas estructurales como la cristalografía de rayos X o la resonancia magnética nuclear, es extremadamente eficaz a la hora de determinar la estructura de complejos macromoleculares de elevado peso molecular que se purifican en cantidades limitantes (Abbott, 2002). Además posee la potencialidad de alcanzar resolución atómica (Henderson, 1995; Zhang et al., 2008), si bien las estructuras que se resuelven son normalmente de baja o media resolución (5-30 Å) y por ello, con el objetivo de reducir en

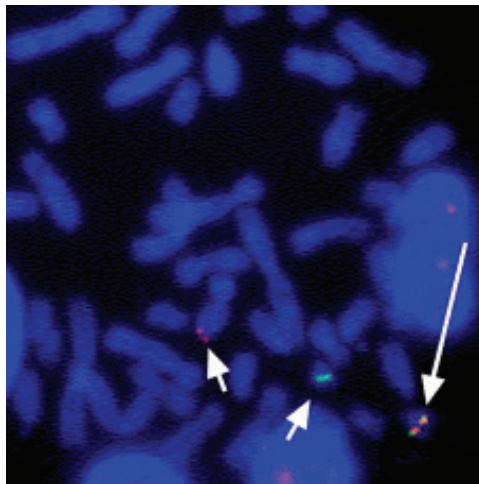
parte esta limitación, se recurre al ajuste de estructuras atómicas en los volúmenes resueltos mediante microscopía electrónica (Chiu et al., 2005; Jiang y Ludtke, 2005). Por tanto la combinación de información estructural obtenida utilizando diferentes técnicas es clave para conocer el mecanismo característico de los distintos procesos biológicos.

## 1.1 Mantenimiento de la estabilidad genómica

La supervivencia y perpetuación de los seres vivos depende del mantenimiento y la conservación de su información genética. La integridad del genoma está continuamente amenazada por factores externos al propio organismo (radiación, productos químicos), pero también por factores internos (productos del propio metabolismo) incluida la inestabilidad intrínseca del ADN (Hoeijmakers, 2001). Dentro de estas amenazas destacan por su letalidad aquellas que provocan la rotura de las dos cadenas de la molécula de ADN dando lugar a la aparición de nuevos extremos. Este tipo de roturas conlleva la pérdida de la continuidad del genoma y puede tener efectos devastadores en la supervivencia celular (Rich et al., 2000). Existen distintos mecanismos por los que pueden generarse roturas de doble cadena en el ADN (de ahora en adelante DSB del inglés “**D**ouble-**s**trand **D**N**A** **b**reaks”). Se sabe que el metabolismo celular normal y la radiación ionizante provocan la aparición de especies altamente reactivas que contienen oxígeno. Estas especies (peróxido de hidrógeno, aniones superóxido, etc.) reaccionan con el ADN dando lugar a roturas de cadena sencilla que, al aparecer próximas en hebras complementarias, generan DSB (Wiseman y Halliwell, 1996; Lieber et al., 2003). Además roturas de este tipo ocurren durante el metabolismo normal del ADN. Por ejemplo si la horquilla de replicación se topa con una rotura de cadena sencilla durante su progresión pueden generarse DSB (Kuzminov, 1995). También se sabe que el mal funcionamiento o desregulación de enzimas que utilizan el ADN como sustrato provoca DSB (Neale et al., 2005). Un metabolismo deficiente de los telómeros puede dar lugar igualmente a la aparición de este tipo de roturas (Riha et al., 2006). Curiosamente las células han desarrollado la capacidad de generar DSB de manera controlada para

llevar a cabo procesos celulares esenciales. Esas roturas aparecen por ejemplo al inicio de la recombinación entre cromosomas homólogos en la meiosis (Dresser, 2000), y también durante las reordenaciones génicas que tienen lugar en los linfocitos y que son responsables de la diversidad característica de los receptores de antígenos (Gellert, 2002).

Existe una correlación directa entre la aparición de DSB y el aumento de la inestabilidad genómica (van Gent et al., 2001). Esta inestabilidad es una característica esencial de las células tumorales en las que se manifiesta en muchos casos como aberraciones cromosómicas que conllevan pérdida, amplificación o translocación de fragmentos génicos (*Figura 1.1.1*). Estas reordenaciones génicas pueden provocar la activación de protooncogenes o la pérdida de genes supresores de tumores. Las células tumorales poseen alteraciones en los mecanismos celulares de respuesta a daño en el ADN y, de esta manera, favorecen la inestabilidad genómica y por ende la acumulación de



**Figura 1.1.1. Ejemplo de alteraciones genéticas en células tumorales.**

Análisis por FISH (del inglés “**F**luorescent **i**n **s**itu **h**ybridization”) de una célula cancerosa obtenida a partir de un paciente enfermo de leucemia mielocítica crónica. En las células cancerosas de estos pacientes aparece una proteína oncogénica que resulta de la fusión del gen *BCR* (del inglés “**B**reakpoint **c**luster **r**egion”) con el gen *ABL1* (del inglés “v-abl Abelson murine leukaemia viral oncogene homologue 1”). La fusión de estos dos fragmentos da lugar a un cromosoma de aspecto característico llamado cromosoma Filadelfia. En la figura la señal correspondiente al gen fusión *BCR-ABL* presente en uno de los cromosomas 22 se indica mediante una flecha de mayor tamaño. La señal roja pertenece al gen *ABL* localizado en uno de los cromosomas 9, mientras que la señal verde corresponde al gen *BCR* localizado en el cromosoma 22 restante (ambas señales se indican mediante flechas de menor tamaño). La imagen se ha extraído de Batista et al., 2005.

alteraciones en el genoma (Lengauer et al., 1998; Khanna y Jackson, 2001). Las alteraciones en los mecanismos de respuesta a daño en el ADN son fundamentales en la progresión del fenotipo tumoral, pero pueden utilizarse también como dianas en los tratamientos antitumorales (Helleday et al., 2007).

## 1.2 Sistemas de reparación de daño en el ADN

Los factores que amenazan la integridad del genoma son inherentes a la vida, por ello los organismos han desarrollado sistemas que permiten lidiar con los daños que provocan en el ADN. Estos sistemas de reparación son específicos para cada tipo de lesión, están conservados a lo largo de la evolución y alteraciones en ellos contribuyen al deterioro general de los organismos, a su envejecimiento prematuro y a la adquisición de fenotipos tumorales (Cromie et al., 2001; Cline y Hanawalt, 2003; Lieber et al., 2003).

Cuando el daño queda restringido a una de las cadenas del ADN es posible utilizar la hebra complementaria como molde para reparar la lesión. Ejemplos de este tipo de lesiones son la oxidación y alquilación de las bases, las deaminaciones espontáneas de citosinas, la liberación hidrolítica de purinas del esqueleto azúcar-fosfato o la aparición de aductos químicos voluminosos (Lindahl, 1993). Para determinados tipos de lesiones existen enzimas capaces de catalizar la reacción inversa de aquella que generó el daño en el ADN. En cualquier caso la mayoría de las lesiones mencionadas anteriormente son reparadas por mecanismos que introducen cortes en la cadena dañada. Ejemplos de este tipo de mecanismos son los procesos BER (del inglés “**B**ase **e**xcision **r**epair”) (Memisoglu y Samson, 2000), NER (del inglés “**N**ucleotide **e**xcision **r**epair”) (Cline y Hanawalt, 2003) y MMR (del inglés “**M**ismatch **r**epair”) (Hsieh, 2001). La aparición de roturas de cadena sencilla en el ADN es también un tipo de lesión restringida a una de las hebras del ADN, y su reparación se lleva a cabo mediante el proceso denominado SSBR (del inglés “**S**ingle **s**trand **b**reak **r**epair”) (Caldecott, 2007).



En el caso de las DSB las dos hebras están dañadas, por ello no es posible utilizar la cadena complementaria de la misma molécula de ADN como molde durante la reparación. Este tipo de roturas son por tanto más difíciles de reparar y, además, son más peligrosas puesto que pueden dar lugar a aberraciones cromosómicas que afecten a muchos genes simultáneamente (Richardson y Jasin, 2000). Una única rotura de doble cadena en el genoma es suficiente para provocar la muerte celular por apoptosis (Rich et al., 2000). Sin embargo se estima que en cada célula de una persona de aproximadamente 70 años de edad, a lo largo de su existencia, han tenido lugar alrededor de 2.000 DSB (Lieber et al., 2003). Las células disponen de dos tipos de mecanismos para reparar este tipo de roturas: aquellos que utilizan la recombinación homóloga (Cromie et al., 2001), y los que unen los extremos generados tras la rotura sin la necesidad de un molde (Lieber, 2008). Ambos mecanismos no son mutuamente exclusivos y participan de manera conjunta en los procesos de reparación, aunque la contribución relativa de cada uno depende del estadio del ciclo celular (Takata et al., 1998). Estos mecanismos de reparación son una parte de la respuesta coordinada, compleja y global que las células generan como consecuencia de la aparición de DSB (Khanna y Jackson, 2001; van Gent et al., 2001; Jackson, 2002). Determinadas moléculas detectan el daño y generan una señal que, mediante cascadas de transducción, se transmite hasta moléculas efectoras las cuales ejercen su función en diferentes procesos celulares. Esa respuesta global conlleva la activación transcripcional y post-transcripcional de diversos genes incluyendo aquellos implicados directamente en la reparación de la rotura. Es además una respuesta coordinada con la progresión del ciclo celular, puesto que las DSB son especialmente peligrosas si aparecen durante la replicación del genoma o durante la segregación de los cromosomas en la mitosis. Finalmente, en caso de que el daño no pueda ser reparado, la célula debe ser capaz de iniciar el programa de muerte celular por apoptosis. Por tanto, sin la existencia de mecanismos que permitan detectar y reparar DSB además de coordinar esos procesos con las distintas etapas del ciclo celular, la viabilidad de los organismos se vería seriamente comprometida.

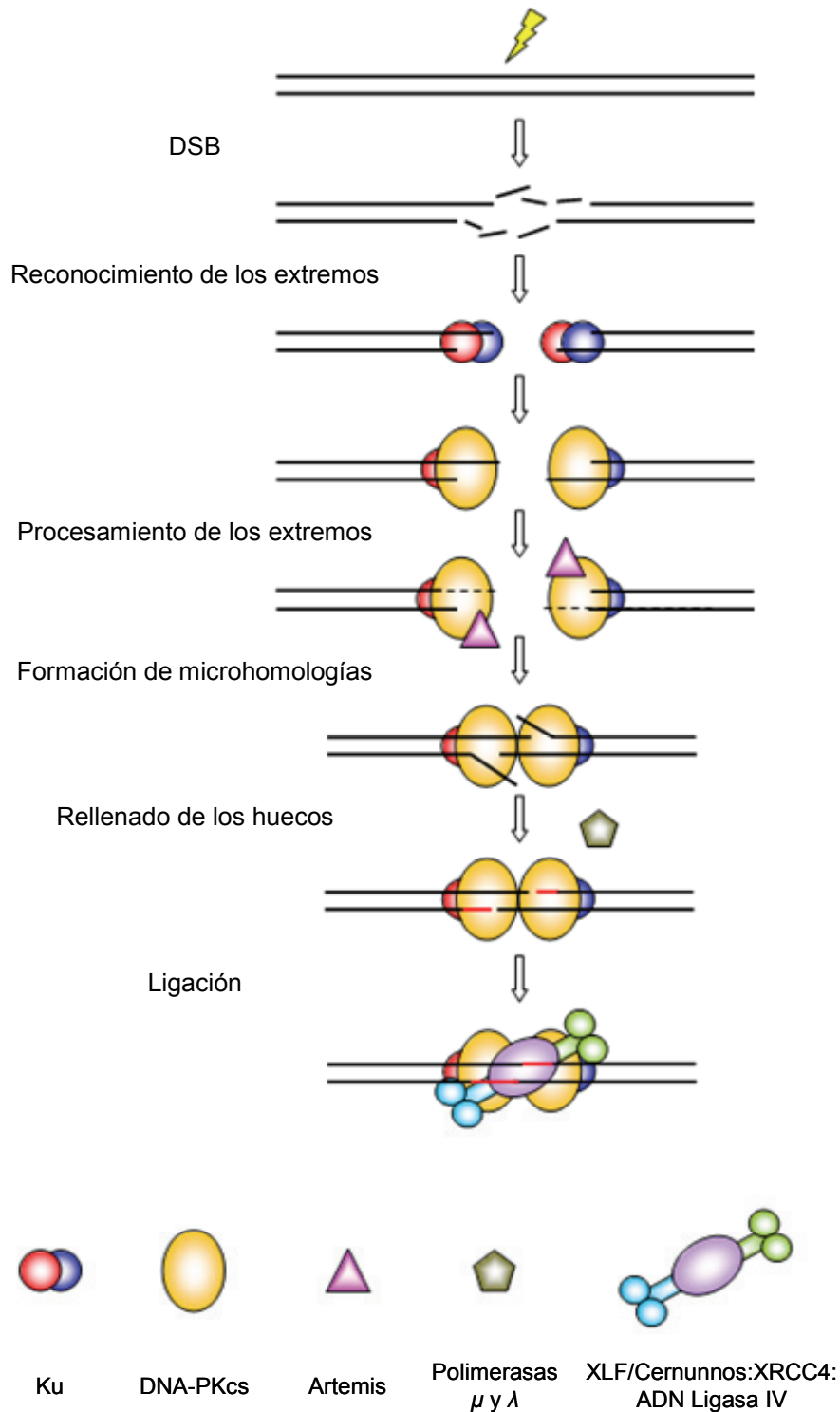
### 1.3 Sistemas de reparación de DSB basados en recombinación homóloga

Este tipo de mecanismos de reparación permite restaurar fielmente la continuidad del genoma gracias a secuencias homólogas que preferentemente se localizan en la cromátida hermana. Por ello la reparación por recombinación homóloga está activa principalmente durante las fases S y G2 del ciclo celular (se pueden encontrar revisiones extensas sobre este sistema de reparación en Cromie et al., 2001; West, 2003; Helleday et al., 2007). La primera etapa del proceso de reparación consiste en la detección y protección de los extremos que aparecen tras la rotura. A continuación los extremos se procesan eliminándose nucleótidos en dirección  $5' \rightarrow 3'$  a partir del extremo. Se genera por tanto un fragmento de cadena sencilla que contiene el extremo  $3'$  al que se unen diversas proteínas dando lugar a un filamento nucleoproteico. Este filamento invade la cromátida hermana y, tras desplazar una de las hebras, forma un heterodúplex de ADN en aquella región que sea complementaria dando lugar a un lazo D (del inglés “Displacement-loop”). Después tiene lugar la síntesis de ADN en el heterodúplex a partir del extremo  $3'$  del sitio original de rotura. La cromátida hermana proporciona por tanto un molde ideal para que tenga lugar la reparación libre de error. Además en la transición entre el homodúplex y el heterodúplex se forma una estructura típica con forma de X que se denomina unión Holliday (del inglés “Holliday junction”). Esta estructura es el resultado de la invasión y posterior entrecruzamiento de las cromátidas hermanas. La unión Holliday puede desplazarse en el mismo sentido en el que se da la síntesis de ADN favoreciendo así la liberación de la hebra recién sintetizada. Esta hebra aparea de nuevo con secuencias complementarias que se encuentren al otro lado del sitio de rotura en la cromátida donde tuvo lugar la lesión. La etapa final consiste en la ligación de las roturas de cadena sencilla gracias a proteínas implicadas en procesos generales de reparación y síntesis.

## 1.4 Reparación de DSB por unión de extremos no homólogos

Los primeros indicios de la existencia de un sistema de reparación no dependiente de recombinación homóloga datan de finales de los años 70 (Resnick y Moore, 1979). Posteriormente se identificaron células de mamíferos que presentaban defectos en la reparación de DSB cuyo fenotipo era diferente al de mutantes caracterizados de *Escherichia coli* y *Saccharomyces cerevisiae* (Jeggo, 1990). Estas y otras evidencias llevaron a proponer la existencia de un mecanismo de reparación denominado de recombinación ilegítima o recombinación no homóloga (de ahora en adelante NHEJ del inglés “Nonhomologous end joining”) (Moore y Haber, 1996). Este sistema de reparación consiste en la unión de los extremos generados tras la rotura sin la ayuda de moléculas accesorias de ADN que funcionen como molde (Lieber et al., 2003; Ma et al., 2005a; Lieber, 2008; Weterings y Chen, 2008).

La reparación por NHEJ es un proceso flexible en el cual el orden de las diferentes actividades enzimáticas que intervienen está determinado por la configuración de los extremos generados tras la rotura (*Figura 1.4.1*). Esta flexibilidad conlleva que los productos que se obtienen tras la reparación sean muy variables (Ma et al., 2004). El primer paso en el proceso de reparación por NHEJ consiste en el reconocimiento de los extremos generados. Tanto en el caso de roturas que aparecen por causas fisiológicas como por causas patológicas, es frecuente que el esqueleto azúcar-fosfato y las bases próximas a los extremos presenten alteraciones. Por ello debe darse un procesamiento previo que permita recuperar la configuración típica de extremos 3' OH y 5' fosfato necesaria para llevar a cabo el proceso de ligación. Las uniones más frecuentes son aquellas en las que se establece algún tipo de complementariedad entre las bases de ambos extremos. Estos apareamientos de entre 1 y 4 nucleótidos se denominan microhomologías y no tienen por qué coincidir con los de la secuencia original, pueden ser apareamientos casuales gracias a la nueva configuración que adquieren los extremos tras la rotura. En



**Figura 1.4.1. Representación esquemática del proceso de NHEJ.** Se indican los principales eventos que tienen lugar durante la reparación. Además se muestran los posibles complejos proteicos encargados de llevarlos a cabo. DNA-PKcs (del inglés “**DNA**-dependent protein kinase catalytic subunit”), XRCC4 (del inglés “**X**-ray cross-complementation group 4”) y XLF (del inglés “**X**RCC4-like factor”) también llamada Cernunnos.

ese caso habrá nucleótidos que no aparezcan correctamente y que deban ser eliminados. Durante el proceso de reparación también puede darse la extensión de los extremos por adición de nucleótidos sin necesidad de un molde. Una vez que los extremos están alineados en la configuración definitiva la etapa de ligación permite recuperar la continuidad de la molécula de ADN. Finalmente participan polimerasas y nucleasas rellenando los huecos y eliminando los nucleótidos sobrantes respectivamente. El proceso descrito implica una serie de complicaciones topológicas que merece la pena mencionar. Durante la reparación ambos extremos han de mantenerse próximos evitando su difusión, lo que se conoce como sinapsis. Los extremos han de estar cercanos para que puedan alinearse y establecer microhomologías en el caso de que las haya. Además han de estar accesibles a la acción de las distintas enzimas implicadas en su procesamiento y ligación. Finalmente mencionar que debe existir algún tipo de comunicación entre los procesos que están teniendo lugar en cada uno de los extremos de las cuatro hebras presentes en la sinapsis.

Como se comentó anteriormente, durante el proceso de reparación por NHEJ participan proteínas que modifican los nuevos extremos que se generan como consecuencia de la rotura. Además la aparición de dos DSB próximas puede provocar la pérdida de fragmentos de ADN intermedios. Por tanto la reparación por NHEJ conlleva en muchos casos modificaciones irreversibles de la información genética. Estas alteraciones podrían contribuir al desarrollo de tumores y al deterioro general del organismo que caracteriza al envejecimiento. A pesar de ello es el mecanismo principal en eucariotas multicelulares para reparar DSB (Sonoda et al., 2006). Es funcional durante todo el ciclo celular incluso durante las fases S, G2 y M en las que la reparación por recombinación homóloga muestra una mayor actividad. Este predominio quizá se deba a la gran cantidad de secuencias repetidas presentes en el genoma de los eucariotas multicelulares (aproximadamente el 40% del genoma humano) (Lieber et al., 2003). En un genoma de estas características, la búsqueda de complementariedad en aquellas roturas que se den en regiones ricas en secuencias repetidas puede prolongarse excesivamente. Además, debido a la

abundancia de secuencias repetidas, las regiones complementarias podrían no localizarse en la zona equivalente de la cromátida hermana.

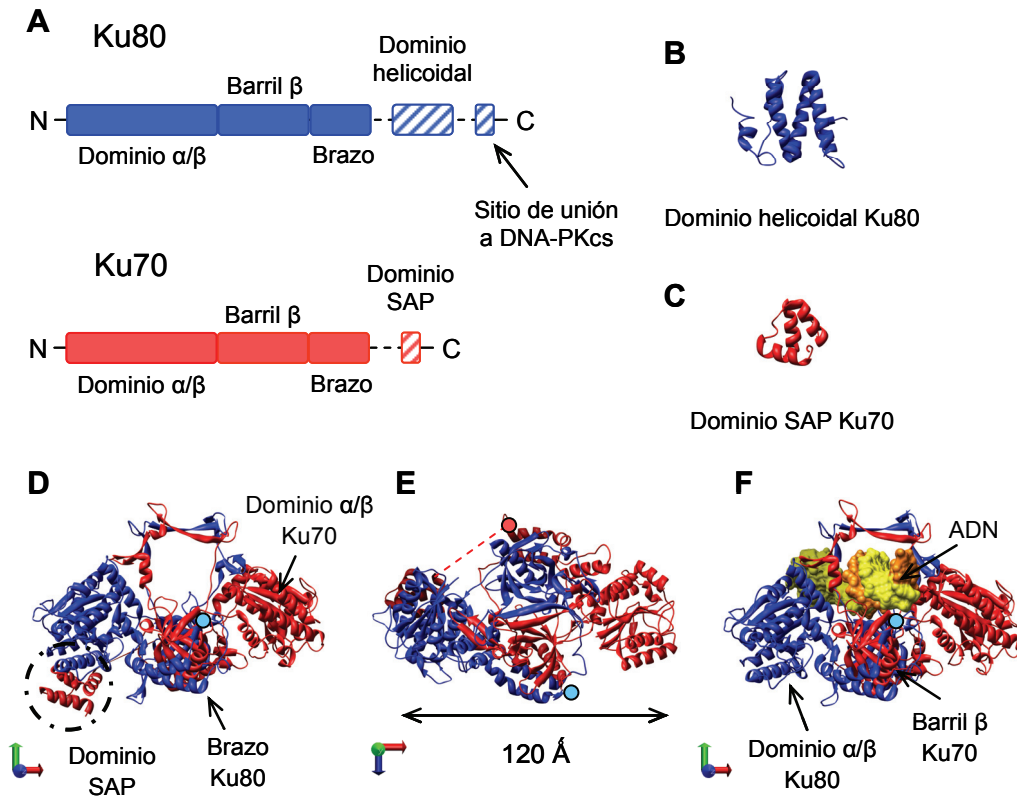
En el caso de los vertebrados los factores principales encargados de llevar a cabo la reparación por NHEJ corresponden a las siguientes proteínas: Ku (Mimori et al., 1981), DNA-PKcs (del inglés “DNA-dependent protein kinase catalytic subunit”) (Dvir et al., 1992; Gottlieb y Jackson, 1993; Chu, 1997), Artemis (Moshous et al., 2001), XRCC4 (del inglés “X-ray cross-complementation group 4”) (Li et al., 1995), XLF (del inglés “XRCC4-like factor”) también llamada Cernunnos (Ahnesorg et al., 2006; Buck et al., 2006), y ADN Ligasa IV (Wei et al., 1995). En muchos procariotas también se ha identificado un sistema de NHEJ más sencillo. En este proceso interviene una proteína homóloga a la proteína Ku que aparece en eucariotas y, además, una ligasa de ADN multifunción (LigD) que puede presentar dominios nucleasa y polimerasa auxiliares (Brissett et al., 2007). Muchas de las bacterias que poseen este sistema de NHEJ se caracterizan por permanecer en un estado no replicativo durante largos períodos de tiempo. Además en el caso de las bacterias disponer de un sistema de reparación mutagénico supone una ventaja añadida en determinadas condiciones de crecimiento *in vivo* (Pitcher et al., 2007). A continuación se describen las proteínas clave del proceso de NHEJ en vertebrados haciendo especial hincapié en aquellas que han sido objeto de estudio durante el desarrollo de esta tesis.

#### 1.4.1 Ku

Esta proteína es una de las más abundantes en el núcleo de las células y lleva a cabo el reconocimiento inicial de los extremos que aparecen como consecuencia de las DSB (Downs y Jackson, 2004). Ku es un heterodímero constituido por dos proteínas que comparten una gran homología y que se denominan Ku70 (69,8 kDa) y Ku80 (82,7 kDa) (Figura 1.4.2). El heterodímero Ku posee una gran afinidad por extremos de ADN que pueden presentar configuraciones muy diversas (Mimori y Hardin, 1986; Arosio et al., 2002). Además la unión a los extremos de ADN induce cambios conformacionales en Ku que permiten reclutar a otros componentes del proceso de reparación por

NHEJ (Lehman et al., 2008). Se ha descrito la interacción de Ku con DNA-PKcs (Hammarsten y Chu, 1998; West et al., 1998), con los complejos DNA-PKcs:Artemis (Drouet et al., 2006) y XRCC4:ADN Ligasa IV (Nick McElhinny et al., 2000), y también con XLF/Cernunnos (Yano et al., 2008) y con polimerasas de la familia pol X asociadas al proceso de reparación por NHEJ (Mahajan et al., 2002; Ma et al., 2004). Por tanto se ha propuesto que Ku podría jugar un papel clave reclutando y estabilizando en el sitio de rotura las distintas actividades enzimáticas que participan en el proceso de NHEJ. Una vez localizado en el extremo Ku puede deslizarse hacia el interior de la molécula de ADN sin consumir energía (de Vries et al., 1989; Yaneva et al., 1997). Además las moléculas de Ku presentes en cada extremo de la rotura interaccionan entre ellas facilitando el correcto alineamiento de las hebras afectadas (Cary et al., 1997; Pang et al., 1997).

En el año 2001 se resolvió la estructura atómica del heterodímero Ku libre (fichero PDB, del inglés “Protein data bank”, 1JEQ) y unido a una molécula de ADN (fichero PDB 1JEY) mediante cristalografía de rayos X (de ahora en adelante “CRX”) (*Figura 1.4.2*) (Walker et al., 2001). Ku70 y Ku80 comparten una organización común que consiste en un dominio  $\alpha/\beta$  amino terminal (de ahora en adelante N-terminal) seguido de un dominio barril- $\beta$  central y, por último, un dominio denominado brazo (*Figura 1.4.2A*). La secuencia correspondiente a los tres dominios comunes de Ku70 y Ku80 es lo que de ahora en adelante se ha denominado núcleo del heterodímero Ku. Por el contrario la región más carboxilo terminal (C-terminal de ahora en adelante) de ambas proteínas ha divergido evolutivamente dando lugar a la aparición de dominios específicos para cada proteína (Aravind y Koonin, 2001; Doherty y Jackson, 2001). La estructura atómica del heterodímero Ku consiste en un anillo preformado en el que se enhebra la molécula de ADN (*Figuras 1.4.2D y 1.4.2E*). Este anillo pseudosimétrico presenta una base más ancha sobre la que descansan aproximadamente dos vueltas de la molécula de ADN enhebrada. La parte superior del anillo sin embargo abraza únicamente los 3 ó 4 pares de bases más centrales respecto a la porción de ADN que apoya sobre la base (de ahora en adelante pares de bases se indica como “pb”). El heterodímero Ku establece contactos únicamente con el esqueleto azúcar-



**Figura 1.4.2. Características estructurales del heterodímero Ku.** (A) Representación esquemática de los dominios que componen Ku80 y Ku70. Las regiones C-terminales divergentes se muestran rayadas. (B) Estructura atómica correspondiente al dominio helicoidal presente en la región C-terminal de Ku80 (fichero PDB 1Q2Z) (Harris et al., 2004). (C) Estructura atómica del dominio SAP (del inglés “SAF-A/B, Acinus and PIAS”) característico de Ku70 (fichero PDB 1JJR) (Zhang et al., 2001). (D) y (E) Vistas frontal y cenital respectivamente de la estructura atómica obtenida para el núcleo del heterodímero Ku. En esta estructura también se pudo resolver el dominio SAP (fichero PDB 1JEQ) (Walker et al., 2001). (F) Vista frontal de la estructura atómica correspondiente al núcleo del heterodímero Ku unido a ADN (fichero PDB 1JEY) (Walker et al., 2001). En todos los casos Ku70 se representa de color rojo y Ku80 de color azul. Los puntos de color azul y rojo que se aprecian en (D), (E) y (F) corresponden a los extremos C-terminales de los dominios brazo de Ku80 y Ku70 respectivamente. A partir de esos puntos emergen las regiones C-terminales divergentes características de Ku80 y Ku70. En (E) se indica mediante una línea discontinua roja la conexión que debe existir entre el dominio brazo de Ku70 y el dominio SAP.

fosfato de la molécula de ADN que se ajustan perfectamente a los contornos de los surcos mayor y menor (Figura 1.4.2F). Este modo de interacción es totalmente compatible con la falta de especificidad de secuencia que caracteriza la interacción entre el heterodímero Ku y la molécula de ADN. La unión al extremo de ADN presenta además una polaridad definida en la que Ku70 se encuentra más próxima al extremo de ADN.



Las estructuras atómicas descritas anteriormente se obtuvieron a partir de una molécula truncada de Ku que carece del dominio C-terminal de Ku80 o CTD (del inglés “**C**-terminal domain”). Esta región abarca los 19 kDa más C-terminales (aminoácidos 545-732) (Walker et al., 2001). El análisis del CTD por resonancia magnética nuclear (de ahora en adelante “RMN”) permitió resolver la estructura atómica de un dominio helicoidal plano que incluye los residuos 592-709 (*Figura 1.4.2B*) (Harris et al., 2004; Zhang et al., 2004). Sin embargo la porción más C-terminal (residuos 720-732) implicada en el reclutamiento de la proteína DNA-PKcs a los sitios de rotura parece estar desordenada. Se ha propuesto que este pequeño fragmento podría adquirir un plegamiento específico como consecuencia de la interacción con DNA-PKcs (Zhang et al., 2004). La región C-terminal específica de Ku70 consiste en un dominio denominado SAP (del inglés “**S**AF-A/B, **A**cinus and **P**IAS”). Este dominio tiene un peso molecular de 5 kDa y su estructura atómica se ha determinado por separado y también formando parte del heterodímero Ku (*Figuras 1.4.2C y 1.4.2D*) (Walker et al., 2001; Zhang et al., 2001). Sin embargo en la estructura atómica correspondiente al complejo formado por Ku y ADN este dominio no se pudo resolver (*Figura 1.4.2F*), por ello se ha propuesto que el dominio SAP sufriría un desplazamiento como consecuencia de la unión de Ku al extremo de ADN (Walker et al., 2001). El dominio SAP se ha definido como un posible dominio de unión a ADN y podría estar implicado en aumentar la afinidad del heterodímero Ku por determinadas secuencias de ADN (Aravind y Koonin, 2000).

Existe por tanto información estructural a resolución atómica para la mayor parte del heterodímero Ku, sin embargo no se ha resuelto la estructura completa de esta proteína. Además tampoco se han determinado los cambios conformacionales que afectan a los dominios más C-terminales de Ku70 y Ku80 como consecuencia de la interacción con los extremos de ADN. Estos cambios conformacionales son esenciales para el subsiguiente reclutamiento del resto de componentes que participan en el proceso de reparación por NHEJ.

### 1.4.2 DNA-PKcs

La proteína DNA-PKcs (469 kDa) es una quinasa que fosforila serinas o treoninas y consiste en una única cadena polipeptídica de 4.128 aminoácidos (*Figura 1.4.3*). La característica principal de DNA-PKcs es que requiere como cofactor para ser activa extremos de ADN. Además su actividad quinasa se ve estimulada especialmente por la presencia de nucleótidos desapareados en el extremo (Leuther et al., 1999; Hammarsten et al., 2000). DNA-PKcs puede unirse directamente a los extremos de ADN si bien, en condiciones fisiológicas, es el heterodímero Ku quien recluta a DNA-PKcs a los sitios donde se produce el daño (Hammarsten y Chu, 1998; West et al., 1998; Falck et al., 2005). La interacción entre Ku y DNA-PKcs es estable cuando tiene lugar sobre extremos de ADN. Esta interacción conlleva el desplazamiento de Ku hacia el interior de la molécula de ADN de forma que DNA-PKcs se localiza más próxima al punto de rotura (Yaneva et al., 1997; Yoo y Dynan, 1999). El extremo de ADN debe tener una longitud mínima de 26 pb para que se dé el ensamblaje simultáneo de estas dos proteínas y la subsiguiente activación de la quinasa. Una longitud menor resulta en una inhibición de DNA-PKcs por parte del heterodímero Ku ya que ambas proteínas compiten por unirse al extremo y se impide la formación de un complejo estable (Hammarsten y Chu, 1998; West et al., 1998). El complejo formado por Ku, DNA-PKcs y el extremo de ADN es lo que se denomina holoenzima DNA-PK o simplemente DNA-PK, y desempeña un papel importante en el reconocimiento inicial y protección de los extremos generados tras la rotura. Además los complejos holoenzima DNA-PK ensamblados en cada extremo interaccionan entre ellos evitando así su difusión y dando lugar a la formación de un complejo sináptico (DeFazio et al., 2002; Weterings et al., 2003; Weterings y van Gent, 2004). El complejo DNA-PK interacciona con otros componentes de la maquinaria de reparación por NHEJ (Chen et al., 2000; Drouet et al., 2006), y podría actuar a modo de andamio facilitando la localización y participando en la regulación de estos componentes adicionales en el sitio donde se produjo la rotura. DNA-PKcs fosforila diversos sustratos entre los que se incluyen XRCC4, Ku70, Ku80 y Artemis (Meek et al., 2004; Collis et al., 2005). Sin embargo en la mayor parte de los casos se desconoce el efecto que tienen esas fosforilaciones en la regulación de las distintas

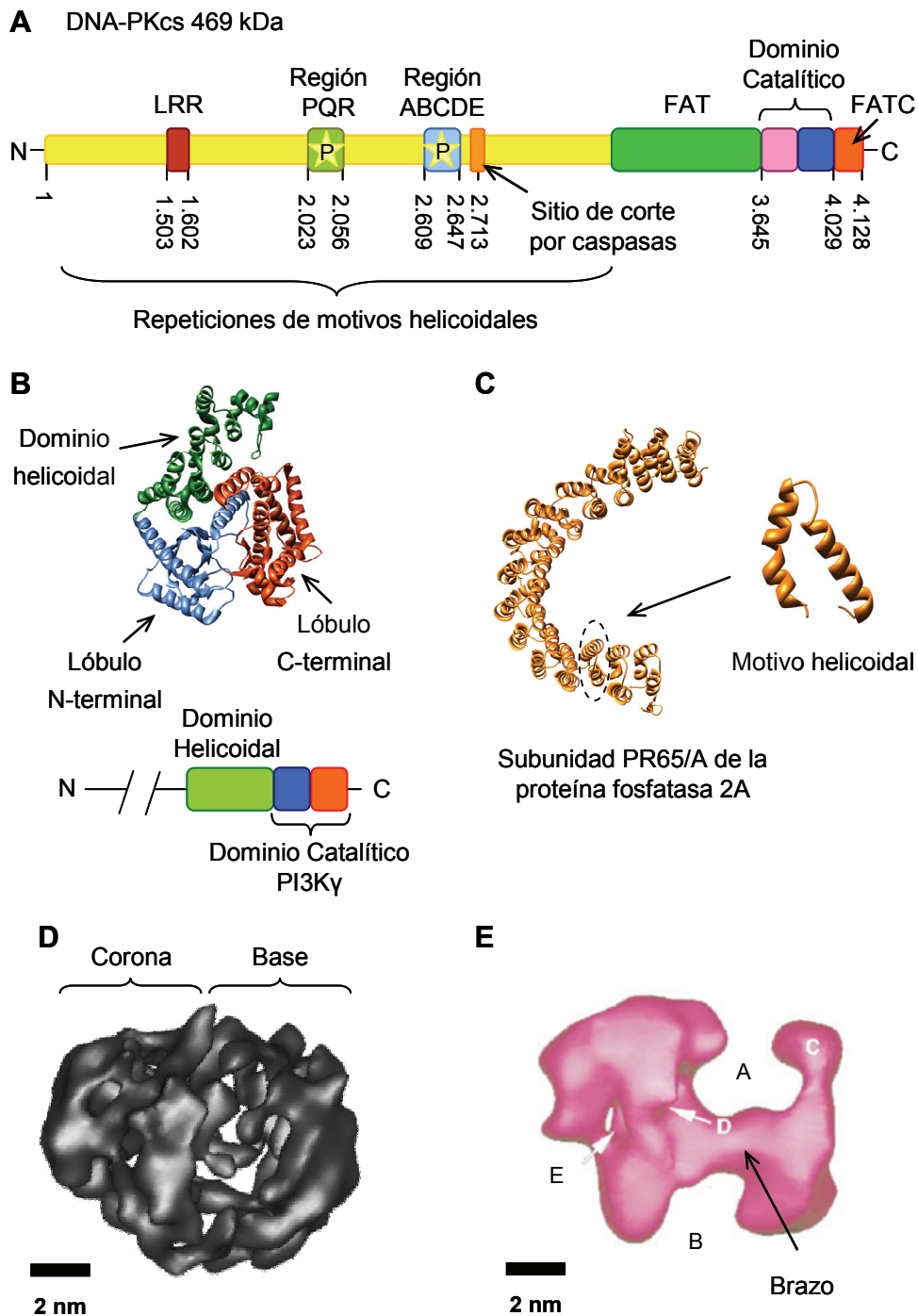
actividades enzimáticas. DNA-PK fosforila también a las histonas H2AX y H1 (Park et al., 2003; Kysela et al., 2005), además se ha visto que la fosforilación de la histona H1 favorece la ligación de nucleosomas reconstituidos *in vitro* (Kysela et al., 2005). El plegamiento de la cromatina en un filamento de 30 nm dificulta la difusión de los extremos que aparecen como consecuencia de la rotura, sin embargo también restringe el acceso de la maquinaria de reparación. Es posible por tanto que el holoenzima DNA-PK altere la conformación local de la cromatina facilitando el acceso de otros complejos implicados en la reparación de la rotura. De hecho la fosforilación de la histona H2X es uno de los primeros eventos que tienen lugar durante la reparación de DSB (van Attikum y Gasser, 2005). En cualquier caso el sustrato principal de DNA-PKcs es ella misma ya que durante la reparación por NHEJ se autofosforila intensamente en *trans* (Meek et al., 2007). Hasta la fecha se han identificado 16 residuos que se autofosforilan y que se localizan principalmente en dos regiones denominadas ABCDE y PQR. Existen sitios de autofosforilación adicionales en la región C-terminal de DNA-PKcs pero su importancia funcional *in vivo* no está bien caracterizada (Douglas et al., 2007). El papel de la autofosforilación durante el proceso de NHEJ no está claro, quizá podría ser una forma de comunicación entre los dos extremos de la rotura. En cualquier caso los distintos eventos de autofosforilación favorecen cambios conformacionales en los complejos DNA-PK y sináptico que permiten el desensamblaje de DNA-PK y, en último término, la correcta reparación de las DSB (Block et al., 2004; Goodarzi et al., 2006; Uematsu et al., 2007).

DNA-PKcs pertenece a la familia de las PIKKs (del inglés “Phosphatidylinositol-3-kinase related protein kinases”) (Hartley et al., 1995). Los miembros de esta familia son transductores que informan a la célula eucariótica sobre asuntos clave tales como el suministro de nutrientes o el estado del genoma y de sus transcritos (Abraham, 2004; Takai et al., 2007). Para ello fosforilan otras proteínas en serinas o treoninas gracias al dominio catalítico homólogo al de la PI3K (del inglés “Phosphatidylinositol-3-kinase”) localizado en la región C-terminal (*Figura 1.4.3A*) (Lavin et al., 1995). Hasta la fecha se han identificado seis PIKKs en mamíferos: ATM (del inglés “Ataxia telangiectasia mutated”), ATR (del inglés “ATM and Rad3 related”), DNA-PKcs,

mTOR (del inglés “**M**ammalian **t**arget **o**f **r**apamycin”), SMG1 (del inglés “**S**uppressor with **m**orphological **e**ffect on **g**enitalia **1**”) y TRRAP (del inglés “**T**ransformation/transcription domain **a**ssociated **p**rotein”) que no posee actividad quinasa. Las PIKKs poseen dos dominios conservados característicos denominados FAT (del inglés “**F**RAP, **A**TM, and **T**RAAP”) y FATC (del inglés “**F**RAP, **A**TM, **T**RAAP and **C**-terminal”) que flanquean el dominio catalítico (*Figura 1.4.3A*) (Keith y Schreiber, 1995; Bosotti et al., 2000). El dominio FATC, fundamental para la actividad catalítica de DNA-PKcs, corresponde a una extensión C-terminal que se dispone a continuación del dominio quinasa. Se ha resuelto la estructura atómica para el dominio FATC de la proteína de levaduras homóloga a mTOR (Dames et al., 2005), y consiste en una  $\alpha$ -hélice seguida de un bucle que contiene un puente disulfuro. Las PI3Ks (del inglés

**Figura 1.4.3. Descripción estructural de DNA-PKcs (página siguiente).** (A)

Representación esquemática de la proteína DNA-PKcs. La región C-terminal incluye los dominios conservados FAT, catalítico y FATC característicos de las PIKKs. La región N-terminal estaría formada principalmente por repeticiones de motivos helicoidales. Dentro de esta región se distinguen las zonas de autofosforilación ABCDE y PQR implicadas en la regulación de DNA-PKcs. Se indica también la localización de una región rica en leucinas (LRR del inglés “**L**eucine rich region”) y el sitio de corte predicho para la casapasa-3. (B) Estructura atómica correspondiente al dominio catalítico de la proteína PI3K $\gamma$  (fichero PDB 1E8X) (Walker et al., 2000). Los lóbulos N-terminal y C-terminal que componen el dominio catalítico se representan en color azul y rojo respectivamente. Se muestra también el dominio helicoidal N-terminal importante para la estabilidad del dominio catalítico. (C) Estructura atómica de la subunidad PR65/A perteneciente a la proteína fosfatasa 2A. Esta subunidad está constituida por repeticiones helicoidales de tipo HEAT que adquieren estructuras terciarias características con forma de tubos aplanados curvos (fichero PDB 1B3U) (Groves et al., 1999). (D) Primera estructura obtenida para DNA-PKcs mediante criomicroscopía electrónica a una resolución aproximada de 21 Å (Chiu et al., 1998). (E) Estructura de DNA-PKcs resuelta mediante microscopía electrónica de cristales bidimensionales teñidos negativamente a una resolución de 22 Å (Leuther et al., 1999). La estructura se representa a un umbral correspondiente al 72% del volumen esperado para una proteína de 469 kDa. DNA-PKcs se compone de una región principal con forma de pirámide invertida que encierra una cavidad accesible al exterior por dos agujeros (se indican como D y E en la figura). A partir de esta región principal emerge una densidad denominada brazo que termina en un dominio aplanado con forma rectangular. El extremo de este dominio (se indica como C en la figura) se conecta con el resto de la región brazo por una zona de baja densidad que podría presentar cierta flexibilidad. La región brazo da lugar a dos canales (se indican como A y B en la figura) con unas dimensiones suficientes para acomodar ADN de cadena doble. Las figuras que se muestran en (D) y (E) se han extraído de Chiu et al., 1998, y Leuther et al., 1999 respectivamente.



“Phosphatidylinositol-3-kinases”) presentan una región inmediatamente N-terminal al dominio catalítico que se compone de varios motivos helicoidales. Esta región se organiza en un dominio helicoidal que proporciona soporte al dominio catalítico y, junto con él, forma el armazón central a partir del cual emergen el resto de dominios (*Figura 1.4.3B*) (Walker et al., 2000; Djordjevic y Driscoll, 2002; Huang et al., 2007). El dominio FAT de las PIKKs abarca una

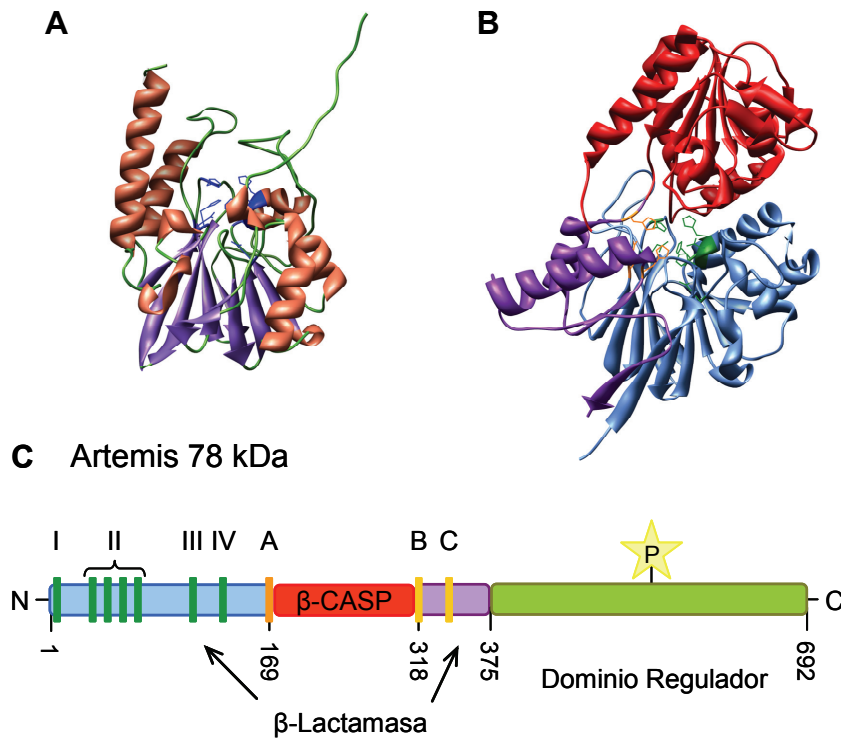
región de aproximadamente 500 aminoácidos inmediatamente N-terminal al dominio catalítico (*Figura 1.4.3A*) (Bosotti et al., 2000). Esta región estaría constituida también por motivos helicoidales repetidos (Perry y Kleckner, 2003; Brewerton et al., 2004), por tanto es posible que el dominio FAT se organice de manera similar al dominio helicoidal de las PI3Ks. Las PIKKs tienen un elevado peso molecular (270-469 kDa), pero la conservación entre los miembros de esta familia se restringe a la región C-terminal correspondiente a los dominios FAT, catalítico y FATC. Por el contrario la región N-terminal de las PIKKs muestra diferencias considerables de tamaño y no presenta homología aparente. Experimentos de análisis de secuencia han permitido determinar que la región N-terminal de todas estas proteínas estaría formada mayoritariamente por regiones ricas en motivos helicoidales (Perry y Kleckner, 2003; Brewerton et al., 2004). Estos motivos se componen de  $\alpha$ -hélices antiparalelas conectadas por un giro pronunciado y pueden ser de tipo HEAT (del inglés “**H**untingtin, **e**longation factor 3, **A** subunit of PP2A, **T**OR1), PFT (del inglés “**P**rotein **f**arnesy transferase”) y ARM (armadillo) (Boguski et al., 1992; Groves y Barford, 1999). La repetición de estos motivos conlleva el apilamiento de sus  $\alpha$ -hélices originando estructuras tubulares aplanadas características cuyo grado de curvatura varía según el motivo helicoidal repetido (*Figura 1.4.3C*) (Groves et al., 1999). Es posible por tanto que las distintas PIKKs compartan una estructura similar a pesar de la baja conservación de secuencia que presentan en la región N-terminal.

El tamaño de DNA-PKcs dificulta su análisis estructural por CRX o RMN pero, por otro lado, lo convierte en un espécimen idóneo para ser analizado mediante microscopía electrónica. En el momento de comenzar esta tesis existían dos estructuras publicadas para la proteína DNA-PKcs humana. La primera de ellas fue resuelta por el grupo de P. L. Stewart en 1998 y se obtuvo utilizando criomicroscopía electrónica (*Figura 1.4.3D*) (Chiu et al., 1998). Esta estructura consistía en dos regiones principales denominadas corona y base separadas por un hueco. En 1999 se publicó otra estructura de DNA-PKcs resuelta mediante microscopía electrónica de cristales bidimensionales teñidos negativamente (*Figura 1.4.3E*) (Leuther et al., 1999). En este caso se distinguían más claramente las regiones características de DNA-PKcs. La

región corona correspondía a una densidad con forma de pirámide invertida que encerraba en su interior una cavidad accesible al exterior por dos agujeros. A partir de esta región emergía una única densidad denominada brazo. El brazo conectaba con un dominio con forma aplanada y rectangular que correspondía a la región base de la estructura resuelta por el grupo de P. L. Stewart. En esta nueva estructura la región base tenía forma aplanada y rectangular (*Figura 1.4.3E*). Recientemente el grupo de P. L. Stewart ha resuelto la estructura de DNA-PKcs a una resolución aproximada de 7 Å utilizando también criomicroscopía electrónica (Williams et al., 2008). Esta estructura se analiza más en detalle en el capítulo correspondiente a Discusión a la vista de los resultados presentados en esta tesis.

### 1.4.3 Artemis

Esta metalo- $\beta$ -lactamasa específica de vertebrados posee actividad exonucleasa 5'→3' *in vitro* sin embargo, en presencia de la proteína DNA-PKcs, se convierte en una endonucleasa que actúa sobre multitud de sustratos (Moshous et al., 2001; Ma et al., 2002). Los diferentes sustratos se caracterizan por presentar transiciones entre regiones de doble cadena y cadena sencilla (Ma et al., 2005c). Dentro de estos sustratos destacan las horquillas que aparecen en los extremos codificantes durante la recombinación V(D)J, por ello se ha propuesto que esta nucleasa puede ser la responsable de abrir este tipo de estructuras *in vivo* (Ma et al., 2002). La actividad endonucleasa de Artemis es dependiente de la interacción con DNA-PKcs y también de la actividad catalítica de esta quinasa. Artemis forma un complejo estable con DNA-PKcs en ausencia del heterodímero Ku y de extremos de ADN en el que DNA-PKcs fosforila eficientemente a Artemis (Ma et al., 2002). Sin embargo, a concentraciones de sal fisiológicas, la formación del complejo DNA-PKcs:Artemis puede requerir del ensamblaje previo del holoenzima DNA-PK (Drouet et al., 2006). A continuación DNA-PKcs se autofosforilaría dando lugar a una reorganización del holoenzima DNA-PK que permite el reclutamiento de Artemis a los sitios de rotura (Goodarzi et al., 2006). En ausencia de ese cambio conformacional las proteínas Ku y Artemis quizás compitan por la unión a DNA-PKcs. Artemis, a diferencia de otros componentes fundamentales del



**Figura 1.4.4. Características estructurales de la familia de metalo-β-lactamasas β-CASP y esquema de la proteína Artemis.** **(A)** Estructura atómica de la metalo-β-lactamasa L1 de *Stenotrophomonas maltophilia* (fichero PDB 1SML) (Ullah et al., 1998). El plegamiento metalo-β-lactamasa consiste en dos hojas β flanqueadas por α-hélices que se disponen a modo de sándwich αβ/βα. Las proteínas de esta familia comparten un patrón característico de histidinas y ácidos aspárticos implicados en la catálisis que se localizan en cinco motivos separados (I-V). Estos motivos se muestran de color azul utilizando la representación de “bolas y palos”. **(B)** Los miembros de la familia β-CASP presentan la particularidad de tener una inserción importante a continuación del motivo IV en la que se han identificado tres residuos característicos (A, B y C) (Callebaut et al., 2002). Recientemente se han resuelto varias estructuras atómicas para diversos miembros de esta familia (Ishikawa et al., 2006; Mandel et al., 2006). En ellas se observa que el residuo C equivale al motivo V característico del plegamiento metalo-β-lactamasa. En **(B)** se muestra la estructura atómica correspondiente a CPSF-73 (del inglés “73 kDa subunit of the cleavage and polyadenylation specificity factor”) (fichero PDB 2I7V) (Mandel et al., 2006). El dominio β-CASP se representa en color rojo y los motivos característicos del plegamiento metalo-β-lactamasa y β-CASP se muestran utilizando la representación de “bolas y palos” de color verde y naranja respectivamente. **(C)** Representación esquemática de la proteína Artemis. Se indican los motivos típicos del plegamiento metalo-β-lactamasa y β-CASP. El código de colores utilizado en la representación es equivalente al que se muestra en **(B)**. El dominio regulador de Artemis es fosforilado (P) eficientemente por DNA-PKcs.

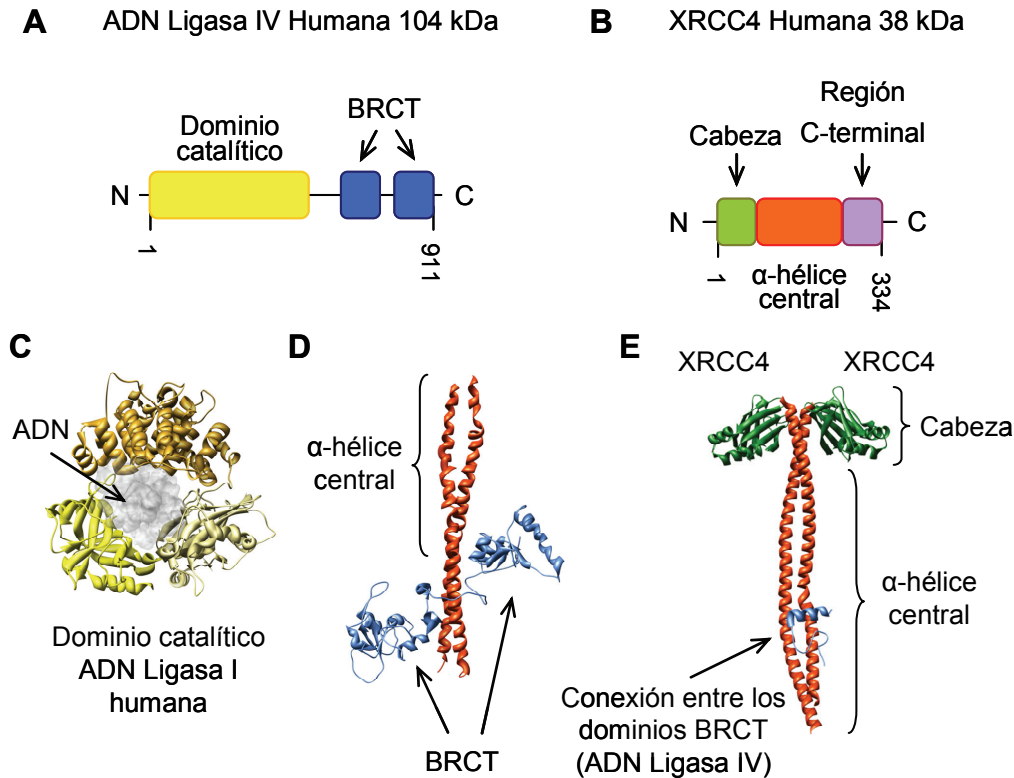


proceso de reparación por NHEJ, se ha propuesto que podría participar en la reparación de determinados tipos de DSB en los que fuese necesario modificar los extremos generados de la rotura (Kurosawa et al., 2007).

La nucleasa Artemis se compone de 692 aminoácidos y pertenece a la familia de metalo- $\beta$ -lactamasas  $\beta$ -CASP (del inglés “Metallo- $\beta$ -lactamase-associated **C**PSF **A**rtemis **S**NM1/**P**SO2”) (*Figuras 1.4.4A y 1.4.4C*) (Callebaut et al., 2002). Los miembros de esta familia comparten la característica de utilizar como sustratos ácidos nucleicos y, a nivel estructural, presentan la particularidad de tener un dominio  $\beta$ -CASP insertado entre los motivos 4 y 5 típicos del plegamiento metalo- $\beta$ -lactamasa (*Figura 1.4.4B*) (Ishikawa et al., 2006; Mandel et al., 2006). La mitad C-terminal de Artemis no se parece a ninguna otra proteína conocida. Esta región presenta una fosforilación basal dependiente de ATM (Goodarzi et al., 2006) e *in vitro* es fosforilada eficientemente por DNA-PKcs (Ma et al., 2005b; Goodarzi et al., 2006). Se ha propuesto que la mitad C-terminal regula negativamente la actividad endonucleasa de Artemis y que la fosforilación dependiente de DNA-PKcs libera esta inhibición (Ma et al., 2005b).

#### 1.4.4 XRCC4:ADN Ligasa IV

Este complejo es clave en el proceso de reparación por NHEJ puesto que lleva a cabo la etapa final de ligación permitiendo restaurar la continuidad de la molécula de ADN (Frank et al., 1998). Es un complejo específico de este proceso de reparación y se caracteriza por su versatilidad ya que puede ligar extremos que presenten configuraciones muy diversas (Ma et al., 2004; Gu et al., 2007). La proteína ADN Ligasa IV (104 kDa) se compone de un dominio ADN ligasa canónico seguido de dos dominios BRCT (del inglés “**B**reast **C**ancer gene 1 **c**arboxy **t**erminal”) C-terminales uno a continuación del otro (*Figuras 1.4.5A, 1.4.5C y 1.4.5D*) (Tomkinson et al., 2006). Cada molécula de XRCC4 (38 kDa) se organiza en un dominio globular N-terminal seguido de una  $\alpha$ -hélice de aproximadamente 120 Å de longitud (*Figuras 1.4.5B y 1.4.5E*) (Junop et al., 2000; Sibanda et al., 2001). No se ha resuelto la estructura atómica de la porción más C-terminal de XRCC4, sin embargo experimentos de



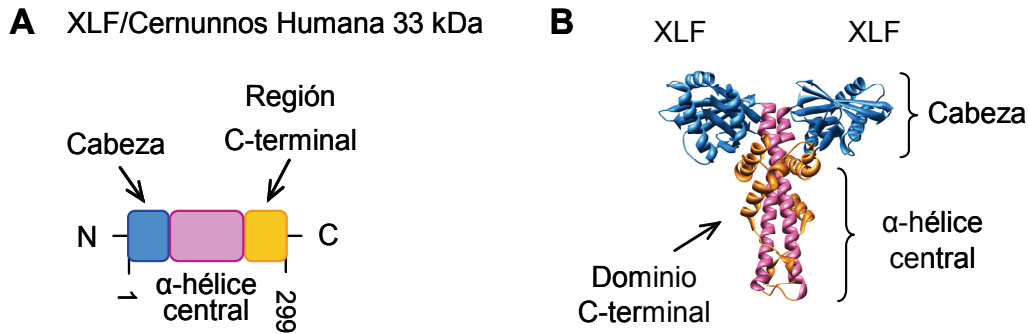
**Figura 1.4.5. Descripción estructural de los componentes que forman el complejo XRCC4:ADN Ligasa IV.** (A) y (B) Representaciones esquemáticas de las proteínas ADN Ligasa IV y XRCC4 respectivamente. (C) Estructura atómica del complejo formado por una molécula de ADN y el dominio catalítico de ADN Ligasa I (archivo PDB 1X9N) (Pascal et al., 2004) homólogo al de la proteína ADN Ligasa IV. (D) Estructura atómica del complejo formado por Lif1p (del inglés “Ligase interacting factor 1”) y los dominios BRCT de la proteína Lig4p (archivo PDB 1Z56) (Dore et al., 2006). Lif1p y Lig4p son los ortólogos de XRCC4 y ADN Ligasa IV respectivamente en *Saccharomyces cerevisiae*. (E) Estructura atómica del complejo humano formado por dos moléculas de XRCC4 truncadas y la región de la proteína ADN Ligasa IV que conecta los dos dominios BRCT (archivo PDB 1IK9) (Sibanda et al., 2001). No se ha resuelto la estructura de la región C-terminal de XRCC4, sin embargo experimentos de análisis de secuencia predicen que podría organizarse en un dominio globular. El código de colores utilizado para representar los dominios de las diferentes estructuras atómicas coincide con el de las representaciones esquemáticas de cada proteína.

análisis de secuencia sugieren que podría plegarse en un dominio independiente. XRCC4 forma homotetrámeros y homodímeros aunque es en esta última configuración en la que se genera un complejo estable con la proteína ADN Ligasa IV (Robins y Lindahl, 1996; Mizuta et al., 1997; Modesti et al., 1999). El ensamblaje de homodímeros conlleva la interacción de dos moléculas XRCC4 por sus respectivas  $\alpha$ -hélices centrales dando lugar a una superhélice (Figura 1.4.5E). La formación de esta superhélice implica la

aparición de un nuevo motivo que interacciona específicamente con la región de la proteína ADN Ligasa IV que une los dos dominios BRCT (*Figuras 1.4.5D y 1.4.5E*) (Sibanda et al., 2001; Dore et al., 2006). XRCC4 se asocia fuertemente a la proteína ADN Ligasa IV estabilizándola y estimulando su actividad tanto *in vitro* como *in vivo* y también su adenilación (Grawunder et al., 1997; Modesti et al., 1999). El complejo XRCC4:ADN Ligasa IV puede interaccionar con ADN pero durante el proceso de NHEJ sería reclutado por otros factores a los sitios de rotura. Estos factores podrían ser Ku y DNA-PKcs puesto que se ha descrito la interacción entre estas proteínas y el complejo XRCC4:ADN Ligasa IV (Chen et al., 2000; Nick McElhinny et al., 2000).

#### 1.4.5 XLF o Cernunnos

La búsqueda de proteínas capaces de interaccionar con el complejo XRCC4:ADN Ligasa IV por un lado, y la caracterización de genes mutados en pacientes inmunodeficientes por otro lado, llevaron al descubrimiento de este nuevo componente del proceso de reparación por NHEJ (Ahnesorg et al., 2006; Buck et al., 2006). Recientemente se ha resuelto la estructura atómica del complejo homodimérico que forma la proteína humana XLF (*Figuras 1.4.6A y 1.4.6B*) (Andres et al., 2007; Li et al., 2008). Esta estructura se compone de una región N-terminal globular desde la que emerge una  $\alpha$ -hélice implicada en la dimerización con otra molécula de XLF. Es por tanto una organización similar a la que se observa en XRCC4. Sin embargo en el caso de XLF la  $\alpha$ -hélice que emerge del dominio N-terminal globular no forma una superhélice, si no que se dobla de nuevo hacia el dominio N-terminal dando lugar a una estructura más compacta (*Figura 1.4.6B*). XLF al igual que XRCC4 se encuentra formando dímeros y su capacidad de unión a ADN depende de la existencia de regiones largas y accesibles en los extremos (más de 80 pb) (Hentges et al., 2006; Lu et al., 2007). En presencia del heterodímero Ku XLF interacciona con el complejo XRCC4:ADN Ligasa IV estimulando su capacidad de ligar extremos de ADN que no aparecen perfectamente (Tsai et al., 2007). Se desconoce la manera en que XLF ejerce esta función, aunque se sabe que los dominios amino y carboxilo terminales de esta proteína interaccionan con XRCC4 y el ADN respectivamente (Andres et al., 2007).



**Figura 1.4.6. Descripción estructural de la proteína XLF/Cernunnos.** (A) Representación simplificada de los dominios presentes en la proteína XLF/Cernunnos. (B) Estructura atómica del homodímero de XLF/Cernunnos (fichero PDB 2QM4) (Li et al., 2008). El código de colores utilizado para la representación esquemática de XLF/Cernunnos coincide con el que se muestra en la estructura atómica.

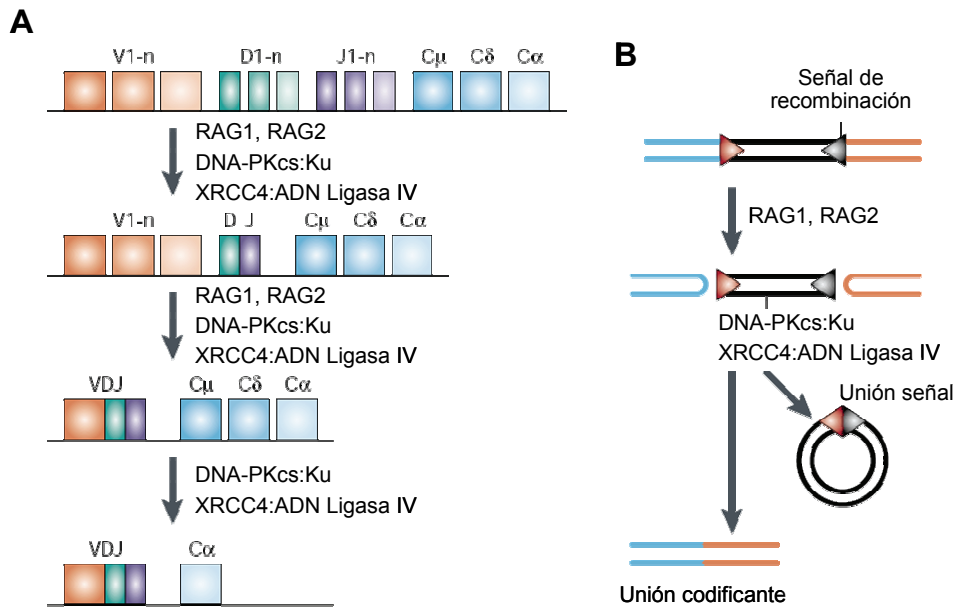
#### 1.4.6 Componentes adicionales del proceso de reparación por NHEJ

En el proceso de reparación de DSB por NHEJ participan además otros componentes adicionales algunos de los cuales se describen a continuación. La proteína del síndrome de Werner (WRN) es una helicasa de tipo RecQ cuya actividad es dependiente de ATP (Hickson, 2003). Presenta además actividad exonucleasa 3'→5', es fosforilada por DNA-PK y podría eliminar grupos fosfato presentes en el extremo 3' de la rotura (Yannone et al., 2001). PNK (del inglés "Polynucleotide kinase") también tienen capacidad de eliminar grupos fosfato del extremo 3'. Participa por tanto en la restauración de los extremos a la configuración 3' OH y 5' fosfato necesaria para que tenga lugar la ligación (Chappell et al., 2002; Bernstein et al., 2005). La actividad de polimerización asociada al proceso de reparación por NHEJ la llevan a cabo las polimerasas pol  $\mu$  y pol  $\lambda$  pertenecientes a la familia pol X (Ma et al., 2004). Estas dos polimerasas rellenan los huecos presentes en la molécula de ADN en regiones próximas al sitio donde se produjo la rotura. La polimerización se lleva a cabo utilizando un molde aunque en el caso de pol  $\mu$  también puede ser de manera independiente. Además pol  $\mu$  y pol  $\lambda$  son propensas a introducir repeticiones como consecuencia de los "patinazos" que tienen lugar durante la polimerización (Nick McElhinny et al., 2005; Moon et al., 2007). La adición de nucleótidos aleatoriamente podría ser beneficiosa puesto que favorecería la

aparición de microhomologías en los casos en los que no se den espontáneamente tras la rotura.

#### 1.4.7 Efectos de las alteraciones en los componentes del proceso de NHEJ

La capacidad mutagénica del proceso de NHEJ es aprovechada por el sistema inmune de los vertebrados para proporcionar una respuesta específica y eficaz frente a cada antígeno. Las características de esta respuesta dependen en gran medida de procesos concretos que tienen lugar en los linfocitos tales como la recombinación V(D)J, la hipermutación somática y el cambio de clase que se da en la cadena pesada de las inmunoglobulinas. Estos procesos conllevan la introducción de DSB de manera controlada y regulada y en ellos, además de proteínas características de linfocitos, participan también los distintos componentes del proceso de NHEJ que están presentes en todas las células eucariotas (*Figura 1.4.7*) (Gellert, 2002). Alteraciones en esos componentes generan fenotipos inmunodeficientes o SCID (del inglés “**S**evere **c**ombined **i**mmune **d**eficiency”) acompañados de una mayor sensibilidad a la radiación ionizante (fenotipo RS-SCID del inglés “**R**adiosensitive-**SCID**”) (Revy et al., 2005). Estos fenotipos han sido fundamentales para identificar los factores clave que llevan a cabo la reparación por NHEJ (Schwarz et al., 2003). Recientemente la caracterización de pacientes RS-SCID que presentaban además retardo en el crecimiento, microcefalia y ligación de extremos de ADN defectuosa permitió identificar la proteína XLF/Cernunnos (Ahnesorg et al., 2006; Buck et al., 2006). Aproximadamente el 15% de los fenotipos RS-SCID en humanos se asocian a defectos en componentes del proceso de NHEJ que, en la mayoría de los casos, corresponden a una deficiencia en la proteína Artemis (Moshous et al., 2001; Lieber et al., 2003). Por el momento no se han descrito pacientes deficientes en Ku o DNA-PKcs, quizás la ausencia de cualquiera de estas dos proteínas sea incompatible con la vida. A partir del análisis de distintas líneas celulares se ha visto que la pérdida del heterodímero Ku impide la formación de



**Figura 1.4.7. Esquema del proceso de recombinación V(D)J.** (A) Las regiones variables de las inmunoglobulinas y del TCR son codificadas a partir de un exón que está constituido por los segmentos génicos V, D y J gracias al proceso de recombinación V(D)J. Este exón no existe como tal en el genoma de las células y se genera a partir de recombinaciones específicas que tienen lugar durante el proceso de maduración de los linfocitos. Los distintos procesos de recombinación los llevan a cabo proteínas específicas de linfocitos como las proteínas RAG1 y RAG2 (del inglés “**R**ecombinase-**a**ctivating **g**enes”), y también componentes del sistema de NHEJ. (B) Los segmentos génicos están delimitados por secuencias conservadas denominadas señales de recombinación (RSS del inglés “**R**ecombination-**s**ignal **s**equence”). Estas señales se representan mediante triángulos en (B). El complejo RAG (RAG1 y RAG2) se une a las RSSs (del inglés “**R**ecombination-**s**ignal **s**equences”) adyacentes de los dos segmentos génicos que van a recombinar y, a continuación, introduce un corte de cadena sencilla entre el segmento génico y la señal RSS adyacente. Los extremos 3’ OH generados reaccionan con la hebra complementaria dando lugar a una rotura de doble cadena. Se forman por tanto dos extremos codificantes adyacentes a los segmentos génicos y dos extremos señal adyacentes a las RSSs. Los extremos codificantes se caracterizan por estar cerrados covalentemente. Los extremos señal pueden unirse directamente dando lugar a uniones señal, sin embargo los extremos codificantes sufren un procesamiento adicional que consiste en la apertura de la horquilla terminal. Después puede darse la eliminación o adición de nucleótidos sin la necesidad de un molde en las regiones de cadena sencilla que sobresalen del extremo tras la apertura de la horquilla. Por último los extremos codificantes se ligan dando lugar a uniones codificantes. Figura adaptada de van Gent et al., 2001.

uniones señal y codificantes durante la recombinación V(D)J (Featherstone y Jackson, 1999). Sin embargo la falta de DNA-PKcs o Artemis afecta principalmente a la formación de uniones codificantes (Rooney et al., 2002). Esta observación es compatible con el papel propuesto para el complejo DNA-PKcs:Artemis consistente en llevar a cabo la apertura de las horquillas presentes en los extremos codificantes. Además se ha visto que células deficientes en DNA-PKcs o Artemis son significativamente menos sensibles a la radiación ionizante que células que no expresan Ku o XRCC4 (Gao et al., 1998a; Rooney et al., 2002; Rooney et al., 2003). Esto sugiere que DNA-PKcs y Artemis participan en la reparación de determinados tipos de DSB mientras que otros componentes del proceso de NHEJ como Ku, XRCC4 y ADN Ligasa IV serían claves en la reparación de la mayoría de DSB. También se han descrito pacientes que presentan mutaciones en la proteína ADN Ligasa IV que no impiden por completo su actividad (mutaciones hipomórficas) (O'Driscoll et al., 2001). En ratones se ha visto que los mutantes que no expresan XRCC4 o ADN Ligasa IV, a diferencia de lo que sucede con aquellos que son deficientes en DNA-PKcs o Ku, son inviables como consecuencia de la masiva muerte neuronal por apoptosis que tiene lugar en el embrión (Frank et al., 1998; Gao et al., 1998b). Posiblemente aquellas mutaciones que impidan por completo la expresión de componentes clave del proceso de NHEJ conlleven fenotipos severos que dificulten la viabilidad del organismo. Por tanto las variaciones en la eficiencia del proceso de NHEJ podrían deberse principalmente a mutaciones hipomórficas, o bien a la presencia de determinados alelos para los distintos componentes de este proceso. Estas pequeñas alteraciones quizá favorezcan el desarrollo de tumores. De hecho pacientes con mutaciones hipomórficas en la proteína Artemis tienen una mayor tendencia a desarrollar linfomas en el timo (Moshous et al., 2003; Woo et al., 2007). En ratones se ha visto que la pérdida de componentes clave del proceso de NHEJ junto con la de p53 favorece la inestabilidad genómica y la aparición de tumores y, por tanto, estos componentes podrían considerarse como guardianes del genoma (Difilippantonio et al., 2000; Gao et al., 2000). Sin embargo la presencia de un sistema de NHEJ defectuoso en células tumorales puede suponer una importante diana terapéutica; aquellos fármacos que permitan inhibir este

proceso aumentarán la sensibilidad de las células tumorales a los tratamientos que generan DSB (Kim et al., 2002).

Paradójicamente se ha visto que componentes del proceso de NHEJ participan también en el mantenimiento de la estabilidad telomérica (Riha et al., 2006). DNA-PKcs y Ku se localizan en los telómeros donde juegan un papel clave en la regulación de su longitud y en la protección de los extremos de los cromosomas (Bailey et al., 1999; Espejel et al., 2004). Estudios citogenéticos muestran que células deficientes en cualquiera de estas dos proteínas presentan una alta incidencia de fusiones cromosómicas (Espejel et al., 2002a). Se ha visto que tanto Ku como DNA-PKcs interactúan con la telomerasa si bien en el caso de esta última la interacción podría estar mediada por la proteína KIP (del inglés “Kinase interacting protein”) (Wu y Lieber, 1997; Chai et al., 2002; Espejel et al., 2002b; Lee et al., 2004). Ku y DNA-PKcs podrían intervenir de manera independiente en la formación de la estructura protectora característica que aparece en los extremos de los cromosomas. Se ha propuesto que la formación de esta estructura conlleva una respuesta de NHEJ incompleta debida al reconocimiento transitorio de los extremos del ADN como DSB (Riha et al., 2006).

#### **1.4.8 Conclusión**

El proceso de NHEJ es un mecanismo de reparación que se caracteriza por su flexibilidad, lo cual le permite hacer frente a las diversas configuraciones que adquieren los extremos tras la rotura. Esa característica es consecuencia de la versatilidad de las enzimas que participan, pero también de la variabilidad en cuanto al orden en el que éstas intervienen durante la reparación. La configuración de los extremos generados condiciona las actividades enzimáticas implicadas y por ello, en determinados casos, la reparación puede consistir en un único paso de ligación. Por el contrario en roturas más complejas será necesaria la intervención iterativa del conjunto de actividades enzimáticas características del proceso de NHEJ. Esa flexibilidad permite en último término recuperar la continuidad de la molécula de ADN, pero es



también responsable de la diversidad de posibles soluciones que se pueden generar a partir de un único evento de reparación.

## 1.5 Microscopía electrónica y procesamiento digital de partículas individuales

### 1.5.1 Introducción al microscopio electrónico y generación de contraste

El primer microscopio electrónico de transmisión fue desarrollado en 1931 por Ernst Ruska, lo que le valió el premio Nobel de Física compartido en 1986. Los principios básicos que rigen el funcionamiento del microscopio electrónico de transmisión (de ahora en adelante microscopio electrónico) son análogos a los de un microscopio óptico convencional. En el caso del microscopio electrónico se utilizan electrones como fuente de iluminación que se generan a partir de un cátodo con un voltaje de aceleración determinado. Los electrones interactúan con la muestra y a continuación, gracias a bobinas electromagnéticas que aprovechan la carga de los electrones, forman una imagen del espécimen a mayor magnificación. Las bobinas electromagnéticas actúan de manera equivalente a las lentes de vidrio de un microscopio óptico convencional focalizando el haz de electrones (Bozzola, 1992).

Los electrones tienen una longitud de onda considerablemente menor que la luz visible y por tanto permiten alcanzar una resolución subatómica. Electrones acelerados a un voltaje de 100 kV poseen una capacidad de resolución teórica de 0,002 nm. En la práctica y en el caso de muestras biológicas la resolución que se obtiene para voltajes de aceleración similares es significativamente menor. Estas limitaciones se deben a la propia naturaleza del espécimen y también a los diferentes tipos de aberraciones (astigmatismo, aberración cromática y aberración esférica) presentes en las lentes del microscopio electrónico (Frank, 2006).

Los electrones al interactuar con la muestra son dispersados dando lugar a la aparición del contraste en la imagen. Las interacciones que se establecen

pueden ser de tipo inelástico o elástico. Las interacciones de tipo inelástico conllevan intercambio de energía entre el espécimen y los electrones. Las interacciones de tipo elástico son las principales responsables de la generación del contraste que se aprecia en las imágenes de microscopía electrónica. El contraste puede ser de dos tipos, amplitud o fase, y su importancia relativa varía según la técnica escogida para preparar la muestra (Frank, 2006).

### **1.5.2 Formación de la imagen en el microscopio electrónico**

Las imágenes que se obtienen en el microscopio electrónico equivalen a imágenes de proyección de un objeto tridimensional en el plano perpendicular a la dirección del haz. Esto se debe a que la profundidad de campo del microscopio electrónico es significativamente mayor que el grosor de las muestras que se observan. Por tanto en las imágenes de microscopía electrónica la información tridimensional del espécimen queda compactada en dos dimensiones. La interpretación directa de este tipo de imágenes no es evidente y por ello, para recuperar la información tridimensional presente en las imágenes, es necesario recurrir al procesamiento digital de partículas individuales. Las imágenes de microscopía electrónica pueden representarse de manera alternativa calculando la transformada de Fourier de la imagen, lo cual permite evaluar fácilmente las posibles alteraciones presentes en la imagen. Este tipo de representación utiliza funciones matemáticas sinusoidales sencillas de diferentes frecuencias. Las frecuencias de las funciones utilizadas para calcular la transformada de Fourier de una imagen se denominan frecuencias espaciales. De manera muy simplificada se puede decir que las funciones de mayor frecuencia corresponden al detalle de alta resolución presente en la imagen de microscopía. Por el contrario las funciones de baja frecuencia correlacionan con la información de baja y media resolución (Frank, 2006).

### **1.5.3 Corrección de CTF**

El proceso de formación de la imagen es un proceso complejo que está modulado además por las propias características del microscopio electrónico.

La coherencia del haz de electrones así como los distintos tipos de aberraciones presentes en las lentes electromagnéticas afectan al proceso de formación de la imagen. Es necesario corregir estas alteraciones de manera que las imágenes obtenidas representen lo más fielmente posible imágenes de proyección del espécimen. La función de transferencia de contraste o CTF (del inglés “**C**ontrast **t**ransfer **f**unction”) describe las alteraciones presentes en las imágenes de microscopía electrónica para cada frecuencia espacial. Estas alteraciones son de dos tipos (van Heel et al., 2000; Frank, 2006):

**a)** En las imágenes que se obtienen en el microscopio electrónico la información de media y baja resolución del espécimen está mejor representada que la información de alta resolución. El decaimiento de la transmisión de la información en función de la frecuencia espacial se describe mediante la función envolvente de la CTF. Este efecto depende principalmente de la aberración cromática de las lentes electromagnéticas y de la coherencia de los electrones emitidos por el filamento. Cuanto más coherente es la fuente de electrones más homogénea es la transferencia de información en la imagen del objeto. Por eso para el análisis estructural de alta resolución mediante microscopía electrónica se utilizan microscopios de cañón de emisión de campo o FEG (del inglés “**F**ield **e**mission **g**un”). El aumento de la coherencia del haz de electrones en este tipo de microscopios mejora la transmisión de la información de alta resolución (Mouche et al., 1999).

**b)** El contraste no es uniforme para las distintas frecuencias espaciales. La información presente en la imagen correspondiente a determinadas frecuencias espaciales sufre una inversión del contraste. Este efecto provoca cambios en la distribución de masa de la estructura tridimensional resuelta y por ello es necesario llevar a cabo su corrección.

La estimación de la CTF es el requisito previo para corregir las alteraciones descritas anteriormente. Esta corrección es especialmente importante en imágenes obtenidas a partir de muestras preparadas utilizando la técnica de criomicroscopía electrónica, puesto que contienen información del espécimen de alta resolución. La corrección más sencilla consiste en restaurar el contraste

apropiado para aquellas frecuencias que han sufrido una inversión. Existe un paso adicional de corrección que permite aumentar la señal en la imagen correspondiente a las frecuencias espaciales atenuadas. Este tipo de corrección, si no se realiza adecuadamente, conlleva el riesgo de introducir artefactos como consecuencia de una amplificación desproporcionada del ruido asociado a las frecuencias espaciales atenuadas.

#### 1.5.4 Técnicas de preparación del espécimen

El interior del microscopio electrónico se encuentra a alto vacío ( $10^{-4}$ - $10^{-7}$  Pa), de esta manera los electrones pueden avanzar sin colisionar con ninguna otra molécula que no pertenezca al espécimen. Esta característica condiciona el modo de preparar la muestra para su observación posterior. Las diferentes técnicas de preparación han de evitar que el espécimen colapse cuando se introduce en el microscopio electrónico y, en la medida de lo posible, han de protegerlo del daño causado por los electrones. La preparación de muestras biológicas para su análisis por microscopía electrónica y análisis de partículas individuales se lleva a cabo mediante dos técnicas principales (*Figura 1.5.1*) (Renau Piqueras, 1998).

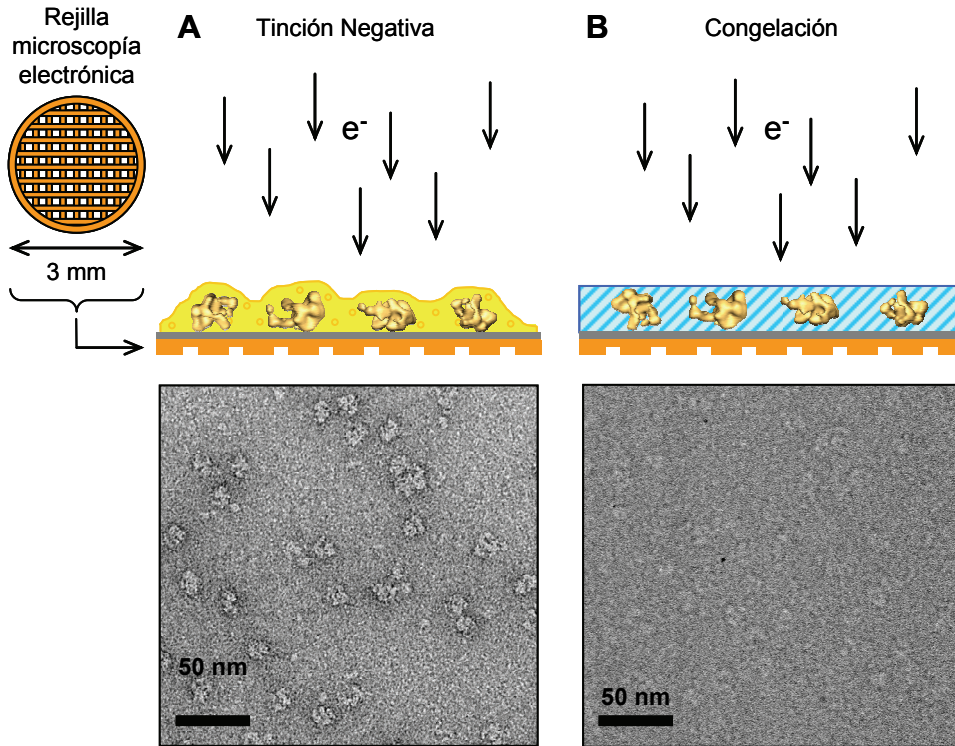
##### Tinción negativa

Esta técnica comenzó a utilizarse en los años 50 para el estudio de virus, y consiste en remplazar la solución en la que se encuentra embebido el espécimen por una sal de un metal pesado (acetato de uranilo por ejemplo). A continuación la muestra se seca y la sal del metal pesado forma una matriz amorfa que encierra y protege la muestra (*Figura 1.5.1A*). Las partículas que se observan en micrografías de tinción negativa corresponden a proyecciones de los huecos que el espécimen deja en la matriz amorfa, de ahí el nombre de tinción negativa. Los átomos pesados van a provocar una alta dispersión de los electrones que justifica la diferencia de contraste tan marcada entre las zonas de matriz y los huecos que alojan al espécimen (*Figura 1.5.1A*). El componente mayoritario del contraste introducido por esta técnica corresponde a contraste de amplitud. Es una técnica sencilla, rápida y que puede aplicarse a

especímenes de muy diverso tamaño, desde pequeñas proteínas a virus. Sin embargo la resolución que permite alcanzar está limitada por el tamaño del grano de la sal utilizada en la tinción. Esa resolución está comprendida normalmente entre 1,5 y 2 nm.

### Congelación y criomicroscopía electrónica

Esta técnica fue desarrollada por J. Dubochet y colaboradores durante los años 80 (Adrian et al., 1984; Dubochet et al., 1988). Consiste en la congelación de la muestra a la temperatura del nitrógeno líquido (en torno a  $-180^{\circ}\text{C}$ ) a alta velocidad. Gracias al modo en que se lleva a cabo la congelación se consigue que el agua presente en la muestra forme hielo amorfo y, de esta manera, se evita el daño irreversible que provocaría la formación de cristales de hielo en la muestra (*Figura 1.5.1B*). Se obtienen por tanto imágenes de proyección del espécimen hidratado en su conformación nativa. Esta técnica conlleva ciertas dificultades técnicas puesto que la muestra una vez congelada ha de mantenerse en ese estado incluso durante su observación en el microscopio electrónico. El problema principal en la observación de este tipo de muestras radica en la baja diferencia de contraste entre el espécimen y el medio en el que está embebido. Esto se debe a que los átomos que componen la materia orgánica así como el tampón son de peso molecular parecido y, por tanto, también es parecida la manera en que dispersan los electrones (*Figura 1.5.1B*). La preparación y posterior observación de muestras congeladas en el microscopio electrónico es lo que se denomina criomicroscopía electrónica. El contraste al observar muestras congeladas por criomicroscopía electrónica se consigue mediante variaciones en el grado de foco durante la toma de micrografías electrónicas. De esta manera se incrementa el contraste de fase que es el componente principal del contraste en criomicroscopía electrónica. Existe otra técnica denominada criotinción en la que se añade un agente de tinción durante el proceso de congelación y así se favorece el contraste de amplitud. La ventaja principal de la criomicroscopía electrónica consiste en que la integridad estructural de la muestra se preserva a resolución atómica y por tanto permite resolver estructuras a esa resolución.



**Figura 1.5.1. Técnicas de preparación de muestras para su análisis posterior mediante microscopía electrónica y procesamiento digital de imágenes individuales.** (A) En el caso de tinción negativa el espécimen queda embebido en una matriz amorfa correspondiente al agente de tinción deshidratado. (B) Cuando la muestra se congela las moléculas individuales del espécimen quedan atrapadas en una fina capa de hielo amorfo que preserva su conformación nativa. Las diferencias de contraste al comparar micrografías electrónicas de un mismo espécimen obtenidas utilizando una u otra técnica son evidentes.

### 1.5.5 Toma de micrografías

Utilizando cualquiera de las técnicas descritas anteriormente es posible observar el espécimen en el interior del microscopio electrónico. Sin embargo estas técnicas protegen de forma limitada del daño causado por la radiación electrónica como consecuencia de interacciones de tipo inelástico. Por tanto la toma de micrografías se debe hacer utilizando la técnica denominada “mínima dosis” en la cual el área de interés se expone a los electrones una sola vez a alta magnificación (50.000 aumentos) (Williams y Fisher, 1970). En ese momento el espécimen sufre una dosis de electrones igual o menor a  $10 \text{ e}^-/\text{Å}^2$  y, de esta manera, se consigue preservar la mayor cantidad de información estructural posible. Este requisito impone una característica principal en las

imágenes de proyección que se utilizan durante el análisis de partículas individuales: la baja relación señal-ruido presente en la imagen. En las proyecciones que se observan en las micrografías electrónicas, la relación entre la intensidad de la señal correspondiente al espécimen y el ruido de fondo presente en la micrografía es baja. Esto es especialmente dramático cuando la muestra se prepara mediante criomicroscopía electrónica. Esta característica condiciona todo el análisis posterior de las imágenes de proyección y supone, además, una dificultad inherente al proceso de resolución de estructuras mediante microscopía electrónica y procesamiento digital de partículas individuales.

Las micrografías contienen por tanto proyecciones de moléculas individuales del espécimen. Al preparar la muestra para su observación en el microscopio electrónico es frecuente que las moléculas del espécimen queden atrapadas en orientaciones diferentes. Según esto las imágenes de proyección o partículas individuales que se utilizan en el análisis tridimensional posterior representan las diferentes orientaciones del espécimen en la muestra. Estas imágenes han de corresponder a un espécimen homogéneo tanto desde un punto de vista conformacional como bioquímico, puesto que en la mayor parte de los métodos de procesamiento se asume que la mayoría de las partículas individuales representan una única estructura.

### **1.5.6 Digitalización**

Las imágenes que se obtienen en el microscopio electrónico se registran principalmente en formato digital directamente utilizando una CCD (del inglés “**C**harged **c**oupled **d**evice”), o bien en películas fotográficas. En este último caso es necesario llevar a cabo un paso previo de digitalización para poder realizar el procesamiento digital de partículas individuales. La digitalización implica convertir la información analógica presente en la micrografía en una matriz bidimensional finita de píxeles. En esa matriz cada píxel tiene un valor de gris determinado que se relaciona linealmente con los niveles de grises de la micrografía. Por tanto los valores de grises de los distintos píxeles reflejan las diferencias de masa presentes en las imágenes de proyección del

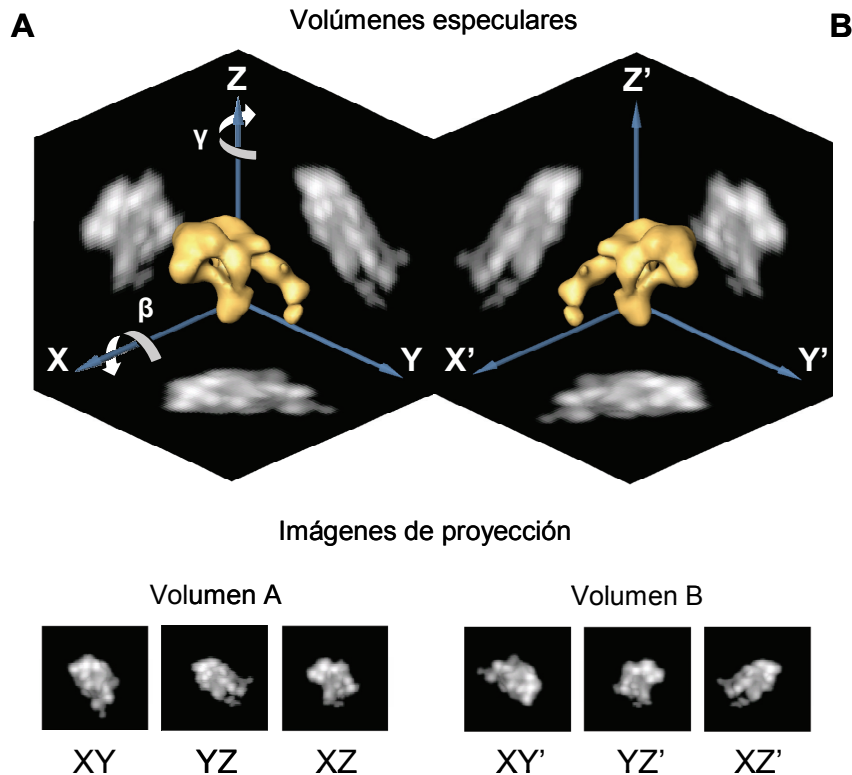
espécimen. El número de tipos de grises que se utilizan durante la digitalización debe representar de manera adecuada las diferentes intensidades presentes en la micrografía. Normalmente se utilizan 256 (8 bit/píxel) ó 65.536 (16 bit/píxel) tipos de grises. Las dimensiones del fragmento de la micrografía que son convertidas en un único píxel vienen definidas por la magnificación a la que se toma la micrografía y por el tamaño del paso del digitalizador (del inglés “Step size”). Según esto, una micrografía tomada a una magnificación de 50.000 aumentos y digitalizada a un tamaño del paso de 10,5  $\mu\text{m}$  tendrá aproximadamente unas dimensiones finales de 2,12 Å/píxel.

### 1.5.7 Ángulos de Euler

Resolver una estructura mediante microscopía electrónica y procesamiento digital de partículas individuales implica encontrar aquella estructura que, al proyectarse en determinadas orientaciones, genera las imágenes de proyección que se observan experimentalmente. Este es el objetivo último de las diferentes estrategias que se utilizan para llevar a cabo el procesamiento de partículas individuales. Los ángulos que relacionan la orientación de la estructura con las imágenes de proyección se denominan ángulos de Euler (*Figura 1.5.2*). Estos tres ángulos se refieren a las rotaciones que hay que llevar a cabo en los ejes X, Y y Z de la estructura correspondiente para obtener una imagen de proyección dada. El eje Z sería perpendicular al plano de proyección definido por la micrografía electrónica (*Figura 1.5.2*).

Dos estructuras especulares producen imágenes de proyección idénticas en determinadas orientaciones (*Figura 1.5.2*). Por ello definir la “mano” o quiralidad correcta de estructuras resueltas mediante microscopía electrónica conlleva realizar experimentos adicionales (Rosenthal y Henderson, 2003). Estos experimentos consisten en analizar imágenes de proyección obtenidas con el brazo portamuestras del microscopio electrónico inclinado. Además la “mano” correcta también se puede determinar a partir del ajuste de estructuras resueltas mediante RMN o CRX en el volumen de microscopía electrónica. Las estructuras obtenidas utilizando estas técnicas reflejan la “mano” real del espécimen.

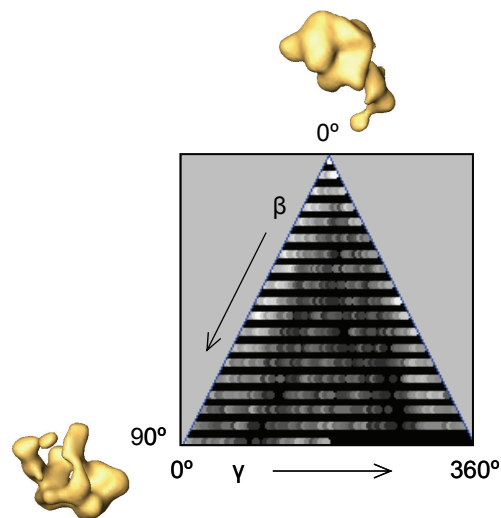




**Figura 1.5.2. Ángulos de Euler y problema de determinación de la “mano” en volúmenes resueltos por microscopía electrónica y procesamiento digital de imágenes individuales.** Las imágenes de proyección que se obtienen experimentalmente se relacionan con rotaciones específicas en los ejes X, Y y Z de la estructura tridimensional correspondiente. El paquete informático EMAN (Ludtke et al., 1999) utiliza los ángulos azimutal ( $\gamma$ ), de elevación ( $\beta$ ) y  $\phi$  para describir esta relación. El ángulo azimutal ( $\gamma$ ) corresponde a la rotación que tendría lugar en el eje Z. El ángulo de elevación ( $\beta$ ) describe la rotación correspondiente en el eje X. Por último el ángulo  $\phi$  describe la rotación que tendría lugar en el nuevo eje Z que se obtiene tras llevar a cabo la rotación en el eje X. Las partículas que varían únicamente en el ángulo  $\phi$  corresponden al mismo tipo de proyección pero alineadas de diferente manera. En la figura se representan dos volúmenes especulares (A) y (B). Estos volúmenes comparten la misma relación que existe entre la mano derecha y la mano izquierda. Imágenes de proyección generadas a partir de este tipo de volúmenes en determinadas orientaciones son idénticas.

Para resolver una estructura a partir de sus imágenes de proyección es necesario contar con un número suficiente de tipos o clases de proyecciones. Durante la preparación de las muestras para su observación en el microscopio electrónico los especímenes pueden adoptar orientaciones preferentes. En estos casos se obtiene un tipo de proyección mayoritario y por tanto, para aumentar la diversidad de imágenes de proyección, es necesario obtener micrografías con el brazo portamuestras inclinado. El procesamiento de partículas individuales que no contienen una diversidad de ángulos de Euler

adecuada genera estructuras incorrectas (Boisset et al., 1998). Estas estructuras presentan anisotropía o pérdida de resolución en las direcciones del espacio correspondientes a los ángulos de Euler ausentes. En algunos casos la falta de información es tan acusada que se manifiesta como deformaciones evidentes en la estructura obtenida. Por ello, con el objetivo de evaluar la calidad de las estructuras resueltas a partir del procesamiento digital de imágenes individuales, los valores de los ángulos de Euler de las imágenes utilizadas en la reconstrucción se representan gráficamente (*Figura 1.5.3*).



**Figura 1.5.3. Representación de los ángulos de Euler del conjunto de partículas individuales utilizadas en la reconstrucción tridimensional.**

La representación de los ángulos de Euler en el caso del paquete informático EMAN (Ludtke et al., 1999) se lleva a cabo mediante un triángulo isósceles. Los lados del triángulo reflejan un incremento del valor del ángulo correspondiente. Según esto los lados de mayor tamaño representan las variaciones del ángulo de elevación ( $\beta$ ) desde  $0^\circ$  en el vértice superior hasta  $90^\circ$  en los vértices inferiores. El lado que constituye la base representa las variaciones del ángulo azimutal ( $\gamma$ ) desde  $0^\circ$  en el vértice izquierdo hasta  $360^\circ$  en el vértice derecho. Cada punto gris dentro del triángulo isósceles representa una clase con unos ángulos de Euler determinados. El valor de gris para cada clase se asigna mediante una escala logarítmica según el número de imágenes presentes en la clase. Por tanto los puntos más claros corresponden a aquellas clases que contienen un mayor número de partículas. Con el objetivo de hacer más evidente el significado de esta representación se muestra el aspecto del volumen (A) (véase la *Figura 1.5.2*) para unos valores ( $\beta, \gamma$ ) de  $(0^\circ, 0^\circ)$  y  $(90^\circ, 0^\circ)$ .

### 1.5.8 Estrategias generales de procesamiento de partículas individuales

Como se comentó anteriormente el objetivo del análisis tridimensional de partículas individuales consiste en determinar los ángulos de Euler de las imágenes de proyección. Surge por tanto una situación un tanto paradójica: cómo determinar estos ángulos cuando se carece de información tridimensional del espécimen. Existe otro problema añadido que dificulta aún más la determinación de los ángulos de Euler: la baja relación señal-ruido de las imágenes de proyección. Esto implica que sean necesarios varios miles de partículas individuales para resolver la estructura de un espécimen dado. El análisis de partículas individuales recurre al promediado de imágenes para sobreponerse a esta limitación. Esta aproximación permite hacer más evidente aquello que es constante en la imagen (la señal del espécimen), mientras que el ruido que se distribuye de manera aleatoria en la imagen se promedia. El promediado de imágenes conlleva dos requisitos básicos. En primer lugar las imágenes que se promedian han de corresponder al mismo tipo de proyección del espécimen. En segundo lugar las imágenes a promediar han de tener la misma alineación. La obtención de imágenes medias requiere por tanto clasificar el conjunto inicial de partículas y alinear después las partículas presentes dentro de cada clase. Existen diferentes estrategias para determinar los ángulos de Euler de las imágenes de proyección:

Inclinación cónica al azar o RCT (del inglés “Random Conical Tilt”) (Radermacher et al., 1986). Esta estrategia permite obtener la estructura tridimensional correspondiente a un único tipo de imagen de proyección. Para ello se toman pares de micrografías de la misma zona de interés, una inclinando el brazo portamuestras (entre 40° y 60°) y otra sin inclinar. A partir de las micrografías no inclinadas se extraen las partículas que posteriormente se clasifican y alinean. Las imágenes dentro de una clase determinada corresponden al mismo tipo de imagen de proyección pero inicialmente no tienen por qué estar alineadas de la misma manera. Según esto para cada partícula de la clase la imagen inclinada correspondiente será diferente. El volumen se obtiene a partir de las imágenes inclinadas a las que se les asigna la alineación calculada durante la clasificación de las partículas no inclinadas.

Esta estrategia es especialmente útil cuando el espécimen a analizar presenta un tipo de imagen de proyección mayoritaria.

Método de reconstitución angular (van Heel, 1987). Este método utiliza comúnmente algoritmos basados en la técnica de líneas comunes (Crowther et al., 1970). Las imágenes de proyección de una estructura tridimensional, cuando se calcula su transformada de Fourier, presentan una línea común con el resto de imágenes de proyección de esa estructura tridimensional. Encontrando las líneas comunes se puede saber la orientación relativa de las diferentes imágenes de proyección. La técnica de líneas comunes produce resultados satisfactorios cuando se utilizan imágenes que presenten una relación señal-ruido alta. Por ello esta técnica se aplica a las imágenes medias de las clases que se obtienen tras clasificar y alinear el conjunto inicial de partículas individuales.

Métodos de alineamiento multirreferencia respecto a imágenes de proyección de una estructura conocida (Harauz y Ottensmeyer, 1984; van Heel, 1984). A partir de un volumen referencia determinado se generan imágenes de proyección con ángulos de Euler conocidos. El volumen se proyecta en múltiples direcciones y se obtienen imágenes de proyección que representan adecuadamente la diversidad de posibles orientaciones del espécimen. A continuación el conjunto de partículas individuales se compara con las imágenes de proyección generadas. A cada partícula individual se le asignan por tanto los ángulos de Euler de la imagen de proyección con la que presentan un mayor parecido. Por último se genera una nueva estructura utilizando las partículas individuales con los ángulos de Euler asignados. También es posible que las partículas individuales asignadas a la misma clase o referencia se promedien para obtener una imagen media. En ese caso las imágenes medias de las diferentes clases serán las que se utilicen en la obtención del nuevo volumen. Este volumen será una representación más fiel de la estructura tridimensional del espécimen ya que se construye utilizando las partículas individuales obtenidas experimentalmente.

La determinación de los ángulos de Euler utilizando cualquiera de las estrategias descritas anteriormente presenta limitaciones. Por ello es necesario llevar a cabo un proceso de refinamiento angular que permita mejorar de manera iterativa la asignación de los ángulos de Euler. Un ejemplo de estrategia de refinamiento angular es el método de emparejamiento con la proyección (del inglés “Projection Matching”). Este método consiste en repetir el proceso de alineamiento multirreferencia descrito anteriormente de forma iterativa (Penczek et al., 1994). El volumen que se obtiene tras cada iteración se utiliza en la siguiente iteración como volumen referencia. El procesamiento evoluciona y los volúmenes generados son cada vez más coherentes con las imágenes experimentales. El proceso de clasificación y alineamiento será más preciso tras cada iteración puesto que las referencias utilizadas representan más fielmente la estructura tridimensional del espécimen. El proceso termina estabilizándose cuando las variaciones en la asignación de las partículas a las diferentes clases son mínimas.

El método de emparejamiento con la proyección conlleva utilizar un volumen inicial que permita generar proyecciones que sirvan como referencias. Partir de un volumen inicial determinado introduce un sesgo que puede condicionar la evolución del proceso de refinamiento angular iterativo (Grigorieff, 2000; Yang et al., 2003). Por tanto la elección del volumen inicial es un paso importante a la hora de llevar a cabo el procesamiento digital de partículas individuales. Además la capacidad de evolución del proceso de refinamiento para converger en la estructura final correcta permite utilizar volúmenes iniciales que no tienen nada que ver con la solución. Este tipo de estrategias aún no estandarizadas se han explorado y puesto a punto durante el desarrollo de esta tesis doctoral. Concretamente los volúmenes iniciales utilizados han correspondido a:

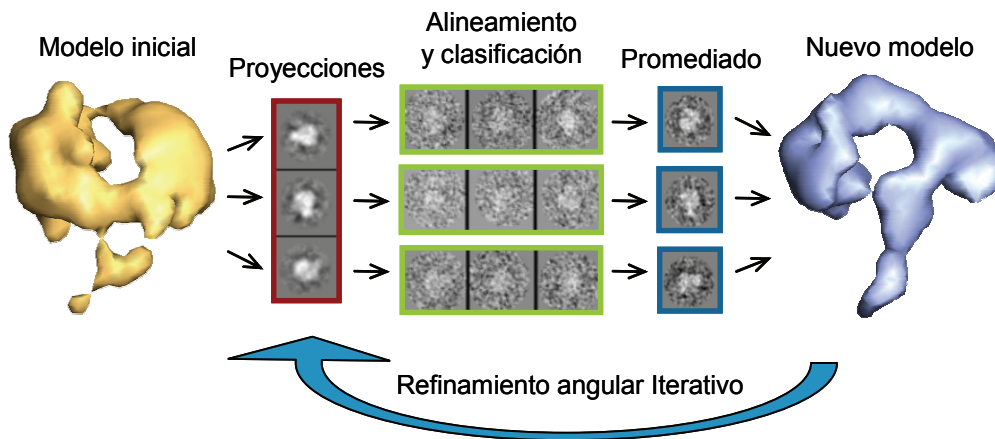
Elipsoides artificiales: El volumen inicial se genera artificialmente con la forma de una elipse o una esfera de dimensiones similares a las que se observan en las partículas individuales. Los volúmenes artificiales tienen una distribución de masa gaussiana. Además se les añade ruido artificialmente de

manera que las imágenes de proyección tengan un aspecto similar al de las partículas individuales a procesar.

Ruido: El volumen inicial consiste simplemente en una distribución homogénea de ruido. Esta estrategia es especialmente útil cuando el volumen que se genera tras cada iteración se construye utilizando las imágenes medias de las diferentes clases.

### 1.5.9 Estrategia de procesamiento del paquete informático EMAN

Existen diversos paquetes informáticos para realizar el procesamiento digital de partículas individuales. Algunos de los más empleados son EMAN (Ludtke et al., 1999), SPIDER (Frank et al., 1996), XMIPP (Marabini et al., 1996; Sorzano et al., 2004), IMAGIC (van Heel y Keegstram, 1981) y Bsoft (Heymann, 2001; Heymann y Belnap, 2007). Los distintos paquetes se componen de gran cantidad de programas que permiten aplicar cualquiera de las estrategias de procesamiento descritas anteriormente. Por ello es frecuente que durante el procesamiento digital de imágenes individuales se combinen programas presentes en distintos paquetes informáticos. EMAN está especialmente diseñado para llevar a cabo el refinamiento angular según el método de emparejamiento con la proyección. El programa que engloba las distintas etapas características de este método de procesamiento se denomina *refine* (Figura 1.5.4). Este programa presenta determinadas particularidades que favorecen la capacidad de evolución y convergencia del proceso de refinamiento y, además, minimizan el sesgo introducido por el volumen referencia. Al igual que se ha descrito anteriormente el programa *refine* consta también de una primera etapa en la que se generan imágenes de proyección a partir de un volumen inicial (Figura 1.5.4). Las proyecciones generadas se utilizan para clasificar y realizar un primer alineamiento del conjunto inicial de partículas (Figura 1.5.4). Después las imágenes dentro de cada clase se someten a 8 ciclos iterativos de alineamiento y promediado (Figura 1.5.4). El primer alineamiento se realiza respecto a la imagen de referencia característica de la clase. Tras este primer alineamiento se genera una imagen media que de nuevo sirve como referencia para alinear las imágenes dentro de la clase. Se

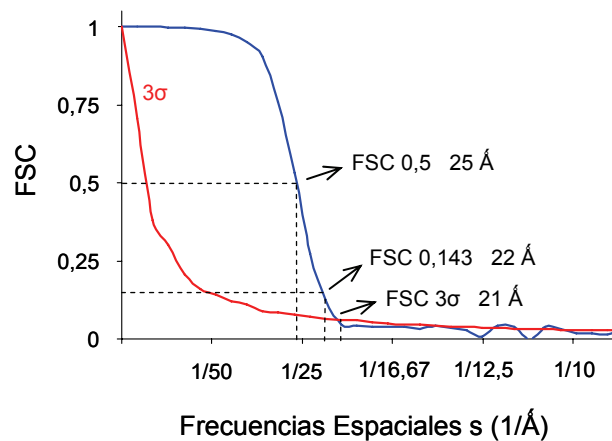


**Figura 1.5.4. Esquema general del proceso de refinamiento angular característico del paquete informático EMAN.** El comando *refine* engloba las siguientes etapas. Primero el volumen referencia se proyecta en diferentes orientaciones. Las proyecciones se utilizan como referencias para clasificar el conjunto inicial de partículas individuales. A continuación las imágenes dentro de cada clase se alinean y promedian de forma iterativa. Se obtiene por tanto una nueva imagen media para cada clase cuyos ángulos de Euler se conocen. Las nuevas imágenes medias se utilizan para construir un nuevo volumen que servirá como volumen referencia en la siguiente iteración. El refinamiento angular iterativo termina estabilizándose generando la estructura final.

obtiene por tanto otra imagen media repitiéndose de nuevo el proceso un total de 8 veces. Existen parámetros que permiten controlar qué imágenes participan en la generación de la imagen media tras cada iteración y el número de ciclos iterativos que se realizan. Esta manera de generar las imágenes medias que construirán el nuevo volumen es una característica importante del paquete informático EMAN. Además disminuye significativamente el sesgo introducido por el volumen referencia y favorece la evolución del proceso de refinamiento. Después las imágenes medias de las diferentes clases se utilizan para construir un nuevo volumen (*Figura 1.5.4*). Estas imágenes medias tienen la información correspondiente a los ángulos de Euler de la clase a la que pertenecen. En esta etapa existen también parámetros que permiten controlar la coherencia de las imágenes medias que contribuyen a la generación del nuevo volumen. A partir de este nuevo volumen se generan proyecciones que servirán para volver a clasificar y alinear el conjunto inicial de partículas (*Figura 1.5.4*).

### 1.5.10 Cálculo de la resolución

Una vez concluido el proceso de refinamiento angular iterativo se lleva a cabo la estimación de la resolución para la estructura resuelta. El objetivo de esta estimación consiste en definir de manera cuantitativa el grado de detalle que se aprecia en el volumen obtenido. Para ello el conjunto inicial de partículas se divide en dos grupos y se calcula la estructura correspondiente para cada grupo (comando *eotest* del paquete informático EMAN). Después las dos estructuras se comparan en el espacio de Fourier y se calcula la correlación entre ambas en función de la frecuencia espacial o FSC (del inglés “Fourier shell correlation”) (*Figura 1.5.5*) (Harauz y van Heel, 1986). Existen diferentes criterios para definir qué valor de correlación es el que mejor refleja la resolución del volumen obtenido: 0,5 (Bottcher et al., 1997),  $3\sigma$  (Orlova et al., 1997) y 0,143 (Rosenthal y Henderson, 2003). A partir de la frecuencia espacial correspondiente a un valor de FSC de 0,5, 0,143 o bien al punto de corte entre la curva  $3\sigma$  y la FSC se estima la resolución (*Figura 1.5.5*). Por último el volumen resuelto se filtra a la resolución estimada de manera que se eliminan la información no sustentada por los datos y se obtiene la estructura final.



**Figura 1.5.5. Estimación de la resolución en estructuras resueltas por microscopía electrónica y procesamiento digital de imágenes individuales.** La gráfica que se muestra representa la variación de la FSC en función de la frecuencia espacial. En ella se aprecia cómo al aumentar la frecuencia espacial y mejorar por tanto la resolución disminuye la correlación entre los volúmenes utilizados para calcular la FSC. Se indican los diferentes criterios empleados para estimar la resolución. El valor máximo de resolución según el criterio 0,5 es aquel en el que la relación señal-ruido es uno. El criterio  $3\sigma$  implica calcular la desviación estándar de la FSC obtenida para volúmenes que contengan únicamente ruido y multiplicarla por tres. El criterio 0,143 se ha establecido experimentalmente a partir del análisis de estructuras resueltas mediante CRX. De ahora en adelante la función  $3\sigma$  se representa de color rojo y la FSC de color azul.





## 2. Objetivos



El objetivo de esta tesis doctoral consiste en profundizar en el conocimiento estructural del proceso de reparación de roturas de doble cadena en el ADN mediante NHEJ. Para ello se pretende:

- Caracterizar estructuralmente el heterodímero Ku libre y unido a ADN.
- Caracterizar estructural y funcionalmente la proteína DNA-PKcs libre y unida a ADN.
- Caracterizar estructural y funcionalmente el complejo DNA-PKcs:Ku:ADN.
- Analizar bioquímicamente la interacción entre las proteínas Artemis y DNA-PKcs como paso previo al abordaje estructural del complejo DNA-PKcs:Artemis.
- Combinar la información estructural obtenida con los datos bioquímicos descritos en la bibliografía para, en último término, proponer un modelo funcional que explique las etapas iniciales del proceso de NHEJ.

La caracterización estructural de estos componentes celulares se llevará a cabo utilizando microscopía electrónica y análisis de partículas individuales. El estudio de los especímenes mencionados anteriormente mediante esta técnica estructural conlleva una serie de retos importantes. Por ello, además de los objetivos biológicos descritos previamente, es posible distinguir los siguientes objetivos metodológicos:

- Aplicar el análisis de partículas individuales a especímenes de bajo peso molecular.
- Aplicar el análisis de partículas individuales a especímenes que no presentan simetría.
- Aplicar el análisis de partículas individuales a mezclas de especímenes.
- Desarrollar estrategias que permitan evitar el sesgo introducido por el volumen inicial en los procesos de refinamiento angular iterativo.





### 3. Materiales y Métodos



### 3.1 Obtención de moléculas de ADN de doble cadena

Las moléculas de ADN utilizadas en los ensayos de formación de complejos proteína:ADN se obtuvieron a partir de la hibridación de diferentes oligonucleótidos (*Tabla 3.1*). Los oligonucleótidos se resuspendieron inicialmente en agua mQ. Posteriormente se tomaron cantidades equimolares de ambos oligonucleótidos complementarios y se añadió tampón. El tampón utilizado variaba en función de la molécula de ADN generada. Por último la mezcla de reacción se incubó durante 1 hora a 37°C. Los oligonucleótidos ya hibridados se conservaron a -80°C.

La reacción de hibridación de la molécula de ADN de 54 pb y extremos romos biotinilada y sin biotinilar se llevó a cabo en tampón 500 mM Tris-HCl, pH 7,5, 10 mM DTE, 1 M NaCl y 100 mM MgCl<sub>2</sub> (10X SuRE/Cut Buffer H, Roche Applied Science). El volumen final de la mezcla de reacción fue 445 µl y la concentración final de la molécula de ADN generada fue de 3,7 µM. El laboratorio del Profesor L. H. Pearl, del “Institute of Cancer Research” en Londres, llevó a cabo la valoración de la calidad de la hibridación para la molécula no biotinilada. Para ello se midió la hipercromicidad tras la digestión con fosfodiesterasa 1 de veneno de serpiente (Amersham).

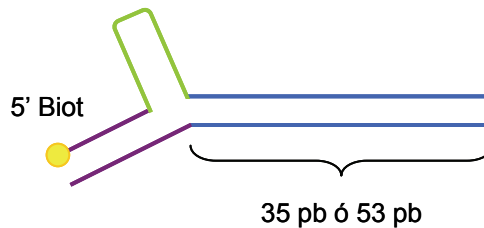
Para generar el complejo DNA-PKcs:Ku:ADN se utilizaron dos tipos de moléculas de ADN con unas características muy concretas. Ambas moléculas presentaban un bloqueo en forma de Y en uno de sus extremos (*Tabla 3.1* y *Figura 3.1.1*). Este tipo de bloqueo había sido utilizado previamente en la obtención de la estructura atómica del núcleo del heterodímero Ku unido a ADN

**Tabla 3.1. Moléculas de ADN de doble cadena utilizadas en el ensamblaje de complejos proteína:ADN (página siguiente).** Algunos de los oligonucleótidos descritos están biotinilados (Biot) en su extremo 5'. La obtención de las moléculas de ADN en las que posteriormente se ensambló el complejo DNA-PKcs:Ku:ADN fue realizada por el laboratorio del Profesor L. H. Pearl. En la secuencia de los oligonucleótidos utilizados para generar estas moléculas, los nucleótidos en minúscula corresponden a la secuencia que anilla consigo misma dentro del oligonucleótido y que por tanto forma el bloqueo (*Figura 3.1.1*). En estos oligonucleótidos el código de colores es equivalente al utilizado posteriormente en la *Figura 3.1.1*.

Complejo		Descripción	Oligonucleótidos hibridados	Purificación y casa comercial
Ku:ADN DNA-PKcs:ADN Artemis:ADN	54 pb Extremos romos Biotinilada en 5'	5' Biot <u>GGCCGCACGGCTCCACCATGGGGTACAACACTAC</u> <u>GATCTAGCTTCATGCACCCGGAC</u> 3'	FPLC  (Oswell, Southampton, Reino Unido)	
		5' <u>GTCCGGTGCATGAAGCTAGATCGTAGTTGTACCCCA</u> <u>TGGTGGACGGGTGCGGGCC</u> 3'		
DNA-PKcs:ADN Ku:ADN (mediante RCT)	54 pb Extremos romos	5' <u>GGCCGCACGGCTCCACCATGGGGTACAACACTACGAT</u> <u>CTAGCTTCATGCACCCGGAC</u> 3'	FPLC  (Oswell, Southampton, Reino Unido)	
		5' <u>GTCCGGTGCATGAAGCTAGATCGTAGTTGTACCCCA</u> <u>TGGTGGACGGGTGCGGGCC</u> 3'		
DNA-PKcs:Ku:ADN (ADN tipo corto)	Bloqueo tipo Y Extremos romos Biotinilada en 5'	5' Biot <u>CGCGCCCaactttcccactAATAAACTAAAAACTATTA</u> <u>TTATGGCCGCACGGCGI</u> 3'	HPLC  (IBA GmbH, Goettingen, Alemania)	
		5' <u>ACGGGTGGCGGCCATAATAATAGTTTTAGTTTATTGG</u> <u>GGCGG</u> 3'		
DNA-PKcs:Ku:ADN (ADN tipo largo)	Bloqueo tipo Y Extremos romos Biotinilada en 5'	5' Biot <u>CGCGCCCaactttcccactAATAAACTAAAAACTATTA</u> <u>TTATGGCCGCACGGCTCCACCATGGGGTACAACI</u> 3'	HPLC  (IBA GmbH, Goettingen, Alemania)	
		5' <u>GTGTACCCCAITGGTGGACGGGTGCGGGCCATAATAA</u> <u>TAGTTTTAGTTTATTGGCGCG</u> 3'		



(Walker et al., 2001). Además en los dos tipos de moléculas de ADN el extremo 5' presente en el bloqueo estaba iminobiotinilado (Biot). Las dos moléculas se diferencian en la longitud del extremo sin bloqueo. En el caso de la molécula de ADN de menor longitud o de tipo corto el extremo sin bloqueo está formado por 35 pb. Para la molécula de ADN de mayor tamaño o de tipo largo la longitud del extremo sin bloquear era de 53 pb.



**Figura 3.1.1. Diseño esquemático de las moléculas de ADN utilizadas en el ensamblaje del complejo DNA-PKcs:Ku:ADN.** El código de colores es equivalente al que se observa en la *Tabla 3.1* al describir la secuencia de los diferentes oligonucleótidos.

## 3.2 Purificación de proteínas humanas

Las proteínas Artemis, Ku, DNA-PKcs, XRCC4 y el complejo DNA-PKcs:Ku:ADN fueron proporcionados por el Profesor L. H. Pearl. Los protocolos que se describen a continuación únicamente pretenden dar una visión general del proceso de purificación.

### 3.2.1 Purificación del heterodímero Ku

La proteína Ku se purificó como un complejo heterodimérico a partir de extractos nucleares de células HeLa (CilBiotech, Mons, Bélgica). El proceso de purificación combinó diferentes cromatografías (GE Healthcare). Primero los extractos nucleares se fraccionaron utilizando una resina de heparina seguida de una columna de celulosa de ADN de doble cadena. Después se realizó un paso de lavado en resina Mono S y a continuación la muestra recogida se concentró mediante una resina de heparina. Posteriormente se utilizó una resina fenil-sefarosa seguida de una columna Mono Q de la que se recuperó la muestra por medio de un gradiente lineal de 0,05 M a 1 M de NaCl. Durante

todo el proceso de purificación se utilizó el siguiente tampón: 20 mM HEPES, pH 7,5, 10% glicerol, 1 mM DTT, 1 mM EDTA, 0,001% BOG y 100 mg/l Pefabloc (Roche Applied Science). La concentración de NaCl varió según el estadio del proceso de purificación. El heterodímero Ku purificado se obtuvo a una concentración final de 0,45  $\mu$ M.

### 3.2.2 Purificación de la proteína DNA-PKcs

DNA-PKcs se purificó siguiendo el protocolo que se detalla en Gell y Jackson, 1999, pero incluyendo algunas modificaciones descritas en Boskovic et al., 2003. Básicamente los extractos nucleares de células HeLa (CilBiotech, Mons, Bélgica) se fraccionaron primero utilizando una columna Q-Sepharose (Pharmacia), y a continuación una columna de heparina-agarosa (Sigma). Las fracciones que contenían la proteína DNA-PKcs se juntaron y se les añadió sulfato amónico hasta alcanzar una concentración de 0,75 M. La mezcla resultante se aplicó a una columna de fenil-sefarosa (Pharmacia) y se recuperó mediante un gradiente lineal de sulfato amónico de 0,75 a 0 M. Las fracciones enriquecidas en la proteína DNA-PKcs se purificaron de nuevo utilizando una columna Mono S (Pharmacia). Por último, las fracciones que contenían la proteína DNA-PKcs pura se juntaron y se dializaron frente al siguiente tampón: 25 mM HEPES, pH 7,6, 50 mM KCl, 1 mM  $MgCl_2$ , 1 mM DTT, 0,5 mM EDTA y 20% glicerol. Posteriormente se hicieron alícuotas que se guardaron a  $-80^{\circ}C$ . La concentración final de la proteína DNA-PKcs purificada fue de 2,2  $\mu$ M.

### 3.2.3 Purificación del complejo DNA-PKcs:Ku:ADN

El complejo DNA-PKcs:Ku:ADN se purificó siguiendo el protocolo descrito en Spagnolo et al., 2006. Los extractos nucleares de células HeLa (CilBiotech, Mons, Bélgica) se fraccionaron en una resina de heparina seguida de una columna de celulosa de ADN de doble cadena (GE Healthcare). Las fracciones que contenía las proteínas DNA-PKcs, Ku70 y Ku80 se aplicaron a una columna de 1 ml ResourceQ (GE Healthcare) y se recuperaron mediante un gradiente lineal de NaCl de 0,05 M a 1 M. La unión de la proteína DNA-PKcs al ADN está regulada negativamente por autofosforilación. Por ello las fracciones

que contenían las tres proteínas se incubaron con fosfatasa- $\lambda$  durante 1 hora a 4°C con el objetivo de aumentar la homogeneidad de la preparación. Además esta incubación restauraba la proteína DNA-PKcs a un estado de alta afinidad por extremos de ADN similar al que se daría en las etapas iniciales de la reparación por NHEJ. La fosfatasa- $\lambda$  fue amablemente proporcionada por la Dra. C. Vaughan del “Institute of Cancer Research” en Londres. Tras la incubación con fosfatasa- $\lambda$  la muestra se dividió en tres partes. Una de ellas no se trató más y las otras dos partes se incubaron a 4°C durante otra hora adicional con ADN de tipo largo o corto respectivamente (véase el apartado 3.1). Después las muestras se aplicaron a gradientes de glicerol del 18% al 60% y se centrifugaron a 25.000 rpm durante 72 horas a 4°C utilizando un rotor SW28 (Beckman). El NaCl se mantuvo a una concentración constante de 200 mM en el gradiente. Las fracciones se recogieron desde la base y se analizaron mediante electroforesis en condiciones desnaturizantes. Para ello se utilizaron geles de poliacrilamida desnaturizantes en gradiente del 4% al 12% (NuPAGE NOVEX, Invitrogen) con tampón Bis-Tris como electrolito. Los complejos DNA-PKcs:Ku:ADN se analizaron adicionalmente por electroforesis en condiciones nativas para determinar su estado oligomérico. Para ello se utilizaron geles de poliacrilamida no desnaturizantes en gradiente del 3% al 8% (NuPAGE NOVEX, Invitrogen) con tampón Tris-Acetato como electrolito. Durante el proceso de purificación se utilizó el siguiente tampón: 20 mM HEPES, pH 7,5, 1 mM DTT, 0,5 mM EDTA, 0,001% BOG y 10% glicerol. La concentración de NaCl variaba en un rango 50 mM a 1 M según el paso cromatográfico correspondiente.

### 3.2.4 Purificación de la proteína Artemis

La proteína Artemis se expresó en células de insecto SF9 infectadas con baculovirus recombinantes. A partir de extractos de células de insecto se llevó a cabo la purificación de la proteína recombinante Artemis marcada con una etiqueta de histidinas. Para ello se utilizó una columna afinidad que contenía un metal inmovilizado (Clontech). A continuación se eliminó la etiqueta mediante incubación con la proteasa TEV (del inglés “Tobacco etch virus”). Después se realizó una cromatografía de intercambio iónico en 5 ml de resina HiTrap S (GE

Healthcare). Las fracciones que contenían la proteína Artemis se concentraron mediante Vivaspins (Sartorius) de 20 ml de capacidad y 10 kDa de límite de separación. La muestra concentrada se aplicó a la columna de exclusión molecular S200 26/60 (GE Healthcare). El volumen al que se recuperaba la proteína Artemis sugería que la proteína se encontraba como un homodímero. Tras la cromatografía las fracciones de interés se concentraron utilizando Vivaspins (Sartorius) de 20 ml de capacidad y 10 kDa de límite de separación. La proteína Artemis se purificó en tampón 20 mM HEPES, pH 7,5, 1 mM DTT, 1 mM EDTA, 0,5 M KCl y 10% glicerol a una concentración de 28,6  $\mu$ M.

### 3.2.5 Purificación de la proteína XRCC4

El cDNA de la proteína XRCC4 de riñón humano (Clontech) se clonó en el plásmido pRSET-B (Invitrogen) y se expresó en células BL21-CodonPlus(DE3) (Stratagene). La proteína recombinante marcada con una etiqueta de seis residuos de histidina se purificó utilizando una columna de afinidad que contenía un metal inmovilizado (Clontech). A continuación se realizó una cromatografía de intercambio iónico en una columna HiTrapQ (Pharmacia). Por último se llevó a cabo una cromatografía de exclusión molecular en una columna Superdex 200 (Pharmacia). La proteína XRCC4 se obtuvo a una concentración final de 25  $\mu$ M y se guardó a  $-80^{\circ}\text{C}$  en 20 mM Tris-HCl pH 8,0, 10% glicerol, 1 M NaCl, 1 mM EDTA, 2 mM DTT y 0,1 mM PMSF.

## 3.3 Electroforesis en condiciones desnaturizantes

La electroforesis en condiciones desnaturizantes se llevó a cabo en geles discontinuos de poliacrilamida (Laemmli, 1970) (acrilamida:bisacrilamida 40,6:1). El gel concentrador se preparó a una concentración final de acrilamida del 3,2% en 285 mM Tris-HCl y pH 6,8. La concentración final de acrilamida en el gel separador fue del 6% en 370 mM Tris-HCl y pH 8,8. En ambos geles la concentración de SDS fue del 0,1%. Como catalizadores de la reacción de polimerización se utilizaron TEMED y PSA. La electroforesis se llevó a cabo en cubetas Mini-PROTEAN 3 Cell (BIO-RAD) y SE 260 (GE Healthcare). Los geles

verticales que se ensamblaron en las cubetas Mini-PROTEAN 3 Cell (BIO-RAD) tenían una anchura de 8 cm y una altura de 7,3 cm. En el caso de las cubetas SE 260 (GE Healthcare) la anchura de los geles verticales era de 10 cm y la altura de 10,5 cm. En ambos casos el espesor de los geles era de 0,75 mm. Antes de aplicar las muestras en el gel se les añadió tampón de carga 4X (62 mM Tris-HCl, pH 6,8, 8% glicerol, 2% SDS, 720 mM  $\beta$ -mercaptoetanol, 0,001% azul de bromofenol), y se desnaturalizaron calentándolas a 95°C durante 3 min. La electroforesis se llevó a cabo en tampón 25 mM Tris-HCl, pH 8,3, 190 mM glicina y 0,1% SDS. Tras aplicar las muestras al gel la electroforesis se inició a un voltaje constante de 80 V hasta que las muestras atravesaron el gel concentrador. A partir de ese momento el voltaje se aumentó a 180 V y se mantuvo constante. El tiempo del ensayo de electroforesis varió según los especímenes que se estuvieran analizando. Los estándares de peso molecular utilizados corresponden al “HMW-SDS Marker Kit” (GE Healthcare).

### 3.4 Electroforesis en condiciones nativas

La electroforesis en condiciones nativas se llevó a cabo de manera similar a como se ha descrito en el apartado 3.3 pero con las siguientes modificaciones. En primer lugar no se utilizó gel concentrador. Las proteínas y complejos estudiados presentan un elevado peso molecular. Esta característica complica su avance a través de la transición entre el gel concentrador y el gel separador cuando se analizan por electroforesis en condiciones nativas. La concentración de acrilamida (acrilamida:bisacrilamida 40,6:1) del gel separador varió entre el 4% y el 6% en 370 mM Tris-HCl y pH 8,8. Para el análisis de la interacción del heterodímero Ku con ADN se utilizaron geles al 6%. Para analizar la proteína Artemis y las interacciones entre las proteínas DNA-PKcs y Artemis y también entre Artemis y el ADN se utilizaron geles al 5%. En los experimentos en los que se comprobó la capacidad de un anticuerpo monoclonal (DNA-PKcs G-4 sc-5282, Santa Cruz) de reconocer a la proteína DNA-PKcs en condiciones nativas se utilizaron geles al 4% (véase el apartado 3.9.3). Como catalizadores de la reacción de polimerización se utilizaron TEMED y PSA. En este caso el gel separador no contenía SDS. Como tampón de electroforesis se utilizó 25

mM Tris-HCl, pH 8,3 y 190 mM glicina. La electroforesis se llevó a cabo en cubetas Mini-PROTEAN 3 Cell (BIO-RAD). Las dimensiones de los geles verticales fueron idénticas a las descritas anteriormente en el apartado 3.3. Los geles se sometieron a una electroforesis previa durante 30 min a un voltaje constante de 90 V y a una temperatura de 4°C. A continuación se aplicaron las muestras a las que previamente se les había añadido tampón de carga 10X (125 mM Tris-HCl, pH 8,8, 67% glicerol y 0,002% azul de bromofenol). La electroforesis se llevó a cabo a un voltaje constante de 90 V y a una temperatura de 4°C. El tiempo del ensayo variaba dependiendo del espécimen que se estuviera analizando. Los estándares de peso molecular utilizados corresponden al “HMW Native Marker Kit” (GE Healthcare).

### 3.5 Tinción de geles nativos con bromuro de etidio

Para evidenciar la presencia del ADN en los ensayos de interacción proteína:ADN los geles nativos se tiñeron con tampón de electroforesis nativo 1,27  $\mu$ M bromuro de etidio. Después de 30 min de incubación a temperatura ambiente los geles se expusieron a luz UV para visualizar el ADN.

### 3.6 Tinción reversa de geles nativos

Este método de tinción inmoviliza de manera reversible las proteínas a la matriz del gel (Ortiz et al., 1992; Hardy y Castellanos-Serra, 2004). Por tanto tras teñir de esta manera es posible extraer las proteínas del gel por difusión. El método de tinción reversa conlleva la formación de precipitados blancos de imidazol y zinc en aquellas regiones en las que no hay proteína. Las zonas que contienen proteína quedan transparentes. Este método permite detectar entre 1 y 10 ng de proteína por banda, lo que significa una sensibilidad similar al método de tinción con plata. Para geles de las dimensiones descritas en el apartado 3.3 la tinción reversa se realizó de la siguiente manera. Tras la electroforesis en condiciones nativas el gel se incubó durante 10 min con 50 ml de una solución 0,2 M imidazol y 0,5% SDS. A continuación el gel se incubó

con 50 ml de una solución 0,3 M de sulfato de zinc durante 1 min. Este paso conlleva la aparición de los precipitados blancos mencionados anteriormente.

### 3.7 Ensayos de “Western-blot”

Las proteínas se transfirieron a una membrana de PVDF (del inglés “Polyvinylidene difluoride”, BIO-RAD). Esta membrana se incubó previamente con metanol al 100% para hacerla más hidrofílica y posteriormente se equilibró con tampón de transferencia (25 mM Tris-HCl, pH 8,3, 124 mM glicina, 20% etanol y 10% SDS). El gel también se equilibró con tampón de transferencia antes de llevar a cabo el ensayo de “Western-blot”. La transferencia se realizó en módulos Mini Trans-Blot (BIO-RAD) a 400 mA durante una hora y a 4°C. Una vez transferidas las proteínas, la membrana se incubó con solución de bloqueo PBS 1X, 0,5% Tween 20 y 5% leche en polvo Sveltesse (Nestlé) durante una hora a temperatura ambiente con agitación. A continuación la membrana se incubó durante una hora con 1 ml de solución de bloqueo que contenía el anticuerpo primario. La incubación se realizó a temperatura ambiente con agitación. Después se realizaron 3 lavados con solución de lavado PBS 1X, 0,05% Tween 20 y 1% leche en polvo Sveltesse (Nestlé). Cada lavado tuvo una duración de 10 min y se llevó a cabo con 100 ml de solución de lavado. Tras los lavados la membrana se incubó con 5 ml de solución de bloqueo que contenía el anticuerpo secundario conjugado a HRP (del inglés “Horseradish peroxidase”). La incubación se prolongó durante una hora a temperatura ambiente con agitación. A continuación se realizaron 3 lavados como los descritos anteriormente tras la incubación con el anticuerpo primario. Por último la membrana se reveló utilizando el sistema de quimioluminiscencia ECL o ECL Plus (GE Healthcare) siguiendo las recomendaciones del fabricante. Para detectar las bandas luminiscentes se utilizaron películas CURIX RP2 Plus (Agfa) que se revelaron en un AGFA Curix 60. Los anticuerpos utilizados en los diferentes ensayos de “Western-blot” se describen en la *Tabla 3.2*.

Anticuerpo	Tipo	Naturaleza	Nombre comercial	Dilución
Anti-Ku70 humana	Primario	Monoclonal	2C3.11, ab87 Abcam	1:1.000
Anti-DNA-PKcs humana	Primario	Monoclonal	DNA-PKcs (G-4) sc-5282 Santa Cruz	1:1.000
Anti-DNA-PKcs humana	Primario	Monoclonal	DNA-PK (Ab-2) Oncogene	1:1.000
Anti-IgG ratón	Secundario	Policlonal	ab6728 Abcam	1:5.000

**Tabla 3.2. Anticuerpos utilizados en los diferentes ensayos de “Western-blot”.** Detalle de los anticuerpos utilizados en los diferentes ensayos de “Western-blot”. Se indican también las diluciones a las que se emplearon,

### 3.8 Tinción con plata

Los geles obtenidos tras realizar la electroforesis, tanto en condiciones desnaturalizantes como nativas, se tiñeron con plata a temperatura ambiente según el siguiente protocolo:

- a) Incubación durante 30 min con 25 ml de solución de fijación (10% ácido acético y 38,4% etanol).
- b) Incubación durante 30 min con 25 ml de solución de sensibilización (28,8% etanol, 0,125% glutaraldehído, 8 mM tiosulfato de sodio y 500 mM acetato de sodio).
- c) A continuación se realizaron 3 lavados utilizando 100 ml de agua mQ en cada lavado. La duración de cada lavado fue de 5 min.
- d) Incubación durante 20 min con 25 ml de solución de plata (5,9 mM nitrato de plata y 0,015% formaldehído)



- e) A continuación se llevó a cabo un lavado rápido con agua mQ.
- f) Incubación con 25 ml de solución de revelado (236 mM carbonato de sodio y 0,007% formaldehído) hasta que aparecieron las bandas de interés.
- g) Incubación con 25 ml de solución de parada (40 mM EDTA) durante 10 min.

## 3.9 Ensayos bioquímicos

### 3.9.1 Interacción del heterodímero Ku con ADN

Se analizó la capacidad de la proteína Ku de interaccionar con una molécula de ADN de 54 pb y extremos romos biotinilada en su extremo 5' (véase el apartado 3.1). Para ello 2  $\mu$ l de la proteína Ku purificada (concentración del heterodímero 0,45  $\mu$ M), 4  $\mu$ l de la molécula de ADN (3,7  $\mu$ M) y 1  $\mu$ l de tampón (50 mM Tris-HCl, pH 7,5, 20% glicerol, 1 mM MgCl<sub>2</sub>, 50 mM KCl y 10 mM EDTA) se incubaron a 25°C durante 20 min. Después la mezcla se analizó por electroforesis en condiciones nativas. Posteriormente el gel se tiñó con bromuro de etidio y se reveló con luz UV. Tras eliminar el bromuro de etidio del gel se llevó a cabo un experimento de "Western-blot" utilizando como anticuerpo primario un anticuerpo monoclonal frente a Ku70 (2C3.11, ab87; Abcam, Cambridge, Reino Unido).

### 3.9.2 Ensayo de la actividad quinasa de la proteína DNA-PKcs

Este experimento fue realizado por el laboratorio del Profesor L. H. Pearl. La actividad quinasa de DNA-PKcs dependiente de ADN se cuantificó valorando la fosforilación de XRCC4 en presencia o ausencia de ADN. Para ello 20  $\mu$ l de una mezcla de reacción que contenía 1  $\mu$ g de XRCC4, 10 ng/ $\mu$ l de ADN de 54 pb y extremos romos no biotinilado (véase el apartado 3.1), 10 mM Tris-HCl, pH 7,5, 50 mM KCl, 10 mM MgCl<sub>2</sub>, 1 mM DTT, 2  $\mu$ l de [ $\gamma$ -<sup>32</sup>P] ATP (6.000 Ci/mmol, 10 mCi/ml) y 200 ng de DNA-PKcs se incubaron durante 30 min a 30°C. Los ensayos de actividad quinasa también se llevaron a cabo utilizando un estándar de ADN de 100 pb (New England BioLabs). Los productos de la

reacción se analizaron por electroforesis en condiciones desnaturalizantes. Después el gel se tiñó con azul de Coomassie, se secó en papel 3MM y se reveló por autorradiografía. Las bandas correspondientes a las proteínas DNA-PKcs y XRCC4 se recortaron y la radiactividad se midió con un contador de Cherenkov.

### **3.9.3 Formación de complejos anticuerpo:DNA-PKcs**

Los inmunocomplejos se prepararon mezclando en una relación molar 3:1 anticuerpo monoclonal (DNA-PKcs G-4 sc-5282, Santa Cruz) y DNA-PKcs respectivamente. La reacción se llevó a cabo en tampón 50 mM Tris-HCl, pH 7,5, 100 mM KCl y 20 mM EDTA, y se incubó durante 30 min a temperatura ambiente. El anticuerpo monoclonal utilizado se generó a partir de una proteína recombinante que incluía la región más N-terminal de la proteína DNA-PKcs humana (aminoácidos 2.965 a 4.127). Las especificaciones del fabricante indicaban que este anticuerpo era apto para inmunoprecipitar la proteína DNA-PKcs. Aun así se comprobó la capacidad del anticuerpo de reconocer a DNA-PKcs en condiciones nativas. Para ello la proteína DNA-PKcs primero se analizó mediante electroforesis en condiciones nativas. Después se realizó un “Western-blot” en el que se eliminó del tampón de transferencia el SDS y el etanol (véase el apartado 3.7). Como anticuerpo primario se utilizó el anticuerpo escogido para generar los complejos anticuerpo:DNA-PKcs.

### **3.9.4 Interacción de la proteína DNA-PKcs con ADN**

Este experimento fue realizado por el laboratorio del Profesor L. H. Pearl. Los experimentos de cambio de movilidad electroforética del complejo DNA-PKcs:ADN se realizaron esencialmente como se describe en Hammarsten y Chu, 1998, pero introduciendo pequeñas modificaciones. Las reacciones de unión se llevaron a cabo en tampón B (10 mM Tris-HCl, pH 7,5, 10 mM NaCl, 1 mM EDTA, 1 mM DTT y 5% glicerol). A este tampón se le añadieron 0,5 ng de ADN de 54 pb y extremos romos no biotinilado (véase el apartado 3.1). Este ADN estaba marcado radiactivamente con  $^{32}\text{P}$ . En último lugar se añadió la proteína DNA-Pkcs y a continuación la mezcla de reacción se incubó durante

30 min a temperatura ambiente. Los complejos proteína:ADN se analizaron mediante experimentos de electroforesis en condiciones nativa a 4°C. Para ello se utilizaron geles de poliacrilamida al 4% con tampón Tris-Glicina-EDTA 0,25X como electrolito.

### 3.9.5 Marcaje con oro de complejos proteína:ADN

Para visualizar la molécula de ADN presente en el complejo DNA-PKcs:ADN se llevó a cabo el siguiente experimento de marcaje. Primero estreptavidina unida covalentemente a oro coloidal de 5 nm de diámetro (BB International) se incubó con ADN biotinilado de 54 pb y extremos romos (véase el apartado 3.1). La incubación se llevó a cabo en tampón 50 mM Tris-HCl, pH 7,5, 1 mM MgCl<sub>2</sub>, 50 mM KCl y 20% glicerol durante 40 min a temperatura ambiente. Después se añadió DNA-PKcs en una relación molar 1:1 de proteína DNA-PKcs y complejo ADN:estreptavidina:oro y, a continuación, la mezcla de reacción se incubó durante 15 min adicionales.

Para obtener complejos DNA-PKcs:Ku:ADN:estreptavidina:oro los conjugados estreptavidina:oro (BB International) se diluyeron previamente en tampón 20 mM HEPES, pH 7,4, 200 mM KCl, 0,5 mM EDTA, 0,01% BOG y 0,1 mM DTT. Después los conjugados diluidos se incubaron con el complejo DNA-PKcs:Ku:ADN ensamblado sobre ADN de tipo corto. La incubación se prolongó durante 20 min a temperatura ambiente.

### 3.9.6 Ensayo de defosforilación de Artemis

La defosforilación de Artemis se llevó a cabo en un volumen final de 10 µl de una solución que contenía 50 mM Tris-HCl, pH 7,5, 100 mM NaCl, 2 mM DTT, 0,01 mM EGTA, 2 mM MnCl<sub>2</sub>, 200 unidades de fosfatasa-λ y 6 µg de Artemis (28,6 µM). La fosfatasa-λ fue amablemente proporcionada por la Dra. C. Vaughan del "Institute of Cancer Research" en Londres. La solución se incubó a 4°C toda la noche con agitación. Posteriormente la mezcla se analizó por electroforesis en condiciones desnaturizantes.

### 3.9.7 Análisis de la muestra de Artemis mediante cromatografía de exclusión molecular

La proteína purificada Artemis se analizó mediante cromatografía de exclusión molecular en una columna Superdex 200 PC3.2/30 (GE Healthcare). Primero la columna se equilibró con tampón 10 mM Tris-HCl, pH 7,5, 100 mM NaCl, 0,5 mM EDTA y 10 mM MgCl<sub>2</sub>. Después se llevó a cabo la calibración de la columna utilizando un estándar de exclusión molecular que contenía tiroglobulina (670 kDa),  $\gamma$ -globulina (158 kDa), ovoalbúmina (44 kDa), mioglobulina (17 kDa) y vitamina B-12 (1,35 kDa) (BIO-RAD). Para preparar la muestra de Artemis que se aplicó a la columna 3  $\mu$ l de la proteína Artemis (28,6  $\mu$ M) se diluyeron 1:10 en tampón 10 mM Tris-HCl, pH 7,5, 100 mM NaCl, 0,5 mM EDTA, 10 mM MgCl<sub>2</sub> y 20% glicerol. A continuación la muestra se centrifugó a 4°C durante 15 min a 13.000 rpm (centrífuga EPPENDORF 5804R, rotor F45-30-11) y se aplicó a la columna. Durante el análisis se recogieron fracciones de 50  $\mu$ l y se midió la absorbancia a 280 nm.

### 3.9.8 Interacción de Artemis con ADN

La muestra de Artemis se diluyó a una concentración final de 4,08  $\mu$ M en tampón 10 mM Tris-HCl, pH 7,5, 100 mM NaCl, 0,5 mM EDTA, 10 mM MgCl<sub>2</sub> y 20% glicerol. A continuación 3  $\mu$ l de la muestra de Artemis diluida (4,08  $\mu$ M) se incubaron con 3  $\mu$ l de ADN biotinilado de 54 pb y extremos romos (véase el apartado 3.1). El análisis de la capacidad de interacción con ADN también se llevó a cabo con la proteína Artemis que había sido tratada previamente con fosfatasa- $\lambda$ . En este caso 1,8  $\mu$ l de la mezcla obtenida tras el ensayo de defosforilación se incubaron con la misma cantidad del fragmento de ADN descrito anteriormente. En todos los casos las diferentes mezclas de proteína y ADN se incubaron a 25°C durante 20 min y a continuación se analizaron por electroforesis en condiciones nativas. Una vez llevada a cabo la electroforesis el gel se tiñó con bromuro de etidio para detectar la presencia del ADN. Posteriormente se eliminó el bromuro de etidio y el gel se tiñó de nuevo con plata para visualizar las proteínas.

### 3.9.9 Interacción entre DNA-PKcs y Artemis

Para llevar a cabo este análisis 3  $\mu\text{l}$  de DNA-PKcs (2,2  $\mu\text{M}$ ) se incubaron bien con 2,5  $\mu\text{l}$  Artemis (4,08  $\mu\text{M}$ ), bien con 1,5  $\mu\text{l}$  de la solución obtenida tras llevar a cabo el ensayo de defosforilación de Artemis. La relación molar de DNA-PKcs y Artemis en las dos mezclas de reacción correspondería aproximadamente a una relación 1:1,5 respectivamente. En ambos casos la incubación se prolongó durante 1 hora a 25°C sometiendo las distintas mezclas a agitación constante y, a continuación, se analizaron por electroforesis en condiciones nativas.

Para evidenciar el cambio de movilidad que sufría DNA-PKcs en presencia de Artemis el gel nativo en el que se aplicaba la mezcla de DNA-PKcs y Artemis se analizó por "Western-blot". Como anticuerpo primario se utilizó un anticuerpo monoclonal frente a DNA-PKcs (Ab-2, Oncogene). En este caso la mezcla de DNA-PKcs y Artemis se preparó a partir de 0,75  $\mu\text{l}$  de la proteína DNA-PKcs (2,2  $\mu\text{M}$ ) que se incubaron con 1  $\mu\text{l}$  de la proteína Artemis (4,08  $\mu\text{M}$ ). La mezcla se incubó durante 1 hora a 25°C sometiéndola a agitación constante.

### 3.9.10 Interacción entre DNA-PKcs y Artemis mediante tinción reversa de geles nativos

Para realizar este ensayo 5,5  $\mu\text{l}$  de DNA-PKcs (2,2  $\mu\text{M}$ ) se incubaron con 4,5  $\mu\text{l}$  de Artemis (4,08  $\mu\text{M}$ ) durante 1 hora a 25°C sometiendo la mezcla a agitación constante. A continuación la mezcla se analizó por electroforesis en condiciones nativas y, tras la electroforesis, las proteínas se visualizaron mediante tinción reversa del gel. Después se identificaron las bandas de interés y se recortaron del gel. Cada banda se sometió primero a dos lavados de 3 min de duración cada uno con 100  $\mu\text{l}$  PBS 1X y 100 mM EDTA. A continuación se llevaron a cabo otros dos lavados de 1 min de duración cada uno con 100  $\mu\text{l}$  de agua mQ. Después las bandas se trocearon concienzudamente y a cada una de ellas se le añadieron 17  $\mu\text{l}$  de tampón 40 mM Tris-HCl, pH 8,3 y 310 mM glicina. Las distintas mezclas que contenían las bandas troceadas se incubaron durante 30 min a 25°C. Simultáneamente se sometieron a agitación para

favorecer la difusión de las proteínas desde la matriz del gel al tampón. Tras la incubación las distintas mezclas se centrifugaron y se recuperaron los sobrenadantes que se analizaron por electroforesis en condiciones desnaturalizantes.

### 3.10 Preparación de muestras mediante tinción negativa para su observación en el microscopio electrónico

Se han analizado diversas muestras en el microscopio electrónico mediante tinción negativa. En todos los casos la preparación de las muestras se realizó de la siguiente manera. Primero sobre rejillas de cobre-paladio de 400 agujeros (Gilder Grids) se depositó una película de carbón por flotación. Posteriormente e inmediatamente antes de utilizarse la rejilla se sometió a una descarga iónica durante 30 segundos. Con esto la película de carbón se hizo más hidrofílica y se favoreció por tanto la adsorción del espécimen. A continuación la rejilla se incubó con la muestra durante aproximadamente 2 min. El volumen y concentración de la muestra con la que se incubó la rejilla han variado dependiendo de cada espécimen. En todos los casos el volumen ha estado comprendido entre 3 y 5  $\mu\text{l}$ , y la concentración de la muestra ha variado entre 0,1 e 1  $\mu\text{M}$ . Tras la incubación con la muestra la rejilla se secó ligeramente. Para ello se utilizó papel de filtro (Albet) de manera que únicamente quedara una película de muestra sobre la rejilla. A continuación la rejilla se lavó dos veces con agua mQ y de nuevo se retiró el exceso de solución con papel de filtro. Después la rejilla se tiñó negativamente con acetato de uranilo al 1 ó 2% (p/v) durante 1 min aproximadamente. Por último se retiró el exceso de agente de tinción y la rejilla se dejó secar al aire. Las diferentes muestras que se analizaron mediante tinción negativa se prepararon de la siguiente manera:

Heterodímero Ku y complejo Ku:ADN: Las rejillas del heterodímero Ku se prepararon a partir 2,5  $\mu\text{l}$  de la proteína purificada que se diluyeron en 6,5  $\mu\text{l}$  de tampón 50 mM Tris-HCl, pH 7,5, 1 mM  $\text{MgCl}_2$ , 50 mM KCl y 10 mM EDTA. Este tampón contenía bolas de látex de un diámetro aproximado de 0,1  $\mu\text{m}$  ("Latex beads" LB1, SIGMA) diluidas 1:20.000 respecto a la concentración comercial.

Las rejillas del complejo Ku:ADN se prepararon a partir de la mezcla de reacción descrita en el apartado 3.9.1.

Complejo Ku:ADN analizado mediante RCT: La rejilla se preparó a partir de una mezcla que contenía 2  $\mu$ l de ADN de 54 pb y extremos romos no biotinilado (véase el apartado 3.1), 2  $\mu$ l del heterodímero Ku purificado, 1,1  $\mu$ l de la proteína DNA-PKcs purificada y 2  $\mu$ l de tampón 50 mM Tris-HCl, pH 7,5, 1 mM  $MgCl_2$ , 50 mM KCl, 10 mM EDTA y 20% glicerol.

DNA-PKcs: Las rejillas se prepararon a partir de 2  $\mu$ l de la proteína DNA-PKcs purificada que se diluyeron en 8  $\mu$ l de tampón 10 mM Tris-HCl, pH 7,5, 100 mM NaCl, 10 mM  $MgCl_2$  y 20% glicerol.

Anticuerpo:DNA-PKcs: Las rejillas se prepararon a partir de la mezcla de reacción descrita en el apartado 3.9.3.

DNA-PKcs:ADN: Para ensamblar este complejo la proteína DNA-PKcs se incubó con ADN de 54 pb y extremos romos no biotinilado (véase el apartado 3.1). La relación molar de proteína:ADN en la mezcla de reacción fue 1:6 respectivamente. A continuación la mezcla se diluyó a una concentración final de proteína de 0,2 mg/ml en 50 mM Tris-HCl pH 7,5, 50 mM KCl, 1 mM  $MgCl_2$  y 20% glicerol.

DNA-PKcs:ADN:estreptavidina:oro: Las rejillas se prepararon a partir de la mezcla de reacción descrita en el apartado 3.9.5.

DNA-PKcs:Ku:ADN: Las rejillas del complejo DNA-PKcs:Ku:ADN (ensamblado sobre ADN de tipo largo o corto), se prepararon a partir de 2,5  $\mu$ l de la muestra de complejo correspondiente más 2,5  $\mu$ l de tampón 50 mM HEPES, pH 7,2 y 90 mM KCl. Las partículas del complejo DNA-PKcs:Ku:ADN y del complejo dimérico de DNA-PKcs:Ku:ADN se obtuvieron a partir de la misma rejilla de microscopía electrónica.

DNA-PKcs:Ku:ADN:estreptavidina:oro: Las rejillas del complejo DNA-PKcs:Ku:ADN:estreptavidina:oro ensamblado sobre ADN de tipo corto se prepararon a partir de la mezcla de reacción descrita en el apartado 3.9.5. Las partículas del complejo DNA-PKcs:Ku:ADN y del complejo dimérico de DNA-PKcs:Ku:ADN marcadas con oro se obtuvieron a partir de la misma rejilla de microscopía electrónica.

DNA-PKcs:Ku:ADN autofosforilado: 2,5  $\mu\text{l}$  del complejo DNA-PKcs:Ku:ADN ensamblado sobre ADN de tipo largo se incubaron con 0,25  $\mu\text{l}$  de ATP 20 mM y 0,2  $\mu\text{l}$  de  $\text{MgCl}_2$  50 mM durante 20 min a temperatura ambiente. Tras la incubación se preparó la rejilla a partir de la mezcla de reacción.

Artemis: Para preparar la rejilla se utilizaron 5  $\mu\text{l}$  de una dilución 1:100 de la proteína Artemis en tampón 20 mM HEPES, pH 7,5, 0,5 M KCl y 1 mM EDTA.

Las rejillas de tinción negativa de los diferentes especímenes se observaron en un microscopio electrónico de transmisión JEOL 1230 estabilizado a 100 kV. El microscopio estaba equipado con un filamento de wolframio y un goniómetro eucéntrico que permitía la inclinación de la muestra. En general las micrografías se tomaron a una magnificación de 50.000 aumentos sin inclinar el brazo portamuestras, aunque en aquellos casos en los que fue necesario también se tomaron micrografías con inclinación. Esto permitió aumentar la diversidad de imágenes de proyección presentes en las micrografías. Para resolver la estructura del complejo Ku:ADN mediante RCT fue necesario tomar pares de micrografías. Las dos micrografías de cada par correspondían a la misma zona de la rejilla, pero una de ellas se tomó inclinando el brazo portamuestras  $35^\circ$  mientras que la otra micrografía se tomó sin ninguna inclinación. El análisis estructural del complejo DNA-PKcs:ADN se llevó a cabo en un microscopio electrónico de transmisión Philips CM100 estabilizado a 100 kV. Este microscopio también estaba equipado con un filamento de wolframio. Las micrografías correspondientes al complejo DNA-PKcs:ADN se tomaron a 39.000 aumentos de magnificación. En todos los casos las micrografías se tomaron en condiciones de mínima dosis (aproximadamente  $10 \text{ e}^-/\text{\AA}^2$ ) y se registraron en películas Kodak SO-163.



### 3.11 Preparación de muestras mediante criomicroscopía electrónica para su observación en el microscopio electrónico

Para los ensayos de criomicroscopía electrónica se utilizaron las mismas rejillas descritas en el apartado 3.10. La rejilla se sometió primero a una descarga iónica. A continuación se sujetó mediante una pinza que se colocó en un aparato similar a una guillotina. Este aparato permite que la pinza con la rejilla sujeta descienda de manera extremadamente rápida. La muestra se aplicó sobre la rejilla y se dejó en contacto durante aproximadamente 2 min. Después se retiró el exceso de muestra con papel de filtro (Nº 1, Whatman) dejando una fina película de solución en la superficie de la rejilla. Inmediatamente después se accionaba el mecanismo de la guillotina. De esta manera la rejilla se sumergió a gran velocidad en etano líquido enfriado a la temperatura del nitrógeno líquido (-180°C). El espécimen queda por tanto incluido en una capa de hielo amorfo que conserva su estado nativo. A partir de este momento la rejilla se mantuvo a la temperatura del nitrógeno líquido incluso durante su observación en el microscopio electrónico.

En el caso de la muestra correspondiente al complejo DNA-PKcs:Ku:ADN se incluyó un paso de diálisis previo a la congelación. De esta manera se eliminaron componentes del tampón que interferían con el proceso de congelación. Un volumen de 10 µl de la muestra correspondiente al complejo DNA-PKcs:Ku:ADN se dializó frente a 200 ml de tampón 20 mM HEPES, pH 7,4, 200 mM KCl, 0,1 mM DTT y 0,5 mM EDTA durante 20 min a temperatura ambiente. Para llevar a cabo la diálisis se utilizó una membrana de diálisis de 3,5 kDa de límite de separación (Slide-A-Lyzer/Mini Dialysis Units, PIERCE). Tras la diálisis se recuperó la muestra y se congeló. La congelación en este caso se llevó a cabo utilizando un dispositivo comercial de inmersión rápida (Cryoplunge, GATAN).

La proteína DNA-PKcs se purificó en tampón 25 mM HEPES, pH 7,6, 50 mM KCl, 1 mM MgCl<sub>2</sub>, 1 mM DTT, 0,5 mM EDTA y 20% glicerol. La presencia

de glicerol en el tampón interfiere con el proceso de congelación. Por ello, con el objetivo de eliminar el glicerol de la muestra, se llevaron a cabo dos lavados como los que se realizan de manera rutinaria durante la preparación de muestras mediante tinción negativa (véase el apartado 3.10). Un volumen de 4  $\mu$ l de la proteína DNA-PKcs purificada se aplicaron a una rejilla que había sido sometida a una descarga iónica. A continuación se realizaron dos lavados con tampón 25 mM HEPES, pH 7,6, 50 mM KCl, 1 mM  $MgCl_2$ , 1 mM DTT y 0,5 mM EDTA. Por último se llevó a cabo la congelación de la muestra.

Las diferentes muestras congeladas se observaron en un microscopio electrónico de transmisión Tecnai G<sup>2</sup> FEG 200 (FEI) estabilizado a 200 kV. El microscopio estaba equipado además con un anticontaminador GATAN. Para poder introducir las rejillas en el microscopio electrónico se utilizó como brazo portamuestras un “criobrazo” GATAN modelo 626.53P50 y el dispositivo de “criotransferencia” asociado. Las micrografías de los diferentes especímenes se tomaron a una magnificación de 50.000 aumentos. El rango de desenfoque con el que se tomaron las micrografías de los diferentes especímenes congelados varió entre 3 y 6  $\mu$ m. Limitaciones debidas a la cantidad de muestra disponible y su concentración impidieron la obtención de más imágenes del complejo DNA-PKcs:Ku:ADN. Las micrografías se tomaron en condiciones de mínima dosis y se registraron en películas Kodak SO-163.

### 3.12 Digitalización de micrografías electrónicas

Para digitalizar las micrografías obtenidas a partir de especímenes teñidos negativamente se utilizó un escáner MINOLTA Dimage Scan Multi Pro. Las micrografías se digitalizaron a un tamaño del paso de 10  $\mu$ m. Esto, combinado con la magnificación con la que se toman las micrografías, define las dimensiones del píxel a nivel del espécimen. En el análisis estructural del complejo DNA-PKcs:ADN las micrografías se digitalizaron en un escáner Leafscan45 (Leaf Systems Inc.).

Las micrografías obtenidas a partir de la muestra de DNA-PKcs congelada se digitalizaron en un escáner Photoscan TD (Zeiss-Intergraph) a un tamaño del paso de 7  $\mu\text{m}$ . Para digitalizar las micrografías correspondientes al complejo DNA-PKcs:Ku:ADN congelado se utilizó un escáner MINOLTA Dimage Scan Multi Pro. Estas micrografías se digitalizaron a un tamaño del paso de 10  $\mu\text{m}$ .

Antes de llevar a cabo la selección de partículas, tanto en micrografías obtenidas a partir de especímenes teñidos negativamente como congelados, se calculó la transformada de Fourier para cada micrografía y de esta manera fue posible descartar las micrografías defectuosas.

### 3.13 Procesamiento digital de imágenes

Durante el desarrollo de esta tesis doctoral se han estudiado diferentes especímenes utilizando microscopía electrónica y procesamiento digital de partículas individuales. Las diferentes muestras se han preparado mediante tinción negativa o criomicroscopía electrónica, pero en todos los casos el esquema general de procesamiento para las diferentes partículas ha sido muy similar. Este esquema coincide con la estrategia de procesamiento definida en el paquete informático EMAN (<http://blake.bcm.tmc.edu/eman/>) (Ludtke et al., 1999) que se ha explicado previamente en el capítulo de Introducción. Por ello los diferentes comandos a los que se hace referencia corresponden al paquete informático EMAN a no ser que se especifique lo contrario. Además también se han utilizado herramientas disponibles en los paquetes informáticos SPIDER ([http://www.wadsworth.org/spider\\_doc/spider/docs/spider.html](http://www.wadsworth.org/spider_doc/spider/docs/spider.html)) (Frank et al., 1996), XMIPP (<http://xmipp.cnb.csic.es/>) (Marabini et al., 1996; Sorzano et al., 2004) y BSOFT (Heymann, 2001; Heymann y Belnap, 2007) (<http://lsbr.niams.nih.gov/bsoft/bsoft.html>) así como el programa CTFFIND3b (Mindell y Grigorieff, 2003) que se encuentra disponible en: [http://emlab.rose2.brandeis.edu/grigorieff/download\\_ctf.html](http://emlab.rose2.brandeis.edu/grigorieff/download_ctf.html). Este programa fue amablemente proporcionado por el Dr. J. J. Fernández de la Universidad de Almería. Cada paquete informático presenta sus ventajas y sus inconvenientes

respecto a los demás, por ello en muchos casos es necesario combinar las herramientas disponibles en los diferentes paquetes informáticos.

### 3.13.1 Selección de partículas individuales

Las partículas correspondientes a moléculas individuales del espécimen se seleccionaron y extrajeron de las micrografías utilizando el programa *boxer*. En el caso de las estructuras del complejo Ku:ADN resueltas mediante RCT se recurrió al programa *micrograph\_mark* del paquete informático XMIPP. Este programa permite el marcaje simultáneo de pares de micrografías tomadas con diferente grado de inclinación. La selección de partículas correspondientes al complejo DNA-PKcs:ADN también se llevó a cabo con el programa *micrograph\_mark*.

### 3.13.2 Corrección de CTF

En determinados casos se ha llevado a cabo la corrección de CTF. Esta corrección es fundamental en el caso de las partículas de criomicroscopía electrónica. Primero se calcularon los parámetros de la CTF utilizando el programa CTFFIND3b. Estos parámetros iniciales se ajustaron de manera más precisa con el programa *ctfit* del paquete informático EMAN. Después se llevó a cabo la corrección de la CTF invirtiendo el signo de aquellas fases que tenían signo negativo. Para ello se utilizó bien el comando *ctfit* bien el comando *bctf* del paquete informático BSOFT. A partir de las micrografías con la CTF corregida se extrajeron las partículas.

### 3.13.3 Preparación de las imágenes

Antes de comenzar el análisis de partículas individuales es necesario preparar las imágenes. Esta etapa previa de preparación consta de diferentes pasos. Primero las partículas se escalaron a la mitad del tamaño original (comando *proc2d*). A continuación las imágenes se normalizaron a media 0 y desviación típica 1 (comando *proc2d*). La normalización permite que los distintos niveles de grises que reflejan diferencias de masa sean equivalentes

entre partículas de diferentes micrografías. Después las imágenes se centraron y se les aplicó una máscara (comando *cenalignint*) de manera que se restringía el área de análisis en cada partícula a la región donde se encontraba la proyección del espécimen. Por último las partículas se filtraron utilizando un filtro pasa-bajas y pasa-altas (comando *proc2d*). El filtro pasa-bajas elimina altas frecuencias correspondientes a resoluciones que no pueden alcanzarse bien mediante tinción negativa, bien por criomicroscopía electrónica. El filtro pasa-altas elimina bajas frecuencias correspondientes a los fenómenos de dispersión inelástica que se producen en el microscopio electrónico. En el caso de las partículas correspondientes al complejo DNA-PKcs:ADN, la etapa previa de preparación se llevó a cabo utilizando diferentes comandos disponibles en el paquete informático XMIPP.

### 3.13.4 Análisis bidimensional de partículas individuales

El análisis bidimensional constituye la primera aproximación al análisis de un conjunto de partículas dado. En el caso de la proteína Artemis y del complejo DNA-PKcs:Ku:ADN ensamblado sobre ADN de tipo largo únicamente se ha realizado un análisis bidimensional. Se han llevado a cabo dos tipos principales de análisis bidimensionales:

#### Análisis de la simetría rotacional

El análisis rotacional de las imágenes correspondientes a dímeros del complejo DNA-PKcs:Ku:ADN permitió determinar el tipo de simetría rotacional mayoritario presente en estas imágenes. Este análisis se realizó utilizando diversos comandos presentes en el paquete informático XMIPP. Primero se calcularon los espectros rotacionales de las 2.152 imágenes correspondientes a dímeros del complejo DNA-PKcs:Ku:ADN teñidas negativamente. A continuación se llevó a cabo la clasificación de los 2.152 espectros rotacionales. Para ello se utilizó el método de clasificación denominado KerDenSOM (del inglés “**K**ernel **d**ensity **s**elf **o**rganizing **m**ap”) (Kohonen, 1990; Pascual-Montano et al., 2001). Este método de clasificación se basa en algoritmos tipo “mapa auto-organizativo” que funcionan a modo de red neuronal. La complejidad del

conjunto inicial de datos se reduce y se obtiene una matriz bidimensional de vectores que condensan la variabilidad presente en la muestra. Los vectores que se disponen más próximos en la matriz son más parecidos entre sí que los que se sitúan más alejados. Este tipo de clasificación se lleva a cabo sin asumir de antemano un número determinado de clases presentes en los datos iniciales. Por tanto la clasificación de los espectros rotacionales permite evidenciar el tipo de simetría rotacional mayoritario presente en las partículas iniciales.

#### Alineamiento, clasificación y promediado bidimensional libres de patrón

Los algoritmos de alineamiento y clasificación libres de patrón permiten obtener imágenes medias de las clases más representativas presentes en el conjunto inicial de partículas. Las imágenes medias se obtienen independientemente de cualquier conocimiento previo de la estructura tridimensional del espécimen. Por ello la comparación entre este tipo de imágenes medias y los resultados obtenidos tras el proceso de refinamiento angular iterativo sirve como criterio para evaluar la calidad del procesamiento.

La baja relación señal-ruido en las imágenes de partículas individuales obtenidas mediante microscopía electrónica justifica la necesidad de estos algoritmos. Esto es especialmente significativo en el caso de las imágenes que se obtienen en criomicroscopía electrónica. Para solventar este problema se recurre al promediado de imágenes. Existen diferentes métodos que, combinando de manera iterativa el alineamiento y la clasificación de imágenes en dos dimensiones, permiten obtener imágenes medias de cada clase (Frank, 2006). Dentro de estos métodos destacan aquellos que llevan a cabo un refinamiento utilizando múltiples referencias. Estos métodos basados en refinamientos multirreferencia son ideales para lidiar con las mezclas que frecuentemente aparecen en el conjunto inicial de partículas individuales. En los refinamientos multirreferencia el conjunto inicial de partículas se compara con un número definido de imágenes referencia. Las imágenes referencia representan la diversidad presente en las partículas iniciales. Cada partícula se alinea respecto a todas las referencias y se asigna a aquella con la que presenta un mayor parecido. Se generan por tanto tantas clases como

imágenes referencia se utilicen. Después se promedian las partículas dentro de cada clase y se obtiene una nueva imagen media que servirá como referencia mejorada en la siguiente iteración. El proceso se detiene cuando no hay cambios en la asignación de las partículas a las diferentes clases, y por tanto las imágenes medias de cada clase no cambian más.

Para llevar a cabo los refinamientos multirreferencia de los diferentes conjuntos de partículas individuales se han utilizado los siguientes métodos:

- *Máxima verosimilitud*

Muchos algoritmos de refinamiento multirreferencia buscan maximizar el valor de cross-correlación entre la partícula y la referencia durante los procesos de alineamiento y clasificación. Sin embargo, el alto nivel de ruido en las imágenes de microscopía electrónica hace que la asignación y el alineamiento de las partículas siguiendo criterios de cross-correlación puedan ser incorrectos. Además el resultado final es muy dependiente de las referencias iniciales que se utilicen, especialmente cuando las partículas que se refinan tienen una baja relación señal-ruido. El alineamiento y la asignación de las partículas a las diferentes clases pueden hacerse también utilizando criterios de máxima verosimilitud. El programa *ml\_align2d* (del inglés “**Maximum-likelihood alignment in 2D**”) (Scheres et al., 2005a; Scheres et al., 2005b) disponible en el paquete informático XMIPP, permite llevar a cabo un refinamiento multirreferencia utilizando estos criterios. El algoritmo implementado en el comando *ml\_align2d* maximiza la probabilidad de observar los datos experimentales dado un modelo estadístico. Este modelo incluye una descripción formal del ruido además de las imágenes medias. La maximización se lleva a cabo mediante el algoritmo de “expectation-maximization” (Dempster et al., 1977) en el que las decisiones de alineamiento y asignación no se toman en función de pequeñas diferencias en el valor de cross-correlación, sino que se calcula la probabilidad para todas las posibles asignaciones. Cada imagen media que se genera tiene en cuenta la probabilidad de asignación de cada partícula a esa clase. Para generar el conjunto inicial de imágenes medias que sirven como referencias, las partículas iniciales no alineadas se dividen

aleatoriamente en grupos y se calcula la imagen media para cada grupo. Los distintos conjuntos de partículas que han sido analizados con el programa *ml\_align2d* se normalizaron previamente utilizando la opción *Ramp* del comando *normalize*.

- *Métodos de alineamiento y clasificación libres de patrón disponibles en el paquete informático EMAN*

EMAN permite obtener imágenes medias libres de referencia utilizando distintos métodos implementados en los comandos *refine2d.py* y *startnrclasses* (Ludtke et al., 1999; Chen et al., 2006). Las etapas iniciales del programa *refine2d.py* las lleva a cabo el programa *startnrclasses*. El primer paso del proceso de refinamiento multirreferencia consiste en calcular los armónicos radiales y rotacionales para cada una de las partículas. A continuación las partículas se clasifican en función de sus armónicos. Se obtienen entre 10 y 20 clases que representan de manera muy simplificada la variabilidad presente en el conjunto inicial de datos. Después las diferentes partículas dentro de cada clase se alinean y promedian. Este esquema de clasificación repetido es el proceso de clasificación que lleva a cabo el programa *startnrclasses*.

El programa *refine2d.py* refina la clasificación realizada por el programa *startnrclasses* mediante métodos de clasificación adicionales. Las imágenes medias obtenidas con el programa *startnrclasses* se alinean entre sí y se lleva a cabo un proceso equivalente al análisis estadístico multivariante que se denomina SVD (del inglés “**S**ingular **v**alue **d**ecomposition”). Este proceso permite determinar las variables más significativas (imágenes “Eigen”) que condensan la variabilidad presente en el conjunto de imágenes medias. Se genera por tanto un espacio multidimensional en el que cada dimensión corresponde a una de las variables determinadas anteriormente. La característica principal del comando *refine2d.py* radica en que el SVD se aplica a las imágenes medias de las clases y no al conjunto de partículas iniciales. Después cada partícula inicial se alinea respecto a la imagen media más parecida y se representa en el espacio multidimensional definido por las imágenes “Eigen”. A continuación se aplica una clasificación denominada k-



medias al conjunto de partículas iniciales representadas en el espacio multidimensional. Esta clasificación es mucho más precisa que la clasificación previa basada en los armónicos de las partículas. Las partículas se separan entonces en un número final de clases determinado por el usuario. Por último las partículas dentro de cada clase se promedian y se genera un nuevo conjunto de imágenes medias. Estas imágenes medias se analizan de nuevo mediante SVD comenzando así la siguiente iteración. El proceso se repite normalmente entre 8 y 10 veces, produciendo finalmente un conjunto de imágenes medias con una alta relación señal-ruido que representan las diferentes clases que existen en el conjunto inicial de partículas.

### 3.13.5 Reconstrucción tridimensional de partículas individuales

El proceso de refinamiento angular iterativo se llevó a cabo utilizando el comando *refine* del paquete informático EMAN (véase el apartado 1.5.9 de Introducción). Uno de los pasos clave del proceso de refinamiento radica en la elección del volumen inicial, puesto que puede influir significativamente en la evolución del refinamiento y condicionar por tanto la estructura final que se obtiene. El paquete informático EMAN está especialmente diseñado para minimizar el sesgo introducido por la referencia. Aun así durante el desarrollo de esta tesis se han utilizado diversas estrategias para minimizar todavía más si cabe este problema. Para cada estructura siempre se han realizado diversos procesos de refinamiento angular iterativo. En cada proceso de refinamiento se partió de un volumen inicial diferente. Cuando se comprobaba que los diferentes procesamientos evolucionaban hacia la misma solución, se desarrollaba aquel refinamiento que venía de un volumen referencia lo más aséptico posible para obtener la estructura final. Los volúmenes iniciales se generaron de diferentes maneras:

Líneas comunes: Primero se obtuvieron imágenes medias a partir del conjunto inicial de partículas utilizando métodos de clasificación libres de patrón (comando *startnrclasses*). Después las imágenes medias se utilizaron para construir diversos volúmenes modelo mediante algoritmos basados en la técnica de líneas comunes (comando *startAny*).

Elipsoides artificiales: Estos volúmenes correspondían a elipsoides con una distribución de masa gaussiana a los que se les añadía una cantidad variable de ruido.

Ruido: Se empleó además una estrategia completamente novedosa. Esta estrategia consistió en utilizar como volumen inicial ruido aleatorio de características similares al fondo que se observa en las imágenes de tinción negativa. Los resultados que se obtuvieron eran totalmente comparables a los resultados logrados partiendo de otros volúmenes iniciales. La eficacia de esta estrategia se debe a las características del esquema de refinamiento angular iterativo definido en el paquete informático EMAN (comando *refine*). Además, la alta relación señal-ruido de las imágenes de tinción negativa así como el peso molecular de la proteína DNA-PKcs (460 kDa), probablemente también contribuyan a su aplicabilidad.

Durante los diferentes procesamientos, en ningún momento se han utilizado como volúmenes de referencia estructuras resueltas previamente para los diferentes especímenes. De esta manera las reconstrucciones obtenidas no estaban influenciadas por ningún tipo de conocimiento estructural anterior. Además el procesamiento de los diferentes conjuntos de partículas se ha llevado a cabo sin asumir ningún tipo de simetría. Solamente durante las etapas finales del procesamiento del complejo dimérico de DNA-PKcs:Ku:ADN se ha tenido en cuenta la simetría. En estas etapas se aplicó simetría rotacional de orden dos con el objetivo de mejorar el aspecto general de la reconstrucción.

Durante el análisis de determinados especímenes se han llevado a cabo pasos adicionales que escapan de este esquema general de procesamiento y que merecen una explicación más concreta:

### *Análisis del complejo Ku:ADN mediante el método de inclinación cónica al azar o RCT*

Las partículas no inclinadas se utilizaron para generar imágenes medias libres de referencia (comando *startnrclasses* del paquete informático EMAN) y, a continuación, se seleccionaron un total de 7 clases. Las imágenes medias de estas clases recordaban a imágenes de proyección típicas obtenidas previamente durante el refinamiento del complejo Ku:ADN. El proceso de alineamiento y clasificación necesario para obtener imágenes medias conlleva la asignación de determinados ángulos a las partículas no inclinadas. Esos ángulos, junto con el valor del ángulo de inclinación con el que se tomaron las micrografías inclinadas, se aplicaron a las partículas inclinadas correspondientes. Para ello se utilizaron los comandos *align2d* y *centilt* disponibles en el paquete informático XMIPP. Después se reconstruyó el volumen correspondiente a cada clase a partir de las partículas inclinadas (comando *art* del paquete informático XMIPP).

### *Estrategia de procesamiento del complejo DNA-PKcs:ADN*

En los ensayos de unión entre la proteína DNA-PKcs y el ADN la relación molar proteína:ADN era 1:6 respectivamente, por tanto la formación de complejos DNA-PKcs:ADN estaba favorecida. Ensayos bioquímicos similares realizados por otros grupos indicaban la presencia en la mezcla de reacción de moléculas de DNA-PKcs libre no unida a ADN (Yaneva et al., 1997; Hammarsten y Chu, 1998). Por ello, durante el proceso de refinamiento angular iterativo definido en el paquete informático EMAN (comando *refine*), se establecieron criterios restrictivos de manera que los volúmenes generados se obtuvieran a partir de un conjunto congruente de partículas. Esos criterios se aplicaron a dos niveles gracias a la modulación de diferentes parámetros del comando *refine* (véase el apartado 1.5.9 de Introducción). El primer nivel corresponde al paso de generación de imágenes medias para cada clase (opción *classkeep* del comando *refine*). Dentro de cada clase sólo las partículas más parecidas se utilizaron para calcular la imagen media. El segundo nivel corresponde a la etapa de generación del nuevo volumen que se utilizaría

como volumen referencia en la siguiente iteración (opción *hard* del comando *refine*). Sólo las clases más coherentes entre sí se utilizaban para construir este nuevo volumen.

### **3.13.6 Cálculo de la resolución del volumen final**

Una vez obtenida la estructura final se llevó a cabo el cálculo de la resolución (véase el apartado 1.5.10 de Introducción). Para ello se utilizó el comando *eotest* del paquete informático EMAN. Este comando divide las partículas dentro de cada una de las clases obtenidas en la última iteración del proceso de refinamiento en dos grupos, par e impar, y a continuación genera una imagen media para cada grupo. Después construye dos volúmenes independientes uno a partir de las imágenes medias pares y otro a partir de las impares y, finalmente, calcula la correlación entre ambos volúmenes en función de la frecuencia espacial. La resolución se ha estimado a partir de los diferentes criterios para el valor de FSC establecidos en la literatura. Por último el volumen resuelto correspondiente se filtró a la resolución establecida.

### **3.13.7 Procesamiento tridimensional multirreferencia**

En determinados casos también se analizó la posible existencia de más de una solución para un conjunto de datos determinado. Para ello se empleó el comando *multirefine* del paquete informático EMAN. Este comando permite realizar un refinamiento angular iterativo equivalente al del comando *refine* pero utilizando varios volúmenes modelo simultáneamente. De esta manera el conjunto inicial de partículas puede distribuirse entre los diferentes volúmenes modelo en cada iteración. El proceso de refinamiento puede por tanto evolucionar generando diferentes estructuras en el caso de que se parta de un conjunto heterogéneo de partículas.

El esquema general de procesamiento para los diferentes especímenes ha sido muy similar aunque cada procesamiento presenta características propias. Algunos de los datos principales correspondientes a los distintos procesamientos llevados a cabo se detallan en la *Tabla 3.3*.

**Tabla 3.3. Resumen de los principales parámetros asociados a los diferentes procesamientos de partículas individuales (página siguiente).**

La técnica de preparación de la muestra puede ser tinción negativa (Tn) o criomicroscopía electrónica (Crio). Los volúmenes referencia iniciales utilizados en los procesos de refinamiento angular iterativo pueden corresponder a elipsoides artificiales (E), ruido (R), o bien generarse mediante algoritmos basados en la técnica de líneas comunes (L). El análisis bidimensional previo se realizó utilizando los comandos *ml\_align2d* (MI), *refine2d.py* (Ref) y *startnrclasses* (St). Las filas de cuadros sombreados se refieren a aquellos especímenes en los que únicamente se realizó un análisis bidimensional.

### 3.14 Visualización

El principal programa utilizado en la visualización y representación de los volúmenes obtenidos así como de las diferentes estructuras atómicas ha sido Chimera (<http://www.cgl.ucsf.edu/chimera/>) (Pettersen et al., 2004). Además en determinados casos puntuales también se utilizaron los programas VMD (<http://www.ks.uiuc.edu/Research/vmd/>) (Humphrey et al., 1996), Pymol (DeLano Scientific LLC, San Carlos, CA, USA, <http://www.pymol.org/>) y Blender (<http://www.blender.org/>). Para representar los diferentes volúmenes resueltos mediante microscopía electrónica se escogió el umbral que correspondía al 100% de la masa de la proteína o complejo proteico. El cálculo de este umbral se llevó a cabo utilizando el comando *volume* del paquete informático EMAN que asume un valor de densidad media para las proteínas de 1,35 g/ml. En la representación de algunas estructuras se han utilizado también umbrales diferentes con el objetivo de resaltar determinadas características. La “mano” escogida para representar los volúmenes correspondientes al heterodímero Ku y al complejo Ku:ADN fue aquella que mejor se ajustaba a la estructura atómica del núcleo de Ku. La “mano” utilizada para representar todos los volúmenes que contienen la proteína DNA-PKcs coincide con la “mano” absoluta de esta proteína. Esta “mano” ha sido definida a partir de la estructura de DNA-PKcs obtenida mediante microscopía electrónica de cristales bidimensionales. Algunos de los volúmenes resueltos se han depositado en la “Macromolecular Structure Database” (<http://www.ebi.ac.uk/msd/>) (Tabla 3.4).

Proteína o complejo	Técnica de preparación	Número de partículas	Corrección CTF	Resolución y criterio	Volumen inicial	Análisis 2D previo	Ápíxel finales	Inclinación del brazo portamuestras
Ku	Tn	3.419	No	25 Å-0,5	L, E	Mi, Ref	2,12	0°
Ku:ADN	Tn	8.285	No	25 Å-0,5	L, E	Mi, Ref	2,12	0°
Ku:ADN (mediante RCT)	Tn	1.360 pares	No	-	-	St	2,12	0° y 35°
DNA-PKcs	Tn	4.812 (1.727 inclinadas)	Sí	24,4 Å-0,5	E, R	Mi, Ref	4,24	0° y 40°
DNA-PKcs:ADN	Tn	10.468 (3.049 DNA-PKcs) (3.280 DNA-PKcs:ADN)	No	30 Å-0,5 DNA-PKcs 30 Å-0,5 DNA-PKcs:ADN	L	St	4,9	0°
Ab:DNA-PKcs	Tn	3.561	No	30 Å-0,5	L	St	4,24	0°
DNA-PKcs:Ku:ADN (ADN tipo corto)	Tn	14.239	No	25 Å-0,5 21 Å-3 σ	L, E, R	Mi, Ref, St	4,24	0° y 40°
Dímero DNA-PKcs:Ku:ADN (ADN tipo corto)	Tn	2.151	No	33 Å-0,5 28 Å-3 σ	L, E, R	Mi, Ref, St	4,24	0° y 40°
DNA-PKcs	Crio	6.857	Sí	12,46 Å-0,14 13,30 Å-3σ 19,78 Å-0,5	L	St	2,8	0°
DNA-PKcs:Ku:ADN (ADN tipo corto)	Crio	3.864	Sí	32 Å-0,5 28 Å-3 σ	L, E	St	4,24	0°
DNA-PKcs:Ku:ADN (ADN tipo largo)	Tn	1.052	No	-	-	Mi, Ref	4,24	0°
DNA-PKcs:Ku:ADN (ADN tipo largo) autofosforilado	Tn	842	No	-	-	Mi, Ref	4,24	0°
Artemis	Tn	734	No	-	-	Mi, Ref	2,12	0°

“Accesion Code”	“Sample name”
1271	“Ku70-Ku80 heterodimer purified from HeLa cell nuclear extracts bound to DNA”
1270	“Ku70-Ku80 heterodimer purified from HeLa cell nuclear extracts”
1210	“Dimer containing two DNA-bound DNAPKcs-Ku70-Ku80 complexes”
1209	“DNA-bound DNAPKcs-Ku70-Ku80 complex”
1208	“Macromolecular Complex containing DNA, DNA-PKcs, Ku70 and Ku80”
1102	“Human DNA-dependent protein kinase catalytic subunit, DNA-PKcs”

**Tabla 3.4. Datos de las diferentes estructuras depositadas en la “Macromolecular Structure Database”.** Esta tabla recoge el “Accesion Code” y el “Sample name” con el que se pueden identificar las estructuras resueltas depositadas.

### 3.15 Análisis de la varianza

Las estructuras que se resuelven mediante microscopía electrónica y análisis de partículas individuales se generan a partir del promediado de varios miles de imágenes. Según esto la estructura que se considera final refleja la conformación promedio presente en el conjunto de imágenes de proyección utilizadas en el refinamiento. El objetivo de calcular la varianza de una estructura determinada consiste precisamente en obtener información sobre la homogeneidad conformacional del espécimen. La varianza también informa sobre posibles errores en el alineamiento y clasificación de las partículas iniciales durante el proceso de refinamiento. Las zonas más flexibles del espécimen serán aquellas que en la estructura obtenida muestren una mayor varianza. Además las partículas que participen en la generación del volumen y se alineen de forma incorrecta también incrementarán la varianza

tridimensional de la estructura. Por el contrario, si el espécimen exhibe una conformación homogénea y el refinamiento angular iterativo es consistente, la varianza será baja en toda la estructura.

El cálculo de la varianza tridimensional en los volúmenes resueltos para el heterodímero Ku y el complejo Ku:ADN se realizó según el método descrito por P. A. Penczek (Grob et al., 2006; Penczek et al., 2006b). En primer lugar se generaron varios cientos de volúmenes para el heterodímero Ku y para el complejo Ku:ADN. La generación de estos volúmenes se lleva a cabo mediante la técnica de muestreo conocida como “bootstrap”. Aplicando esta técnica a las imágenes utilizadas previamente en el proceso de refinamiento angular iterativo se obtienen varios cientos de conjuntos de partículas. Cada uno de ellos se genera seleccionando imágenes de forma aleatoria y con remplazamiento del total de partículas iniciales. Por tanto todos los conjuntos tienen igual número de partículas, sin embargo las partículas que hay en cada uno son diferentes a las de los demás. También es posible que dentro de un conjunto haya partículas que estén repetidas. Cada uno de los distintos conjuntos de partículas se utiliza para generar un volumen. Finalmente se obtienen varios cientos de volúmenes muy parecidos entre sí pero que no son idénticos. Después los volúmenes se filtran y se utilizan para calcular la varianza tridimensional. La varianza así calculada incluye la variabilidad introducida por otros muchos factores además de la variabilidad debida a la flexibilidad conformacional del espécimen. Algunos de estos factores son el propio agente de tinción, los algoritmos de procesamiento, el proceso de registro de la imagen en la micrografía, etc. Con el objetivo de eliminar la variabilidad presente en las imágenes debida al ruido se lleva a cabo una etapa adicional de corrección. Esta etapa consiste también en calcular la varianza tridimensional para varios cientos de volúmenes. En este caso los volúmenes se generan a partir de imágenes que contienen únicamente el fondo de la micrografía próximo a cada partícula. Estas partículas no incluyen por tanto imágenes de proyección del espécimen. Posteriormente la varianza calculada a partir de estas partículas se resta de la varianza obtenida previamente a partir de las imágenes del espécimen. De esta manera la varianza que se obtiene refleja



mayoritariamente la variabilidad debida a la flexibilidad conformacional del espécimen.

Los ficheros necesarios para realizar el análisis de la varianza descrito anteriormente fueron amablemente proporcionados por P. A. Penczek. El cálculo de la varianza se llevó a cabo utilizando el paquete informático SPIDER. Para ello las imágenes correspondientes al heterodímero Ku y al complejo Ku:ADN se transformaron en formato SPIDER utilizando el paquete informático XMIPP. Estas imágenes habían sido procesadas previamente según el esquema general de procesamiento característico del paquete informático EMAN. Después se llevó a cabo un refinamiento angular de una sola iteración con las partículas en formato SPIDER (comando *projection\_matching* del paquete informático XMIPP). Como volumen modelo se utilizó el volumen final obtenido previamente a partir del procesamiento de cada tipo de imagen. Este refinamiento permitió asignar las imágenes en formato SPIDER a determinadas proyecciones del volumen modelo y, de esta manera, las partículas recuperaron la información correspondiente a los ángulos de Euler. Después se aplicó la técnica de “bootstrap” a las imágenes correspondientes al heterodímero Ku y también a las del complejo Ku:ADN (comando *VA 3Q* del paquete informático SPIDER). Se generaron 500 nuevos conjuntos de imágenes para cada tipo de partícula. Cada conjunto correspondiente a las partículas del heterodímero Ku tenía 2.151 imágenes, mientras que los conjuntos correspondientes a partículas del complejo Ku:ADN estaban formados por 5.841 imágenes. A partir de cada conjunto de datos se generó el volumen correspondiente y, a continuación, los 500 volúmenes de cada tipo de partícula se filtraron a 30 Å de resolución y se calculó la varianza tridimensional (comando *VA 3R* del paquete informático SPIDER).

El paso de corrección se llevó a cabo como se ha descrito anteriormente. Imágenes que contenían únicamente fondo se extrajeron directamente a partir de micrografías correspondientes bien al heterodímero Ku, bien al complejo Ku:ADN. A continuación se aplicó también la técnica de “bootstrap” a estas imágenes y se obtuvieron 500 conjuntos de datos para cada tipo de imagen. El número de imágenes en cada tipo de conjunto de datos fue de 2.151 ó 5.841.

Esto dependió de que las imágenes de fondo se hubieran extraído de micrografías pertenecientes al heterodímero Ku o bien al complejo Ku:ADN respectivamente. Para reconstruir los volúmenes a partir de los distintos conjuntos se asignó a las imágenes de fondo los ángulos de Euler calculados previamente para las imágenes del espécimen. Los volúmenes obtenidos se filtraron a 30 Å de resolución y se calculó su varianza. Por último, para calcular la varianza debida mayoritariamente a la variabilidad conformacional del espécimen, el valor medio de la varianza calculada a partir de imágenes de fondo se restó a la varianza correspondiente obtenida a partir de partículas del espécimen.

### 3.16 Ajuste computacional de estructuras atómicas en volúmenes de microscopía electrónica

Los experimentos de ajuste se realizaron utilizando el comando *colores* (del inglés “**C**orrelation based **l**ow **r**esolution docking”) del paquete informático SITUS (<http://situs.biomachina.org/index.html>) (Wriggers et al., 1999; Chacon y Wriggers, 2002). Este comando permite ajustar estructuras atómicas en volúmenes en los que la estructura atómica no rellena toda la densidad. La estructura atómica se transforma en un cuerpo rígido que se ajusta de la manera más precisa posible en el volumen. Para ello se exploran exhaustivamente todas las posibles translaciones y todas las posibles rotaciones para cada translación de la estructura atómica en el volumen. La principal ventaja de este comando es que permite aplicar un filtro laplaciano. De esta forma se tiene en cuenta la información del contorno de la estructura atómica durante el proceso de ajuste. El resultado es un ajuste más preciso en volúmenes de media y baja resolución.

#### 3.16.1 Ajuste de estructuras atómicas en los volúmenes obtenidos para el heterodímero Ku y el complejo Ku:ADN

Las estructuras atómicas que se emplearon en los experimentos de ajuste corresponden a los ficheros depositados en el “Protein Data Bank”

(<http://www.rcsb.org>) 1JEQ y 1JEY (Walker et al., 2001). El fichero 1JEQ contiene la estructura atómica del núcleo del heterodímero Ku junto con el dominio SAP de Ku70. El fichero 1JEY contiene la estructura atómica del núcleo de Ku unido a una molécula de ADN. El fichero 1JEQ se ajustó en el volumen obtenido para el heterodímero Ku, mientras que el fichero 1JEY se ajustó en el volumen resuelto para el complejo Ku:ADN. En el caso del fichero 1JEY no se tuvo en cuenta la información correspondiente a la molécula de ADN durante los procesos de ajuste. Los ajustes con mayor valor de correlación se reajustaron manualmente para generar el ajuste final. El valor de correlación obtenido para el ajuste final del fichero 1JEQ es 0,66. En el caso del ajuste final del fichero 1JEY el valor de correlación obtenido es 0,75.

Además se llevó a cabo el cálculo de volúmenes diferencia entre los mapas de microscopía electrónica y las estructuras atómicas ajustadas en ellos. Para ello las estructuras atómicas primero se filtraron a 25 Å de resolución (comando *pdb2mrc* del paquete informático EMAN) y, a continuación, se calcularon los volúmenes diferencia (comando *proc3d* del paquete informático EMAN). Se obtuvieron por tanto dos volúmenes diferencia; uno a partir del volumen correspondiente al heterodímero Ku y otro a partir del volumen obtenido para el complejo Ku:ADN. Se realizaron experimentos de ajuste con el objetivo de localizar el dominio helicoidal del CTD de Ku80 en los volúmenes diferencia. En estos experimentos se utilizó una de las estructuras atómicas resueltas para este dominio disponible en el fichero 1Q2Z (Harris et al., 2004). El ajuste se llevó a cabo utilizando el programa *colores* del paquete informático SITUS.

### **3.16.2 Ajuste de estructuras atómicas en el volumen obtenido para la proteína DNA-PKcs mediante criomicroscopía electrónica**

Para llevar a cabo estos experimentos se utilizaron estructuras atómicas de determinadas proteínas que corresponden a dominios homólogos presentes en la proteína DNA-PKcs. Las estructuras atómicas utilizadas se extrajeron de los siguientes ficheros depositados en el “Protein Data Bank”: 1E8X, 1B3U y 1UKL. El fichero 1E8X corresponde a la estructura atómica del dominio catalítico de la proteína PI3Ky (Walker et al., 2000). Utilizando técnicas de “threading” (Kelley

et al., 2000; McGuffin et al., 2000) se combinó la información estructural del fragmento catalítico de la PI3K $\gamma$  con la secuencia de la región homóloga de DNA-PKcs y, de esta manera, se definieron las regiones del fichero PDB 1E8X que se utilizarían posteriormente en la generación de un modelo atómico. El modelo contenía el dominio PIKK mínimo (residuos 727 a 1.092 de la proteína PI3K $\gamma$ ), y la porción de repeticiones helicoidales N-terminales (residuos 546 a 726 de la proteína PI3K $\gamma$ ). Posteriormente este modelo se utilizó en los experimentos de ajuste. A partir de los ficheros 1B3U y 1UKL se obtuvieron estructuras atómicas que contenían repeticiones de motivos helicoidales como las descritas para la proteína DNA-PKcs. El fichero 1B3U corresponde a la estructura atómica de un dominio regulador de la proteína humana PP2A (del inglés “**P**rotein **p**hosphatase **2A**”) (Groves et al., 1999). El fichero 1UKL contiene la estructura atómica del complejo formado por la importina- $\beta$  y SREBP-2 (del inglés “**S**terol **r**egulatory **e**lement-**b**inding protein **2**”) (Lee et al., 2003). En los experimentos de ajuste se utilizó únicamente la estructura atómica correspondiente a la importina- $\beta$ . También se llevaron a cabo experimentos de predicción de estructura secundaria que permitieron generar un modelo atómico para el dominio FATC de DNA-PKcs (McGuffin et al., 2000). Además se generó un modelo atómico para el dominio FAT de esta proteína a partir de varias copias del dominio helicoidal presente en la estructura atómica de la proteína PI3K $\gamma$  (fichero PDB 1E8X). Las estructuras y modelos atómicos descritos se ajustaron en el volumen obtenido para la proteína DNA-PKcs mediante criomicroscopía electrónica. Para ello se utilizó el comando *colores* del paquete informático SITUS. A continuación el ajuste se mejoró manualmente utilizando los programas Pymol y O (<ftp://xray.bmc.uu.se>) (Jones et al., 1991). La obtención de los distintos modelos atómicos y los experimentos de ajuste posteriores fueron realizados por el laboratorio del Profesor L. H. Pearl.

### **3.16.3 Ajuste de una molécula de ADN en la estructura obtenida mediante tinción negativa del complejo DNA-PKcs:ADN**

A partir de estructuras atómicas resueltas para oligonucleótidos de ADN de tipo B se generó una molécula de ADN de 36 pb. Esta molécula se escaló a

una dimensiones finales de 4,9 Å/voxel y se ajustó manualmente en el volumen resuelto mediante tinción negativa para el complejo DNA-PKcs:ADN.

#### **3.16.4 Ajuste de la estructura atómica de Ku en los volúmenes obtenidos para el complejo DNA-PKcs:Ku:ADN**

Para llevar a cabo este experimento se utilizó la estructura atómica correspondiente al fichero PDB 1JEY descrita anteriormente. Durante los ensayos de ajuste no se incluyó la información correspondiente a la molécula de ADN presente en la estructura. El ajuste se llevó a cabo en los volúmenes resueltos para el complejo DNA-PKcs:Ku:ADN mediante tinción negativa y criomicroscopía electrónica. Para ello se utilizó el comando *colores* del paquete informático SITUS. En ambos volúmenes primero se seleccionó la solución que tenía un mayor valor de correlación y, a continuación, esta solución se reajustó manualmente para obtener el ajuste final. En el caso del ajuste en el volumen obtenido mediante criomicroscopía electrónica para el complejo DNA-PKcs:Ku:ADN, primero se calculó el volumen diferencia entre este volumen y el volumen de criomicroscopía electrónica correspondiente a la proteína DNA-PKcs libre. Para ello la estructura de la proteína DNA-PKcs libre se escaló, filtró y alineó adecuadamente y a continuación se calculó el volumen diferencia. Este nuevo volumen evidenciaba la masa adicional presente en el volumen del complejo DNA-PKcs:Ku:ADN. A continuación se llevó a cabo el ajuste de la estructura atómica en el volumen diferencia. Además también se calculó el volumen diferencia entre la estructura de tinción negativa del complejo DNA-PKcs:Ku:ADN y el volumen de criomicroscopía electrónica de la proteína DNA-PKcs libre.

#### **3.16.5 Ajuste del volumen de tinción negativa del complejo DNA-PKcs:Ku:ADN en la estructura del complejo dimérico de DNA-PKcs:Ku:ADN**

Para ello la estructura del complejo DNA-PKcs:Ku:ADN primero se convirtió en un fichero tipo PDB (comando *vol2pdb* del paquete informático SITUS). A continuación la nueva estructura generada se ajustó en la densidad

correspondiente al complejo dimérico de DNA-PKcs:Ku:ADN (comando *colores* del paquete informático SITUS).

### 3.17 Predicción de movimientos macromoleculares utilizando análisis vibracional

El primer paso de este análisis consiste en generar un modelo elástico a partir de la estructura obtenida por microscopía electrónica utilizando redes neuronales (Chacon et al., 2003). De esta manera los volúmenes quedan representados de manera simplificada en el espacio real por una serie de puntos característicos. Estos puntos se conectan por muelles armónicos en función de su distancia relativa para generar un modelo "elastomecánico". A partir del modelo se lleva a cabo el análisis de los modos normales de vibración. Este análisis de mecánica clásica descompone los posibles movimientos del modelo en un conjunto ortonormal de desplazamientos o modos ordenados por frecuencia. Los movimientos que se observan experimentalmente en complejos macromoleculares pueden describirse eficientemente mediante la combinación lineal de modos (Ming et al., 2002).

Se llevó a cabo el análisis vibracional de la estructura resuelta para DNA-PKcs a partir del conjunto de imágenes que también permitió obtener la estructura del complejo DNA-PKcs:ADN. Este análisis fue realizado con la ayuda del Dr. P. Chacón del Centro de Investigaciones Biológicas (CSIC) en Madrid. A partir del volumen de microscopía electrónica se generó un modelo elástico utilizando un total de 1.000 puntos característicos. Este modelo "elastomecánico" se construyó conectando puntos que estuvieran a una distancia menor de 14 Å. Con este modelo elástico se llevó a cabo el análisis vibracional. Los desplazamientos del primer modo se aplicaron al volumen mediante técnicas de interpolación. Las amplitudes de los movimientos se escalaron para visualizar mejor el desplazamiento asociado al modo.

### 3.18 Predicción de la estructura terciaria de Artemis

La secuencia de la proteína Artemis humana (se identifica con el código Q96SD1 en los servidores UniProtKB/Swiss-Prot) se obtuvo a partir del servidor EXPASY (<http://www.expasy.org>). Esta secuencia constituida por 692 aminoácidos se analizó en el meta-servidor de predicción de estructura terciaria Phyre 0.2 (del inglés “**P**rotein **h**omology/**a**nalogy **r**ecognition **e**ngine”) (<http://www.sbg.bio.ic.ac.uk/~phyre/>, Imperial College London) (Kelley et al., 2000; Bennett-Lovsey et al., 2008). El meta-servidor Phyre genera un modelo de estructura terciaria para una secuencia problema gracias al alineamiento con secuencias de las que se conoce su estructura atómica. De esta manera fue posible obtener un modelo atómico del dominio catalítico de la proteína Artemis que incluía los primeros 346 aminoácidos. Este modelo se construyó a partir del alineamiento con la secuencia de la proteína TTHA0252 de *Thermus thermophilus* cuya estructura atómica corresponde al fichero PDB 2DKF (Ishikawa et al., 2006). El valor E del alineamiento fue de  $5,1 e^{-22}$ . El modelo atómico se filtró a una resolución de 25 Å y a continuación se proyectó en diferentes orientaciones (comandos *proc2d* y *project3d* respectivamente disponibles en el paquete informático EMAN).







## 4. Resultados



## 4.1 Análisis estructural del heterodímero Ku

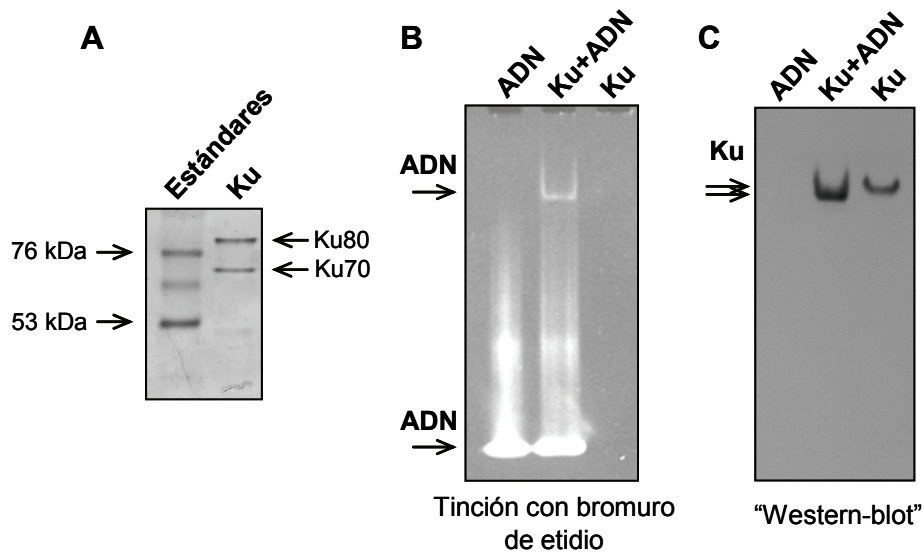
Como se comentó en el capítulo de Introducción existe información estructural a resolución atómica para la mayor parte de la proteína Ku. Esa información proviene de diferentes estructuras que abarcan distintos dominios pero en ningún caso la secuencia completa de la proteína (Walker et al., 2001; Zhang et al., 2001; Harris et al., 2004; Zhang et al., 2004). Se desconoce por tanto la organización tridimensional del heterodímero Ku completo. Además la interacción de la proteína Ku con los extremos del ADN conlleva cambios en su conformación (Lehman et al., 2008). Estos cambios afectan principalmente a los dominios C-terminales presentes únicamente en las proteínas Ku70 y Ku80 de eucariotas superiores (Downs y Jackson, 2004). Los dominios C-terminales de ambas proteínas tienen importancia funcional puesto que participan en la estabilización del heterodímero Ku en el extremo del ADN. Además también intervienen en el reclutamiento posterior de la proteína DNA-PKcs a los sitios donde se produjo la rotura (Gell y Jackson, 1999; Aravind y Koonin, 2000).

Con el fin de conocer la arquitectura global del heterodímero humano Ku completo, libre y unido al ADN, se llevó a cabo su estudio mediante microscopía electrónica y análisis de partículas individuales.

### 4.1.1 Análisis de la interacción de la proteína Ku con ADN

El heterodímero Ku fue purificado por la Dra. L. Spagnolo en el laboratorio del Profesor L. H. Pearl del "Institute of Cancer Research" en Londres. Ku se purificó a partir de extractos nucleares de células HeLa. Al analizar la proteína purificada mediante geles de poliacrilamida en condiciones desnaturizantes, únicamente se distinguían las dos bandas características correspondientes a las proteínas Ku70 y Ku80 (*Figura 4.1.1A*). El heterodímero Ku presenta una gran afinidad por los extremos del ADN (Mimori y Hardin, 1986; Arosio et al., 2002). Con el objetivo de ver si la proteína Ku purificada era funcional en su capacidad de interactuar con extremos de ADN, se llevaron a cabo experimentos de cambio de movilidad electroforética en condiciones nativas. El

heterodímero Ku purificado se incubó con ADN en una relación molar 1:10 de proteína y ADN respectivamente. Después la mezcla se analizó mediante electroforesis en condiciones nativas. El gel primero se tiñó con bromuro de etidio para localizar el ADN (*Figura 4.1.1B*), y posteriormente se llevó a cabo un “Western-blot” utilizando un anticuerpo monoclonal frente a la proteína Ku70 (*Figura 4.1.1C*). Cuando el heterodímero Ku estaba presente en la mezcla de reacción aparecía una nueva banda que contenía ADN en la que también se observó la proteína Ku (*carril “Ku+ADN” de las Figuras 4.1.1B y 4.1.1C*). Esta banda mostraba una migración mayor que la banda correspondiente a la proteína Ku sin ADN (*compárense los carriles “Ku+ADN” y “Ku” de la Figura 4.1.1C*). Además en presencia de ADN se identificó una única banda correspondiente al heterodímero Ku, tanto en el gel nativo teñido con bromuro de etidio (*carril “Ku+ADN” de la Figura 4.1.1B*), como en el “Western-blot” (*carril “Ku+ADN” de la Figura 4.1.1C*). Por tanto, en las condiciones ensayadas es posible ensamblar un complejo homogéneo formado por la proteína Ku y ADN



**Figura 4.1.1. Análisis de la interacción del heterodímero Ku con ADN.** (A)

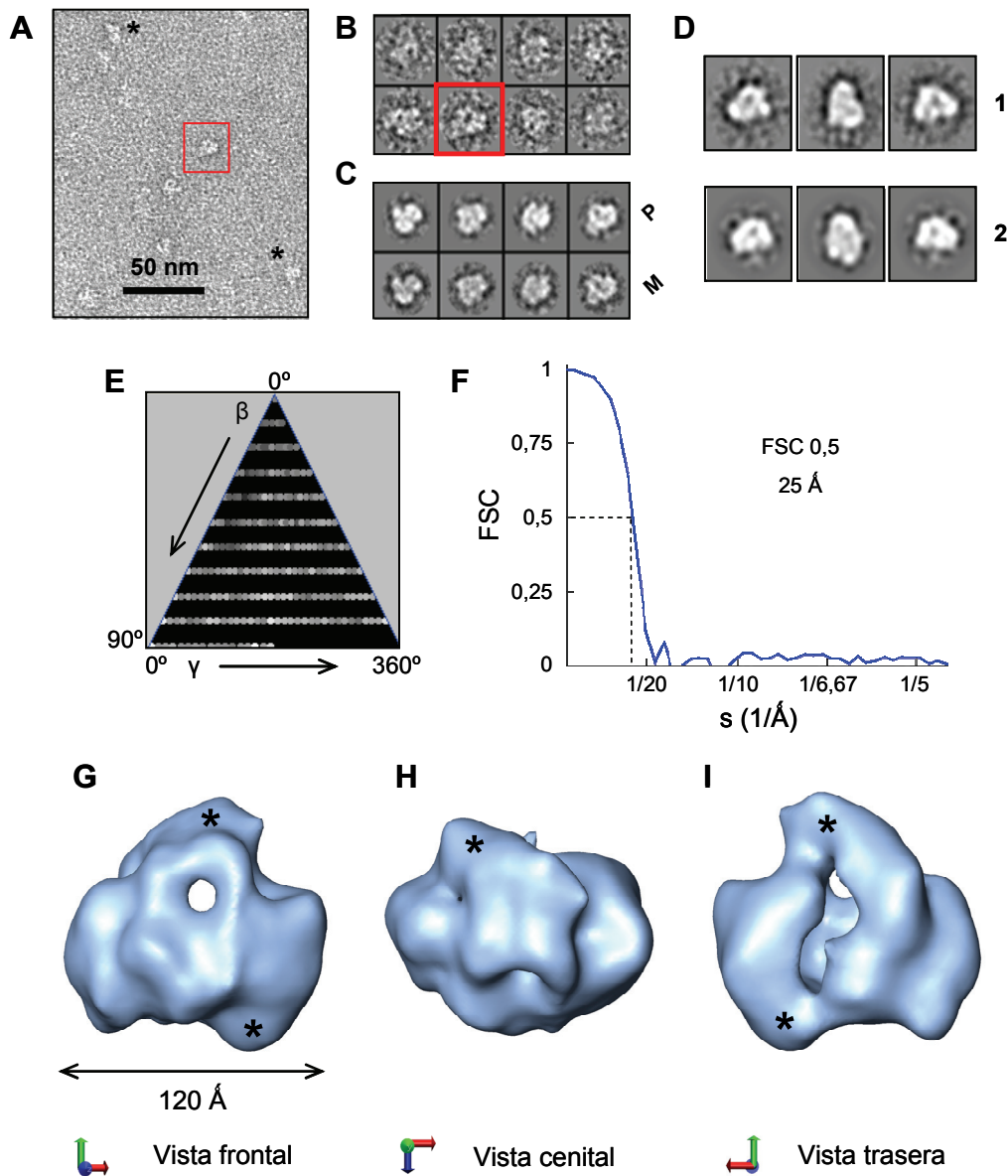
Gel de poliacrilamida en condiciones desnaturizantes de la proteína Ku purificada teñido con plata. (B) Gel de poliacrilamida en condiciones nativas teñido con bromuro de etidio y revelado con luz UV para evidenciar la presencia del ADN. Únicamente cuando la proteína Ku y el ADN están presentes en la mezcla de reacción (*carril “Ku+ADN”*) aparece una nueva banda de ADN. (C) “Western-blot” del gel nativo mostrado en (B) utilizando un anticuerpo monoclonal frente a la proteína Ku70. En presencia de ADN la banda correspondiente al heterodímero Ku presenta un cambio en su movilidad. Los carriles “Ku” y “ADN” contienen las mismas cantidades de Ku y ADN respectivamente que las utilizadas en la mezcla de reacción.

(Ku:ADN). Estas condiciones fueron las mismas que las que se utilizaron en el análisis posterior del complejo Ku:ADN mediante microscopía electrónica.

#### 4.1.2 Reconstrucción tridimensional del heterodímero Ku

El heterodímero Ku tiene un peso molecular aproximado de 160 kDa, lo que limita su análisis estructural mediante criomicroscopía electrónica. Utilizando tinción negativa sin embargo las moléculas individuales del heterodímero Ku se identificaban claramente (*Figura 4.1.2A*). En algunas de estas imágenes era posible reconocer el aspecto general de la estructura atómica del núcleo de Ku (ficheros PDB 1JEQ y 1JEY, Walker et al., 2001, véase el apartado 1.4.1 de Introducción) (*recuadro rojo de las Figuras 4.1.2A y 4.1.2B*). Se seleccionaron un total de 3.419 imágenes correspondientes a moléculas individuales del heterodímero Ku (*Figura 4.1.2B*). Posteriormente se llevó a cabo el análisis de estas imágenes mediante procesos de refinamiento angular iterativo (véase el apartado 3.13.5 de Materiales y Métodos). En la *Figura 4.1.2C* se muestra una galería de proyecciones y medias equivalentes obtenidas en la última iteración del proceso de refinamiento angular iterativo. La semejanza entre proyecciones y medias se debe a la estabilización del proceso de refinamiento. Además, las proyecciones generadas a partir del volumen final son totalmente compatibles con imágenes medias obtenidas utilizando algoritmos de clasificación libres de

**Figura 4.1.2. Reconstrucción tridimensional del heterodímero Ku (página siguiente).** (A) Micrografía obtenida a partir de una rejilla de microscopía electrónica del heterodímero Ku teñida negativamente. Algunas partículas evidentes se indican con “\*”. (B) Galería de partículas individuales correspondientes a moléculas del heterodímero Ku. En (A) y (B) se muestra en un recuadro rojo la misma partícula. El aspecto general de esta partícula concuerda con el de la estructura atómica del núcleo del heterodímero Ku. (C) Galería de proyecciones (P) y medias equivalentes (M) obtenidas en la última iteración del proceso de refinamiento angular iterativo que permitió obtener la estructura final. (D, 1) Galería de imágenes medias generadas utilizando métodos de clasificación libres de patrón presentes en los paquetes informáticos EMAN y XMIPP. (D, 2) Galería de proyecciones obtenidas durante el proceso de refinamiento angular iterativo. (E) Representación de los ángulos de Euler del conjunto de partículas individuales en la última iteración del proceso de refinamiento que dio lugar a la estructura final. Se ha escogido el modo de representación utilizado por el paquete informático EMAN (véase el apartado 1.5.7 de Introducción). (F) Estimación de la resolución para un valor de FSC de 0,5. (G), (H) e (I) Vista frontal, cenital y trasera respectivamente de la reconstrucción tridimensional resuelta para el heterodímero Ku. Las principales regiones de masa que rompen la simetría del modelo se indican con “\*”.

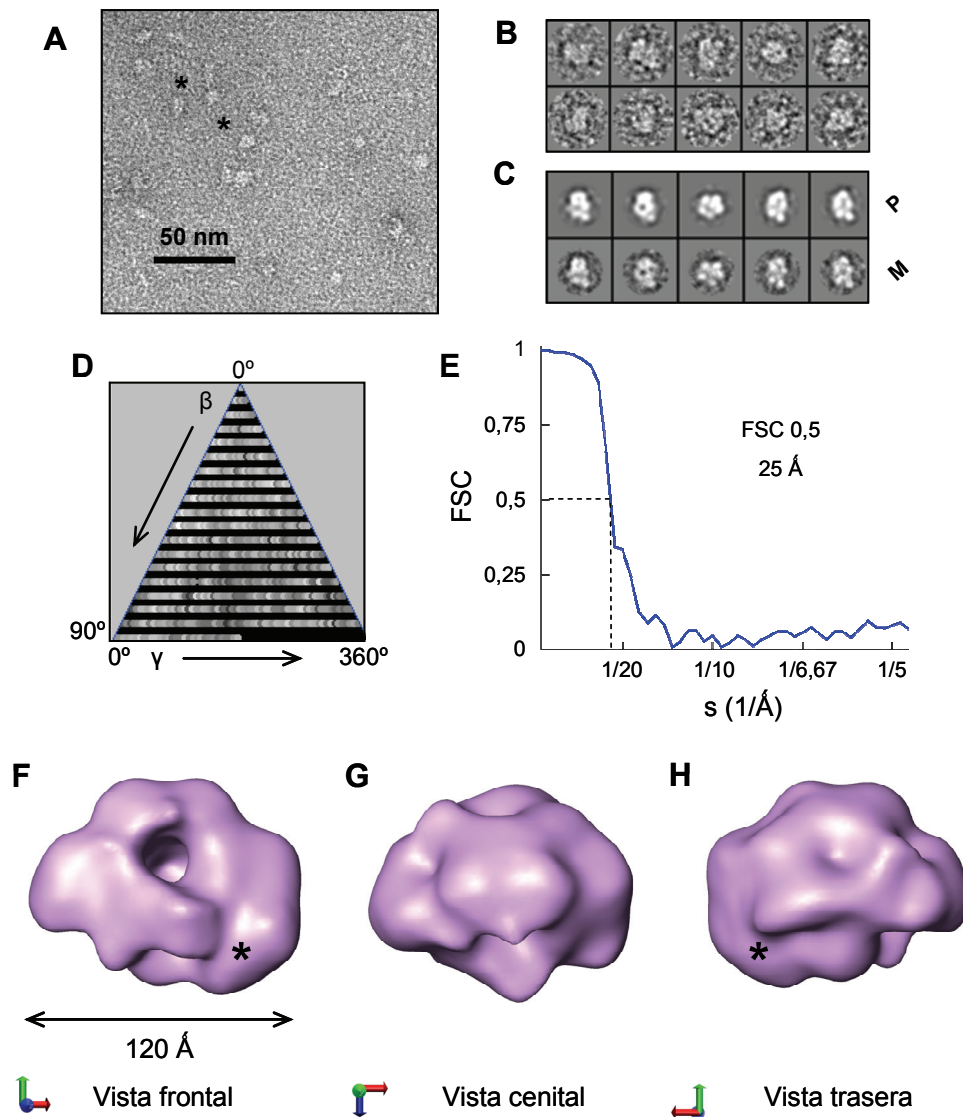


patrón (compárese la galería 1 y la galería 2 de la Figura 4.1.2D). El conjunto inicial de partículas individuales contenía un rango adecuado de ángulos de Euler (Figura 4.1.2E). El volumen obtenido tiene una resolución estimada de 25 Å para un valor de FSC de 0,5 (Figura 4.1.2F). Sus dimensiones aproximadas son 120 Å en el eje longitudinal, 100 Å de altura y un grosor de unos 70 Å (Figuras 4.1.2G, 4.1.2H y 4.1.2I). En la estructura resuelta se aprecian las características principales de la estructura atómica del núcleo de Ku (véase el apartado 1.4.1 de Introducción). Presenta un agujero central limitado por un pequeño arco de densidad en la parte superior (Figura 4.1.2G). El agujero se

asienta sobre la masa principal del volumen que consiste en una masa extendida a lo largo de un eje longitudinal. El volumen obtenido presenta cierta simetría. Masas evidentes que se disponen en el eje longitudinal del volumen y también en la parte trasera del agujero rompen la simetría de la estructura resuelta (se indican con “\*” en las Figuras 4.1.2G, 4.1.2H y 4.1.2I). La presencia de simetría en el volumen obtenido es compatible con las estructuras atómicas resueltas para el núcleo del heterodímero Ku (véase el apartado 1.4.1 de Introducción). Estas estructuras presentan pseudo simetría rotacional que, a una resolución equivalente a la que se obtiene a partir de especímenes teñidos negativamente, es compatible con la simetría que se observa en el volumen resuelto.

#### 4.1.3 Reconstrucción tridimensional del complejo Ku:ADN

Para ensamblar el complejo Ku:ADN se utilizaron las condiciones descritas previamente en el apartado 4.1.1. En estas condiciones toda la proteína estaba unida a ADN formando un complejo homogéneo. La muestra se tiñó negativamente y se observó en el microscopio electrónico (Figura 4.1.3A). A partir de las micrografías electrónicas se extrajeron 8.285 imágenes que se utilizaron en un proceso de refinamiento angular iterativo (Figuras 4.1.3B y 4.1.3C). Los ángulos de Euler estaban convenientemente representados en el conjunto inicial de partículas individuales (Figura 4.1.3D). La estructura tridimensional obtenida tiene una resolución aproximada de 25 Å para un valor de FSC de 0,5 (Figura 4.1.3E). En ella de nuevo se distinguen las características principales de la estructura atómica correspondiente al núcleo de Ku (Figuras 4.1.3F, 4.1.3G y 4.1.3H). Las dimensiones generales de esta nueva estructura son compatibles con las de la estructura resuelta anteriormente para el heterodímero Ku, sin embargo las dos estructuras son diferentes. En el eje longitudinal del nuevo volumen se aprecia una densidad evidente que rompe la simetría de la estructura (se indica con “\*” en las Figuras 4.1.3F y 4.1.3H).



**Figura 4.1.3. Reconstrucción tridimensional del complejo Ku:ADN.** (A) Micrografía electrónica obtenida a partir de una rejilla del complejo Ku:ADN teñida negativamente. Algunas de las partículas más evidentes se indican con “\*”. (B) Galería de partículas individuales correspondientes a moléculas del complejo Ku:ADN. (C) Galería de proyecciones (P) y medias equivalentes (M) obtenidas en la última iteración del proceso de refinamiento angular iterativo. (D) Representación, según el paquete informático EMAN, de los ángulos de Euler del conjunto de partículas individuales en la última iteración del procesamiento. (E) Estimación de la resolución para un valor de FSC de 0,5. (F), (G) y (H) Vista frontal, cenital y trasera respectivamente de la reconstrucción tridimensional resuelta para el complejo Ku:ADN. La región de masa principal que rompe la simetría del volumen se indica con “\*”.

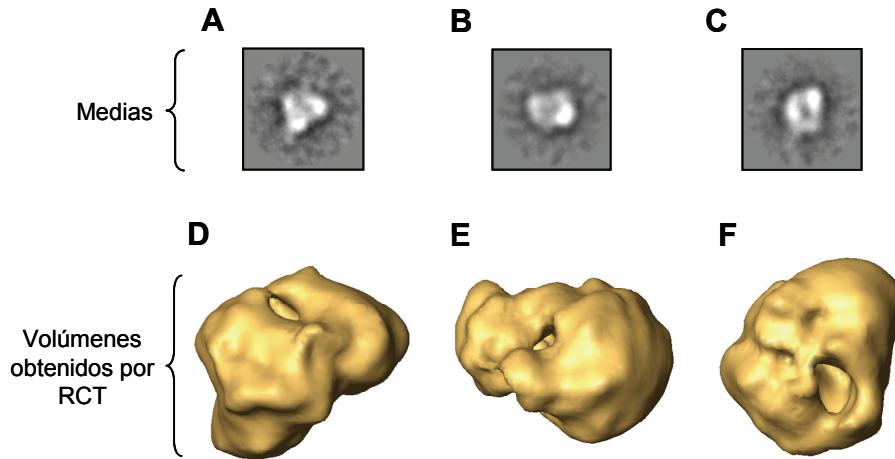


#### 4.1.4 Experimentos control que confirman la consistencia de las estructuras obtenidas para el heterodímero Ku y el complejo Ku:ADN

Durante el procesamiento de las imágenes individuales correspondientes al heterodímero Ku y al complejo Ku:ADN, se realizaron diversos experimentos control con el objetivo de evaluar la homogeneidad y consistencia de las estructuras obtenidas.

El proceso de refinamiento angular iterativo implica la utilización de un volumen de referencia inicial. En el caso de los procesamientos que permitieron obtener las estructuras correspondientes al heterodímero Ku y al complejo Ku:ADN se utilizaron distintos volúmenes iniciales (véase el apartado 3.13.5 de Materiales y Métodos), y todos ellos convergieron a las mismas soluciones. Además en ningún caso se utilizaron volúmenes referencia generados a partir de estructuras atómicas del núcleo de Ku libre o unido a ADN. Las proyecciones y las medias que se obtuvieron en el paso final del procesamiento, tanto para el heterodímero Ku como para el complejo Ku:ADN, eran consistentes con medias obtenidas directamente a partir del conjunto inicial de datos. Estas medias se generaban utilizando métodos de clasificación libres de patrón (véase el apartado 3.13.4 de Materiales y Métodos), y por tanto no estaban influenciadas por el refinamiento de los datos.

Además se utilizó la técnica de inclinación cónica al azar o RCT (Radermacher et al., 1986) con determinadas vistas típicas del complejo Ku:ADN (*Figura 4.1.4*). Esta técnica permite generar la estructura tridimensional de cada grupo de partículas homogéneo obtenido tras clasificar el conjunto inicial de imágenes (*Figuras 4.1.4A, 4.1.4B y 4.1.4C*) (véase el apartado 1.5.8 de Introducción). Las estructuras de RCT conservan las características globales de los volúmenes obtenidos para el heterodímero Ku y el complejo Ku:ADN (*Figuras 4.1.4D, 4.1.4E y 4.1.4F*). Las reconstrucciones de RCT se diferencian principalmente en su orientación relativa, lo que sugiere que corresponden a la misma molécula pero unida a la superficie de la rejilla en diferentes orientaciones. Por tanto las estructuras resueltas para el

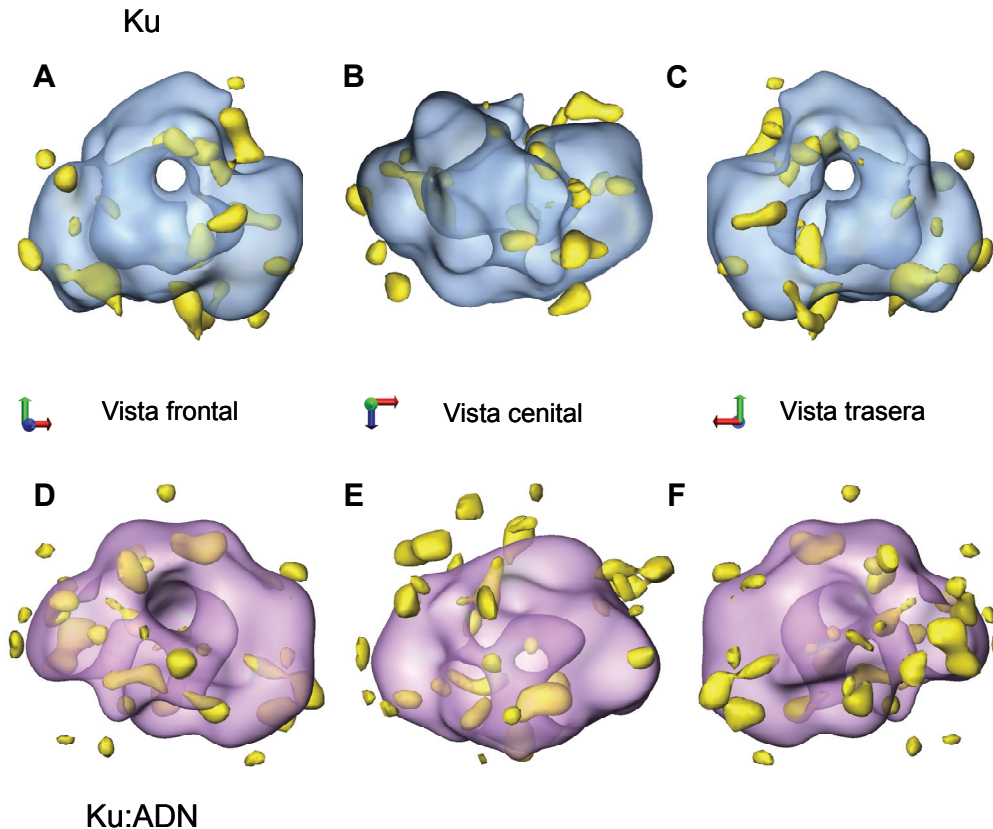


**Figura 4.1.4. Reconstrucciones tridimensionales del complejo Ku:ADN obtenidas mediante RCT. (A), (B) y (C) Imágenes medias obtenidas tras clasificar partículas correspondientes al complejo Ku:ADN utilizando algoritmos de clasificación libres de referencia. (D), (E) y (F) Volúmenes obtenidos mediante la técnica de RCT a partir de las medias (A), (B) y (C) respectivamente. Los volúmenes se muestran en una orientación similar a la de la imagen media correspondiente.**

heterodímero Ku y el complejo Ku:ADN se han obtenido a partir de clases compatibles y coherentes con una única estructura.

También se calcularon las regiones de alta varianza presentes en las estructuras resueltas para el heterodímero Ku y el complejo Ku:ADN. Para ello se utilizó la técnica de muestreo conocida como “bootstrap” propuesta por P. A. Penczek (Penczek et al., 2006a; Penczek et al., 2006b) (*Figura 4.1.5*). Las regiones de alta varianza dentro de los volúmenes finales son minoritarias y no se concentran en zonas específicas (*estas zonas se representan en color amarillo en la Figura 4.1.5*).

Los volúmenes obtenidos para el heterodímero Ku y el complejo Ku:ADN comparten características comunes, pero también presentan diferencias. Para confirmar que estas diferencias eran genuinas se llevaron a cabo diversos procesos de refinamiento. En estos procesamientos los volúmenes finales correspondientes al heterodímero Ku y al complejo Ku:ADN se utilizaron como volúmenes de referencia iniciales frente al conjunto de partículas alternativo. De esta manera, la estructura resuelta para el heterodímero Ku se utilizó como volumen inicial para las partículas del complejo Ku:ADN, mientras que el



**Figura 4.1.5. Análisis de la varianza en las estructuras resueltas para el heterodímero Ku y el complejo Ku:ADN.** (A), (B) y (C) Vista frontal, cenital y trasera respectivamente de la estructura resuelta para el heterodímero Ku. (D), (E) y (F) Vista frontal, cenital y trasera respectivamente de la estructura resuelta para el complejo Ku:ADN. La densidad correspondiente a regiones de alta varianza se representa en color amarillo.

volumen obtenido para el complejo Ku:ADN se utilizó como volumen inicial frente a las imágenes del heterodímero Ku. En ambos casos los procesamientos evolucionaron hacia la estructura final correspondiente para cada conjunto de datos, y por lo tanto no mantuvieron las características del sesgo introducido por la referencia inicial.

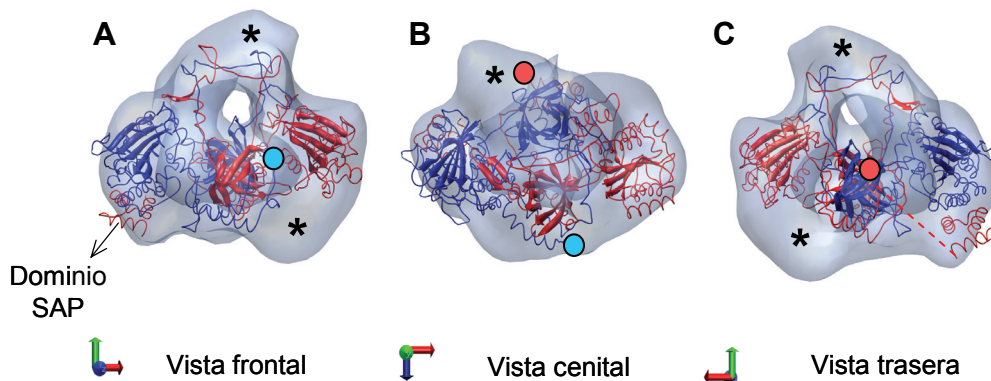
El ajuste de estructuras atómicas en volúmenes obtenidos por microscopía electrónica se utiliza como criterio para comprobar la calidad de los volúmenes resueltos. Además permite definir la “mano” correcta y también ampliar la información estructural que proporcionan las estructuras resueltas mediante microscopía electrónica (Chiu et al., 2005; Jiang y Ludtke, 2005). Por ello se llevó a cabo el ajuste de las estructuras atómicas del núcleo de la proteína Ku libre y unido a ADN (ficheros PDB 1JEQ y 1JEY respectivamente) en los

volúmenes obtenidos. En los experimentos de ajuste que se describen en los apartados 4.1.5 y 4.1.6, la calidad del encaje entre los volúmenes resueltos para el heterodímero Ku y el complejo Ku:ADN y las estructuras atómicas correspondientes es evidente (coeficientes de correlación 0,66 y 0,75 respectivamente). Además esto es más relevante si se tiene en cuenta que, durante los refinamientos, en ningún caso se utilizaron como volúmenes iniciales las estructuras atómicas mencionadas. Por tanto el proceso de refinamiento ha minimizado las posibles distorsiones causadas por el agente de tinción en las imágenes individuales. Todas estas comprobaciones adicionales refuerzan la calidad de los volúmenes obtenidos y confirman, además, la coherencia entre los datos iniciales y los volúmenes finales.

#### **4.1.5 Experimentos de ajuste en el volumen obtenido para el heterodímero Ku**

La estructura atómica que incluye el núcleo de Ku y el dominio SAP de Ku70 (fichero PDB 1JEQ) se ajustó en el volumen obtenido mediante microscopía electrónica para el heterodímero Ku. La solución con mayor coeficiente de correlación (0,66) muestra un ajuste excelente entre ambas estructuras (*Figura 4.1.6*). Prácticamente toda la estructura atómica queda contenida dentro del volumen. El agujero central, la forma alargada en el eje longitudinal, y la presencia del dominio SAP que rompe claramente la simetría de la estructura atómica son características que restringen las soluciones posibles durante el proceso de ajuste (*Figuras 4.1.6A, 4.1.6B y 4.1.6C*). En el volumen resuelto por microscopía electrónica también se observa densidad en la posición equivalente a la del dominio SAP. Además existe masa adicional que no es rellenada por la estructura atómica y que podría corresponder al CTD de Ku80 (se indica con “\*” en la *Figura 4.1.6*). Otras soluciones con un coeficiente de correlación menor dejaban pequeñas porciones de la estructura atómica fuera de la densidad y por lo tanto se descartaron. Los ajustes entre la estructura atómica y el volumen de microscopía electrónica con la “mano” espejular tampoco eran satisfactorios. En todas estas soluciones descartadas, el dominio de la estructura atómica correspondiente al arco se localizaba lejos de su posición evidente en el volumen obtenido por microscopía electrónica.

Por tanto se escogió la “mano” del volumen de microscopía electrónica que mejor encaja con la estructura atómica resuelta.

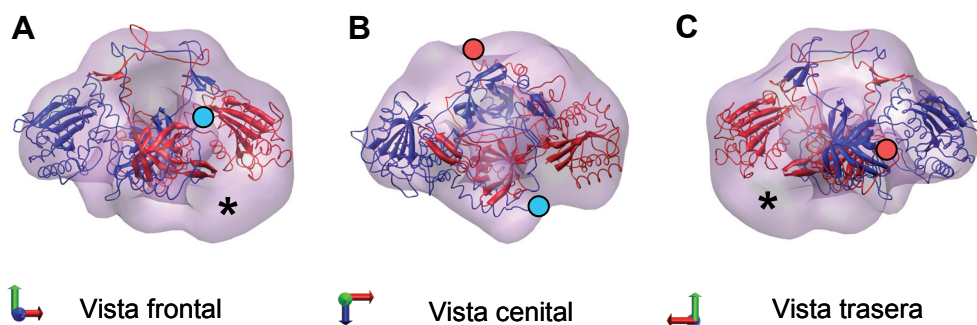


**Figura 4.1.6. Experimentos de ajuste en el volumen obtenido para el heterodímero Ku.** (A), (B) y (C) Vista frontal, cenital y trasera respectivamente del ajuste de la estructura atómica correspondiente al fichero PDB 1JEQ en el volumen resuelto para el heterodímero Ku. Esta estructura atómica incluye el núcleo de Ku y el dominio SAP de Ku70. De ahora en adelante en las estructuras atómicas del heterodímero Ku las proteínas Ku70 y Ku80 se representan en color rojo y azul respectivamente. En (A) se señala el dominio SAP de Ku70. Las principales regiones de densidad que no son rellenadas por la estructura atómica se indican con “\*”. Los extremos C-terminales de los dominios brazo de Ku70 y Ku80 se indican con un punto rojo y azul respectivamente. Este criterio de representación se mantiene en el resto de figuras. En (C) se muestra mediante una línea discontinua roja y de manera muy simplificada la posible conexión entre el dominio brazo de Ku70 y el dominio SAP.

#### 4.1.6 Experimentos de ajuste en el volumen obtenido para el complejo Ku:ADN

El proceso de ajuste se llevó a cabo utilizando la estructura atómica que contiene el núcleo de la proteína Ku unido a una molécula de ADN (fichero PDB 1JEY). En esta estructura a pesar de estar presente en la molécula no se pudo determinar la posición del dominio SAP (Walker et al., 2001). Para realizar el ajuste no se utilizó la información correspondiente a la molécula de ADN presente en la estructura del cristal. De nuevo las características estructurales del heterodímero Ku restringen las soluciones posibles. La calidad del ajuste entre la estructura atómica y el volumen de microscopía electrónica es notable (Figura 4.1.7). En el ajuste con mayor coeficiente de correlación (0,75), el volumen engloba casi por completo a la estructura atómica. El perfil

de ambas estructuras es idéntico y también la localización del agujero central (Figuras 4.1.7A y 4.1.7C). Experimentos de ajuste llevados a cabo con la “mano” espejular del volumen de microscopía electrónica no proporcionaron soluciones satisfactorias. Por tanto se escogió la “mano” que mejor encajaba con la estructura atómica utilizada en los experimentos de ajuste. En el volumen obtenido para el heterodímero Ku el extremo del volumen presenta una región de densidad compatible con el dominio SAP (Figura 4.1.6). En el caso del volumen correspondiente al complejo Ku:ADN no se aprecia densidad en la zona equivalente (compárense las Figuras 4.1.6A y 4.1.7A). Además el agujero que en la estructura atómica encierra a la molécula de ADN ahora muestra algo de densidad (Figuras 4.1.7A y 4.1.7C). El volumen obtenido para el complejo Ku:ADN también presenta regiones que no son rellenadas por la estructura atómica (se indican con “\*” en la Figura 4.1.7). A diferencia de lo que sucede con la estructura atómica de la proteína Ku libre (fichero PDB 1JEQ), la estructura atómica de Ku unida a ADN presenta pseudo simetría dos (fichero PDB 1JEY). Esto se debe a que en la primera estructura atómica el dominio SAP rompe la pseudo simetría, sin embargo en la segunda estructura este dominio no se resolvió. Por ello una rotación de 180° del ajuste con mayor coeficiente de correlación en el volumen resuelto para el complejo Ku:ADN daría una solución que en principio sería igual de válida. Para solucionar esta ambigüedad se escogió la solución que, teniendo el mayor coeficiente de



**Figura 4.1.7. Experimentos de ajuste en el volumen obtenido para el complejo Ku:ADN. (A), (B) y (C) Vista frontal, cenital y trasera respectivamente del ajuste de la estructura atómica del núcleo de Ku unido a ADN (fichero PDB 1JEY) en el volumen obtenido para el complejo Ku:ADN. Las regiones de densidad adicionales que no son rellenadas por la estructura atómica se indican con “\*”.**

correlación, era compatible con la localización de dominios definida previamente para el heterodímero Ku (*Figuras 4.1.7A y 4.1.7C*).

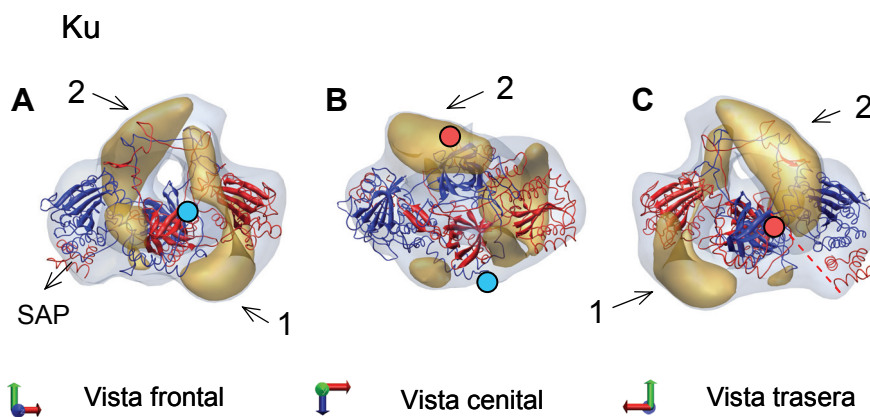
#### **4.1.7 Localización del dominio SAP de Ku70 y del dominio CTD de Ku80**

Los experimentos de ajuste evidencian la consistencia entre las estructuras obtenidas mediante microscopía electrónica y la información a resolución atómica. Por tanto es posible llevar a cabo una aproximación basada en el cálculo de volúmenes diferencia con el objetivo de localizar los dominios C-terminales específicos de Ku70 y Ku80. Estos volúmenes resultan de la diferencia entre los volúmenes de microscopía electrónica y las estructuras atómicas ajustadas en ellos filtradas a la resolución equivalente. Consecuentemente los mapas diferencia muestran únicamente regiones que posiblemente correspondan a los dominios C-terminales ausentes en las estructuras atómicas.

##### *Localización del dominio CTD de Ku80 en el volumen resuelto para el heterodímero Ku*

El volumen que se obtiene a partir del cálculo de la diferencia entre el volumen resuelto para el heterodímero Ku y la estructura atómica del núcleo de Ku (fichero PDB 1JEQ), presenta dos zonas claramente definidas (*se indican como "1" y "2" en la Figura 4.1.8*). La zona 1 se localiza bajo el dominio  $\alpha/\beta$  de Ku70 mientras que la zona 2 se localiza por detrás del agujero central sobresaliendo ligeramente por la parte superior de la estructura (*Figuras 4.1.8A, 4.1.8B y 4.1.8C*). En estas dos zonas podrían localizarse los fragmentos ausentes en la estructura atómica del núcleo de Ku (fichero PDB 1JEQ). El CTD de Ku80 tiene un peso molecular de 19 kDa y consta de varias regiones. La primera región corresponde al fragmento que conecta el dominio brazo de Ku80 con un dominio helicoidal. A continuación del dominio helicoidal viene la porción más C-terminal del CTD que está implicada en el reclutamiento de DNA-PKcs (véase el apartado 1.4.1 de Introducción). En principio cualquiera de las dos zonas mencionadas anteriormente puede alojar el dominio helicoidal. Además la combinación de ambas zonas permite acomodar el CTD de Ku80

completo. Sin embargo las estructuras obtenidas por RMN del dominio helicoidal se ajustaban mejor en la zona 1. Esta zona se sitúa más próxima al extremo C-terminal del dominio brazo de Ku80 (se representa con un punto azul en las Figuras 4.1.8A y 4.1.8B), por lo que posiblemente corresponda al dominio helicoidal. De esta manera el CTD de Ku80 quedaría localizado debajo del dominio  $\alpha/\beta$  Ku70 y el dominio SAP debajo de la región más distal del dominio  $\alpha/\beta$  de Ku80. La zona 2, más próxima al agujero, podría corresponder a la porción más C-terminal del CTD de Ku80 responsable del reclutamiento de DNA-PKcs.



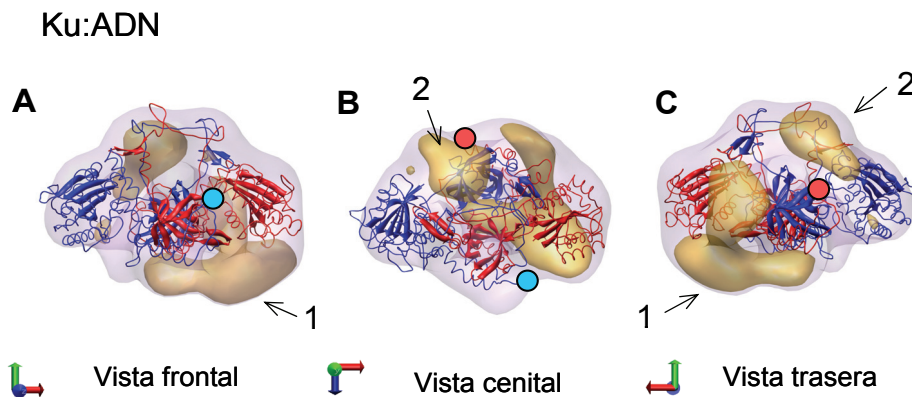
**Figura 4.1.8. Localización del dominio CTD de Ku80 en el volumen obtenido para el heterodímero Ku.** (A), (B) y (C) Vista frontal, cenital y trasera respectivamente del ajuste de la estructura atómica que incluye el núcleo de Ku y el dominio SAP de Ku70 (archivo PDB 1JEQ), en el volumen resuelto para el heterodímero Ku. Las densidades de color marrón corresponden al volumen calculado a partir de la diferencia entre el volumen de microscopía electrónica y la estructura atómica ajustada. Dentro del volumen diferencia se distinguen dos zonas principales denominadas “1” y “2”. En (A) se señala el dominio SAP de Ku70. En (C) se muestra mediante una línea discontinua roja y de manera muy simplificada, la posible conexión entre el dominio brazo de Ku70 y el dominio SAP.

*Localización del dominio SAP de Ku70 y del CTD de Ku80 en el volumen resuelto para el complejo Ku:ADN*

Como en el caso anterior, se obtuvo un volumen a partir de la diferencia entre el volumen del complejo Ku:ADN y la estructura atómica del núcleo de Ku (archivo PDB 1JEY) ajustado en él (Figura 4.1.9). El volumen diferencia también presenta dos zonas claramente definidas (se indican como “1” y “2” en la Figura



4.1.9). La zona 1 se localiza bajo el dominio  $\alpha/\beta$  de Ku70, en una región equivalente a la descrita para la zona 1 del volumen diferencia calculado a partir del heterodímero Ku (*compárense las Figuras 4.1.8 y 4.1.9*). La zona 2 se localiza próxima al agujero, pero en este caso es más compacta y está más cercana a la parte central del núcleo de Ku (*Figura 4.1.9*).



**Figura 4.1.9. Localización del dominio SAP de Ku70 y del CTD de Ku80 en el volumen obtenido para el complejo Ku:ADN. (A), (B) y (C) Vista frontal, cenital y trasera respectivamente del ajuste de la estructura atómica del núcleo de Ku (fichero PDB 1JEY) en el volumen obtenido para el complejo Ku:ADN. El volumen calculado a partir de la diferencia entre el volumen de microscopía electrónica y la estructura atómica ajustada se representa de color marrón. Dentro de este volumen se distinguen dos zonas principales denominadas "1" y "2".**

Las densidades principales de los dos volúmenes diferencia descritos anteriormente se localizan en regiones equivalentes. Esto sugiere que el CTD de Ku80 se mantiene en la misma posición después de que el heterodímero Ku se una al extremo del ADN. En cualquier caso, el mayor cambio conformacional que se produce como consecuencia de la unión a los extremos del ADN afecta al dominio SAP de Ku70. En ausencia de ADN este dominio contacta con el dominio  $\alpha/\beta$  de Ku80 y, tras la unión al ADN, el dominio SAP desaparece de la posición que ocupaba en la estructura del heterodímero Ku libre. Este dominio podría haberse desplazado a una localización próxima al agujero, en el lado del heterodímero Ku que mira hacia la parte continua de la molécula de ADN. Esta localización ha sido sugerida a partir de la estructura atómica correspondiente al núcleo del heterodímero Ku unido a ADN (fichero PDB 1JEY) (Walker et al.,

2001). En esta nueva localización el dominio SAP podría estar implicado en el mantenimiento de la interacción entre el heterodímero Ku y la molécula de ADN.

#### **4.1.8 Conclusión**

Se han resuelto las estructuras correspondientes al heterodímero Ku y al complejo Ku:ADN utilizando microscopía electrónica de especímenes teñidos negativamente. Los volúmenes obtenidos se han combinado con la información disponible a resolución atómica para la proteína Ku y, de esta manera, ha sido posible proponer un modelo de organización tridimensional para el heterodímero humano Ku completo. Este modelo justifica también los cambios conformacionales que tienen lugar en el heterodímero Ku como consecuencia de la unión a los extremos del ADN.

## **4.2 Análisis estructural de la proteína DNA-PKcs mediante tinción negativa**

En el momento de llevar a cabo este análisis existían dos estructuras para la proteína DNA-PKcs que presentaban diferencias significativas (Chiu et al., 1998; Leuther et al., 1999). Por ello, como primera aproximación para caracterizar estructuralmente la proteína DNA-PKcs, se llevó a cabo su análisis mediante microscopía electrónica y análisis de partículas individuales teñidas negativamente.

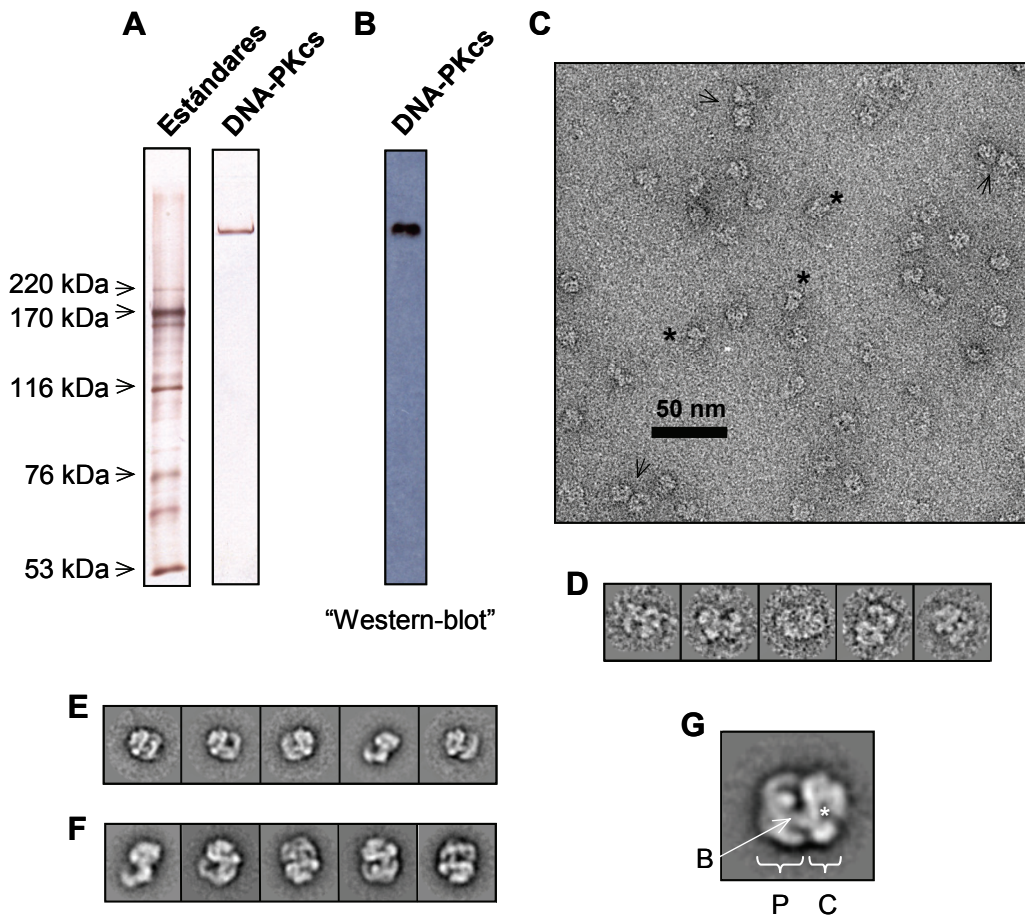
### **4.2.1 Purificación de la proteína DNA-PKcs**

DNA-PKcs consiste en una única cadena polipeptídica de 460 kDa de peso molecular (véase el apartado 1.4.2 de Introducción). Su tamaño supone por tanto una limitación importante a la hora de tratar de incrementar su expresión utilizando técnicas de biología molecular convencionales. Por ello la proteína DNA-PKcs purificada corresponde a la proteína endógena que, de manera constitutiva, está presente en el núcleo de las células HeLa. La purificación de esta proteína fue realizada por el Dr. J. D. Maman en el marco de la

colaboración establecida con el Profesor L. H. Pearl. Al analizar la proteína DNA-PKcs purificada por electroforesis en condiciones desnaturalizantes se observó una única banda (*Figura 4.2.1A*). Esta banda era reconocida específicamente por un anticuerpo monoclonal (DNA-PK Ab-2, Oncogene) frente a la proteína DNA-PKcs humana (*Figura 4.2.1B*).

#### 4.2.2 Reconstrucción tridimensional de la proteína DNA-PKcs

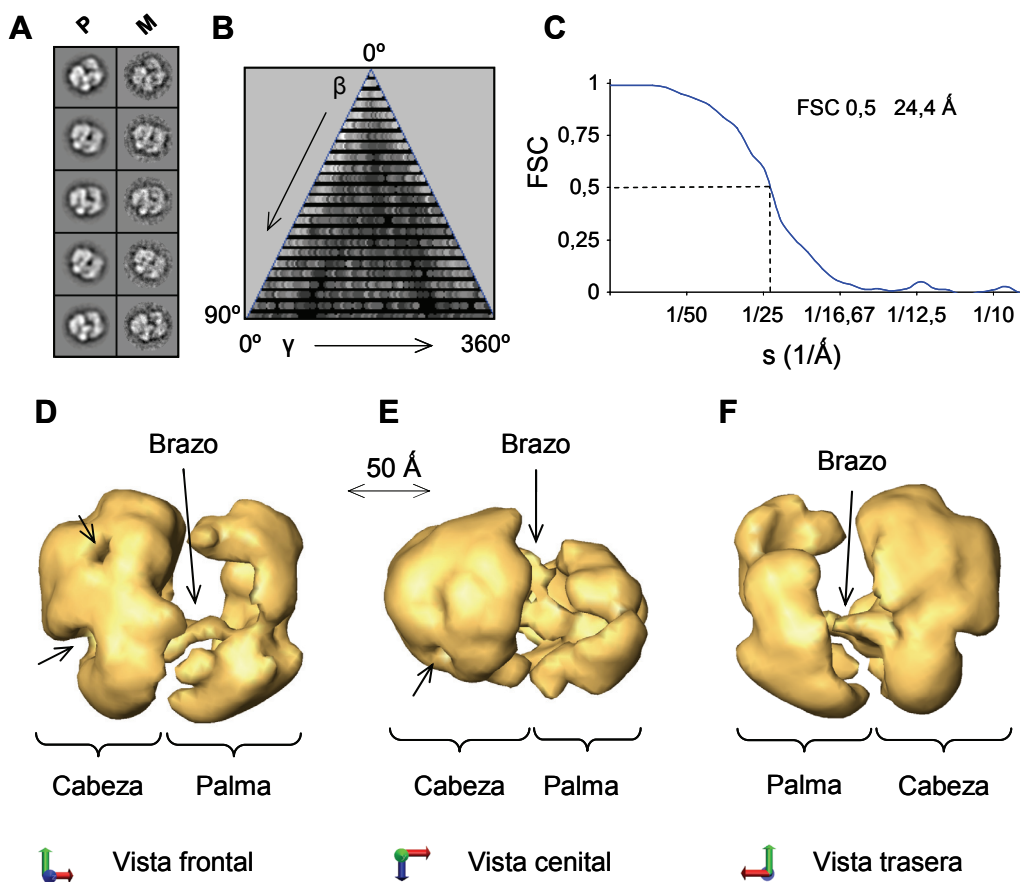
La proteína DNA-PKcs purificada se tiñó negativamente y se observó en el microscopio electrónico (*Figura 4.2.1C*). A partir de micrografías sin inclinar e inclinadas 40° se extrajeron un total de 4.812 imágenes (*Figura 4.2.1D*) que inicialmente se analizaron en dos dimensiones utilizando algoritmos de clasificación libres de patrón. Las imágenes medias obtenidas tienen una forma general rectangular y presentan características muy definidas (*Figuras 4.2.1E y 4.2.1F*). En muchas de estas medias podían distinguirse dos regiones principales denominadas cabeza y palma que se disponen a lo largo de un eje longitudinal común (*regiones C y P respectivamente de la Figura 4.2.1G*). La región cabeza más masiva (*región C de la Figura 4.2.1G*) tiene una forma globular compleja en la que se distingue una cavidad evidente (*se indica con “\*” en la Figura 4.2.1G*). La región palma (*región P de la Figura 4.2.1G*) es más compacta y estrecha y su altura es similar a la de la región cabeza. Las dos regiones principales mencionadas se conectan por una pequeña densidad de masa denominada brazo (*se indica como B en la Figura 4.2.1G*). En las micrografías obtenidas además de moléculas individuales también se distinguían partículas que correspondían a dímeros de DNA-PKcs (*se indican con flechas en la Figura 4.2.1C*). Estas partículas eran minoritarias y su número no alcanzaba el 5% del total de las imágenes. El significado funcional de estos dímeros no es evidente y podría reflejar la capacidad intrínseca de DNA-PKcs de interactuar consigo misma. Esta propiedad sería clave en los procesos de regulación por autofosforilación en *trans* (Meek et al., 2007). Por tanto las imágenes descritas podrían corresponder a un intermediario del proceso de reparación por NHEJ.



**Figura 4.2.1. Purificación y análisis mediante microscopía electrónica de especímenes teñidos negativamente de la proteína DNA-PKcs.** (A) Gel de poliacrilamida en condiciones desnaturalizantes de la proteína DNA-PKcs purificada teñido posteriormente con plata. (B) “Western-blot” de un gel equivalente al que se muestra en (A) en el que se ha utilizado un anticuerpo monoclonal (DNA-PK Ab-2, Oncogene) frente a la proteína DNA-PKcs. (C) Micrografía obtenida a partir de una rejilla de microscopía electrónica de la proteína DNA-PKcs teñida negativamente. Algunas partículas evidentes se indican con “\*”. Partículas que podrían corresponder a dímeros de DNA-PKcs se indican mediante una flecha. (D) Galería de partículas correspondientes a moléculas individuales de la proteína DNA-PKcs. (E) y (F) Galería de imágenes medias generadas utilizando métodos de clasificación libres de patrón presentes en los paquetes informáticos EMAN y XMIPP respectivamente. (G) Detalle de una imagen media generada mediante algoritmos de clasificación libres de patrón. Las dos regiones principales de masa “P” y “C” están conectadas por una pequeña densidad de masa “B”. Dentro de la región “C” se distingue una cavidad evidente que se indica con “\*”.

Las 4.812 partículas correspondientes a moléculas individuales de la proteína DNA-PKcs se utilizaron en un proceso de refinamiento angular iterativo (Figura 4.2.2A). La diversidad de imágenes de proyección presentes en el conjunto inicial de partículas (Figura 4.2.2B) permitió obtener un volumen

a una resolución aproximada de 25 Å para un valor de FSC de 0,5 (Figura 4.2.2C). La organización general de esta nueva estructura concuerda con la observada previamente en la estructura obtenida a partir de cristales bidimensionales teñidos negativamente (véase la Figura 1.4.3 de Introducción) (Leuther et al., 1999). Las dimensiones aproximadas del volumen resuelto para la proteína DNA-PKcs son de 140 Å en su eje longitudinal principal, 115 Å de altura máxima y un grosor que varía entre 100 Å y 40 Å (Figuras 4.2.2D, 4.2.2E y 4.2.2F). Dentro de esta estructura se distinguen tres regiones principales que se han denominado cabeza, palma y brazo. Estas regiones son análogas a las



**Figura 4.2.2 Reconstrucción tridimensional de la proteína DNA-PKcs.** (A)

Galería de proyecciones (P) y medias equivalentes (M) obtenidas en la última iteración del proceso de refinamiento angular iterativo que permitió obtener la estructura final. (B) Representación mediante el paquete informático EMAN de los ángulos de Euler asignados a las partículas individuales en la última iteración del proceso de refinamiento. (C) Estimación de la resolución para un valor de FSC de 0,5. (D), (E) y (F) Vista frontal, cenital y trasera respectivamente de la reconstrucción tridimensional resuelta para la proteína DNA-PKcs. Se indican las regiones características cabeza, brazo y palma en las que se organiza la proteína DNA-PKcs. Además se señalan los agujeros que permiten el acceso a la cavidad interior de la región cabeza.

regiones equivalentes observadas previamente en imágenes medias obtenidas mediante algoritmos de clasificación libres de patrón (*compárense las Figuras 4.1.2E y 4.1.2F con las Figuras 4.2.2D, 4.2.2E y 4.2.2F*). La cabeza es la región más voluminosa y tiene una forma global de pirámide invertida que encierra una cavidad en su interior (*Figuras 4.2.2D y 4.2.2F*). El brazo tiene forma tubular y conecta la región cabeza con la región compacta y aplanada que se ha denominado palma (*Figuras 4.2.2D, 4.2.2E y 4.2.2F*).

### **4.2.3 Conclusión**

El volumen resuelto mediante microscopía electrónica de especímenes teñidos negativamente permite definir la organización general de la proteína DNA-PKcs. Esta proteína se organiza en tres regiones principales denominadas cabeza, brazo y palma. El nuevo volumen es totalmente compatible con la estructura obtenida previamente mediante microscopía electrónica de cristales bidimensionales teñidos negativamente.

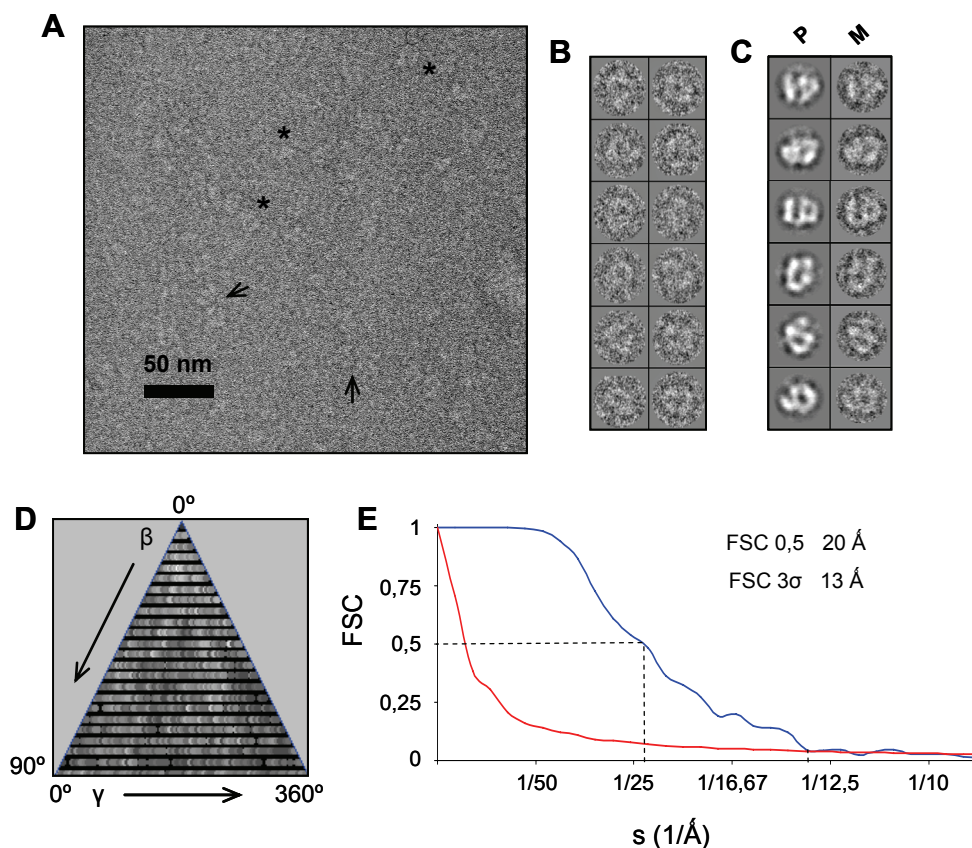
## **4.3 Análisis estructural de la proteína DNA-PKcs mediante criomicroscopía electrónica**

Para profundizar en la organización estructural de esta proteína era necesario mejorar la resolución de la estructura resuelta. Por ello se llevó a cabo el análisis estructural de la proteína DNA-PKcs mediante criomicroscopía electrónica. El objetivo último consistía en obtener una estructura en la que fuese posible distinguir y localizar los dominios principales, especialmente el dominio catalítico.

### **4.3.1 Reconstrucción tridimensional de la proteína DNA-PKcs obtenida mediante criomicroscopía electrónica**

La proteína DNA-PKcs humana se congeló a alta velocidad y se observó en un microscopio electrónico Tecnai G<sup>2</sup> FEG 200 (FEI). A partir de micrografías electrónicas con la CTF corregida se extrajeron 6.857 imágenes que se

utilizaron en diferentes procesos de refinamiento (*Figuras 4.3.1A y 4.3.1B*). Durante los distintos refinamientos no se utilizó ninguna de las estructuras resueltas anteriormente para la proteína DNA-PKcs. De esta manera se evitaba introducir algún sesgo que condicionase la evolución del procesamiento. Como volúmenes iniciales se utilizaron volúmenes generados mediante algoritmos basados en la técnica de líneas comunes. A partir de cada volumen se llevó a cabo un proceso de refinamiento angular iterativo independiente. Los diferentes procesamientos evolucionaron hacia una solución similar, de manera que fue posible obtener la estructura de la proteína DNA-PKcs a una resolución



**Figura 4.3.1. Reconstrucción tridimensional de la proteína DNA-PKcs mediante criomicroscopía electrónica.** (A) Micrografía electrónica obtenida a partir de una rejilla de la proteína DNA-PKcs congelada. Algunas moléculas individuales de DNA-PKcs se indican con “\*”. Partículas que podrían corresponder a dímeros de DNA-PKcs se indican con flechas. (B) Galería de partículas individuales correspondientes a moléculas de la proteína DNA-PKcs. (C) Galería de proyecciones (P) y medias equivalentes (M) generadas en la última iteración del proceso de refinamiento angular iterativo. (D) Representación, según el paquete informático EMAN, de los ángulos de Euler del conjunto de partículas individuales en la última iteración del procesamiento. Como se observa en la figura las imágenes de proyección utilizadas abarcan un rango adecuado de ángulos de Euler. (E) Estimación de la resolución según los criterios 0,5 y  $3\sigma$ .

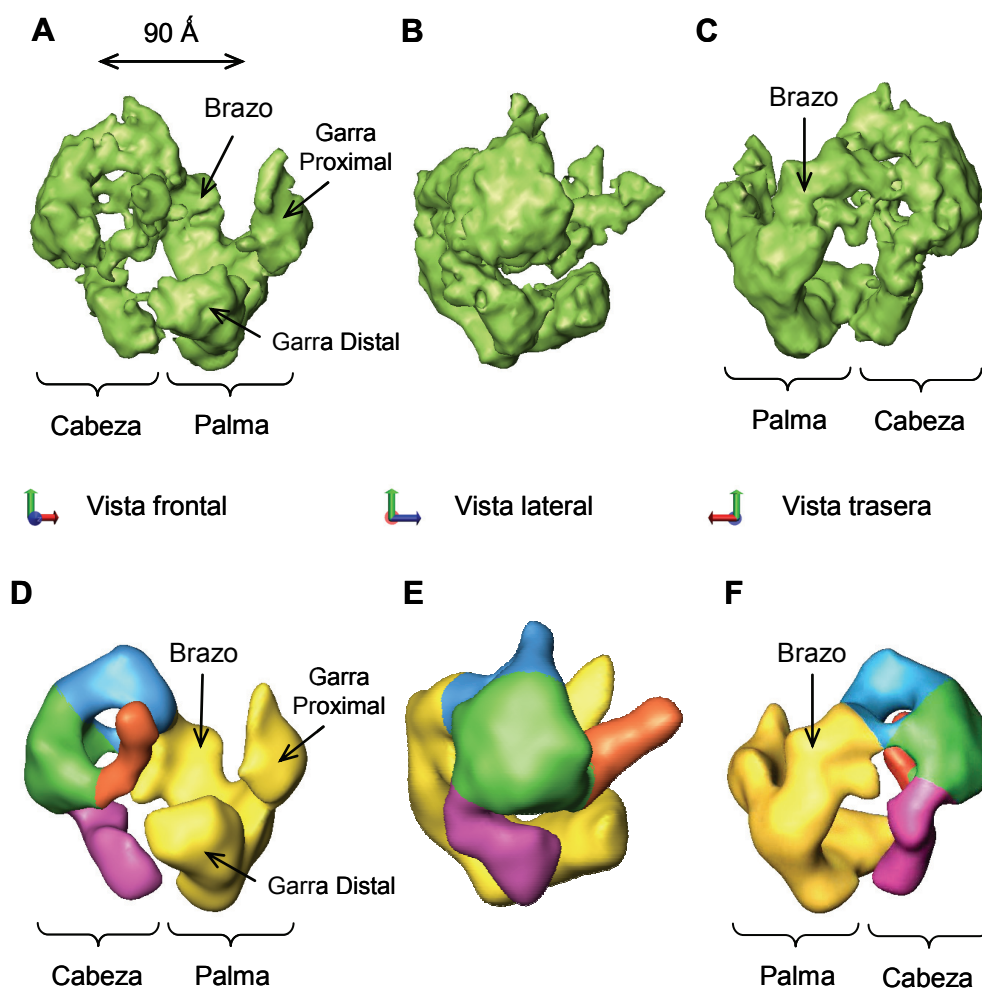
aproximada de 13 Å según el criterio  $3\sigma$  (*Figura 4.3.1E*). En la *Figura 4.3.1C* se muestran proyecciones del volumen final y las medias equivalentes obtenidas en la última iteración del refinamiento. La semejanza entre las proyecciones y las medias evidencia la estabilización del procesamiento. Además la buena correlación entre el conjunto inicial de datos y el volumen final indica que la muestra es estructuralmente homogénea.

La estructura resuelta mediante criomicroscopía electrónica presenta unas dimensiones globales compatibles con las de la estructura obtenida por tinción negativa (*Figuras 4.3.2A, 4.3.2B y 4.3.2C*). Además se organiza también en las regiones características cabeza, palma y brazo, pero la mejora significativa de la resolución permite distinguir los dominios que constituyen cada una de esas tres regiones (*compárense las Figuras 4.2.2 y 4.3.2*). La estructura de criomicroscopía electrónica se muestra filtrada a una menor resolución con el objetivo de hacer más evidentes estos dominios característicos (*Figuras 4.3.2D, 4.3.2E y 4.3.2F*). En este nuevo volumen se han coloreado los diferentes dominios que se describen a continuación. La región cabeza estaría formada por cuatro dominios. El dominio central del que emergen todos los demás se localiza en uno de los laterales de la estructura de DNA-PKcs (*de color verde en las Figuras 4.3.2D, 4.3.2E y 4.3.2F*). Tiene forma de prisma triangular, con unas dimensiones aproximadas de 90 Å x 60 Å x 90 Å y un grosor de unos 20 Å. A partir de los vértices de este dominio central emergen tres dominios diferentes. El menor de estos dominios consiste en un pequeño saliente con forma cilíndrica de unos 80 Å de longitud y un grosor de 10 Å en la zona más estrecha (*de color rojo en las Figuras 4.3.2D, 4.3.2E y 4.3.2F*). En la estructura de criomicroscopía electrónica este pequeño dominio es evidente e independiente del resto de la cabeza (*Figuras 4.3.2A, 4.3.2B y 4.3.2C*). En estructuras obtenidas a partir de especímenes teñidos negativamente también se aprecia un dominio en la posición equivalente. En estos volúmenes sin embargo no aparece como un dominio independiente y posiblemente se encuentre contactando con la masa principal de la región cabeza. Esto sugiere que el dominio descrito podría ser flexible y, en presencia del agente de tinción, quedaría aplastado sobre la estructura. La conformación observada a partir de



especímenes congelados probablemente representa su disposición en condiciones nativas.

A partir del dominio central sale un segundo dominio más masivo que termina en una región globular voluminosa (*de color magenta en las Figuras 4.3.2D, 4.3.2E y 4.3.2F*). Desde el vértice superior del dominio central emerge un tercer dominio que se ha denominado hombro (*de color azul en las Figuras 4.3.2D, 4.3.2E y 4.3.2F*). Es un dominio tubular y conecta la cabeza con el resto de la molécula de DNA-PKcs. A continuación de la región cabeza vendría la



**Figura 4.3.2. Estructura resuelta mediante criomicroscopía electrónica para la proteína DNA-PKcs.** (A), (B) y (C) Vista frontal, lateral y trasera respectivamente de la reconstrucción tridimensional resuelta mediante criomicroscopía electrónica para la proteína DNA-PKcs. (D), (E) y (F) Vistas equivalentes a las mostradas en (A), (B) y (C) pero del volumen de criomicroscopía electrónica filtrado a 30 Å de resolución. Los diferentes dominios que constituyen las regiones principales en las que se organiza la proteína DNA-PKcs se representan con disitintos colores. Se indican las regiones principales cabeza, brazo y palma y también los dominios garra proximal y garra distal que constituyen la región palma.

región brazo (*de color amarillo en las Figuras 4.3.2D, 4.3.2E y 4.3.2F*). Esta región tiene aspecto de tubo aplanado y curvo y abarcaría desde el final del dominio hombro hasta la región palma (*la región palma también se representa de color amarillo en las Figuras 4.3.2D, 4.3.2E y 4.3.2F*).

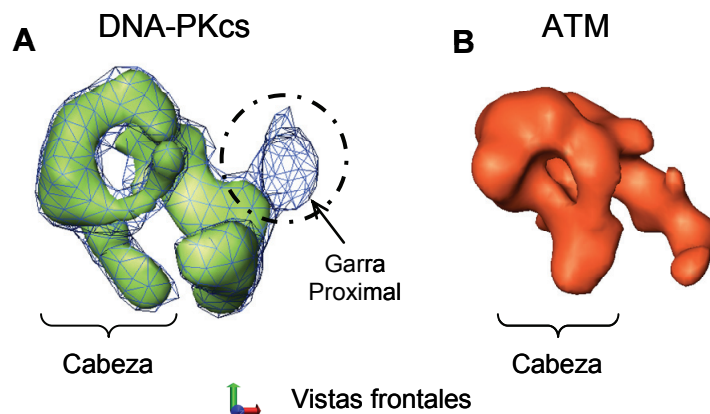
En las estructuras obtenidas previamente utilizando agentes de tinción la región palma tiene un aspecto macizo (*véase la Figura 4.2.2*). En el volumen obtenido por criomicroscopía electrónica sin embargo se distinguen dos dominios (*Figura 4.3.2*). Estos dos dominios se han denominado garras y tienen unas dimensiones y una orientación que es compatible con el aspecto macizo de la región palma visto en estructuras de menor resolución (*compárense las Figuras 4.2.2 y 4.3.2*). Las dos garras salen perpendicularmente desde el borde del brazo que constituiría por tanto la base de la región palma (*Figuras 4.3.2A, 4.3.2C, 4.3.2D y 4.3.2F*). Una de las garras (garra distal) parece ser la continuación del final del brazo, mientras que la otra garra (garra proximal) surge desde el punto medio del brazo (*Figuras 4.3.2A y 4.3.2D*). Por tanto la estructura de DNA-PKcs obtenida mediante criomicroscopía electrónica a 13 Å de resolución permite identificar claramente siete dominios que se organizan en torno a una cavidad central.

#### **4.3.2 Localización de los diferentes dominios de la proteína DNA-PKcs en la estructura obtenida mediante criomicroscopía electrónica**

La estructura resuelta tiene una resolución tal que permite distinguir la distribución espacial de los distintos dominios que componen DNA-PKcs. Actualmente no existen estructuras a resolución atómica para ningún fragmento de la proteína DNA-PKcs. Por ello, para poder ampliar las conclusiones estructurales del volumen obtenido y proponer entonces un modelo pseudo atómico de la proteína DNA-PKcs, ha sido necesario combinar diferentes aproximaciones bioquímicas y bioinformáticas y recurrir también a la comparación de estructuras resueltas mediante microscopía electrónica.

### A) Localización de la región C-terminal conservada

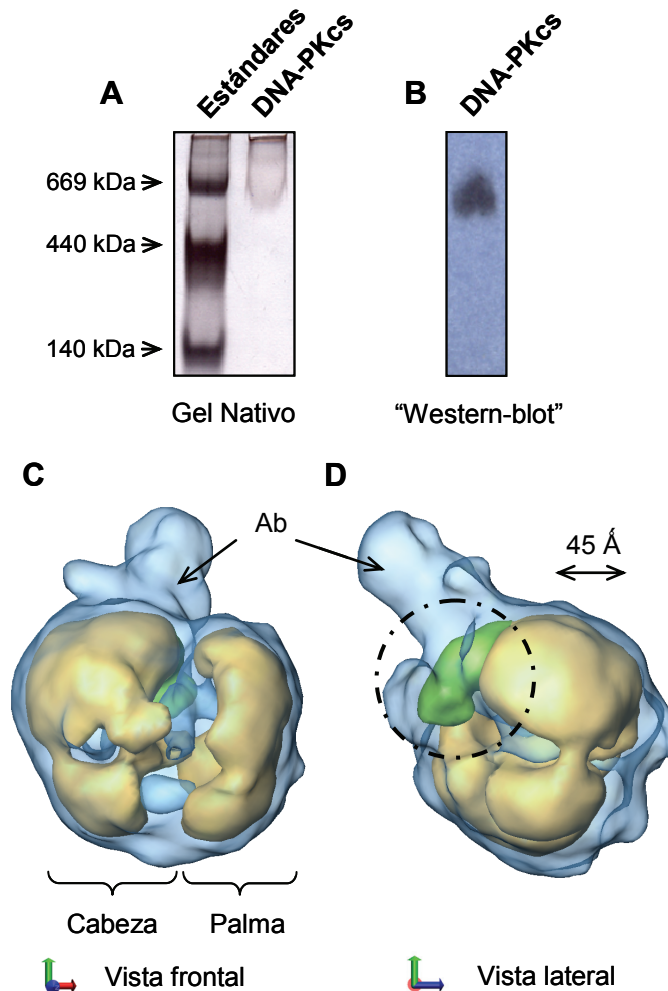
La homología entre los distintos miembros de la familia de las PIKKs está restringida a la región C-terminal y más concretamente a los dominios FAT, catalítico y FATC (véase el apartado 1.4.2 de Introducción). Las regiones C-terminales de las secuencias de las PIKKs pueden alinearse sin que haya inserciones ni eliminaciones significativas (Keith y Schreiber, 1995; Bosotti et al., 2000). La región N-terminal por el contrario es muy variable entre los distintos miembros de esta familia. La mayoría de las diferencias en tamaño que presentan las PIKKs afectan a esta región, y estaría formada principalmente por repeticiones de motivos helicoidales (Perry y Kleckner, 2003; Brewerton et al., 2004). Es probable que la conservación que presentan las PIKKs a nivel de secuencia en la región C-terminal también se refleje a nivel tridimensional. De hecho al comparar la estructura de criomicroscopía electrónica de DNA-PKcs filtrada a 30 Å de resolución con la estructura de tinción negativa de ATM (Llorca et al., 2003), se aprecia que las regiones cabeza de ambas proteínas son muy parecidas en tamaño y forma (*Figura 4.3.3*). La diferencia principal entre ambos volúmenes radica en la región palma. Esto sugiere que la región C-terminal podría corresponder a la región cabeza mientras que la mitad variable N-terminal podría localizarse en la región palma.



**Figura 4.3.3. Comparación de estructuras resueltas mediante microscopía electrónica de las PIKKs ATM y DNA-PKcs.** (A) Vista frontal de la estructura de criomicroscopía electrónica resuelta para la proteína DNA-PKcs filtrada a 30 Å de resolución. La malla representa la proteína completa. En color verde se representa la misma estructura pero sin la densidad correspondiente a la garra proximal. (B) Vista frontal de la estructura resuelta mediante tinción negativa para la proteína ATM (Llorca et al., 2003). En (A) y (B) se indica la región cabeza común a ambas proteínas. Ambas estructuras se representan a la misma escala.

Esta asignación contradice los modelos propuestos por diversos grupos (Brewerton et al., 2004; Williams et al., 2008), por lo que se llevaron a cabo experimentos de marcaje con anticuerpos con la intención de comprobar la nueva asignación predicha.

No se conoce el epítipo exacto que reconoce el anticuerpo monoclonal utilizado en los experimentos de marcaje (DNA-PKcs G-4 sc-5282, Santa Cruz), aunque se sabe que se localizaría en la región comprendida entre los residuos 2.965 y 4.127 de la proteína DNA-PKcs humana. Tras comprobar que el anticuerpo era capaz de reconocer la proteína DNA-PKcs purificada en condiciones nativas (*Figuras 4.3.4A y 4.3.4B*), se realizaron experimentos de formación de complejos anticuerpo:DNA-PKcs en solución (véase el apartado 3.9.3 de Materiales y Métodos). Los complejos anticuerpo:DNA-PKcs se tiñeron negativamente y se observaron en el microscopio electrónico. A partir de imágenes extraídas de las micrografías se llevó a cabo un proceso de refinamiento angular iterativo. El volumen resuelto tiene una resolución aproximada de 30 Å para un valor de FSC de 0,5 (*Figuras 4.3.4C y 4.3.4D*). Este volumen de nuevo comparte la organización descrita anteriormente a partir de estructuras obtenidas previamente para la proteína DNA-PKcs libre (*compárense las Figuras 4.2.2, 4.3.2 y 4.3.4*). En la estructura resuelta para el complejo anticuerpo:DNA-PKcs aparece además una masa adicional clara. Esta densidad emerge de la región cabeza y no de la región palma y correspondería por tanto al anticuerpo unido (*Figuras 4.3.4C y 4.3.4D*). Al disminuir el umbral de visualización esta masa es aún más evidente (*Figuras 4.3.4C y 4.3.4D*). Comparando las estructuras de criomicroscopía electrónica y del complejo anticuerpo:DNA-PKcs parece que el anticuerpo se uniría a la parte del dominio hombro más próxima a la región cabeza. Por tanto el marcaje con anticuerpos y la comparación de estructuras de microscopía electrónica refuerzan la asignación de la porción C-terminal a la región cabeza.



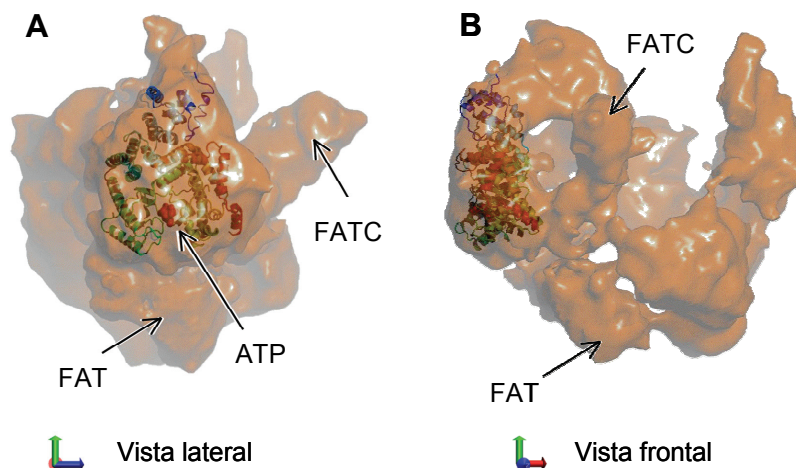
**Figura 4.3.4. Localización de la región C-terminal de la proteína DNA-PKcs mediante marcaje con anticuerpos.** (A) Gel de poliacrilamida en condiciones nativas de la proteína DNA-PKcs purificada teñido con plata. (B) “Western-blot” de un gel nativo en el que se analizó DNA-PKcs. El anticuerpo primario utilizado (DNA-PKcs G-4 sc-5282, Santa Cruz) reconoce a la proteína DNA-PKcs en condiciones nativas. (C) y (D) Vista frontal y lateral respectivamente de la estructura de tinción negativa obtenida para el complejo anticuerpo:DNA-PKcs. La estructura se representa a dos umbrales diferentes. El umbral del volumen interior opaco corresponde al 100% de la masa de la proteína DNA-PKcs. Este volumen contiene una densidad añadida clara que se ha representado en color verde y que se indica además con un círculo discontinuo. Cuando este volumen se representa a un menor umbral mediante la densidad azul transparente, la masa añadida se hace aún más evidente. La densidad adicional emerge desde la región cabeza de la proteína DNA-PKcs. Se indican las regiones principales cabeza y palma que componen la proteína DNA-PKcs y también la masa añadida correspondiente al anticuerpo “Ab”.

## B) Asignación y modelado de los dominios que componen la región cabeza

### *Dominio catalítico*

Por el momento no se ha resuelto la estructura atómica del dominio catalítico de DNA-PKcs ni de ningún otro miembro de la familia de las PIKKs. Sí se han resuelto diversas estructuras atómicas para diferentes isoformas de PI3Ks de tipo I (Walker et al., 1999; Miled et al., 2007). Recientemente se ha obtenido la estructura atómica correspondiente a la subunidad catalítica completa de la PI3K isoforma  $\alpha$  (PI3K $\alpha$ ) (fichero PDB 2RD0) (Huang et al., 2007). Anteriormente se había resuelto la estructura atómica de un fragmento que contiene parte de la subunidad catalítica de la proteína PI3K $\gamma$  (fichero PDB 1E8X) (Walker et al., 2000). Las PI3Ks presentan un dominio inmediatamente N-terminal al dominio catalítico formado por repeticiones no canónicas de motivos helicoidales (Djordjevic y Driscoll, 2002). Este dominio tiene una longitud aproximada de 200 aminoácidos que se organizan en cinco pares de  $\alpha$ -hélices antiparalelas. En las estructuras resueltas para las diferentes PI3Ks el dominio helicoidal actúa a modo de esqueleto sobre el que descansa el dominio quinasa. Los dominios helicoidal y catalítico serían estructuralmente inseparables y forman un núcleo central alrededor del cual se disponen el resto de dominios (Huang et al., 2007). Esta organización característica se mantiene en las subunidades catalíticas de las diferentes isoformas de PI3Ks a pesar de la baja identidad de secuencia que comparten (Djordjevic y Driscoll, 2002). En el caso de las subunidades catalíticas de las proteínas PI3K $\alpha$  y PI3K $\gamma$  la identidad de secuencia es del 35%. La proteína DNA-PKcs también presenta secuencias N-terminales al dominio quinasa homólogas a las del dominio helicoidal de las PI3Ks, por ello se propuso que estas secuencias podrían tener una función similar a la que se observa en las proteínas PI3Ks. Según esto la región formada por repeticiones de motivos helicoidales estaría también presente en DNA-PKcs y por tanto, en el análisis llevado a cabo posteriormente, se asumió que la disposición general de la región catalítica de las proteínas PI3Ks estaría conservada en los dominios catalíticos de las PIKKs.

A partir de la estructura atómica correspondiente al fragmento catalítico de la proteína PI3K $\gamma$  (fichero PDB 1E8X) se generó un modelo atómico que incluía el dominio catalítico mínimo (aminoácidos 727 a 1.092 de la PI3K $\gamma$ ), y el dominio formado por repeticiones helicoidales (aminoácidos 546 a 726 de la PI3K $\gamma$ ). Posteriormente este modelo se ajustó en la estructura de criomicroscopía electrónica resuelta para la proteína DNA-PKcs. Los ajustes con mayor coeficiente de correlación colocaban el modelo en la región cabeza (*Figura 4.3.5*). Concretamente el modelo se ajustaba con una orientación definida en la densidad con forma triangular correspondiente al dominio central de la región cabeza (*de color verde en las Figuras 4.3.2D, 4.3.2E y 4.3.2F*). Cuando el modelo atómico se ajustaba bien en volúmenes más preliminares del proceso de refinamiento, bien en el volumen final pero filtrado a una menor resolución, se obtenía además una segunda solución con un coeficiente de correlación alto. En esta otra solución el modelo atómico también se localizaba en el dominio central de la región cabeza, pero estaba girado 180° en el eje vertical del dominio central triangular respecto al ajuste obtenido en el volumen final. El modelo atómico filtrado a una resolución comparable a la obtenida en el volumen de criomicroscopía electrónica tiene cierta simetría rotacional de orden 2. Esto también sucede en la densidad correspondiente al dominio central de la región cabeza, por tanto el ajuste en ambas orientaciones es



**Figura 4.3.5. Localización del dominio catalítico en la región cabeza.** (A) y (B) Vista lateral y frontal respectivamente del ajuste del modelo atómico construido a partir de la estructura resuelta para el dominio catalítico de la proteína PI3K $\gamma$  (fichero PDB 1E8X) (Walker et al., 2000). Se indica también la localización del ATP en el centro activo y las densidades del volumen asignadas a los dominios FAT y FATC (véase más adelante).

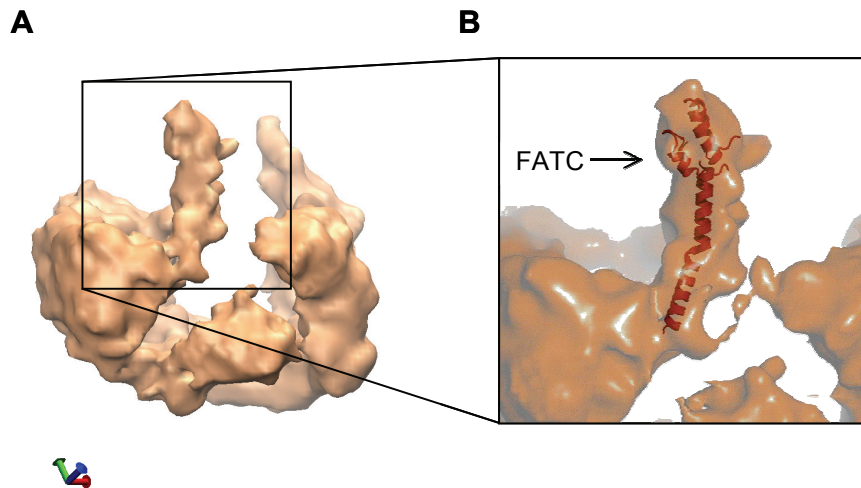
posible. Esta ambigüedad se debe a la resolución limitada del volumen de criomicroscopía electrónica, pero también es consecuencia de la pseudo simetría del modelo atómico y de la densidad donde ajusta. En cualquier caso el modelo construido a partir de la estructura atómica del dominio catalítico de la proteína PI3Ky siempre se ajustaba en el dominio central de la región cabeza. Además la forma triangular característica de este modelo hace que no pueda ajustarse en ninguna otra región de la estructura de criomicroscopía (*Figura 4.3.5*). Este ajuste de nuevo es totalmente compatible con los datos anteriores que indican que la porción C-terminal de la proteína DNA-PKcs se localizaría en la región cabeza.

### *Dominio FATC*

De las dos orientaciones posibles del modelo atómico en el dominio central sólo una justifica de manera consistente la conectividad con el resto de los dominios de la región cabeza. Esta orientación corresponde al ajuste que se obtiene en el volumen final de criomicroscopía electrónica (*Figura 4.3.5*). En esa posición el centro activo se orienta hacia el exterior y por tanto queda accesible a posibles sustratos (*en la Figura 4.3.5A se indica el sitio de unión del ATP en el centro activo*). Además el extremo C-terminal del modelo atómico quedaría próximo a la conexión entre el dominio central y el dominio de menor tamaño con forma tubular curvada. Las dimensiones de este pequeño dominio (*de color rojo en las Figuras 4.3.2D, 4.3.2E y 4.3.2F*) permiten alojar únicamente estructuras poco voluminosas como por ejemplo una  $\alpha$ -hélice. Estructuras secundarias de mayor tamaño no podrían acomodarse dentro de esta densidad, lo que sugiere que corresponde a un dominio terminal. Esto, junto con la proximidad al extremo C-terminal del modelo atómico ajustado en el dominio central, permitió asignar este dominio al dominio más C-terminal FATC. El dominio FATC tiene una longitud aproximada de 100 aminoácidos. La estructura secundaria predicha para este dominio consiste en una  $\alpha$ -hélice N-terminal de unos 35 aminoácidos seguida por dos  $\alpha$ -hélices de menor tamaño separadas por segmentos enrollados. Un modelo atómico con estas características encaja significativamente bien en la densidad asignada previamente a este pequeño dominio (*Figura 4.3.6*). El modelo propuesto se vio



confirmado posteriormente con la obtención de la estructura atómica correspondiente al dominio FATC de la proteína de levaduras TOR1 (fichero PDB 1W1N) (Dames et al., 2005).

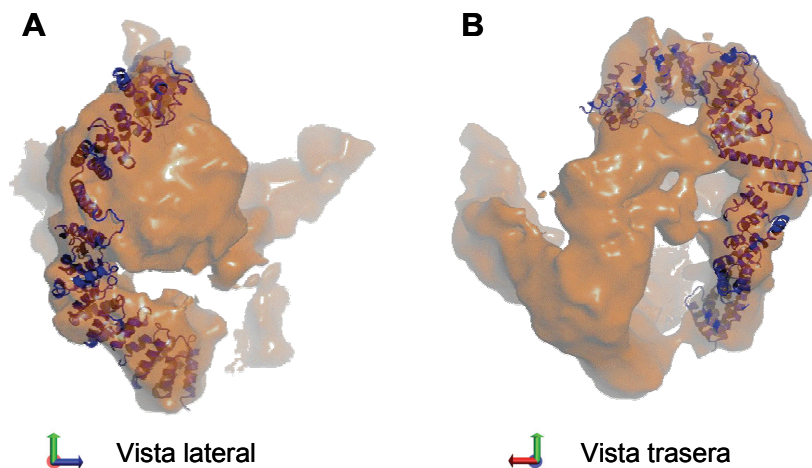


**Figura 4.3.6. Localización del dominio FATC en la región cabeza.** (A) Vista de la reconstrucción tridimensional resuelta mediante criomicroscopía electrónica para la proteína DNA-PKcs. (B) Detalle del ajuste de un modelo generado para el dominio FATC en la orientación que se muestra en (A).

#### *Dominio FAT*

A partir del dominio central sale un segundo dominio de longitud similar al dominio asignado como FATC pero más masivo (*de color magenta en las Figuras 4.3.2D, 4.3.2E y 4.3.2F*). La conexión entre este segundo dominio y el dominio central se localiza, en el modelo atómico ajustado correspondiente al dominio catalítico de la proteína PI3K $\gamma$ , próxima a la conexión entre el lóbulo N-terminal del dominio quinasa y el dominio formado por repeticiones helicoidales. Como se mencionó anteriormente este dominio está presente en todas las PI3Ks y tiene una organización característica que consiste en repeticiones de motivos helicoidales (Djordjevic y Driscoll, 2002). El análisis de la secuencia de las PIKKs sugiere que la región inmediatamente N-terminal al dominio quinasa también estaría formada por repeticiones helicoidales características (Perry y Kleckner, 2003; Brewerton et al., 2004). En el caso de las PIKKs esta región débilmente conservada se denomina dominio FAT (véase el apartado 1.4.2 de Introducción), y tendría una longitud aproximada de 500 aminoácidos (Bosotti et al., 2000). Este dominio podría formar una estructura continua de

repeticiones helicoidales que recorriese la región cabeza. Esta estructura comenzaría en el extremo superior de la región cabeza donde conectaría con la región brazo. A continuación bordearía el dominio catalítico dando lugar a una región equivalente al dominio helicoidal sobre el que descansa el dominio catalítico de las PI3Ks (*Figura 4.3.7*). Esta región de repeticiones helicoidales se localizaría en el lado del dominio central triangular opuesto al dominio FATC. Por último el dominio FAT continuaría hacia el dominio masivo que emerge desde el dominio central (*Figura 4.3.7*).



**Figura 4.3.7. Localización del dominio FAT en la región cabeza.** (A) y (B) Vista lateral y trasera respectivamente del ajuste de un modelo atómico para el dominio FAT en el volumen obtenido. El dominio FAT podría formar una estructura continua que recorriese la región cabeza. El modelo ajustado se construyó a partir de varias copias del dominio helicoidal presente en la estructura atómica del dominio catalítico de la PI3K $\gamma$  (fichero PDB 1E8X).

La región cabeza se organiza por tanto alrededor del dominio catalítico que ocupa la densidad central con forma triangular. Dentro de este dominio central el centro activo se localiza en la base de la densidad triangular entre los dominios FAT y FATC accesible a los distintos sustratos proteicos.

C) Asignación y modelado de los dominios que componen las regiones brazo y palma

Previamente se ha asignado la región cabeza completa a los dominios conservados C-terminales presentes en la secuencia de la proteína DNA-PKcs.

La densidad correspondiente a las regiones brazo y palma debe por tanto justificar el resto de la secuencia.

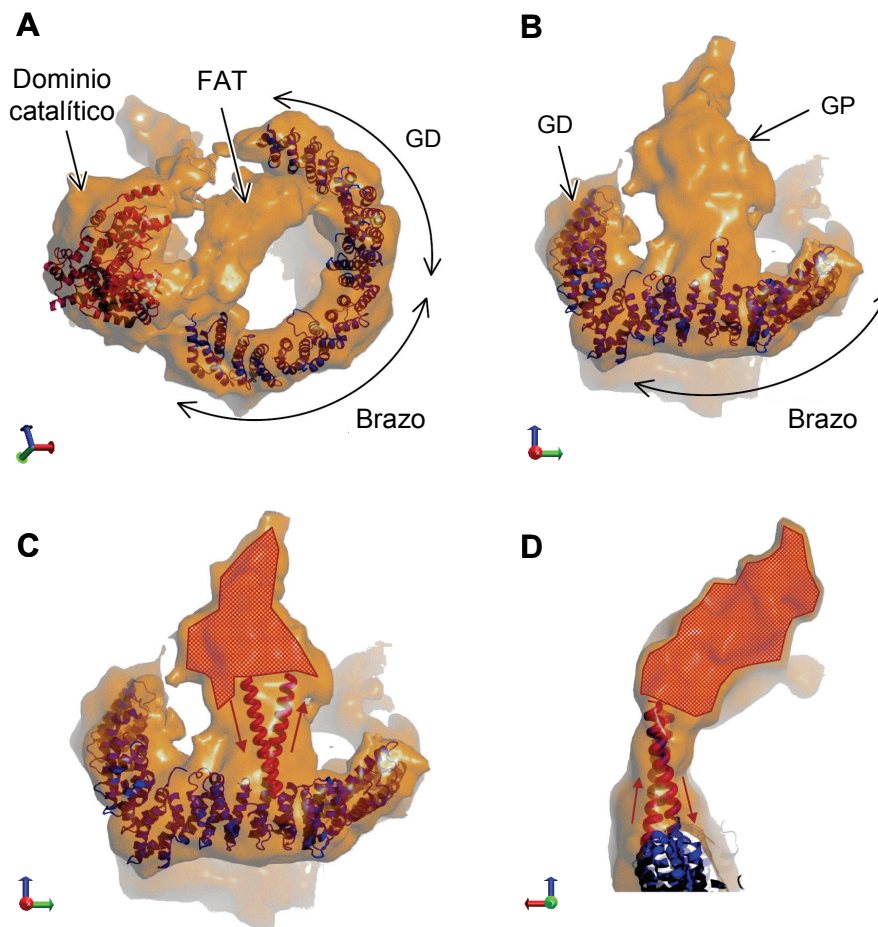
### *Región Brazo*

Como se comentó en el capítulo de Introducción la mayor parte de la secuencia de las PIKKs se compone de repeticiones de motivos helicoidales de tipo HEAT y PFT (Perry y Kleckner, 2003; Brewerton et al., 2004). Estas repeticiones presentan variaciones en su secuencia y por ello la conservación entre las PIKKs se limita a la región C-terminal. A pesar de la diversidad de secuencia, regiones ricas en estas repeticiones adquieren estructuras terciarias características muy similares (Groves y Barford, 1999). Estas estructuras tienen una forma típica de tubo aplanado ligeramente curvado como la que se observa en estructuras atómicas de la fosfatasa PP2A y de la importina  $\beta$  (véase apartado 1.4.2 de Introducción) (Groves et al., 1999; Lee et al., 2003). La región brazo se localiza a continuación del dominio hombro de la región cabeza (véase la Figura 4.3.2). Su forma y dimensiones son totalmente compatibles con las de estructuras generadas a partir de regiones ricas en repeticiones de tipo HEAT. En la mayoría de la densidad correspondiente a la región brazo se observan estas características, aunque también presenta pequeñas alteraciones. Esto podría deberse a que la arquitectura de las repeticiones HEAT no es homogénea ni continua en la región brazo. En cualquier caso, la mayor parte de esta región puede acomodar segmentos ricos en repeticiones HEAT extraídos de estructuras atómicas de otras proteínas (ficheros PDB 1B3U y 1UKL) (Figuras 4.3.8A y 4.3.8B).

### *Región Palma*

Esta región se compone de dos dominios garra que emergen perpendicularmente desde el borde de la región brazo (Figuras 4.3.8A y 4.3.8B). Además por su localización en la secuencia la región palma estaría constituida principalmente por  $\alpha$ -hélices. Las regiones ricas en repeticiones de motivos helicoidales forman estructuras con una polaridad claramente definida en las cuales las zonas N y C-terminales se encuentran en extremos opuestos. La

extensión de este tipo de estructuras tendría lugar por adición de motivos helicoidales (Groves y Barford, 1999). Según esto la garra distal podría aparecer como consecuencia de la extensión de la región brazo, ya que su forma y dimensiones son compatibles con las de un segmento de repeticiones HEAT (*Figuras 4.3.8A y 4.3.8B*). La garra proximal sin embargo implica una alteración en lo que sería una estructura terciaria canónica de un segmento rico en repeticiones HEAT. Esta alteración podría aparecer como consecuencia de una inserción en las repeticiones helicoidales que forman la región brazo. Se

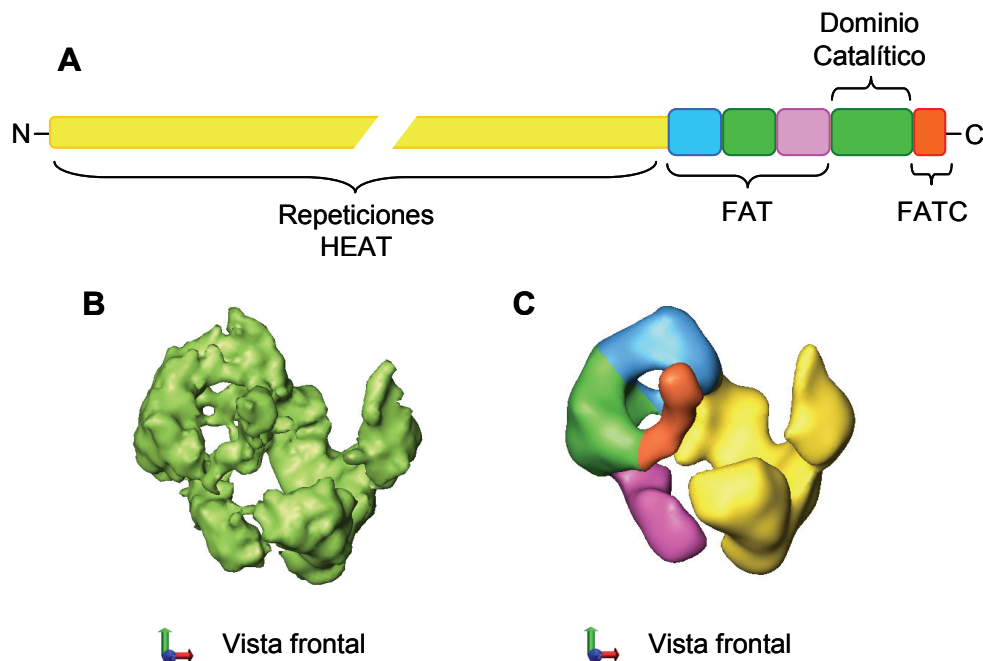


**Figura 4.3.8. Localización de los dominios que componen las regiones brazo y palma. (A) y (B)** Ajuste en el volumen de criomicroscopía electrónica de estructuras atómicas correspondientes a diferentes segmentos de repeticiones HEAT. Los segmentos ajustados se extrajeron de las estructuras resueltas para las proteínas correspondientes a la importina  $\beta 1$  (Lee et al., 2003) y a la subunidad PR65 $\alpha$  de la fosfatasa PP2A (Groves et al., 1999), (archivos PDB 1UKL e 1B3U respectivamente). En (A) se muestra también el ajuste del modelo atómico construido a partir de la estructura resuelta para el dominio catalítico de la proteína PI3K $\gamma$  (archivo PDB 1E8X). Los dominios garra distal y garra proximal se indican como “GD” y “GP” respectivamente. **(C) y (D)** Diferentes vista de un modelo de inserción de la garra proximal en la estructura canónica de un segmento de repeticiones HEAT.

han descrito inserciones en estructuras formadas por repeticiones de motivos helicoidales (Malek et al., 2003). Una inserción que justifique la garra proximal podría aparecer como consecuencia de la sustitución del bucle que conecta dos  $\alpha$ -hélices adyacentes por el nuevo dominio insertado. Si estas  $\alpha$ -hélices sobresalen de la estructura tubular podría darse la inserción sin que se viera afectada la estructura terciaria del segmento rico en repeticiones helicoidales (*Figuras 4.3.8C y 4.3.8D*). Este modelo de inserción es compatible con la forma y dimensiones de la garra proximal. Además la densidad correspondiente a la conexión entre la región brazo y la garra proximal tiene un grosor que podría acomodar un par de  $\alpha$ -hélices (*Figuras 4.3.8C y 4.3.8D*).

### 4.3.3 Conclusión

La estructura resuelta para la proteína DNA-PKcs a 13 Å de resolución permite identificar los distintos dominios en los que se organiza esta PIKK. Además a partir de esta estructura ha sido posible proponer un modelo que



**Figura 4.3.9. Organización tridimensional de la proteína DNA-PKcs.** (A) Esquema de la cadena polipeptídica de la proteína DNA-PKcs. Se indican de manera aproximada las regiones principales identificadas en esta proteína. (B) Vista frontal de la estructura resuelta mediante criomicroscopía electrónica para la proteína DNA-PKcs. (C) Vista frontal de la reconstrucción tridimensional mostrada en (B) pero filtrada a 30 Å de resolución. El código de colores coincide con el que se utiliza en (A), por lo que se representa de manera simplificada la posible organización tridimensional de la proteína DNA-PKcs.

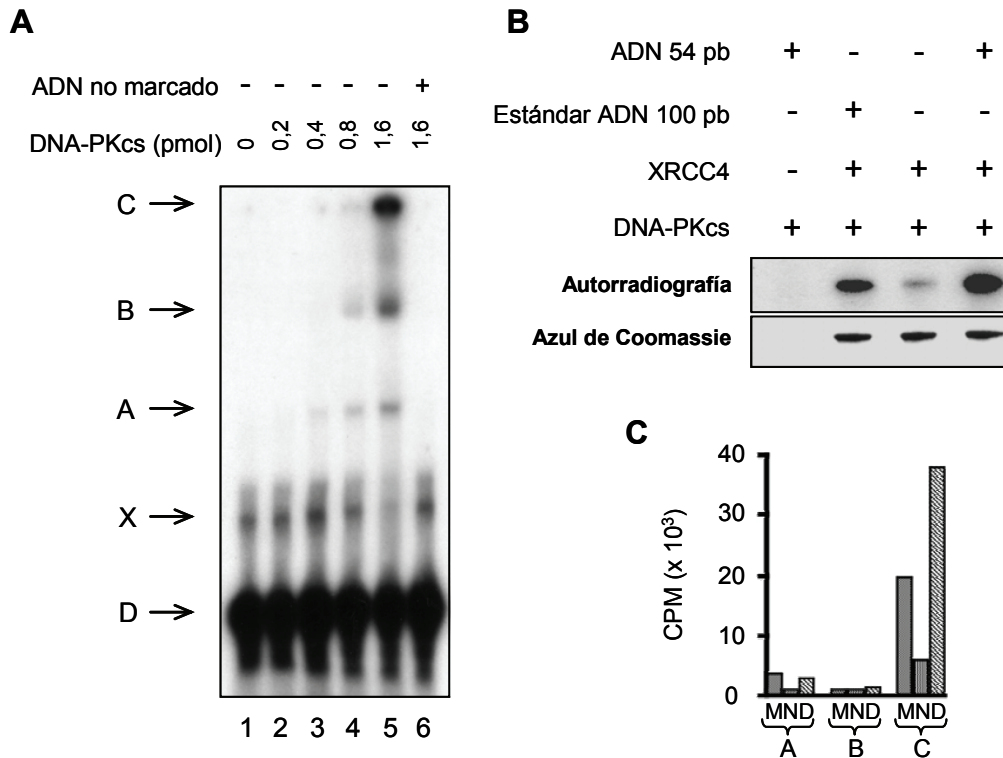
explique la distribución espacial de la cadena polipeptídica completa (*Figura 4.3.9*). La estructura resuelta es totalmente compatible con datos estructurales obtenidos previamente a partir de especímenes teñidos negativamente.

## 4.4 Análisis estructural del complejo DNA-PKcs:ADN mediante tinción negativa

La proteína DNA-PKcs se asocia a los extremos de ADN y esa interacción conlleva la activación de su capacidad quinasa (Hammarsten y Chu, 1998; Leuther et al., 1999). Con el objetivo de estudiar los cambios conformacionales que tienen lugar en la proteína DNA-PKcs al asociarse al ADN se llevó a cabo el análisis estructural del complejo DNA-PKcs:ADN.

### 4.4.1 Caracterización bioquímica de la interacción entre la proteína DNA-PKcs y el ADN

Para determinar la capacidad de la proteína purificada de interactuar con ADN se llevaron a cabo experimentos de cambio de movilidad electroforética en condiciones nativas (*Figura 4.4.1A*). En presencia de ADN de 54 pb marcado radiactivamente la proteína DNA-PKcs forma tres complejos distintos con el ADN denominados A, B y C. Cuando se añade exceso de ADN no marcado radiactivamente a la mezcla de reacción los diferentes complejos desaparecen (*carril 6 de la Figura 4.4.1A*, y por tanto la señal observada es debida a la formación de complejos estables que contienen DNA-PKcs y ADN. La aparición de estos tres complejos se ha descrito previamente (Leuther et al., 1999), de manera que el complejo A contiene una única molécula de DNA-PKcs, el complejo B contiene dos moléculas de DNA-PKcs y el complejo C que aparece a la altura del pocillo corresponde a agregados que se generan como consecuencia de interacciones proteína:proteína. La banda X característica de este análisis aparece independientemente de que la proteína DNA-PKcs esté presente en la mezcla de reacción. Además la intensidad de esta banda disminuye al incrementar la cantidad de DNA-PKcs en la reacción. Por ello se ha propuesto que podría aparecer como consecuencia de la capacidad del



**Figura 4.4.1. Análisis bioquímico de la interacción entre la proteína DNA-PKcs y el ADN.** **(A)** Experimentos de cambio de movilidad electroforética en condiciones nativas de la proteína DNA-PKcs. Cantidades crecientes de DNA-PKcs se incubaron con 20 fmol de un oligonucleótido de 54 pb marcado radiactivamente con  $^{32}\text{P}$  en 5'. La proteína DNA-PKcs forma tres complejos diferentes con el ADN indicados como A, B y C. Estos complejos desaparecían cuando se añadía a la mezcla de reacción 500 ng de ADN competidor no marcado radiactivamente. La banda X corresponde a ADN capaz de adquirir determinada estructura secundaria, mientras que la banda D corresponde al exceso de ADN marcado radiactivamente no unido a DNA-PKcs. **(B)** Autorradiografía en la que se aprecia la fosforilación de la proteína XRCC4 por DNA-PKcs dependiente de ADN. **(C)** A partir de autorradiografías como la que se muestra en (B) se extrajeron determinadas bandas y se midió su radiactividad. En A se analiza la autofosforilación de DNA-PKcs en ausencia de XRCC4. En B se analiza la autofosforilación de DNA-PKcs en presencia de XRCC4. En C se analiza la fosforilación de XRCC4. Dentro de cada experimento M corresponde a la fosforilación en presencia de un estándar de ADN de 100 pb (New England BioLabs). N corresponde a la fosforilación cuando no hay ADN de ningún tipo en la mezcla de reacción. D corresponde a la fosforilación cuando se añade ADN de 54 pb y extremos romos a la mezcla de reacción. CPM viene de "Cuentas por minuto".

oligonucleótido antisentido de formar dímeros y, por tanto, correspondería a ADN capaz de adquirir determinada estructura secundaria. También se valoró la activación de la capacidad quinasa de DNA-PKcs dependiente de ADN utilizando como sustrato la proteína XRCC4 humana (*Figura 4.4.1B*). Cuando se añadía ADN de 54 pb y extremos romos a la mezcla de reacción, la

fosforilación de la proteína XRCC4 dependiente de DNA-PKcs aumentaba aproximadamente seis veces respecto a cuando el ADN no estaba presente en la mezcla (*Figura 4.4.1B* y *experimento C de la Figura 4.4.1C*). En presencia de ADN de 54 pb y extremos romos la fosforilación de la proteína XRCC4 era el doble que la que se obtenía utilizando un estándar de ADN de 100 pb (New England BioLabs). Además DNA-PKcs mantiene la capacidad de autofosforilarse cuando se activa su capacidad quinasa dependiente de ADN (*experimentos A y B de la Figura 4.4.1C*). Por tanto la proteína DNA-PKcs purificada conserva su actividad catalítica dependiente de ADN y puede interaccionar de manera estable con ADN de 54 pb y extremos romos. Los experimentos descritos fueron realizados por el Dr. J. D. Maman en el marco de la colaboración establecida con el laboratorio del Profesor L. H. Pearl.

#### 4.4.2 Reconstrucción tridimensional del complejo DNA-PKcs:ADN

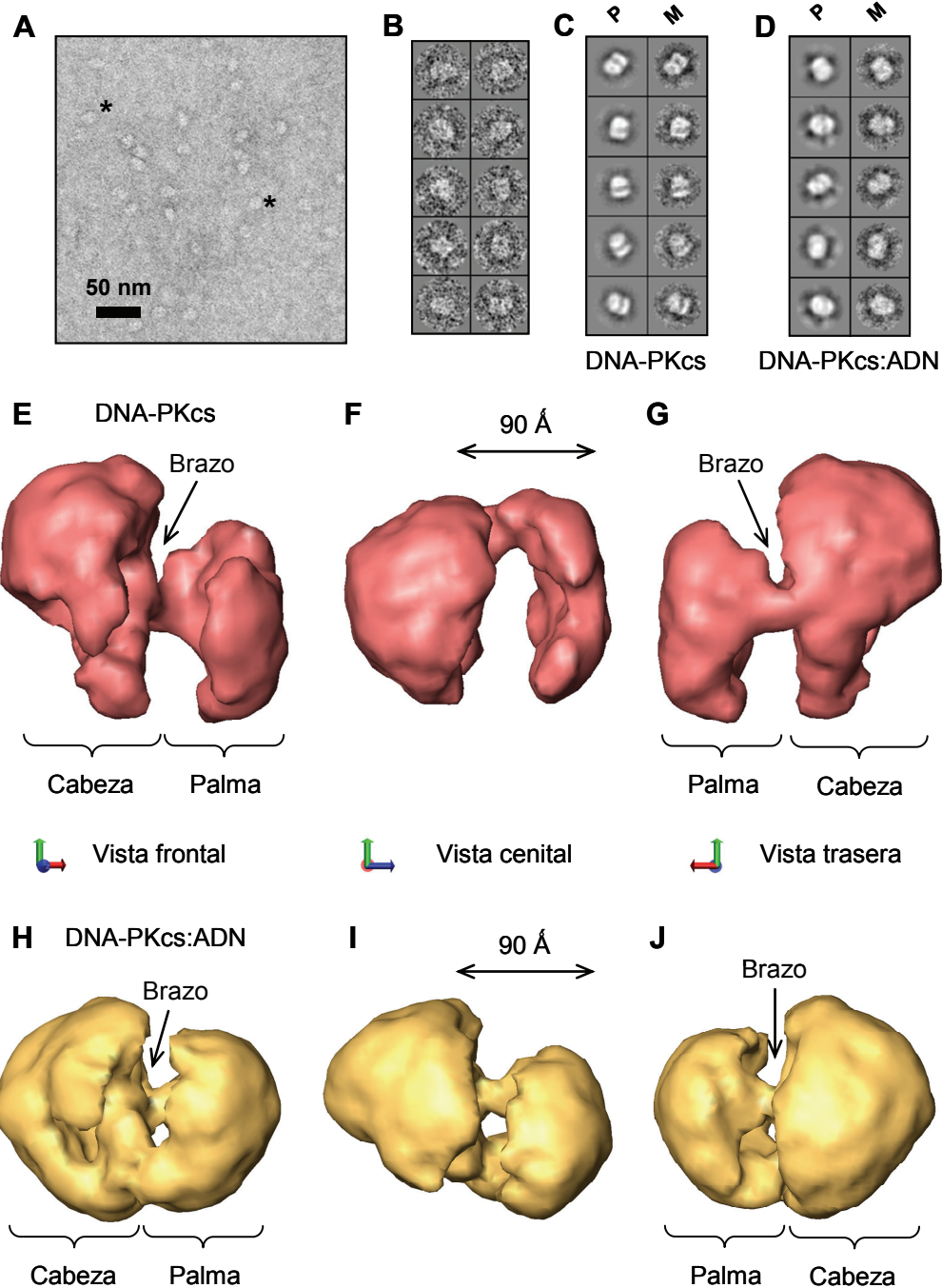
La caracterización bioquímica de la interacción entre la proteína DNA-PKcs y el ADN se llevó a cabo utilizando una molécula de ADN de 54 pb y extremos romos. Estudios previos habían demostrado también que esta molécula era un sustrato idóneo para analizar la unión y activación de DNA-PKcs mediada por ácidos nucleicos (Leuther et al., 1999), por ello el ensamblaje del complejo DNA-PKcs:ADN se llevó a cabo utilizando esta misma molécula de ADN. La proteína DNA-PKcs se incubó con el ADN en una relación molar 1:6 de proteína y ADN respectivamente y, de esta manera, se favorecía la formación de complejos DNA-PKcs:ADN. La mezcla se tiñó negativamente y se observó en el microscopio electrónico (*Figura 4.4.2A*). A partir de las micrografías electrónicas se seleccionaron un total de 10.468 imágenes correspondientes a moléculas individuales que, seguidamente, se utilizaron en un proceso de refinamiento angular iterativo (*Figura 4.4.2B*). Para generar los diferentes volúmenes iniciales primero se obtuvieron imágenes medias libres de referencia a partir del conjunto inicial de datos. A continuación las imágenes medias se combinaron en tres dimensiones mediante la técnica de líneas comunes. Durante el proceso de refinamiento angular iterativo en ningún momento se utilizaron estructuras resueltas previamente para la proteína DNA-PKcs.



El exceso de ADN presente en la mezcla de reacción debe desplazar el equilibrio hacia la formación de complejos DNA-PKcs:ADN. Aun así no se podía descartar la presencia de DNA-PKcs libre no unida a ADN, por ello durante el proceso de refinamiento angular iterativo se consideró que se partía de una mezcla heterogénea de moléculas. Esta asunción implicaba definir una estrategia de procesamiento restrictiva de manera que los volúmenes que se obtuvieran fueran coherentes con un único tipo de moléculas homogéneas (véase el apartado 3.13.5 de Materiales y Métodos). Según esto el conjunto inicial de partículas pudo dividirse en dos grupos cada uno de los cuales era compatible con un único volumen. Un total de 3.049 imágenes generaban un volumen que correlacionaba con la estructura obtenida mediante tinción negativa para la proteína DNA-PKcs no unida a ADN (*Figuras 4.4.2E, 4.4.2F y 4.4.2G*). Por otro lado 3.280 partículas no incluidas en este primer conjunto generaban un volumen que, si bien compartía la organización descrita anteriormente para DNA-PKcs, también presentaba diferencias significativas (*Figuras 4.4.2H, 4.4.2I y 4.4.2J*). Ambos volúmenes tienen una resolución aproximada de 30 Å para un valor de FSC de 0,5 (*Figuras 4.4.3B y 4.4.3D*), además sus dimensiones son totalmente compatibles con estructuras resueltas previamente para la proteína DNA-PKcs. El resto de las 10.468 partículas no podían clasificarse inequívocamente en uno de los volúmenes y se descartaron.

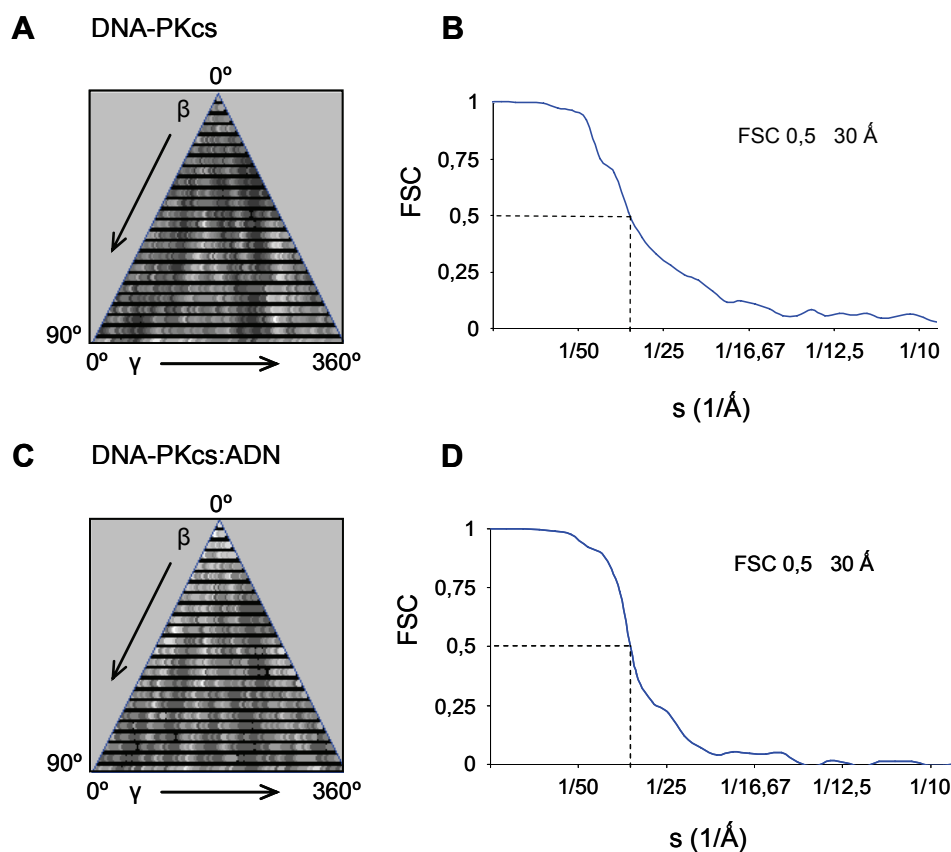
**Figura 4.4.2. Análisis mediante microscopía electrónica de especímenes teñidos negativamente de la muestra de DNA-PKcs incubada con ADN (página siguiente).**

**(A)** Micrografía electrónica obtenida a partir de una rejilla teñida negativamente de la proteína DNA-PKcs incubada con ADN. Algunas moléculas evidentes se indican con “\*”. **(B)** Galería de partículas individuales extraídas de micrografías como la que se muestra en (A). **(C)** y **(D)** Galerías de proyecciones (P) y medias equivalentes (M) obtenidas en la última iteración de los procesos de refinamiento que permitieron generar las estructuras de DNA-PKcs y DNA-PKcs:ADN que se muestran a continuación. En ambos casos el parecido entre proyecciones y medias equivalentes evidencia la estabilización de los procesos de refinamiento correspondientes. **(E)**, **(F)** y **(G)** Vista frontal, cenital y trasera respectivamente de la estructura resuelta que correspondería a la proteína DNA-PKcs libre. **(H)**, **(I)** y **(J)** Vista frontal, cenital y trasera respectivamente de la estructura resuelta correspondiente al complejo DNA-PKcs:ADN. La “mano” escogida se seleccionó por comparación con estructuras previas obtenidas para DNA-PKcs. Se indican también las regiones características cabeza, brazo y palma que componen esta proteína.



Para comprobar la calidad de la asignación de las partículas a cada volumen se llevaron a cabo procesos de refinamiento adicionales. En estos refinamientos se utilizaron como volúmenes modelo las estructuras finales obtenidas a partir de cada conjunto de imágenes. Cada estructura final se utilizó como volumen modelo del conjunto de imágenes alternativo de manera que se favorecía un sesgo durante el proceso de refinamiento. En ambos casos los volúmenes intermedios evolucionaron hacia volúmenes que recordaban a la estructura final obtenida previamente para cada conjunto de datos.

Por tanto a partir de los datos iniciales se resolvieron dos estructuras. Una de ellas era totalmente compatible con estructuras resueltas previamente para la proteína DNA-PKcs. El otro volumen resuelto, aun compartiendo la misma organización estructural descrita para la proteína DNA-PKcs libre, presentaba cambios conformacionales evidentes. Este segundo volumen debía generarse a partir de las imágenes de proyección del complejo DNA-PKcs:ADN presente en la mezcla de reacción.

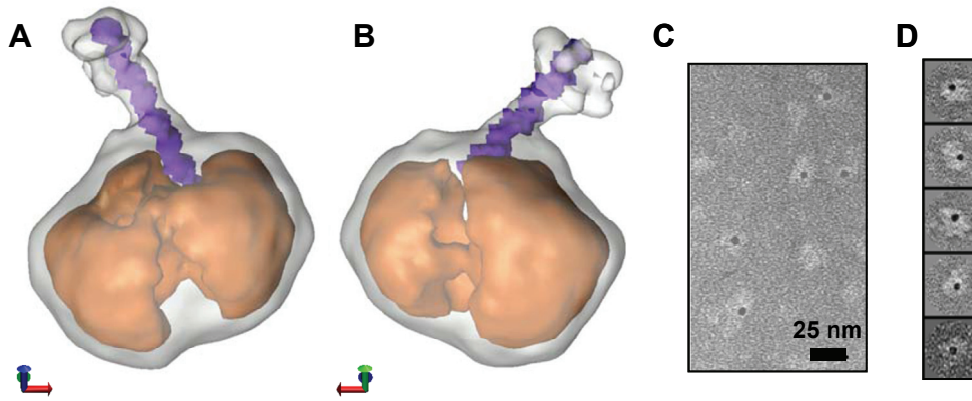


**Figura 4.4.3. Representación de los ángulos de Euler y cálculo de la resolución para las estructuras resueltas correspondientes a DNA-PKcs y al complejo DNA-PKcs:ADN. (A)** Representación de los ángulos de Euler del conjunto de partículas individuales que permitió obtener la estructura correspondiente a DNA-PKcs. **(B)** Estimación de la resolución del volumen correspondiente a la proteína DNA-PKcs para un valor de FSC de 0,5. **(C)** Representación de los ángulos de Euler correspondientes a las partículas que permitieron obtener la estructura del complejo DNA-PKcs:ADN. **(D)** Resolución estimada para la estructura correspondiente al complejo DNA-PKcs:ADN según un valor de FSC de 0,5. (A) y (C) Evidencian que, en ambos casos, las imágenes utilizadas para generar las diferentes estructuras tenían una diversidad de ángulos de Euler adecuada.

#### 4.4.3 Localización del ADN en los complejos DNA-PKcs:ADN

Se realizaron experimentos adicionales para confirmar que la nueva estructura resuelta en la que se apreciaban cambios conformacionales característicos correspondía al complejo DNA-PK:ADN. La localización del ADN en estructuras de complejos proteína:ADN obtenidas a partir de especímenes teñidos negativamente presenta complicaciones (Yu et al., 1996; Miyata et al., 2004; Tidow et al., 2007). Estas dificultades se deben principalmente a la flexibilidad intrínseca del ADN y también al modo en que interaccionan el ADN y el agente de tinción. En el volumen resuelto para el complejo DNA-PKcs:ADN se aprecian cambios conformacionales importantes, sin embargo no hay ninguna densidad evidente que pueda asignarse a la molécula de ADN (*Figuras 4.4.2H, 4.4.2I y 4.4.2J*). No obstante, cuando esta estructura se representa a un menor umbral pero que aún está claramente por encima del umbral del fondo, se observa una densidad tubular larga que se proyecta hacia el exterior de la estructura (*Figuras 4.4.4A y 4.4.4B*). Las dimensiones de esta densidad son compatibles con las de un pequeño fragmento de ADN que tuviera un tamaño aproximado de 36 pb (*Figuras 4.4.4A y 4.4.4B*). En el ensamblaje del complejo DNA-PKcs:ADN se utilizó una molécula de ADN de 54 pb y por lo tanto DNA-PKcs debe estar abrazando los 18 pb restantes. Se ha descrito que la longitud mínima necesaria para que una molécula de ADN active la capacidad quinasa de la proteína DNA-PKcs es de 15 pb (Leuther et al., 1999; Hammarsten et al., 2000; Martensson y Hammarsten, 2002). Esta longitud es totalmente compatible con las dimensiones de la porción de ADN que queda encerrada en el interior de DNA-PKcs. La densidad tubular únicamente se aprecia cuando la estructura del complejo DNA-PKcs:ADN se representa a un menor umbral. Esto podría indicar que posiblemente se estén promediando un pequeño rango de posiciones dentro de la estructura. En cualquier caso el ADN se proyecta hacia el exterior del complejo DNA-PKcs:ADN por un solo lado y en una orientación definida.

Como se mencionó anteriormente, a partir del conjunto inicial de datos también se obtuvo un volumen totalmente compatible con estructuras resueltas previamente para la proteína DNA-PKcs. Cuando este volumen se representa



**Figura 4.4.4. Localización del ADN presente en el complejo DNA-PKcs:ADN.** (A) y (B) Diferentes vistas de la estructura resuelta para el complejo DNA-PKcs:ADN. La estructura se muestra a dos umbrales diferentes. Cuando el volumen se representa a un menor umbral mediante la densidad blanca transparente aparece una densidad evidente en la que es posible encajar una molécula de ADN de 36 pb. (C) Micrografía de una rejilla de tinción negativa preparada a partir de una mezcla de reacción que contenía DNA-PKcs y complejos ADN:estreptavidina:oro. (D) Galería de partículas correspondientes a complejos DNA-PKcs:ADN marcados con estreptavidina:oro.

al umbral equivalente al que se evidencia la densidad tubular presente en el complejo DNA-PKcs:ADN no se observa ninguna densidad. Ambas estructuras se obtuvieron a partir de imágenes extraídas de las mismas micrografías. Todas estas evidencias sugieren por tanto que la densidad que se observa al disminuir el umbral corresponde a ADN presente en el complejo y, además, corroboran que la asignación de las imágenes iniciales a cada uno de los dos volúmenes ha sido correcta.

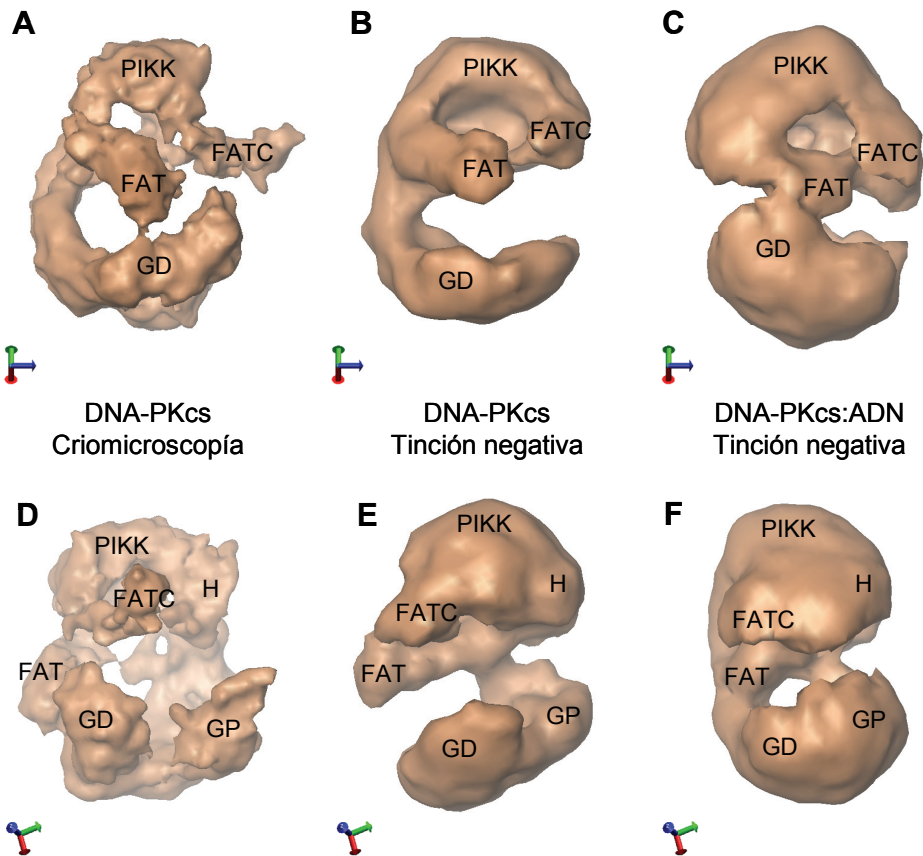
Para confirmar la presencia del ADN en las imágenes de microscopía electrónica se llevaron a cabo experimentos de marcaje. Primero se incubó ADN biotinilado de 54 pb y extremos romos con estreptavidina acoplada a un conjugado de oro de 5 nm de diámetro. Las características del ADN empleado, a excepción de la biotina, eran idénticas a las del ADN que se había utilizado previamente en el ensamblaje del complejo DNA-PKcs:ADN (véase el apartado 3.1 de Materiales y Métodos). Se generaban por tanto complejos ADN:estreptavidina:oro que posteriormente se incubaron con la proteína DNA-PKcs en una relación molar 1:1. La presencia de estreptavidina:oro en uno de los extremos del ADN restringe la unión de DNA-PKcs al extremo libre de la molécula de ADN. A continuación se prepararon rejillas de tinción negativa a

partir de la mezcla de reacción que se observaron en el microscopio electrónico (*Figuras 4.4.4C y 4.4.4D*). Además también se prepararon rejillas control de complejos estreptavidina:oro, de DNA-PKcs más complejos estreptavidina:oro en ausencia de ADN y, por último, de DNA-PKcs más complejos estreptavidina:oro más una molécula de ADN de 54 pb y extremos romos no biotinilada. Únicamente en presencia de complejos ADN:estreptavidina:oro se observaron un gran número de moléculas de DNA-PKcs marcadas con una partícula de oro (*Figuras 4.4.4C y 4.4.4D*). Además se observaron partículas de oro no asociadas a DNA-PKcs que correspondían a complejos ADN:estreptavidina:oro libres. Los experimentos de marcaje con estreptavidina:oro confirman por tanto la presencia de ADN unido a DNA-PKcs en la mezcla de reacción.

#### **4.4.4 Cambios conformacionales en la proteína DNA-PKcs al unirse a ADN**

El volumen resuelto para el complejo DNA-PKcs:ADN también se organiza en las regiones características cabeza, brazo y palma descritas anteriormente (*Figura 4.4.2*). En este caso la orientación relativa de las distintas regiones es diferente a la que se observa en estructuras obtenidas previamente para la proteína DNA-PKcs. Estas diferencias sugieren en último término la existencia de cambios conformacionales en DNA-PKcs como consecuencia de la unión al ADN. El análisis estructural mediante criomicroscopía electrónica permitió identificar y localizar los dominios que componen las diferentes regiones. Por tanto es posible asignar los cambios conformacionales presentes en la estructura del complejo DNA-PKcs:ADN a movimientos específicos de determinados dominios (*Figura 4.4.5*).

El cambio conformacional principal que conlleva la unión del ADN consiste en la interacción que se establece entre la región cabeza y la región palma. Este cambio conformacional es muy evidente al comparar vistas cenitales de las dos estructuras resueltas (*compárense las Figuras 4.4.2F y 4.4.2I*). Como consecuencia de la unión se producen cambios en la orientación de los dominios FATC y de la porción más C-terminal del dominio FAT así como en la garra distal (*Figura 4.4.5*). Estos cambios conformacionales provocan la



**Figura 4.4.5. Cambios conformacionales en la proteína DNA-PKcs provocados por la unión al ADN.** (A), (B), (C), (D), (E) y (F) Vistas equivalentes de las estructuras resueltas para la proteína DNA-PKcs y el complejo DNA-PKcs:ADN. Comparando las diferentes estructuras obtenidas es posible asignar los cambios conformacionales observados a dominios concretos. Se indican los dominios FAT, FATC, catalítico (PIKK), garra distal (GD), garra proximal (GP) y hombro (H).

aparición de un pequeño canal entre las regiones cabeza y palma con unas dimensiones suficientes como para alojar ADN de cadena doble (*Figuras 4.4.4A y 4.4.4B*). La formación del complejo DNA-PKcs:ADN implica que el extremo de ADN sea reconocido específicamente de algún modo. Por tanto deben existir regiones concretas en DNA-PKcs que permitan este reconocimiento y el mantenimiento de la interacción con la molécula de ADN.

#### 4.4.5 Análisis vibracional

El análisis vibracional es una herramienta extremadamente útil para predecir posibles movimientos presentes en complejos macromoleculares (Ming et al., 2002; Chacon et al., 2003). Por ello se llevó a cabo este análisis

con la estructura de DNA-PKcs resuelta a partir del conjunto de partículas que también permitió obtener la estructura del complejo DNA-PKcs:ADN. El análisis vibracional fue realizado en colaboración con el Dr. P. Chacón del Centro de Investigaciones Biológicas (CSIC).

En el caso de la proteína DNA-PKcs el modo de frecuencia más baja corresponde a un movimiento de la región palma dirigido hacia la región cabeza. Esta predicción concuerda con el cambio conformacional principal descrito a partir de la estructura resuelta para el complejo DNA-PKcs:ADN. Además la combinación lineal de este primer movimiento con otros modos de baja frecuencia permitía justificar el resto de cambios conformacionales presentes en el complejo. El análisis vibracional proporciona por tanto un modelo acerca de cómo podría tener lugar la transición entre las conformaciones de la proteína DNA-PKcs libre y unida a ADN.

#### **4.4.6 Conclusión**

Se ha realizado el análisis mediante microscopía electrónica de especímenes teñidos negativamente de la proteína DNA-PKcs incubada previamente con ADN. A partir de este análisis se han resuelto las estructuras correspondientes a la proteína DNA-PKcs libre y al complejo DNA-PKcs:ADN. Comparando ambas estructuras se observa que, en presencia de ADN, la proteína DNA-PKcs sufre cambios conformacionales importantes que consisten principalmente en un movimiento tipo abrazo que encierra el ADN. El modelo topológico propuesto para DNA-PKcs permite definir los dominios principales implicados en estos cambios conformacionales.

### **4.5 Análisis estructural del complejo DNA-PKcs:Ku:ADN mediante tinción negativa y criomicroscopía electrónica**

La purificación del complejo DNA-PKcs:Ku:ADN supuso un logro importante en el marco de la colaboración establecida entre los grupos del Profesor L. H. Pearl y del Dr. Óscar Llorca. A partir de la información acumulada en la



bibliografía se definieron pasos clave que permitieron llevar a cabo el ensamblaje del complejo DNA-PKcs:Ku:ADN *in vitro* (Douglas et al., 2001; Walker et al., 2001; Merkle et al., 2002). Posteriormente la muestra purificada se analizó mediante tinción negativa y criomicroscopía electrónica. La existencia de información estructural previa para DNA-PKcs y Ku (Leuther et al., 1999; Walker et al., 2001; Boskovic et al., 2003; Rivera-Calzada et al., 2005) permitió realizar una interpretación más completa de las estructuras que se resolvieron para el complejo DNA-PKcs:Ku:ADN.

#### **4.5.1 Purificación y caracterización bioquímica del complejo DNA-PKcs:Ku:ADN**

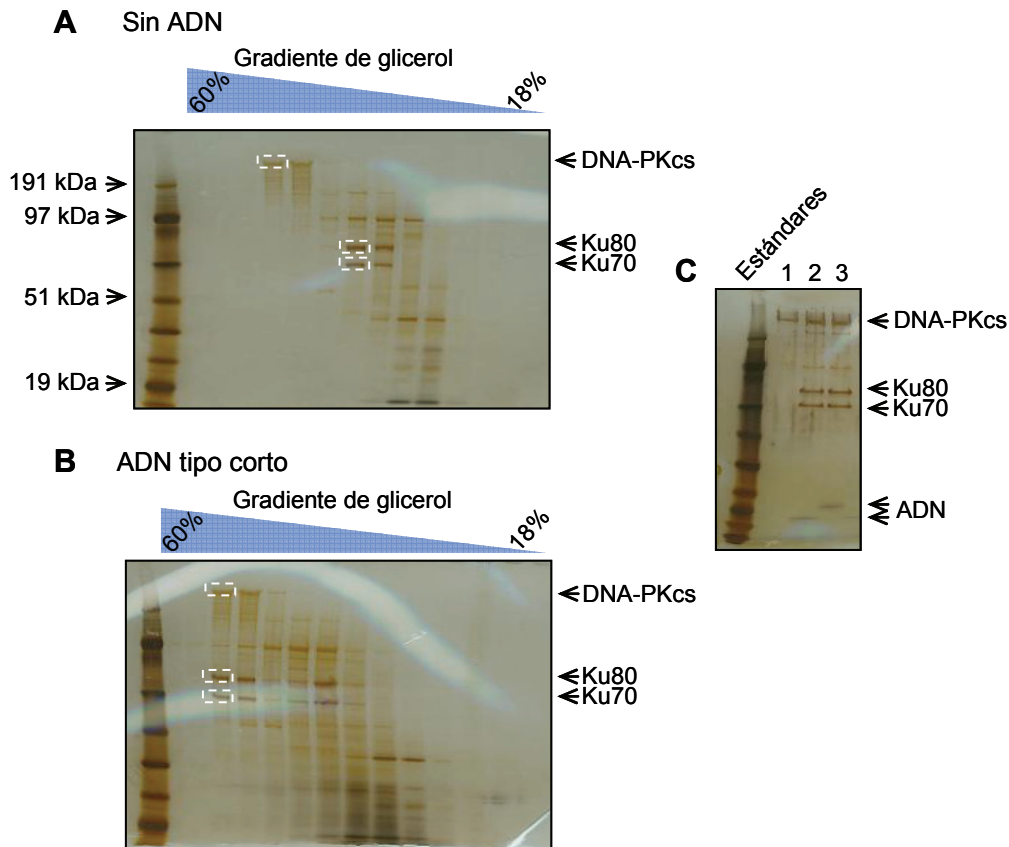
Se sabe que el heterodímero Ku y la proteína DNA-PKcs necesitan de la presencia de extremos de ADN para asociarse (Yaneva et al., 1997). Estudios previos han determinado los requisitos de la unión de ambas proteínas al ADN (Hammarsten y Chu, 1998; West et al., 1998; Walker et al., 2001; Boskovic et al., 2003). En base a esos estudios se diseñaron dos moléculas de ADN que se utilizaron en el ensamblaje del complejo DNA-PKcs:Ku:ADN (véase el apartado 3.1 de Materiales y Métodos). La longitud de ambas moléculas de ADN permitía la unión simultánea de las proteínas Ku y DNA-PKcs. Además el extremo de cada molécula de ADN no sobresalía excesivamente del complejo holoenzima:ADN. De esta manera se redujo la posible flexibilidad conformacional del complejo facilitando así su estudio por microscopía electrónica. Los dos tipos de moléculas tenían bloqueado uno de sus extremos. Esto restringía el acceso de la proteína Ku por un único extremo e impedía su posible liberación al migrar a lo largo del ADN. El bloqueo consistió en una estructura en forma de Y (véase la Figura 3.1.1 de Materiales y Métodos). Este tipo de estructura se había utilizado anteriormente en la obtención de la estructura atómica del complejo Ku:ADN (fichero PDB 1JEY) (Walker et al., 2001). El extremo bloqueado además estaba iminobiotinilado. Los dos tipos de moléculas se diferenciaban únicamente en la longitud del extremo sin bloqueo.

La purificación y ensamblaje del complejo DNA-PKcs:Ku:ADN fue realizado por la Dra. L. Spagnolo. Primero se purificaron las proteínas Ku y DNA-PKcs a

partir de extractos nucleares de células HeLa. Se ha descrito que la afinidad de la proteína DNA-PKcs por los extremos de ADN está regulada negativamente por autofosforilación (Douglas et al., 2001; Merkle et al., 2002), por ello aquellas fracciones que contenían la proteína DNA-PKcs y el heterodímero Ku se incubaron con fosfatasa- $\lambda$  antes de formar el complejo. Con esto se consiguió aumentar la homogeneidad de la muestra cuyo estado de fosforilación tras la purificación se desconocía y, además, se restauró la proteína DNA-PKcs a un estado de alta afinidad por extremos de ADN equivalente al de las etapas iniciales de la reparación por NHEJ. Para ensamblar el complejo las proteínas purificadas y defosforiladas se incubaron con los dos tipos de moléculas de ADN descritos anteriormente. A continuación la mezcla de reacción se centrifugó en gradientes de glicerol. Cuando se centrifugó una muestra a la que no se le había añadido ADN, Ku70 y Ku80 se purificaban a la vez puesto que ambas proteínas se asociaban para formar el heterodímero Ku (*Figura 4.5.1A*). Por el contrario el heterodímero Ku no se purificaba junto con la proteína DNA-PKcs (*Figura 4.5.1A*). Sin embargo, cuando en el gradiente de glicerol se centrifugaron las muestras a las que previamente se les había añadido ADN, se obtuvieron especies de mayor peso molecular (*Figura 4.5.1B*). Estas nuevas especies aparecieron tanto en presencia de ADN de tipo largo como de tipo corto. Posteriormente se comprobó que las especies de mayor peso molecular correspondían a un complejo estequiométrico formado por ADN y por las proteínas DNA-PKcs, Ku80 y Ku70 (*Figura 4.5.1C*). Fue posible por tanto obtener una muestra homogénea del complejo humano DNA-PKcs:Ku:ADN purificado. El análisis de partículas individuales tridimensional se llevó a cabo con la muestra correspondiente al complejo ensamblado sobre ADN de tipo corto. En esta muestra el extremo de la molécula de ADN sobresaldría menos del complejo y por ello debería presentar una mayor homogeneidad estructural.

#### **4.5.2 Reconstrucción tridimensional del complejo DNA-PKcs:Ku:ADN obtenida por tinción negativa**

El complejo purificado ensamblado sobre ADN de tipo corto se tiñó negativamente y se observó en el microscopio electrónico. En las micrografías



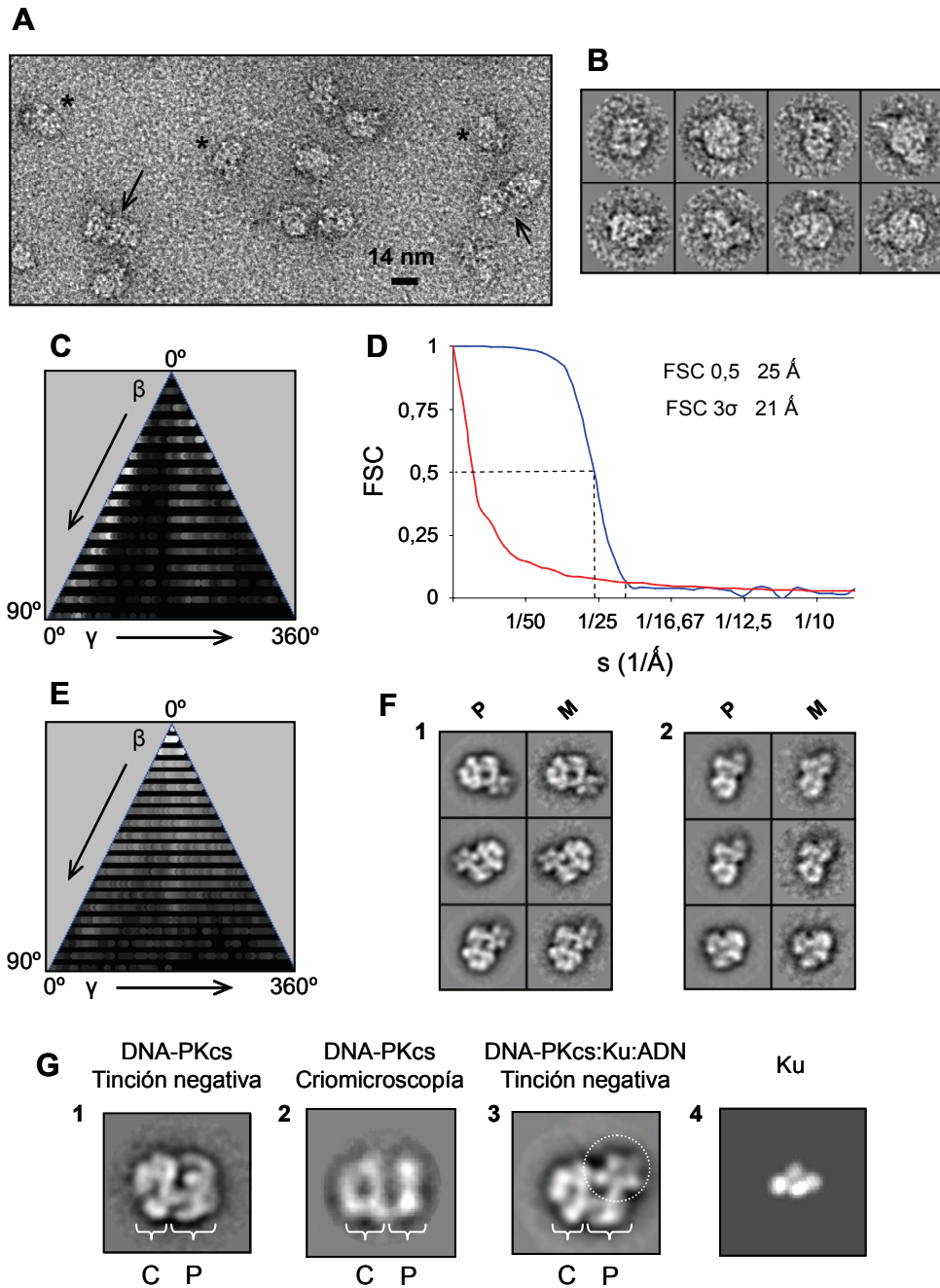
**Figura 4.5.1. Purificación del complejo DNA-PKcs:Ku:ADN.** (A) Análisis, por electroforesis en condiciones desnaturalizantes, de las diferentes fracciones obtenidas tras centrifugar en un gradiente de glicerol una muestra que contenía las proteínas Ku y DNA-PKcs. (B) Análisis como el que se describe en (A), pero en este caso la mezcla de reacción incluía además ADN de tipo corto. (C) Análisis por electroforesis en condiciones desnaturalizantes de la proteína DNA-PKcs (carril 1), y de la muestra del complejo DNA-PKcs:Ku:ADN purificado ensamblado sobre ADN de tipo corto (carril 2) o largo (carril 3).

se distinguían dos tipos de partículas de diferente tamaño (*Figura 4.5.2A*). Las partículas más pequeñas recordaban a vistas típicas de la proteína DNA-PKcs descritas en estudios previos, pero en este caso era evidente una masa adicional añadida (*Figura 4.5.2B*). Sin embargo las partículas más grandes parecían estar formadas por dos de las partículas de menor tamaño y también se llevó a cabo su análisis (véase más adelante el apartado 4.5.7).

Inicialmente se seleccionaron unas 7.000 imágenes correspondientes a partículas pequeñas. Estas imágenes se utilizaron en un proceso de refinamiento angular iterativo en el que se vio que la gran mayoría de partículas correspondían a un único tipo de imágenes de proyección. El complejo por

tanto se unía a la rejilla en una orientación preferente (*Figura 4.5.2C*). Para solucionar este problema el brazo del microscopio electrónico se inclinó 40° y se tomaron nuevas micrografías. A partir de las micrografías inclinadas se extrajeron aproximadamente otras 7.000 imágenes que se combinaron con las imágenes no inclinadas previas. Por lo tanto se utilizaron un total de 14.239 partículas en un proceso de refinamiento angular iterativo. Las diferentes estrategias utilizadas durante el procesamiento progresaron en todos los casos hacia soluciones similares (véase el apartado 3.13.5 de Materiales y Métodos). La estructura del complejo DNA-PKcs:Ku:ADN resuelta tiene una resolución aproximada de 25 Å para un valor de FSC de 0,5 (*Figura 4.5.2D*). En la obtención de la estructura final, además de vistas frontales características (*Figura 4.5.2F1*), también se utilizaron imágenes medias correspondientes a proyecciones en muy diversas orientaciones (*Figura 4.5.2F2*). La combinación de imágenes no inclinadas e inclinadas 40° permitió por tanto representar los

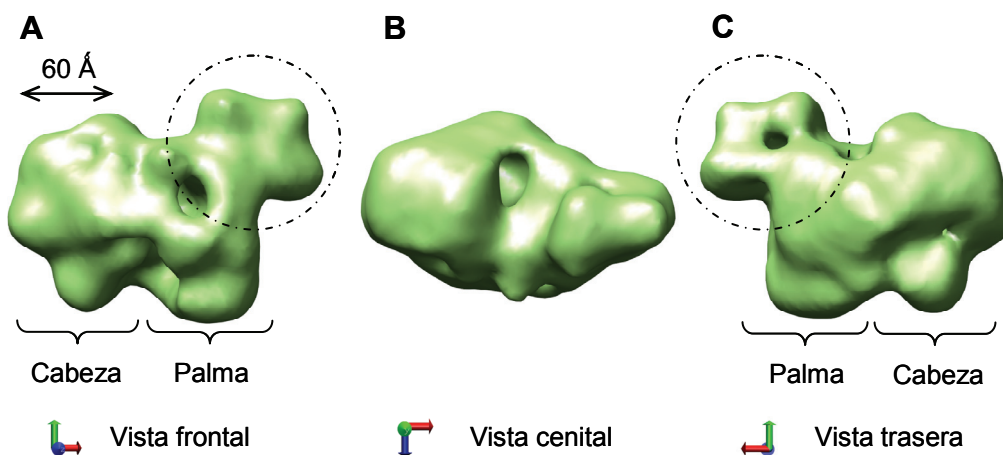
**Figura 4.5.2. Análisis del complejo DNA-PKcs:Ku:ADN mediante microscopía electrónica de especímenes teñidos negativamente (página siguiente).** (A) Micrografía obtenida a partir de una rejilla de tinción negativa del complejo DNA-PKcs:Ku:ADN purificado ensamblado sobre ADN de tipo corto. Dentro de la micrografía se distinguían partículas de dos tamaños diferentes que se indican con “\*” y flechas. (B) Galería de partículas individuales como las que se señalan con “\*” en (A). (C) Representación de los ángulos de Euler del conjunto de imágenes no inclinadas que se procesaron inicialmente. (D) Estimación de la resolución de la estructura resuelta para un valor de FSC de 0,5 y también utilizando el criterio  $3\sigma$ . (E) Representación de los ángulos de Euler del conjunto de imágenes que permitió obtener la estructura final del complejo DNA-PKcs:Ku:ADN. Estas imágenes incluyen partículas inclinadas 40° y partículas sin inclinar, lo que conlleva un aumento significativo de la diversidad de los ángulos de Euler. (F) Galerías de proyecciones (P) y medias equivalentes (M) obtenidas en la última iteración del proceso de refinamiento angular iterativo. La galería que se muestra en (F, 1) corresponde a vistas frontales características de DNA-PKcs. (F, 2) Este tipo de proyecciones, junto con sus medias equivalentes, se obtenían al incluir en el procesamiento partículas extraídas de micrografías que se tomaron con el brazo portamuestras inclinado. (G) Comparación de imágenes de proyección de las proteínas Ku y DNA-PKcs y del complejo DNA-PKcs:Ku:ADN. (G, 1) Detalle de una imagen media de DNA-PKcs obtenida a partir de especímenes teñidos negativamente mediante algoritmos de clasificación libres de patrón. (G, 2) Imagen de proyección de la estructura resuelta mediante criomicroscopía electrónica para la proteína DNA-PKcs. (G, 3) Detalle de una imagen de proyección como las que se muestran en (F, 1). La densidad adicional característica que aparece en este tipo de proyecciones se representa encerrada en un círculo. (G, 4) Proyección generada a partir de la estructura atómica correspondiente al heterodímero Ku (fichero PDB 1JEQ) filtrada a una resolución aproximada de 25 Å. Se indican las regiones cabeza (C) y palma (P) características de la proteína DNA-PKcs.



ángulos de Euler adecuadamente (*compárense las Figuras 4.5.2C y 4.5.2E*). Las proyecciones y medias equivalentes obtenidas en la última iteración del procesamiento presentaban la morfología típica de DNA-PKcs (*Figura 4.5.2F*), pero en este caso además se observaron masas adicionales a nivel de las regiones brazo y palma (*Figuras 4.5.2F y 4.5.2G*). Estas masas recordaban a proyecciones del volumen que se obtenía tras filtrar la estructura atómica del heterodímero Ku a una resolución aproximada de 25 Å (archivo PDB 1JEQ) (*Figura 4.5.2G4*).

El complejo DNA-PKcs:Ku:ADN ensamblado sobre ADN de tipo largo también se analizó mediante microscopía electrónica de especímenes teñidos negativamente. En las micrografías obtenidas se distinguían de nuevo partículas de dos tamaños claramente distintos en una relación similar a la que se había observado previamente (87% de partículas pequeñas y 13% de partículas grandes). A partir de las micrografías se extrajeron imágenes de los dos tipos y se analizaron en dos dimensiones mediante algoritmos de clasificación libres de patrón. Las imágenes medias generadas fueron totalmente compatibles con los resultados obtenidos al analizar por tinción negativa el complejo DNA-PKcs:Ku:ADN ensamblado sobre ADN de tipo corto (véase también el apartado 4.5.7).

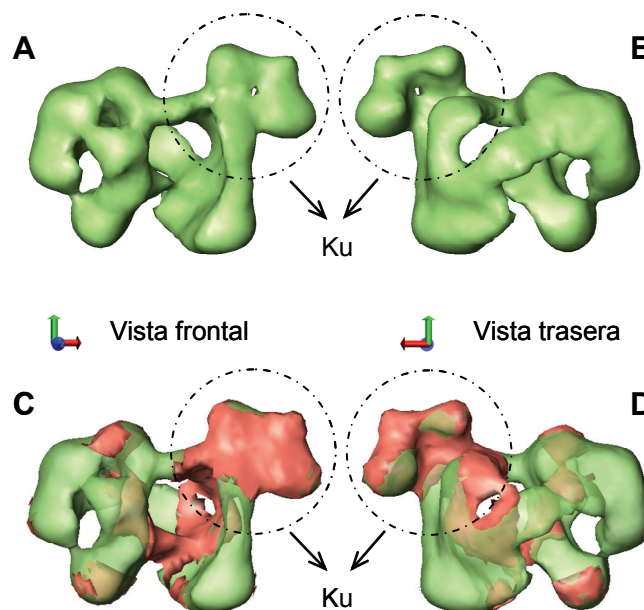
En la estructura obtenida para el complejo DNA-PKcs:Ku:ADN se distinguen de nuevo las diferentes regiones características que forman la proteína DNA-PKcs (*Figura 4.5.3*). Las dimensiones de esta nueva estructura son totalmente compatibles con las de estructuras previas resueltas para la proteína DNA-PKcs (*compárense las Figuras 4.5.3, 4.2.2 y 4.3.2*). Además se aprecia una masa adicional evidente que recuerda a las estructuras del heterodímero Ku y del complejo Ku:ADN obtenidas mediante tinción negativa (*compárense las Figuras 4.5.3, 4.1.2 y 4.1.3*).



**Figura 4.5.3. Reconstrucción tridimensional del complejo DNA-PKcs:Ku:ADN.** (A), (B) y (C) Vista frontal, cenital y trasera respectivamente de la reconstrucción tridimensional obtenida por tinción negativa para el complejo DNA-PKcs:Ku:ADN. Se indican las regiones principales cabeza y palma que componen la proteína DNA-PKcs. La densidad adicional específica de esta nueva estructura se representa encerrada en un círculo discontinuo.

### 4.5.3 Localización del heterodímero Ku en la estructura resuelta mediante tinción negativa para el complejo DNA-PKcs:Ku:ADN

Con el objetivo de confirmar la localización del heterodímero Ku, se calculó el volumen que resulta de la diferencia entre la estructura del complejo y el volumen de criomicroscopía electrónica resuelto para la proteína DNA-PKcs (*Figura 4.5.4*). Para ello el volumen de criomicroscopía electrónica se alineó y filtró a una resolución comparable a la de la estructura correspondiente al ensamblado DNA-PKcs:Ku:ADN. El volumen diferencia resalta la densidad adicional presente en el complejo que no puede ser justificada por la proteína DNA-PKcs (*Figuras 4.5.4C y 4.5.4D*), y por tanto indica la localización del heterodímero Ku en el ensamblado. Además se llevó a cabo el ajuste de la estructura atómica del núcleo de Ku unido a ADN (fichero PDB 1JEY) en la



**Figura 4.5.4. Localización del heterodímero Ku en la estructura resuelta para el complejo DNA-PKcs:Ku:ADN.** (A) y (B) Vista frontal y trasera respectivamente del volumen obtenido mediante tinción negativa para el complejo DNA-PKcs:Ku:ADN. La estructura se muestra a un umbral correspondiente al 55% de la masa del complejo con el objetivo de resaltar los detalles estructurales. (C) y (D) Vistas equivalentes a (A) y (B) respectivamente pero en este caso la estructura se representa transparente. Se muestra también el volumen que se obtiene a partir del cálculo de la diferencia entre la estructura de criomicroscopía electrónica de DNA-PKcs y el volumen resuelto para el complejo DNA-PKcs:Ku:ADN. El volumen diferencia corresponde a la densidad roja opaca. La masa principal de este volumen se indica mediante un círculo discontinuo y revela la localización del heterodímero Ku.

estructura completa del complejo DNA-PKcs:Ku:ADN (*Figura 4.5.5*). En los experimentos de ajuste no se tuvo en cuenta la información correspondiente a la molécula de ADN presente en la estructura del cristal. En estos experimentos tampoco se incluyeron los dominios C-terminales de Ku70 y Ku80, puesto que cambian de posición cuando el heterodímero Ku se une a los extremos del ADN (Walker et al., 2001; Harris et al., 2004; Zhang et al., 2004). Las soluciones con mayor coeficiente de correlación colocaban la proteína Ku dentro de la estructura resuelta para el complejo DNA-PKcs:Ku:ADN y, más concretamente, ocupando por completo la densidad que previamente se había evidenciado mediante el cálculo de volúmenes diferencia (*Figuras 4.5.5A y 4.5.5B*).

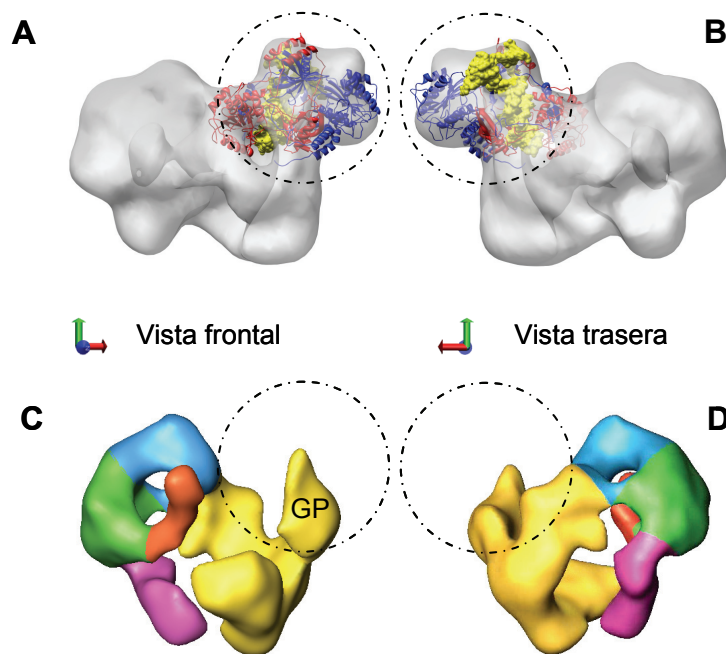
El ajuste del heterodímero Ku en la estructura del complejo DNA-PKcs:Ku:ADN evidencia que, para que se dé el ensamblaje del complejo, DNA-PKcs y Ku han de establecer una superficie de interacción importante (*Figuras 4.5.5A y 4.5.5B*). La proteína Ku contacta fundamentalmente con dos regiones de DNA-PKcs. La región principal engloba la garra proximal de la región palma y la porción de la región brazo desde la que emergen los dominios garra (*se representan en color amarillo en las Figuras 4.5.5C y 4.5.5D*). Además en esta interacción también participa una zona de la región cabeza. Esta zona podría corresponder a la porción del dominio FAT localizada en la transición entre el dominio hombro (*de color azul en las Figuras 4.5.5C y 4.5.5D*) y la región brazo (*de color amarillo en las Figuras 4.5.5C y 4.5.5D*). Estos datos son consistentes con trabajos previos en los que se vio que una región N-terminal al dominio quinasa de DNA-PKcs (residuos 3.002-3.850), era capaz de coprecipitar al heterodímero Ku en un sistema de transcripción-traducción *in vitro* (Jin et al., 1997).

#### **4.5.4 Observación al microscopio electrónico del complejo DNA-PKcs:Ku:ADN autofosforilado**

El complejo DNA-PKcs:Ku:ADN se incubó con ATP en condiciones en las que se favorecía la autofosforilación. Tras la incubación la muestra se tiñó negativamente y se observó en el microscopio electrónico. A partir de las



imágenes individuales pudieron obtenerse dos tipos de medias mayoritarias utilizando algoritmos de clasificación libres de patrón. El primer tipo recordaba a imágenes medias obtenidas previamente durante el análisis por tinción negativa de la proteína DNA-PKcs. El segundo tipo era compatible con medias generadas durante el análisis del heterodímero Ku a partir de especímenes teñidos negativamente. Por tanto, en condiciones en las que se favorece la autofosforilación, el complejo DNA-PKcs:Ku:ADN se desensambla y como consecuencia la densidad añadida desaparece. Estos resultados refuerzan la asignación de la masa adicional al heterodímero Ku.

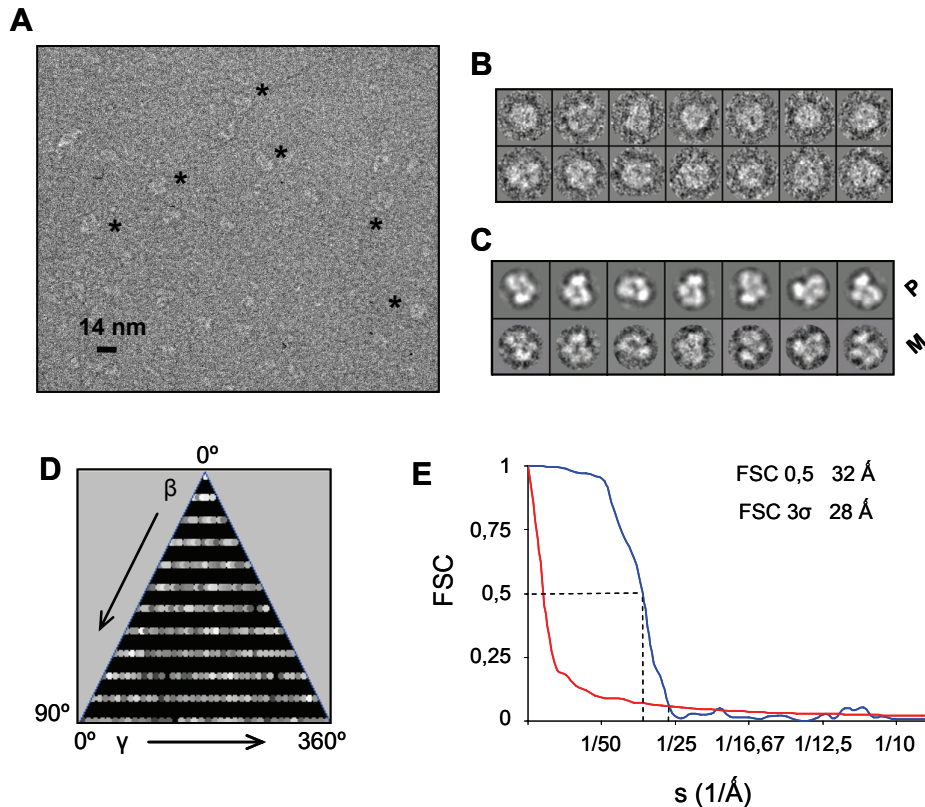


**Figura 4.5.5. Ajuste de la estructura atómica correspondiente al núcleo de Ku en la estructura resuelta para el complejo DNA-PKcs:Ku:ADN. (A) y (B)** Vista frontal y trasera respectivamente del ajuste con mayor coeficiente de correlación obtenido. En la estructura atómica ajustada Ku70 se representa de color rojo y Ku80 de color azul. Se muestra también la molécula de ADN que se utilizó para resolver la estructura atómica del complejo Ku:ADN (archivo PDB 1JEY). A partir de ella sería posible estimar la trayectoria de la molécula de ADN presente en el complejo DNA-PKcs:Ku:ADN. **(C) y (D)** Vista frontal y trasera respectivamente de la estructura de criomicroscopía electrónica resuelta para la proteína DNA-PKcs filtrada a 30 Å de resolución. Los diferentes dominios que constituyen las regiones principales en las que se organiza DNA-PKcs se representan con distintos colores (véanse las Figuras 4.3.2 y 4.3.9). La zona aproximada donde se localizaría el heterodímero Ku en el complejo DNA-PKcs:Ku:ADN se indica mediante un círculo discontinuo. Esta representación permite inferir las regiones de la proteína DNA-PKcs implicadas en la interacción con el heterodímero Ku. Se indica también la garra proximal (GP).

#### 4.5.5 Reconstrucción tridimensional del complejo DNA-PKcs:Ku:ADN obtenida mediante criomicroscopía electrónica

Para confirmar los resultados obtenidos por tinción negativa se llevó a cabo el análisis estructural del complejo DNA-PKcs:Ku:ADN mediante criomicroscopía electrónica. Durante la preparación de la muestra se vio que el detergente presente en el tampón de purificación interfería con el proceso de congelación a alta velocidad. Por ello fue necesario incluir un paso previo de diálisis inmediatamente anterior a la congelación que resultó en un aumento de la inestabilidad de la muestra. Estas dificultades, junto con problemas relacionados con la cantidad de material purificado así como su concentración, limitaron el número de imágenes que se obtuvieron. A partir de micrografías a las que previamente se les había corregido la CTF (*Figura 4.5.6A*), se extrajeron aproximadamente 4.000 imágenes que se utilizaron en un proceso de refinamiento angular iterativo (*Figuras 4.5.6B y 4.5.6C*). El volumen obtenido tiene una resolución aproximada de 32 Å para un valor de FSC de 0,5 (*Figura 4.5.6E*). En este caso el complejo no se unía con una orientación preferente a la rejilla, por tanto no fue necesario tomar micrografías con el brazo portamuestras inclinado (*Figura 4.5.6D*). Durante el proceso de refinamiento no se utilizó la estructura del complejo DNA-PKcs:Ku:ADN obtenida anteriormente por tinción negativa. Tampoco se utilizaron como modelos iniciales estructuras disponibles de las proteínas DNA-PKcs o Ku. Las estrategias empleadas coinciden con las que previamente se habían utilizado en el procesamiento de partículas del complejo teñidas negativamente.

La estructura de criomicroscopía electrónica resuelta para el complejo DNA-PKcs:Ku:ADN es compatible con la estructura obtenida previamente por tinción negativa (*Figura 4.5.7*). De nuevo se distinguen las regiones cabeza, brazo y palma características de la proteína DNA-PKcs. Al comparar esta estructura con el volumen obtenido para DNA-PKcs mediante criomicroscopía electrónica se aprecian cambios conformacionales sustanciales que, en último término, podrían estar inducidos por la interacción con el heterodímero Ku y la molécula de ADN (*compárense las Figuras 4.3.2 y 4.5.7*). La estructura del complejo DNA-PKcs:Ku:ADN resuelta por criomicroscopía electrónica también presenta



**Figura 4.5.6. Reconstrucción tridimensional del complejo DNA-PKcs:Ku:ADN mediante criomicroscopía electrónica.** (A) Micrografía electrónica obtenida a partir de una rejilla congelada a alta velocidad del complejo purificado DNA-PKcs:Ku:ADN ensamblado sobre ADN de tipo corto. Algunas moléculas se indican con “\*”. (B) Galería de partículas individuales extraídas a partir de micrografías como la que se muestra en (A). (C) Galería de proyecciones (P) y medias equivalentes (M) generadas en la última iteración del procesamiento. (D) Representación de los ángulos de Euler asignados a las partículas que permitieron obtener la estructura de criomicroscopía electrónica del complejo DNA-PKcs:Ku:ADN. (E) Resolución estimada para la estructura resuelta según un valor de FSC de 0,5 y también utilizando el criterio  $3\sigma$ .

una masa adicional clara a nivel de la región palma que contacta con la región cabeza (Figura 4.5.7). Sin embargo esta estructura tiene menor resolución que el volumen correspondiente de tinción negativa y por lo tanto la masa adicional está peor definida.

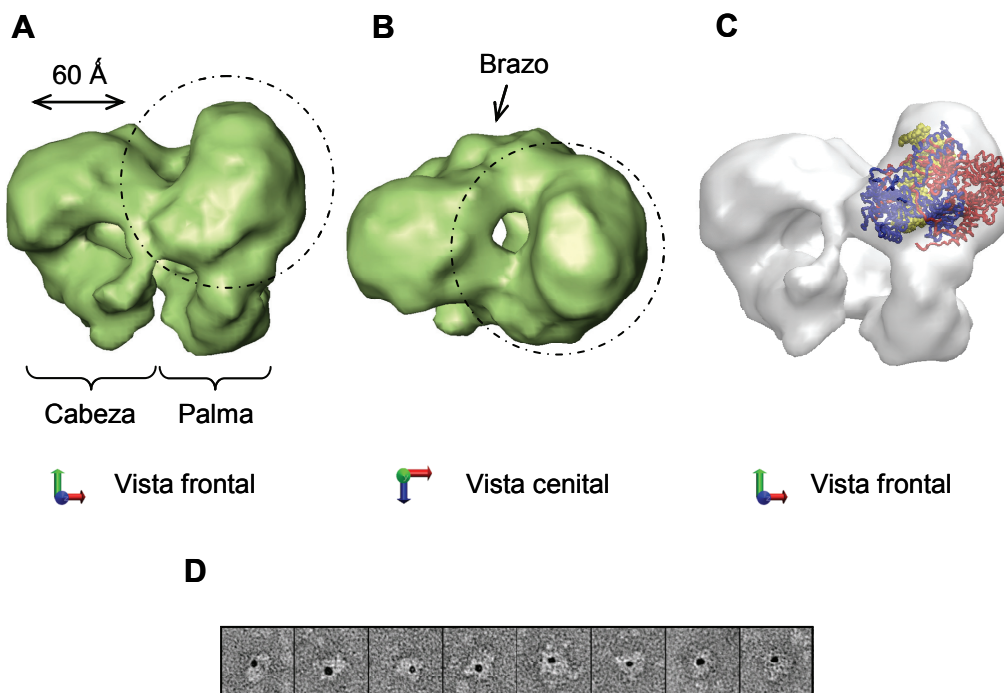
Para confirmar la localización del heterodímero Ku se calculó el volumen diferencia entre las estructuras de criomicroscopía electrónica de DNA-PKcs y del ensamblado DNA-PKcs:Ku:ADN. Después se llevó a cabo el ajuste del heterodímero Ku (archivo PDB 1JEY) de forma análoga a como se ha descrito en el apartado 4.5.3 (Figura 4.5.7C). Las soluciones obtenidas son totalmente compatibles con las orientaciones descritas previamente a partir del ajuste en

la estructura de tinción negativa del complejo (véase el apartado 4.5.3). Por tanto, a pesar de las diferencias en resolución entre las estructuras del complejo resueltas mediante criomicroscopía electrónica y tinción negativa, las dos estructuras presentan una organización parecida y la localización del heterodímero Ku en ambos casos es similar.

#### **4.5.6 Localización de la molécula de ADN en el complejo DNA-PKcs:Ku:ADN**

Ninguna de las dos estructuras resueltas para el complejo DNA-PKcs:Ku:ADN presenta una densidad evidente que pueda asignarse a la molécula de ADN. Ambas estructuras contienen ADN puesto que sin él no se da el ensamblaje del complejo (véase la *Figura 4.5.1*). Es posible que la molécula de ADN se encuentre íntimamente asociada a las proteínas DNA-PKcs y Ku y por tanto no aparezca como una densidad clara e independiente. El ADN utilizado para ensamblar el complejo está iminobiotinilado en el extremo 5' que forma el bloqueo. Por tanto, para demostrar la presencia de la molécula de ADN, el complejo DNA-PKcs:Ku:ADN se incubó con estreptavidina unida covalentemente a oro de 5 nm de diámetro (BB International). En presencia de conjugados estreptavidina:oro las moléculas de complejo aparecen marcadas con una partícula de oro (*Figura 4.5.7D*). El ADN presente en el complejo puede evidenciarse gracias al marcaje con conjugados de estreptavidin:oro, pero no sucede lo mismo con la trayectoria que sigue la molécula de ADN dentro del complejo. Anteriormente se llevó a cabo el ajuste de la estructura atómica del núcleo de Ku (fichero PDB 1JEY) en la densidad adicional presente en las estructuras del complejo DNA-PKcs:Ku:ADN. En estos ajustes no se consideró la molécula de ADN incluida en el fichero PDB. Según esto para estimar la trayectoria del ADN dentro del complejo bastaría con incluir la molécula de ADN en la estructura atómica ajustada (*Figuras 4.5.5A, 4.5.5B y 4.5.7C*). El ADN emerge del anillo de Ku perpendicular al eje longitudinal de la proteína DNA-PKcs, y se dirige hacia la superficie interna de las dos garras que forman la región palma. La molécula de ADN quedaría por tanto encerrada entre las regiones palma y cabeza de DNA-PKcs. Esta trayectoria es totalmente compatible con los resultados descritos anteriormente

a partir de la estructura de tinción negativa del complejo DNA-PKcs:ADN (véase el apartado 4.4.3). En este volumen DNA-PKcs se unía al extremo de la molécula de ADN y el resto del fragmento de ADN, al disminuir el umbral de visualización del volumen, emergía de la estructura dando lugar a una protrusión evidente. Esta protrusión correspondería, en una rotura de doble cadena que tuviera lugar en condiciones fisiológicas, a la parte continua de la molécula de ADN que habría sido reconocida previamente por el heterodímero Ku.



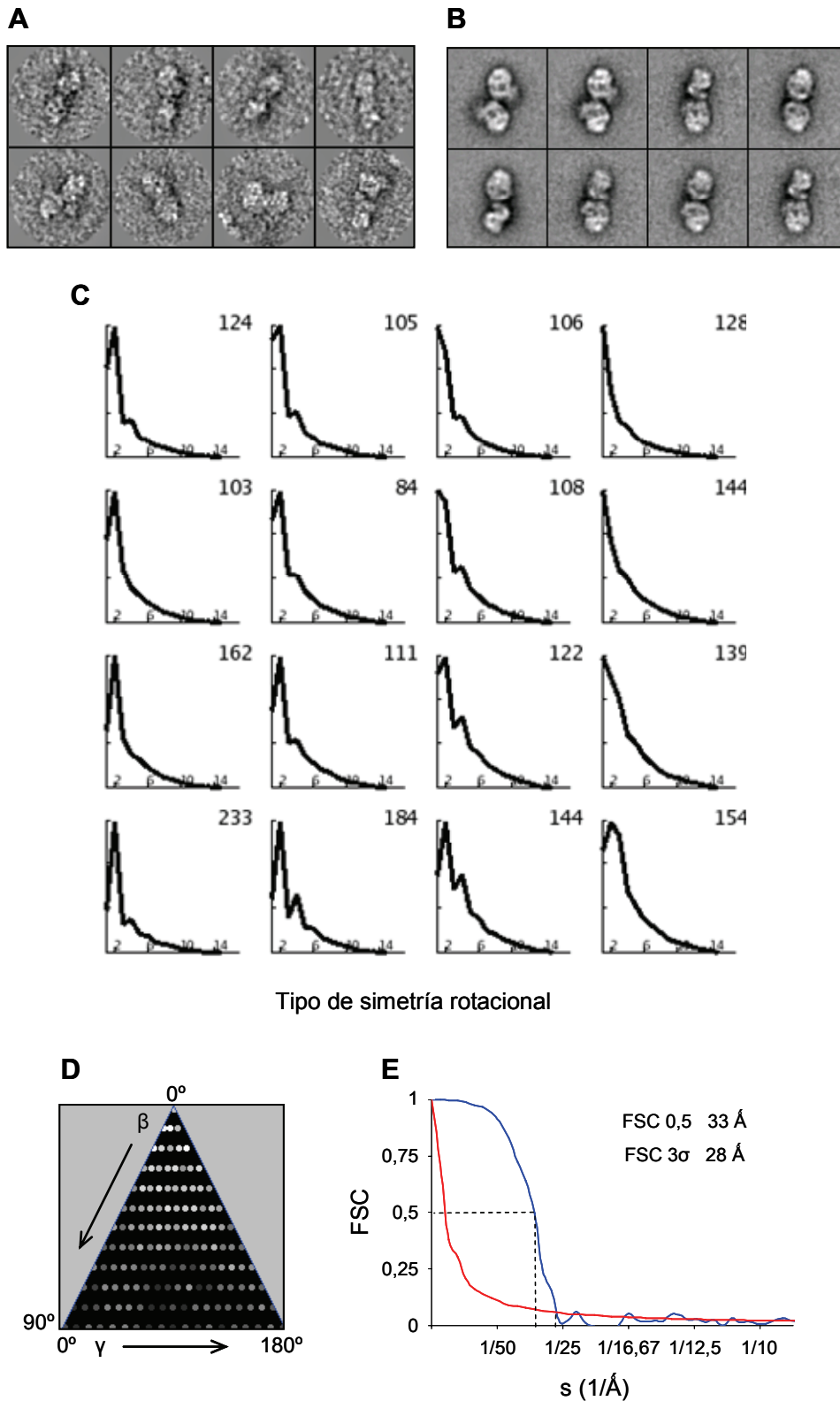
**Figura 4.5.7. Estructura resuelta mediante criomicroscopía electrónica para el complejo DNA-PKcs:Ku:ADN, ajuste de la estructura atómica correspondiente al núcleo de Ku y marcaje del ADN presente en el complejo.** (A) y (B) Vista frontal y cenital respectivamente del volumen obtenido por criomicroscopía electrónica para el complejo DNA-PKcs:Ku:ADN. Se indican las regiones cabeza, palma y brazo en las que se organiza la proteína DNA-PKcs. La densidad adicional evidente se muestra encerrada en un círculo discontinuo. (C) Vista equivalente a la que se muestra en (A), pero en este caso se representa además el ajuste de la estructura atómica correspondiente al núcleo del heterodímero Ku con mayor coeficiente de correlación. En la estructura ajustada Ku70 se representa de color rojo y Ku80 de color azul. Se muestra también la molécula de ADN presente en la estructura del cristal (archivo PDB 1JEY), si bien durante los experimentos de ajuste no se tuvo en cuenta. (D) Galería de moléculas del complejo DNA-PKcs:Ku:ADN marcadas con un conjugado de estreptavidina:oro. Este conjugado interacciona específicamente con el extremo iminobiotinilado de la molécula de ADN presente en el complejo.

#### 4.5.7 Reconstrucción tridimensional del complejo dimérico de DNA-PKcs:Ku:ADN obtenida a partir de especímenes teñidos negativamente

Al observar la muestra del complejo purificado DNA-PKcs:Ku:ADN teñida negativamente se distinguía un segundo tipo de partículas de mayor tamaño (*Figura 4.5.2A*). Estas otras partículas correspondían a moléculas que parecían estar formadas por dos complejos DNA-PKcs:Ku:ADN (*Figura 4.5.8A*). Además eran relativamente abundantes puesto que representaban aproximadamente una quinta parte de las moléculas de DNA-PKcs presentes en las micrografías.

Se seleccionaron un total de 2.151 partículas de mayor tamaño que se procesaron inicialmente en dos dimensiones (véase el apartado 3.13.4 de Materiales y Métodos). Las 2.151 imágenes incluían partículas tomadas con 40° de inclinación y también sin inclinar. El cálculo de imágenes medias libres de referencia mostraba que las partículas estaban formadas por dos complejos DNA-PKcs:Ku:ADN (*Figura 4.5.8B*). La interacción entre ambos complejos no era aleatoria si no que se daba según una orientación claramente definida. Además el análisis rotacional reveló que la mayoría de las partículas tenían simetría de orden 2 (*Figura 4.5.8C*). Estas imágenes de mayor tamaño correspondían por tanto a un ensamblado específico con una organización simétrica evidente.

**Figura 4.5.8. Análisis estructural del complejo dimérico de DNA-PKcs:Ku:ADN a partir de especímenes teñidos negativamente (página siguiente).** (A) Galería de partículas individuales de mayor tamaño presentes en las micrografías obtenidas a partir de rejillas de tinción negativa del complejo DNA-PKcs:Ku:ADN purificado. (B) Imágenes medias generadas utilizando métodos de clasificación libres de patrón disponibles en el paquete informático EMAN. (C) Se llevó a cabo el análisis de la simetría rotacional presente en las imágenes de mayor tamaño (véase apartado 3.13.4 de Materiales y Métodos). Para ello primero se calcularon los espectros rotacionales de las imágenes. A continuación los espectros calculados se clasificaron utilizando algoritmos de tipo “mapa auto-organizativo” (Kohonen, 1990; Pascual-Montano et al., 2001), y de esta manera fue posible agrupar las partículas en función de su simetría. En la figura se representan los vectores obtenidos tras llevar a cabo la clasificación. Para cada vector el eje X indica el tipo de simetría rotacional, mientras que el número presente en la esquina superior derecha corresponde a las partículas asignadas a ese vector. Se aprecia que la mayoría de los grupos poseen un componente mayoritario de simetría rotacional de orden dos. (D) Representación de los ángulos de Euler presentes en las imágenes que permitieron obtener la estructura final del complejo dimérico. (E) Estimación de la resolución para la estructura resuelta según los criterios  $3\sigma$  y 0,5.



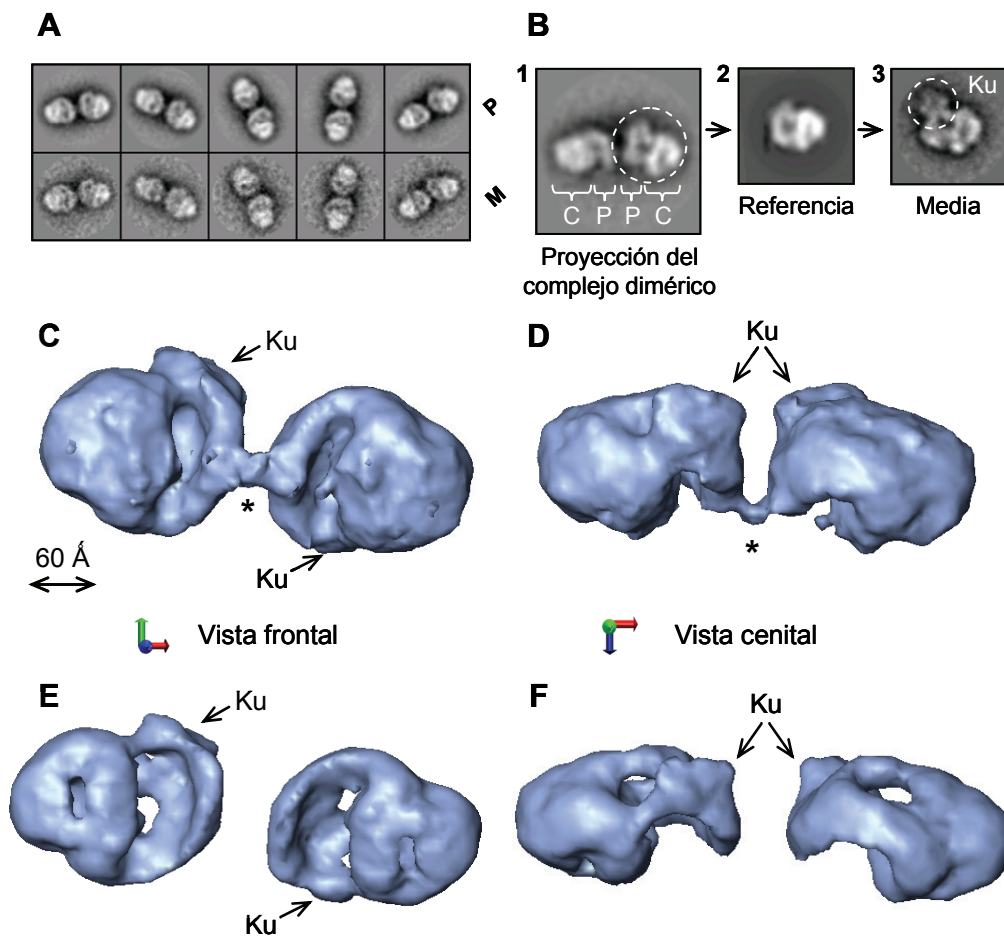
Las partículas se utilizaron en un proceso de refinamiento angular iterativo en el que inicialmente no se impuso ninguna simetría. Tras el procesamiento se obtuvo una estructura con una resolución aproximada de 33 Å para un valor de FSC de 0,5 (Figura 4.5.9). Los volúmenes iniciales de nuevo consistían en

esferas con una distribución de masa gaussiana. Las imágenes medias que se obtenían tras el procesamiento eran totalmente compatibles con medias libres de patrón generadas previamente (*compárense las Figuras 4.5.8B y 4.5.9A*). El procesamiento se llevó a cabo sin imponer ninguna simetría. Sólo cuando convergió se impuso simetría rotacional de orden dos para mejorar el aspecto general de la reconstrucción.

Por tanto las partículas de mayor tamaño eran compatibles con un único volumen que consistía en un complejo dimérico (*Figura 4.5.9*). Si este complejo estaba formado por dos ensamblados DNA-PKcs:Ku:ADN, entonces sería posible alinear imágenes correspondientes al complejo DNA-PKcs:Ku:ADN respecto a proyecciones del ensamblado dimérico (*Figura 4.5.9B*). Para llevar a cabo este control a determinadas proyecciones generadas a partir de la estructura dimérica se les aplicó una máscara. Esta máscara permitía que únicamente quedara visible uno de los complejos DNA-PKcs:Ku:ADN presentes en la imagen de proyección (*Figuras 4.5.9B1 y 4.5.9B2*). Las proyecciones modificadas se utilizaron como referencias para clasificar, alinear y promediar imágenes individuales del complejo DNA-PKcs:Ku:ADN. Las medias obtenidas evidencian que el ensamblado dimérico se componen de dos complejos DNA-PKcs:Ku:ADN que interaccionan por la región palma (*Figura 4.5.9B3*). En la estructura resuelta los dos complejos DNA-PKcs:Ku:ADN interaccionan enfrentados a lo largo de un eje longitudinal común perpendicular al eje de simetría dos (*Figuras 4.5.9C y 4.5.9E*). Además la estructura de los complejos DNA-PKcs:Ku:ADN que forman el ensamblado dimérico es esencialmente idéntica a la que presentan cuando se encuentran libres, aunque con menor resolución (*compárense las Figuras 4.5.3 y 4.5.9*). Cuando el complejo dimérico se representa a un umbral equivalente al 100% de la masa predicha, los dos complejos DNA-PKcs:Ku:ADN contactan por un pequeño puente de densidad (*esta densidad se indica con un “\*” en la Figura 4.5.9*). Esta conexión simétrica conecta las caras externas de las garras distales de las dos moléculas de DNA-PKcs.

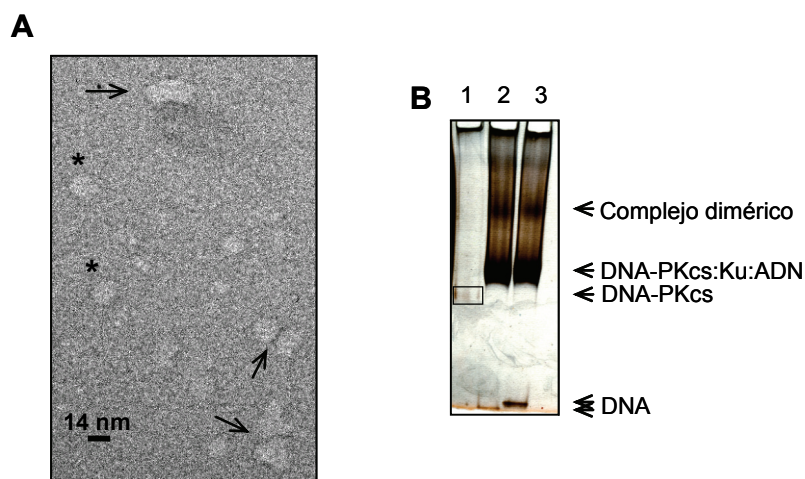
Los complejos diméricos corresponden aproximadamente al 13% del total de partículas seleccionadas a partir de las micrografías electrónicas. Cuando la





**Figura 4.5.9. Reconstrucción tridimensional del complejo dimérico de DNA-PKcs:Ku:ADN.** (A) Galería de proyecciones (P) y medias equivalentes (M) obtenidas en la última iteración del procesamiento. (B) Confirmación de la presencia de dos ensamblados DNA-PKcs:Ku:ADN en el complejo dimérico. A partir de una proyección de la estructura resuelta para el complejo dimérico (B, 1), se generó una imagen que únicamente contenía uno de los complejos DNA-PKcs:Ku:ADN (B, 2). Utilizando esta imagen modificada como referencia era posible alinear imágenes individuales del complejo DNA-PKcs:Ku:ADN. Se obtenía por tanto una nueva imagen media donde se apreciaba la densidad correspondiente a la proteína Ku (B, 3). (C) y (D) Vista frontal y cenital respectivamente de la estructura resuelta para el complejo dimérico de DNA-PKcs:Ku:ADN. La estructura que se muestra corresponde al volumen que se obtuvo tras aplicar simetría de orden dos en las etapas finales del procesamiento. El volumen se representa al umbral correspondiente al 100% de la masa del complejo. A este umbral es posible identificar una pequeña densidad de masa que conecta los dos complejos DNA-PKcs:Ku:ADN (se indica con "\*" en la figura). (E) y (F) Vistas equivalentes a (C) y (D) respectivamente pero en este caso la estructura se muestra al umbral correspondiente al 55% de la masa del complejo. En las diferentes vistas se señala la densidad correspondiente al heterodímero Ku.

proteína DNA-PKcs se observó por si sola en el microscopio electrónico este tipo de partículas era minoritario. Además las partículas de mayor tamaño también aparecían en muestras congeladas del complejo DNA-PKcs:Ku:ADN (*Figura 4.5.10A*), por tanto la formación del complejo dimérico no es dependiente de la presencia del agente de tinción. La existencia del ensamblado dimérico también pudo demostrarse bioquímicamente mediante experimentos de electroforesis en condiciones nativas. Al analizar la muestra del complejo DNA-PKcs:Ku:ADN purificado, se observó una banda con un peso molecular significativamente más elevado que el de la banda correspondiente al complejo DNA-PKcs:Ku:ADN (*Figura 4.5.10B*). Esta banda era consistente con la formación de un complejo dimérico a partir de dos complejos DNA-PKcs:Ku:ADN.



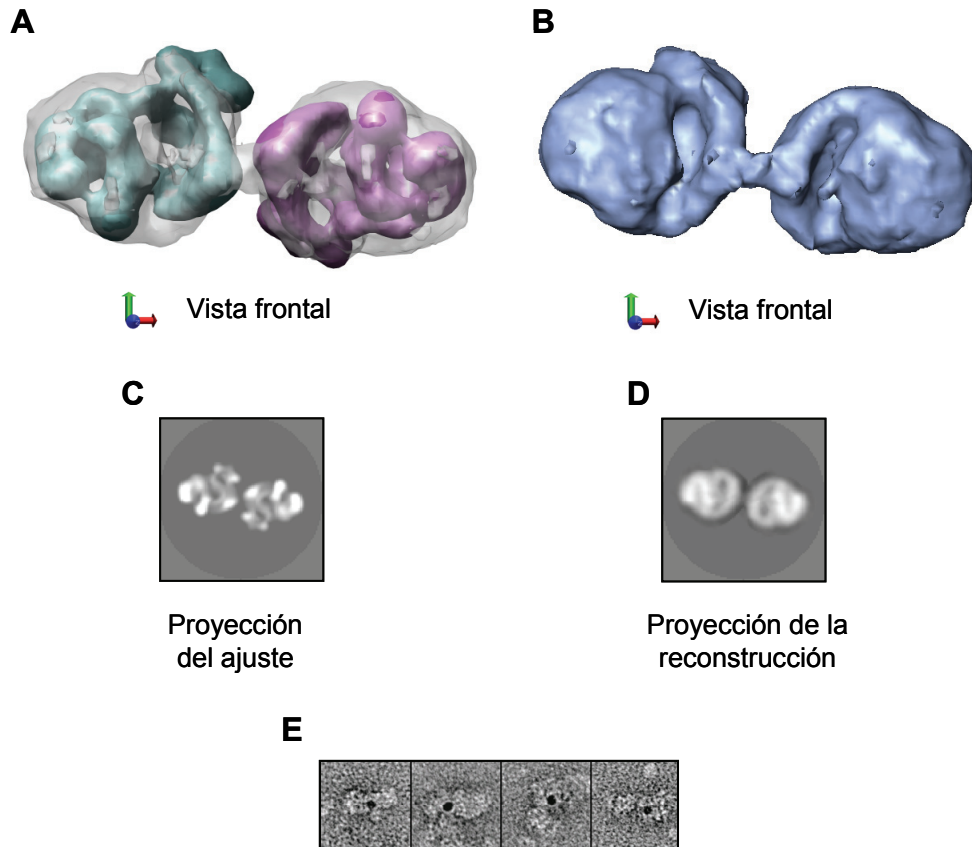
**Figura 4.5.10. Evidencias adicionales que confirman la presencia del complejo dimérico en la muestra del ensamblado DNA-PKcs:Ku:ADN purificado.** (A) Micrografía electrónica obtenida a partir de una rejilla congelada del complejo DNA-PKcs:Ku:ADN purificado. Se indican algunas partículas correspondientes al ensamblado DNA-PKcs:Ku:ADN con "\*" y también se señalan moléculas de complejo dimérico. (B) Análisis por electroforesis en condiciones nativas de la proteína DNA-PKcs (recuadro del carril 1), y de la muestra del complejo DNA-PKcs:Ku:ADN purificado ensamblado sobre ADN de tipo corto (carril 2) o largo (carril 3). En los carriles 2 y 3 se aprecia una banda de alto peso molecular compatible con la existencia de complejos diméricos en las muestras de complejo DNA-PKcs:Ku:ADN.

#### 4.5.8 Orientación de los complejos DNA-PKcs:Ku:ADN en la estructura del complejo dimérico

Con el objetivo de explorar la orientación relativa de cada complejo DNA-PKcs:Ku:ADN dentro del ensamblado dimérico, se llevó a cabo el ajuste de la estructura de tinción negativa del complejo DNA-PKcs:Ku:ADN en la estructura resuelta para el ensamblado dimérico. Los ajustes con mayor coeficiente de correlación colocaban el complejo DNA-PKcs:Ku:ADN a cada lado del eje de simetría dos en orientaciones idénticas, y con la región cabeza dirigida hacia el exterior (*Figura 4.5.11A*). Es posible por tanto construir un complejo dimérico artificial utilizando dos complejos DNA-PKcs:Ku:ADN. Proyecciones generadas a partir de este volumen artificial son similares a las proyecciones equivalentes del complejo dimérico (*Figuras 4.5.11C y 4.5.11D*). Esto confirma de nuevo que el complejo dimérico está formado por dos ensamblados DNA-PKcs:Ku:ADN. Además la orientación relativa de los ensamblados DNA-PKcs:Ku:ADN en el complejo dimérico es similar a la predicha a partir de los experimentos de ajuste.

#### 4.5.9 Localización del ADN en el complejo dimérico

Los complejos estreptavidina:oro descritos en el apartado 4.5.6 también se asociaban a las partículas de mayor tamaño. En este caso el oro se localizaba en el centro de las partículas, en la posición correspondiente al puente de densidad que conectaba los dos complejos DNA-PKcs:Ku:ADN (*Figura 4.5.11E*). El extremo iminobiotinilado de al menos una de las moléculas de ADN presentes en el complejo dimérico quedaría accesible al conjugado de estreptavidina:oro. El ADN se dispone por tanto en el centro del complejo dimérico. Esta localización es totalmente consistente con experimentos previos de localización del ADN en los complejos DNA-PKcs:ADN y DNA-PKcs:Ku:ADN. En estos complejos la molécula de ADN queda encerrada en el canal comprendido entre las regiones cabeza y palma. En la estructura resuelta para el ensamblado dimérico las moléculas de DNA-PKcs se disponen enfrentadas contactando por la cara externa de los dominios garra distal. Los contactos que se establecen entre las dos moléculas de DNA-PKcs son



**Figura 4.5.11. Orientación de los complejos DNA-PKcs:Ku:ADN en la estructura resuelta para el ensamblado dimérico y localización de las moléculas de ADN. (A)** Vista frontal del ajuste en la estructura del complejo dimérico del volumen resuelto mediante tinción negativa para el ensamblado DNA-PKcs:Ku:ADN. Se muestran las dos soluciones con mayor coeficiente de correlación obtenidas. La estructura del complejo dimérico se representa mediante una densidad transparente. **(B)** Vista frontal de la estructura resuelta para el complejo dimérico. **(C)** y **(D)** Imágenes de proyección generadas a partir del ajuste descrito en (A) y del volumen correspondiente al complejo dimérico respectivamente. **(E)** Galería de partículas del complejo dimérico marcadas con un conjugado de estreptavidina:oro (BB International). Este conjugado interacciona específicamente con los extremos iminobiotinilados de las moléculas de ADN presentes en el complejo.

mínimos, pero permiten definir una estructura ordenada en la cual las regiones palma se localizan en el centro y, por tanto, los canales que alojan el ADN quedan también próximos al centro de la estructura.

#### 4.5.10 Conclusión

La purificación del complejo humano DNA-PKcs:Ku:ADN ha sido el paso previo para resolver su estructura tridimensional mediante microscopía

electrónica y análisis de partículas individuales. El análisis por microscopía electrónica ha revelado además la existencia de un ensamblado proteico formado por dos complejos DNA-PKcs:Ku:ADN. Las características estructurales de este ensamblado hacen sospechar que posiblemente corresponda al complejo sináptico que participa en el proceso de reparación por NHEJ.

#### 4.6 Caracterización bioquímica de la interacción entre las proteínas Artemis y DNA-PKcs

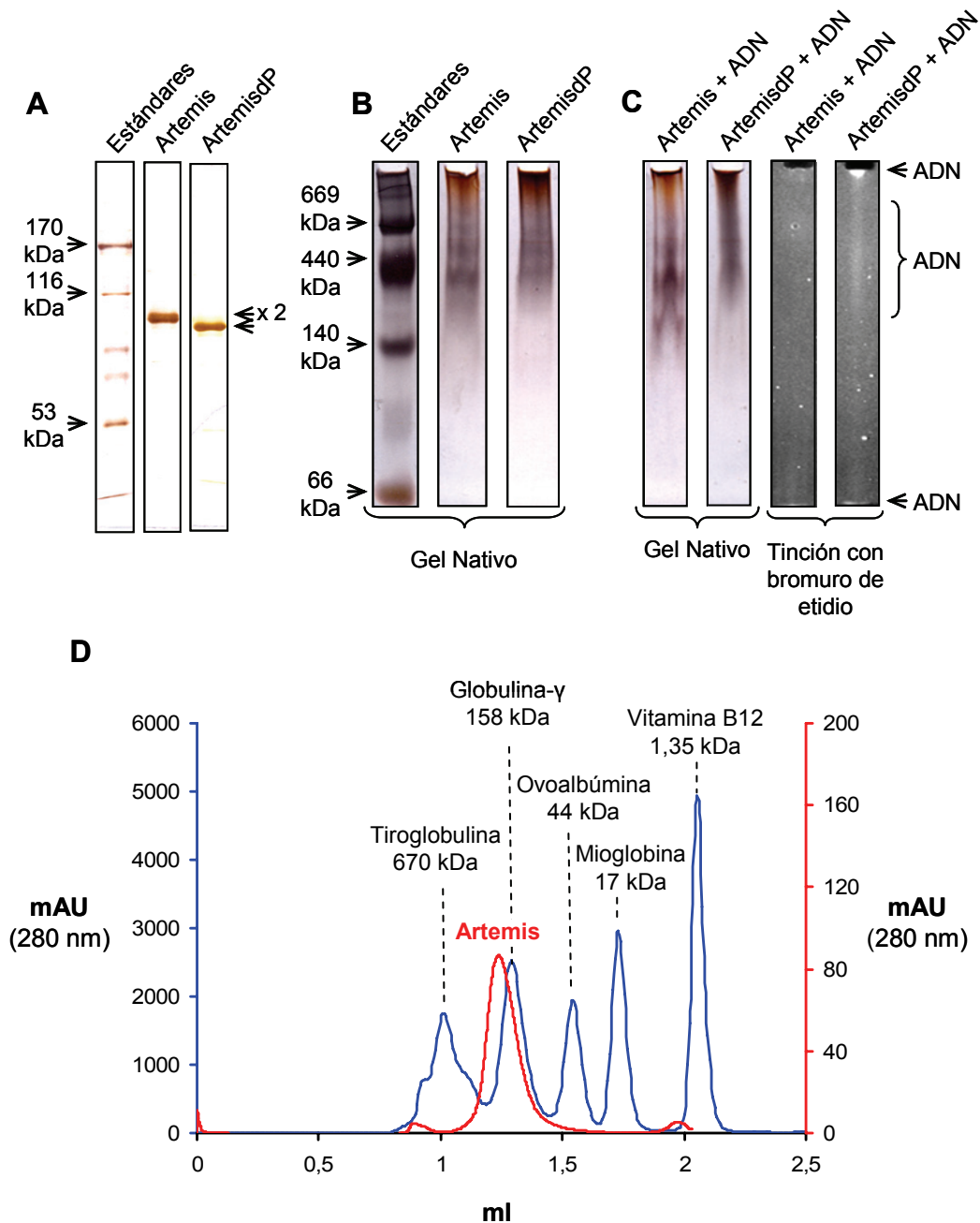
La proteína Artemis es uno de los componentes de la maquinaria de reparación por NHEJ que han sido identificados más recientemente (Moshous et al., 2001). Esta nucleasa de 78 kDa lleva a cabo la apertura de las horquillas generadas en el ADN durante la recombinación V(D)J (Ma et al., 2002). Además, en el proceso de reparación por NHEJ, Artemis actúa como una endonucleasa eliminando regiones de cadena sencilla que sobresalen de los extremos generados tras la rotura del ADN (Ma et al., 2002). La actividad de Artemis está modulada por la proteína DNA-PKcs y depende en primer lugar de la interacción entre ambas proteínas y, en segundo lugar, de la fosforilación por DNA-PKcs (Ma et al., 2002; Ma et al., 2005b; Goodarzi et al., 2006). Parece ser que las dos proteínas forman un complejo estable en ausencia de ADN y del heterodímero Ku (Ma et al., 2002), aunque los resultados respecto a la existencia de este complejo son contradictorios (Drouet et al., 2006). Por tanto, para llevar a cabo la futura caracterización estructural del complejo DNA-PKcs:Artemis, es necesario definir primero las condiciones en las que pueda darse el ensamblaje de este complejo. Con el objetivo de determinar estas condiciones se ha llevado a cabo un primer análisis de la muestra de la proteína Artemis purificada. Además también se ha realizado el análisis bioquímico de la muestra que se obtiene tras incubar las proteínas Artemis y DNA-PKcs.

#### 4.6.1 Caracterización bioquímica de la proteína Artemis

La proteína Artemis fue purificada a partir de células de insecto por N. Rzechorzek en el marco de la colaboración establecida con el grupo del Profesor L. H. Pearl. Tras analizar esta muestra mediante electroforesis en condiciones desnaturalizantes se vio que Artemis se distribuía en dos bandas (*Figura 4.6.1A*). Este comportamiento ha sido descrito previamente por otros autores y se debe a la fosforilación basal que presenta la proteína Artemis en este tipo de células (Ma et al., 2005b). Al incubar la muestra purificada con fosfatasa- $\lambda$  y analizar a continuación la mezcla mediante electroforesis en condiciones desnaturalizantes, se obtenía una única banda y que presentaba una migración mayor (*Figura 4.6.1A*).

La proteína Artemis purificada también se analizó mediante electroforesis en condiciones nativas. En las condiciones ensayadas Artemis daba lugar a bandas de diferente tamaño (*Figura 4.6.1B*). Ese comportamiento no dependía del grado de fosforilación basal puesto que la incubación con fosfatasa- $\lambda$  no impedía la formación de diversas bandas (*Figura 4.6.1B*). Para confirmar la posible presencia de diferentes oligómeros en la muestra purificada se llevaron a cabo experimentos de cromatografía de exclusión molecular (*Figura 4.6.1D*). El pico característico que se obtiene al realizar estos experimentos presenta una migración compatible con un peso molecular próximo a 160 kDa (*Figura 4.6.1D*). Este valor es el doble del peso esperado para el monómero de Artemis, lo cual sugiere que en solución se encuentra principalmente como un homodímero. Los experimentos de cromatografía indican que el comportamiento de Artemis en geles nativos puede ser artefactual y, por tanto, no reflejaría la existencia de diferentes oligómeros en la muestra purificada. Otro ejemplo de metalo- $\beta$ -lactamasa homodimérica es la tRNasa Z (Li de la Sierra-Gallay et al., 2006), por ello es posible que la existencia de homodímeros de Artemis tenga una implicación fisiológica.

Se analizó también la capacidad de la muestra purificada de interactuar con ADN mediante experimentos de cambio de movilidad electroforética en condiciones nativas (*Figura 4.6.1C*). Para ello la proteína Artemis se incubó con



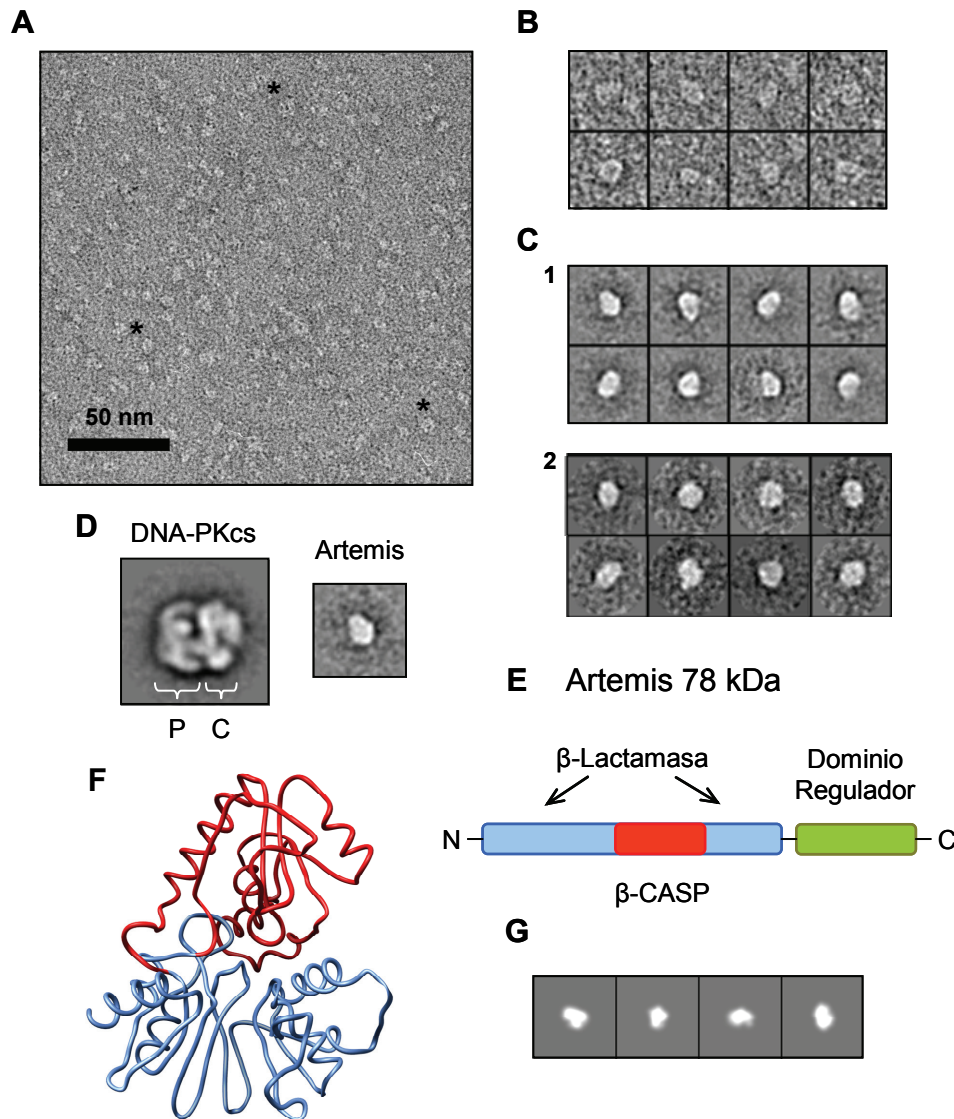
**Figura 4.6.1. Análisis bioquímico de la proteína Artemis purificada.** (A) Gel en condiciones desnaturizantes teñido con plata de la proteína Artemis purificada y también de la muestra de Artemis incubada con fosfatasa- $\lambda$  (carril "ArtemisdP"). (B) Gel de poliacrilamida en condiciones nativas teñido con plata de las muestras que se indican en (A). Las cantidades de Artemis y ArtemisdP utilizadas en este experimento son las mismas que las que se emplearon en los ensayos de interacción con ADN que se muestran en (C) (véase el apartado 3.9.8 de Resultados). (C) Gel en condiciones nativas teñido con plata de la mezcla de reacción que resulta de incubar ADN bien con la proteína Artemis purificada (carril "Artemis + ADN"), bien con la muestra de Artemis incubada previamente con fosfatasa- $\lambda$  (carril "ArtemisdP + ADN"). (D) Análisis mediante cromatografía de exclusión molecular de la muestra correspondiente a la proteína Artemis purificada. Los estándares de peso molecular se muestran en color azul, mientras que el resultado obtenido para la proteína Artemis se representa en color rojo. El pico principal presenta una migración compatible con un peso molecular ligeramente superior a 158 kDa.

una molécula de ADN biotinilada de 54 pb y extremos romos. La molécula de ADN empleada se había utilizado previamente en el ensamblaje de diversos complejos proteína:ADN (véase apartado 3.1 de Materiales y Métodos). En el caso de la muestra de Artemis no tratada previamente con fosfatasa- $\lambda$  no se vio la formación de complejos estables Artemis:ADN (*Figura 4.6.1C*). Sin embargo la muestra de Artemis incubada con fosfatasa- $\lambda$  daba lugar a agregados de proteína y ADN de alto peso molecular que no progresaban en el gel (*compárense las Figuras 4.6.1B y 4.6.1C*). Además también se distinguía una zona difusa en la que las señales correspondientes a la proteína Artemis y al ADN solapaban (*Figura 4.6.1C*). Estos experimentos sugieren que quizás la interacción entre el ADN y Artemis podría depender del grado de fosforilación de esta proteína.

#### **4.6.2 Análisis bidimensional de imágenes de microscopía electrónica correspondientes a la proteína Artemis**

La proteína Artemis purificada se tiñó negativamente y se observó en el microscopio electrónico (*Figura 4.6.2A*). En las micrografías se distinguieron partículas de tamaño variable. A partir de las micrografías se seleccionaron 700 imágenes aproximadamente correspondientes a partículas de menor tamaño (*Figura 4.6.2B*). Estas imágenes se analizaron en dos dimensiones con el objetivo de identificar imágenes de proyección características y, de esta manera, tener una idea preliminar del aspecto y dimensiones generales de la proteína Artemis. Las imágenes medias obtenidas utilizando algoritmos de clasificación libres de patrón indicaban que el conjunto inicial de partículas podía agruparse en clases específicas (*Figura 4.6.2C*). Estas imágenes correspondían por tanto a vistas definidas de la proteína Artemis y presentaban una distribución de masa característica. La longitud y anchura máxima de la densidad presente en las imágenes medias era aproximadamente de 65 Å y 50 Å respectivamente (*Figura 4.6.2C y 4.6.2D*).





**Figura 4.6.2. Análisis por microscopía electrónica y predicción de un modelo atómico para la proteína Artemis.** (A) Micrografía electrónica de una rejilla teñida negativamente de la proteína Artemis purificada. Algunas partículas evidentes se indican con “\*”. (B) Galería de partículas individuales correspondientes a moléculas de la proteína Artemis. (C) Imágenes medias generadas utilizando métodos de clasificación libres de patrón presentes en los paquetes informáticos EMAN (C, 1) y XMIPP (C, 2). (D) Comparación de imágenes medias características correspondientes a las proteínas DNA-PKcs y Artemis. Las dos imágenes se representan a la misma escala. (E) Esquema de la cadena polipeptídica de la proteína Artemis. Se indican de manera esquemática los dominios principales identificados en esta proteína. (F) Modelo atómico para los primeros 346 aminoácidos de la proteína Artemis. Este modelo se generó mediante modelado por homología a partir de la estructura atómica resuelta para la proteína TTHA0252 de *Thermus thermophilus* (archivo PDB 2DKF) (Ishikawa et al., 2006). El código de colores utilizado para los distintos dominios coincide con el que se muestra en (E). (G) El modelo que se muestra en (F) se filtró a una resolución aproximada de 25 Å y se proyectó en diferentes orientaciones. Las proyecciones generadas tienen unas dimensiones compatibles con las de las imágenes medias que se muestran en (C). Las galerías de imágenes que se muestran en (C) y (G) se representan a la misma escala.

### 4.6.3 Generación de un modelo tridimensional para la proteína Artemis

Artemis pertenece a la familia de metalo- $\beta$ -lactamasas denominada  $\beta$ -CASP (Callebaut et al., 2002). Esta familia se caracteriza por presentar un dominio globular  $\beta$ -CASP a continuación del motivo 4 típico del plegamiento metalo- $\beta$ -lactamasa (véase el apartado 1.4.3 de Introducción). Recientemente se han resuelto las estructuras atómicas de varios miembros de la familia  $\beta$ -CASP en las que, efectivamente, se aprecia un dominio globular insertado entre las dos hojas  $\beta$  que forman el núcleo del plegamiento característico  $\beta$ -lactamasa (Ishikawa et al., 2006; Mandel et al., 2006). Los dominios metalo- $\beta$ -lactamasa y  $\beta$ -CASP se encuentran en la mitad N-terminal de la proteína Artemis (*Figura 4.6.2E*). La mitad C-terminal por el contrario no presenta homología con ninguna proteína conocida y apenas contiene elementos de estructura secundaria. Esta región es fosforilada intensamente por la proteína DNA-PKcs y posiblemente tenga un papel regulador de la actividad nucleasa. A partir de las estructuras atómicas resueltas para miembros de la familia  $\beta$ -CASP y utilizando el meta-servidor Phyre (Kelley et al., 2000; Bennett-Lovsey et al., 2008), se pudieron modelar por homología los dominios metalo- $\beta$ -lactamasa y  $\beta$ -CASP de la proteína Artemis (*Figura 4.6.2F*). Este modelo carece de la mitad C-terminal pero aporta información sobre el aspecto y dimensiones generales de la proteína Artemis. El modelo atómico obtenido se filtró a una resolución aproximada de 25 Å, ya que este valor de resolución es similar al que se obtiene rutinariamente a partir del análisis por microscopía electrónica de especímenes teñidos negativamente. Imágenes de proyección de la estructura filtrada son compatibles con medias generadas mediante algoritmos de clasificación libres de patrón (*compárense las Figuras 4.6.2C y 4.6.2G*).

### 4.6.4 Caracterización bioquímica de la interacción entre las proteínas Artemis y DNA-PKcs

La primera aproximación para ver si las proteínas Artemis y DNA-PKcs interaccionaban consistió en incubar las dos proteínas juntas y, a continuación, analizar la mezcla mediante electroforesis en condiciones nativas (*Figura 4.6.3A*). Al llevar a cabo este análisis se observó una banda nueva

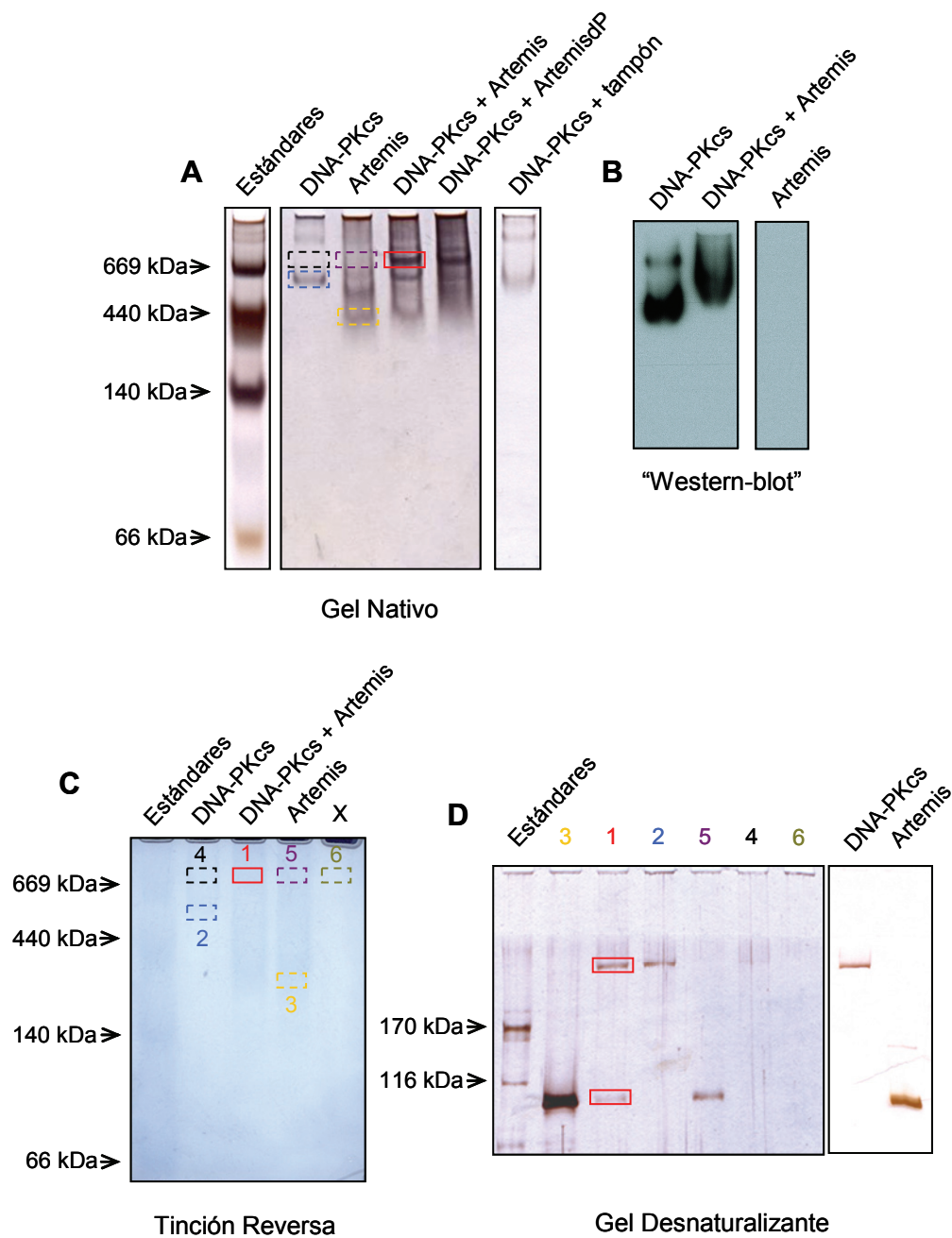
característica que es mayoritaria (*se muestra en un recuadro rojo en la Figura 4.6.3A*). Esta banda no aparecía cuando se realizaba el mismo análisis con las muestras correspondientes a las proteínas Artemis y DNA-PKcs por separado (*Figura 4.6.3A*). La adición del tampón en el que se diluyó Artemis a la proteína DNA-PKcs no indujo la formación de una banda similar (*Figura 4.6.3A*). Por tanto esta nueva banda se originaba específicamente como consecuencia de la presencia de la proteína Artemis en la mezcla de reacción. Además la nueva banda se generaba independientemente de que la proteína Artemis hubiera sido incubada previamente con fosfatasa- $\lambda$  (*Figura 4.6.3A*). Esto sugiere que el grado de fosforilación basal de Artemis no influye en la aparición de esta banda característica.

Para determinar si la nueva banda correspondía a un complejo formado por las proteínas Artemis y DNA-PKcs, se llevó a cabo un ensayo de “Western-blot” utilizando un anticuerpo monoclonal frente a DNA-PKcs (*Figura 4.6.3B*). Este experimento demuestra que, en presencia de la proteína Artemis, DNA-PKcs sufre un cambio de movilidad significativo. Además en las condiciones ensayadas el cambio de movilidad afecta a toda la proteína DNA-PKcs presente en la mezcla de reacción. Estos resultados son compatibles con la formación de un complejo mayoritario y homogéneo entre las proteínas DNA-PKcs y Artemis.

Se realizaron experimentos adicionales para confirmar la presencia de las proteínas Artemis y DNA-PKcs en la banda adicional descrita anteriormente. Para ello las dos proteínas se incubaron de nuevo y la mezcla de reacción se analizó mediante electroforesis en condiciones nativas. A continuación el gel se tiñó utilizando el método de tinción reversa (véase el apartado 3.6 de Materiales y Métodos) (*Figura 4.6.3C*). Este tipo de tinción permite inmovilizar de manera reversible las proteínas presentes en el gel (Ortiz et al., 1992; Hardy y Castellanos-Serra, 2004). Después se identificó la banda adicional (*se muestra en un recuadro rojo en las Figuras 4.6.3A y 4.6.3C*), y se recortó del gel teñido. Posteriormente la banda se incubó con una solución que permitía la movilización de las proteínas presentes. Además la banda se troceó con el objetivo de favorecer la difusión de las proteínas desde el gel hacia la solución

de movilización. Por último la mezcla se centrifugó y se separó el sobrenadante que posteriormente se analizó mediante electroforesis en condiciones desnaturizantes (*carril 1 de la Figura 4.6.3D*). Este análisis reveló que las proteínas Artemis y DNA-PKcs están presentes en la banda característica y mayoritaria que aparece en los ensayos de electroforesis en condiciones nativas (*recuadros rojos de la Figura 4.6.3D*). También se realizó el mismo experimento con fragmentos del gel que estarían a una altura equivalente a la de la banda adicional, pero en los carriles en los que se analizaron por separado las muestras correspondientes a las proteínas Artemis y DNA-PKcs (*recuadros violeta y negro respectivamente de las Figuras 4.6.3A y 4.6.3C*). En el fragmento control de la proteína Artemis se identifica claramente esta proteína, lo cual es compatible con la formación de diferentes bandas de tamaño variable que se ha descrito anteriormente para Artemis (*carril 5 de la Figura 4.6.3D*). En el caso de DNA-PKcs la cantidad de proteína presente en la banda control (*carril 4 de la Figura 4.6.3D*) es insignificante comparada con la

**Figura 4.6.3. Análisis bioquímico de la interacción entre las proteínas Artemis y DNA-PKcs (página siguiente).** **(A)** Gel de poliacrilamida en condiciones nativas teñido con plata en el que se analiza una mezcla de reacción que contiene DNA-PKcs y Artemis (*carril "DNA-PKcs + Artemis"*), y otra mezcla de reacción que contiene también ambas proteínas pero en este caso la proteína Artemis se incubó previamente con fosfatasa- $\lambda$  (*carril "DNA-PKcs + ArtemisdP"*). Además se analizaron como controles y por separado las mismas cantidades de DNA-PKcs y Artemis utilizadas en la mezcla de reacción "DNA-PKcs+Artemis". Otro control adicional consistió en analizar una mezcla de reacción que contiene únicamente DNA-PKcs y el tampón de dilución de la proteína Artemis (*carril "DNA-PKcs + tampón"*). En este último control también se utilizó la misma cantidad de DNA-PKcs que en el caso de la mezcla de reacción "DNA-PKcs+Artemis". Los diferentes recuadros que se indican en la figura corresponden a las regiones que se extrajeron del gel teñido según el método de tinción reversa. El código de colores de los diferentes recuadros se mantiene también en (C) y (D). **(B)** A partir de un gel como el que se muestra en (A) se llevó a cabo un "Western-blot" utilizando un anticuerpo primario frente a la proteína DNA-PKcs (DNA-PK Ab-2, Oncogene). Como controles se analizaron por separado 1,25  $\mu$ l de la proteína DNA-PKcs (2,2  $\mu$ M) y 0,625  $\mu$ l de la proteína Artemis (4,08  $\mu$ M). **(C)** Gel de poliacrilamida como el que se muestra en (A) teñido según el método de tinción reversa. A partir de este gel se extrajeron los fragmentos que se indican mediante recuadros. Los números de los distintos recuadros así como los diferentes colores coinciden con los que se muestran en (A) y (D). Como controles se analizaron por separado las mismas cantidades de DNA-PKcs y Artemis utilizadas en la mezcla de reacción. **(D)** Gel de poliacrilamida en condiciones desnaturizantes teñido con plata de las proteínas extraídas a partir de las bandas recortadas en (C). También se muestran los controles correspondientes a las proteínas purificadas Artemis y DNA-PKcs.



cantidad que se observa en la banda adicional. Como control también se analizaron por electroforesis en condiciones desnaturalizantes las bandas características que generan las proteínas Artemis y DNA-PKcs en gels nativos (*recuadros amarillo y azul respectivamente de las Figuras 4.6.3A y 4.6.3C*). Al llevar a cabo este análisis se obtuvo el patrón típico correspondiente a cada proteína (*carriles 2 y 3 de la Figura 4.6.3D*). Por tanto las dos proteínas interactúan de manera específica formando un complejo DNA-PKcs:Artemis que, en las condiciones ensayadas, es mayoritario y homogéneo. Además este

complejo es estable puesto que puede analizarse mediante electroforesis en condiciones nativas. Ensamblar el complejo DNA-PKcs:Artemis de manera reproducible es esencial para que nuestro grupo, en un futuro próximo, pueda realizar su análisis estructural por microscopía electrónica.

#### **4.6.5 Conclusión**

Se ha llevado a cabo el análisis bidimensional de imágenes de microscopía electrónica de la proteína Artemis teñidas negativamente. Además se ha realizado la caracterización bioquímica de la proteína Artemis purificada y también de la mezcla de reacción que contiene las proteínas Artemis y DNA-PKcs. En las condiciones ensayadas ha sido posible ensamblar un complejo DNA-PKcs:Artemis estable que, en el futuro, podrá ser analizado estructuralmente mediante microscopía electrónica.



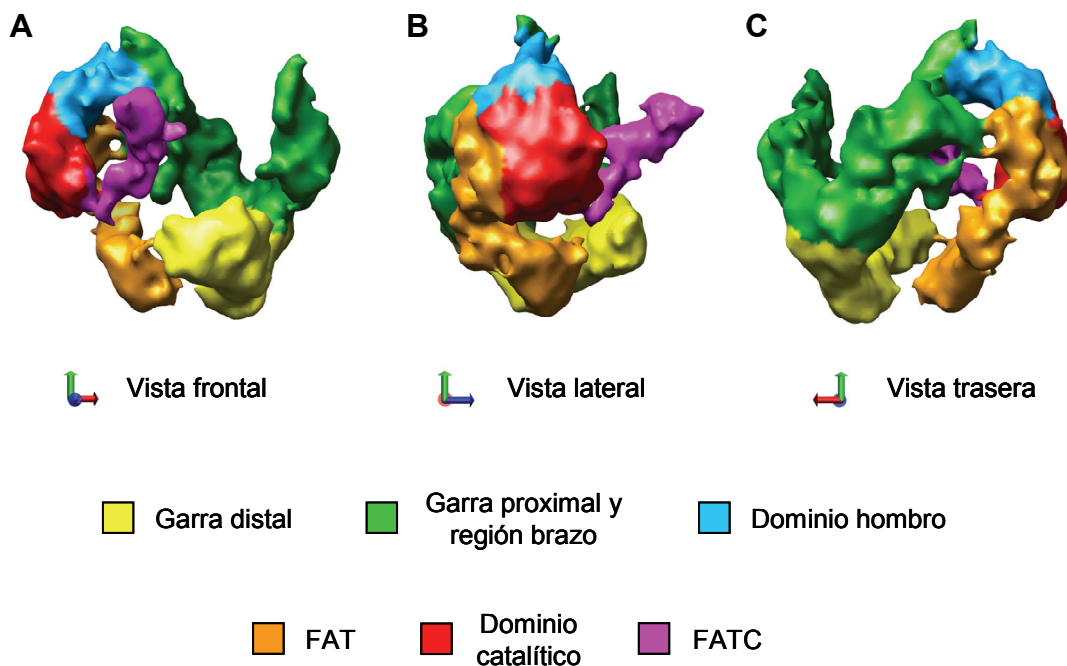
## 5. Discusión





## 5.1 Modelo topológico de DNA-PKcs

La estructura resuelta por criomicroscopía electrónica para DNA-PKcs a 13 Å de resolución permitió proponer un modelo de organización tridimensional para esta proteína (véase el apartado 4.3 de Resultados). Esta estructura describe una superficie continua en la que se identifican claramente siete dominios que se organizan en torno a una cavidad central. La resolución obtenida permite llevar a cabo un mapeo aproximado de la cadena polipeptídica de DNA-PKcs en la estructura resuelta (*Figura 5.1.1*). Según esto la garra distal y la porción de la región brazo asociada podrían alojar un fragmento de la cadena polipeptídica que correspondería a los primeros 800 aminoácidos (*de color amarillo en la Figura 5.1.1*). Los siguientes 1.000 aminoácidos podrían repartirse entre la densidad principal de la región brazo y la garra proximal (*de color verde en la Figura 5.1.1*). La porción de la cadena polipeptídica anterior al dominio FAT se distribuiría entre el final de la región brazo y el comienzo del dominio hombro (*de color azul en la Figura 5.1.1*). La densidad que bordea el dominio catalítico desde el extremo superior de la



**Figura 5.1.1. Modelo de organización tridimensional para DNA-PKcs.** (A), (B) y (C) Vista frontal, lateral y trasera respectivamente de la estructura de criomicroscopía electrónica resuelta para DNA-PKcs. Las regiones de la estructura asignadas a dominios específicos de DNA-PKcs se muestran en diferentes colores.

región cabeza e incluye, además, el dominio más masivo que emerge desde el dominio central estaría formada principalmente por el dominio FAT (*de color naranja en la Figura 5.1.1*). La densidad asignada al dominio quinasa alojaría los residuos 3.700 a 4.000 (*de color rojo en la Figura 5.1.1*) y, finalmente, los últimos 100 aminoácidos correspondientes al dominio FATC rellenarían la densidad de menor tamaño y forma cilíndrica que emerge del dominio central (*de color violeta en la Figura 5.1.1*). Esta distribución, a pesar de ser un modelo hipotético, se fundamenta en experimentos bioquímicos, bioinformáticos así como en el análisis comparativo de estructuras resueltas mediante microscopía electrónica. Estas aproximaciones han permitido localizar el dominio catalítico en la región cabeza de DNA-PKcs y el extremo N-terminal en la región palma.

A partir de este modelo se puede proponer la localización aproximada en la estructura de aquellas regiones que son importantes desde un punto de vista funcional (*véanse las Figuras 1.4.3 y 4.3.9*). DNA-PKcs fosforila otras proteínas implicadas en el proceso de reparación por NHEJ y además se autofosforila eficientemente. Hasta el momento se han identificado dos regiones principales de autofosforilación: ABCDE y PQR (Cui et al., 2005; Meek et al., 2007). La región ABCDE (residuos 2.609-2.647) junto con el sitio de corte de la caspasa-3 (aminoácido 2.713) (Itoh y Horio, 2001), se localizarían próximos al dominio hombro o bien en ese mismo dominio. La región PQR (residuos 2.023-2.056) se ubicaría en la región brazo pero más cercana a la garra proximal. Las regiones ABCDE y PQR se encuentran por tanto alejadas del dominio catalítico. Según esto, para que estas regiones puedan ser fosforiladas por el dominio catalítico de la misma molécula, DNA-PKcs tendría que sufrir una reorganización estructural considerable. El modelo topológico propuesto predice que la autofosforilación en estas regiones sólo podrá ocurrir en *trans* gracias a la interacción con una segunda molécula de DNA-PKcs. Experimentos bioquímicos recientes realizados por el grupo de S. P. Lees-Miller han confirmado que la autofosforilación en las regiones ABCDE y PQR tiene lugar en *trans* (Meek et al., 2007). Las zonas de autofosforilación descritas intervienen en la regulación de la actividad quinasa de DNA-PKcs. Por ello su ubicación en regiones accesibles de la estructura es totalmente compatible con su papel funcional. En cualquier caso la estructura de

criomicroscopía electrónica resuelta para DNA-PKcs presenta una resolución limitada y por ello, para confirmar el modelo topológico propuesto para DNA-PKcs, será necesario disponer de una estructura con más detalle de esta proteína.

### **5.1.1 Comparación entre el modelo estructural propuesto y la información existente para DNA-PKcs obtenida mediante criomicroscopía electrónica**

En 1998 el grupo de P. L. Stewart resolvió la estructura de DNA-PKcs a 21 Å de resolución utilizando criomicroscopía electrónica (Chiu et al., 1998). La densidad de este volumen no presenta una continuidad evidente puesto que está perforada por multitud de cavidades, y por tanto es difícil de conciliar con el resto de estructuras existentes para la proteína DNA-PKcs (*véase la Figura 1.4.3*). Este grupo ha resuelto recientemente la estructura de DNA-PKcs a 7 Å de resolución utilizando también criomicroscopía electrónica (de ahora en adelante esta estructura se va a denominar también DNA-PKcs\_7) (Williams et al., 2008). En este caso DNA-PKcs\_7 muestra la organización típica de DNA-PKcs consistente en las regiones cabeza, palma y brazo que aparece en la mayoría de estructuras resueltas para esta proteína (*Figura 5.1.2A*). A pesar de ello, existen diferencias entre esta nueva estructura y el volumen resuelto previamente por nuestro grupo a 13 Å de resolución utilizando también criomicroscopía electrónica (de ahora en adelante me refiero a nuestra estructura como DNA-PKcs\_13).

#### *Región palma*

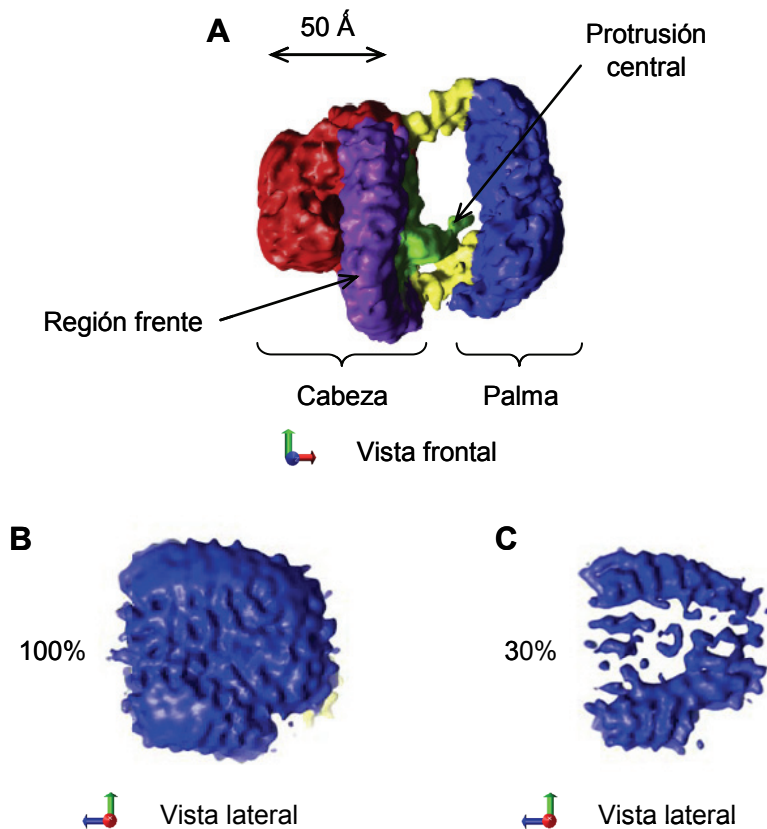
En la estructura obtenida por el grupo de P. L. Stewart esta región parece estar constituida por una densidad continua. Al representar esta estructura a un umbral correspondiente al 30% de la masa del espécimen los autores aprecian dos densidades principales semejantes a los dominios garra descritos por nuestro grupo (*Figuras 5.1.2B y 5.1.2C*).

### *Región brazo*

Una de las diferencias principales entre las dos estructuras se localiza en esta región. En el volumen obtenido por el grupo de P. L. Stewart las regiones cabeza y palma se conectan por dos sitios diferentes (*las dos conexiones se muestran de color amarillo en la Figura 5.1.2A*), sin embargo en nuestra estructura estas dos regiones se conectan únicamente por la región brazo. Cuando se observa DNA-PKcs\_13 se aprecia que las porciones más distales de las regiones cabeza y palma se localizan próximas en el espacio (*véase la Figura 4.3.2*). DNA-PKcs es una molécula flexible cuya conformación está modulada principalmente por el grado de autofosforilación y por la interacción con ligandos. Durante el proceso de NHEJ DNA-PKcs sufre cambios conformacionales asociados a eventos de autofosforilación (Block et al., 2004; Uematsu et al., 2007). Además la estructura del complejo DNA-PKcs:ADN evidenció que la interacción con extremos de ADN provoca importantes cambios conformacionales responsables, en último término, de la interacción entre las regiones cabeza y palma (*véase el apartado 4.4 de Resultados*). La presencia de una segunda conexión entre las regiones cabeza y palma en DNA-PKcs\_7 podría reflejar la flexibilidad intrínseca de DNA-PKcs. Esta segunda conexión puede aparecer por tanto como consecuencia de la interacción que se establece entre las regiones cabeza y palma.

### *Región cabeza*

En la estructura resuelta por el grupo de P. L. Stewart esta región presenta una organización similar a la que se observa en la región equivalente de DNA-PKcs\_13 (*Figura 5.1.2A*). Sin embargo DNA-PKcs\_7 carece de la densidad conspicua que previamente hemos asignado al dominio FATC. Esta densidad, a pesar de ser muy evidente en DNA-PKcs\_13, no se había observado en estructuras resueltas anteriormente a partir de especímenes teñidos negativamente. El dominio FATC corresponde a una extensión C-terminal específica de las PIKKs que se dispone a continuación de la región catalítica (Keith y Schreiber, 1995). Este dominio presenta un plegamiento característico e independiente del dominio catalítico que es totalmente compatible con el



**Figura 5.1.2. Estructura de DNA-PKcs a 7Å de resolución obtenida utilizando criomicroscopía electrónica.** (A) Vista frontal de la estructura de DNA-PKcs resuelta por el grupo de P. L. Stewart (Williams et al., 2008). Se indican las regiones cabeza y palma características de DNA-PKcs así como los dominios frente (de color violeta) y protrusión central (de color verde). Además se muestran en color amarillo las dos densidades por las que conectan las regiones cabeza y palma. (B) y (C) Vistas laterales de la región palma. (B) Corresponde a un umbral de visualización equivalente al 100% de la masa de la proteína. (C) Refleja un umbral equivalente al 30% de la masa de la proteína. En (C) se aprecia que la región palma se compone de dos dominios principales equivalentes a los dominios garra descritos anteriormente. Las diferentes figuras se han extraído de Williams et al., 2008.

aspecto de la densidad asignada en DNA-PKcs<sub>13</sub> (Dames et al., 2005). Es posible por tanto que la conexión entre los dominios FATC y catalítico sea flexible, y como consecuencia de esa plasticidad pueda disponerse en orientaciones diversas. Además en el caso de especímenes teñidos negativamente el agente de tinción podría compactar la proteína haciendo que el dominio FATC contacte con el resto de la región cabeza.

En el volumen resuelto por el grupo de P. L. Stewart, la región que han denominado frente (*de color violeta en la Figura 5.1.2A*) corresponde a una

densidad tubular continua compatible con el aspecto de la densidad asignada al dominio FAT en DNA-PKcs<sub>13</sub>. Sin embargo la región denominada protrusión central (*de color verde en la Figura 5.1.2A*) no presenta equivalencia en DNA-PKcs<sub>13</sub> y por tanto, sin la existencia de información a resolución atómica para DNA-PKcs, resulta imposible aclarar esta ambigüedad.

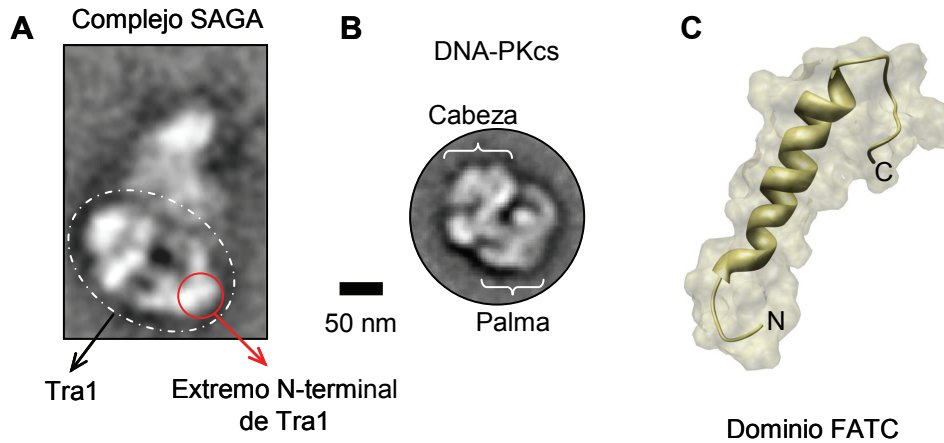
La estructura resuelta por el grupo de P. L. Stewart a 7 Å de resolución comparte una organización global similar a la que se observa en DNA-PKcs<sub>13</sub>. Ambas estructuras sin embargo también presentan diferencias que podrían ser consecuencia de la flexibilidad intrínseca de DNA-PKcs. Como se mencionó anteriormente esta flexibilidad depende principalmente del grado de autofosforilación de DNA-PKcs y de la interacción con ligandos. Una muestra heterogénea desde un punto de vista bioquímico podría generar diferentes conformaciones en solución y, en último término, un conjunto heterogéneo de imágenes de proyección. Procesar este tipo de imágenes como un conjunto homogéneo de partículas individuales da lugar a volúmenes incorrectos. Concretamente en los volúmenes que se obtienen las regiones flexibles aparecen peor representadas e incluso pueden estar ausentes. Esto es aún más acusado en el caso de que la flexibilidad se concentre en dominios de pequeño tamaño como por ejemplo el dominio FATC. En el volumen correspondiente a DNA-PKcs<sub>7</sub> hay ciertas evidencias que sugieren heterogeneidad en las partículas utilizadas durante el procesamiento. Las estructuras existentes para DNA-PKcs presentan una única densidad principal asignada a la región brazo que conecta las regiones cabeza y palma, sin embargo en DNA-PKcs<sub>7</sub> se establecen dos conexiones entre ambas regiones. Además estas dos conexiones presentan una menor densidad de masa respecto a las regiones cabeza o palma de la misma estructura (*las dos conexiones se muestran de color amarillo en la Figura 5.1.2A*). Otro dato adicional que sugiere heterogeneidad en las imágenes de proyección tiene que ver con el aspecto de la región palma a diferentes umbrales de visualización (*Figuras 5.1.2B y 5.1.2C*). Esta región, según el umbral de visualización, tiene aspecto de bloque homogéneo o bien se distinguen dos dominios equivalentes a los dominios garra presentes en DNA-PKcs<sub>13</sub>. Es posible por tanto que el aspecto continuo de la región palma en DNA-PKcs<sub>7</sub> sea consecuencia del

promediado de las diversas orientaciones que adquiere esta región en la molécula. Esta afirmación no es descabellada teniendo en cuenta que la región palma sufre cambios evidentes en su orientación como consecuencia de la interacción de DNA-PKcs con extremos de ADN.

### 5.1.2 Datos adicionales refuerzan el modelo topológico propuesto para DNA-PKcs

Existen experimentos realizados por otros grupos que refuerzan la asignación del extremo N-terminal de DNA-PKcs a la región palma y por tanto apoyan nuestro modelo topológico. Tra1 (del inglés “**T**ranscription-associated protein 1”) (437 kDa) es la proteína ortóloga de TRRAP en *Saccharomyces cerevisiae* (Grant et al., 1998). Esta proteína perteneciente a la familia de las PIKKs y cuyo peso molecular es similar al de DNA-PKcs, forma parte de un complejo multifunción denominado SAGA (del inglés “**S**pt-**A**da-**G**cn5-**A**cetyltransferase”) (1,8 MDa) que es un coactivador de la RNA polimerasa II (Green, 2005). La estructura de este complejo se ha resuelto mediante microscopía electrónica de especímenes teñidos negativamente (Wu et al., 2004). En las proyecciones del complejo SAGA se aprecia una densidad evidente cuyo aspecto es muy parecido al de las imágenes de proyección obtenidas durante el procesamiento de DNA-PKcs (*Figura 5.1.3A*). La proteína Tra1 presente en el complejo SAGA estaba marcada con una etiqueta c-Myc en su extremo N-terminal. Gracias a esta etiqueta fue posible localizar su ubicación dentro del complejo utilizando un anticuerpo frente a c-Myc. El marcaje con anticuerpos reveló que la etiqueta c-Myc se localiza en una zona que sería equivalente a la región palma en las imágenes de proyección de DNA-PKcs (*Figuras 5.1.3A y 5.1.3B*). Estos resultados refuerzan nuestra hipótesis sobre la asignación de la región más N-terminal de DNA-PKcs a la región palma.

Las estructuras existentes para distintas PIKKs también apoyan el modelo topológico propuesto para DNA-PKcs. Utilizando microscopía electrónica y procesamiento digital de partículas individuales se han resuelto las estructuras de la proteína humana ATM y de la proteína TOR1 de levaduras (Llorca et al.,



**Figura 5.1.3. Experimentos adicionales refuerzan el modelo topológico propuesto para DNA-PKcs.** (A) y (B) Comparación de imágenes de tinción negativa correspondientes al complejo SAGA y a DNA-PKcs. (A) Imagen de proyección de la estructura resuelta para el complejo SAGA de *Saccharomyces cerevisiae* a partir de especímenes teñidos negativamente. Se indica la densidad correspondiente a Tra1 dentro de la proyección y también la localización aproximada de su extremo N-terminal. La figura se ha extraído de Wu et al., 2004. (B) Detalle de una imagen media de DNA-PKcs generada mediante algoritmos de clasificación libres de patrón a partir de imágenes de tinción negativa. Se indican las regiones cabeza y palma características. Las dos imágenes se muestran a la misma escala. (C) Estructura atómica correspondiente al dominio FATC de la proteína de levaduras homóloga a mTOR (archivo PDB 1W1N) (Dames et al., 2005). Además se muestra mediante una densidad transparente la superficie de la estructura atómica.

2003; Adami et al., 2007). Como se describe en el apartado 5.3, comparando estos volúmenes es posible proponer un modelo estructural común para las distintas PIKKs análogo al descrito para DNA-PKcs.

Existe además otro dato adicional que apoya nuestro modelo estructural. Durante el análisis de la estructura de DNA-PKcs obtenida mediante criomicroscopía electrónica se generó un modelo atómico para el dominio FATC de esta proteína (véase el apartado 4.3.2 de Resultados). Este modelo encajaba perfectamente en la densidad con forma cilíndrica que emergía del dominio central de la región cabeza (véase la Figura 4.3.6). Además las dimensiones de esta densidad eran compatibles únicamente con las de una estructura terminal. Posteriormente se resolvió la primera estructura atómica de un dominio FATC perteneciente a la proteína de levaduras homóloga a mTOR (Figura 5.1.3C) (Dames et al., 2005). Esta estructura es totalmente compatible con el modelo atómico propuesto previamente y corresponde, en efecto, a una



estructura terminal que se pliega de manera independiente (*compárense las Figuras 4.3.6 y 5.1.3C*). Estos datos sugieren por tanto que la asignación de los diferentes dominios que se observan en la estructura de criomicroscopía electrónica de DNA-PKcs ha sido correcta.

## 5.2 El reconocimiento de extremos de ADN por DNA-PKcs conlleva la activación de la quinasa mediante cambios conformacionales

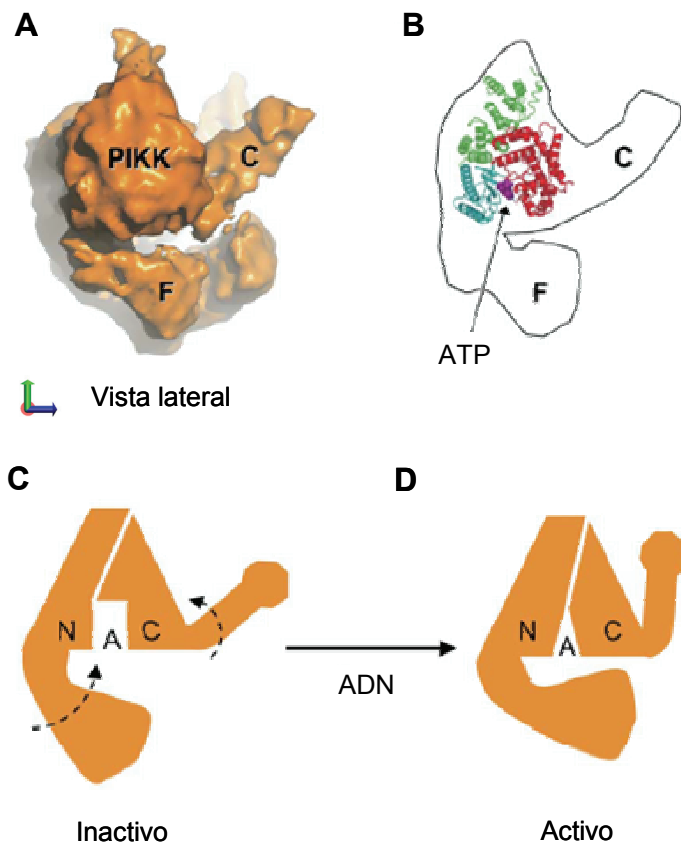
Estudios bioquímicos previos han permitido caracterizar la unión de DNA-PKcs a los extremos de ADN (Hammarsten y Chu, 1998; West et al., 1998; Leuther et al., 1999). Concretamente para que tenga lugar la activación de DNA-PKcs es necesario que la molécula de ADN tenga una longitud mínima de 15 pb. Fragmentos de ADN de menor tamaño pueden unirse a DNA-PKcs con una afinidad comparable pero no inducen la activación de la quinasa. Además aumentar la longitud del fragmento de ADN por encima de 20 pb no conlleva una mayor actividad quinasa de DNA-PKcs (Leuther et al., 1999). Este modo de activación es compatible con un modelo que implique la interacción simultánea del extremo de ADN con al menos dos regiones diferentes de DNA-PKcs. Una de estas regiones participaría en el reconocimiento del extremo de ADN. La segunda región sin embargo interaccionaría con una zona de la molécula de ADN que se encuentre a una distancia del extremo de aproximadamente 15 pb. De esta manera se justifica el requisito de 15 pb como longitud mínima para que tenga lugar la activación. El análisis estructural previo realizado por otros grupos ha permitido proponer diversas hipótesis sobre cómo tendría lugar la unión de DNA-PKcs a los extremos de ADN (Chiu et al., 1998; Leuther et al., 1999). Según estos modelos DNA-PKcs reconoce de forma específica las regiones de cadena sencilla y de cadena doble presentes en los extremos. Esta unión estaría mediada por los diferentes agujeros y cavidades de DNA-PKcs y podría requerir la participación de un dominio flexible (*véase la Figura 1.4.3E*).

El análisis estructural del complejo DNA-PKcs:ADN muestra que DNA-PKcs atrapa parte del fragmento de ADN entre las regiones cabeza y palma. Para ensamblar el complejo DNA-PKcs:ADN se utilizó una molécula de ADN de 54 pb capaz de activar DNA-PKcs y provocar la fosforilación de XRCC4 (véase el apartado 4.4 de Resultados). En la estructura resuelta para el complejo DNA-PKcs:ADN se apreciaba una densidad conspicua que emergía del interior del volumen. Las dimensiones de esta densidad eran semejantes a las de un fragmento de ADN de 36 pb, por lo tanto 18 de los 54 pb presentes en la molécula de ADN quedaban encerrados en el interior de DNA-PKcs (véase la *Figura 4.4.4*). Una longitud de 18 pb es totalmente compatible con la longitud mínima de 15 pb descrita a partir del análisis bioquímico de la interacción entre DNA-PKcs y los extremos de ADN. Además en el complejo DNA-PKcs:ADN se observan cambios conformacionales evidentes que correlacionan, en último término, con la activación de DNA-PKcs. El cambio conformacional principal consiste en la formación de un canal entre las regiones cabeza y palma que encierra la molécula de ADN. Este cambio conformacional podría estar estabilizado por al menos dos puntos de contacto entre DNA-PKcs y la molécula de ADN. El primero de ellos se localizaría en el interior de DNA-PKcs y reconocería el extremo del ADN. El segundo punto de contacto podría localizarse en una zona más externa de DNA-PKcs. Este segundo punto reconocería el esqueleto azúcar-fosfato de la molécula de ADN en su recorrido hacia el exterior del complejo. Ambos puntos se encuentran distantes en el espacio y por ello, para que tenga lugar la activación de DNA-PKcs, el fragmento de ADN debe tener una longitud mínima que permita la interacción simultánea con los dos puntos de contacto descritos. Esa longitud mínima sería clave para estabilizar la conformación cerrada característica del complejo DNA-PKcs:ADN. Una molécula de ADN que no fuese lo suficientemente larga sólo se uniría al sitio que reconoce el extremo del ADN y por tanto sería incapaz de activar a DNA-PKcs. Además fragmentos de ADN que excedan de la longitud mínima no aportarán una mayor estabilización a la conformación cerrada ni provocarán, por tanto, un aumento de la activación de DNA-PKcs. El modelo estructural propuesto es por tanto totalmente consistente con los datos bioquímicos descritos en la literatura.

La estructura de DNA-PKcs resuelta mediante criomicroscopía electrónica y la estructura del complejo DNA-PKcs:ADN presentan una organización similar. Según esto es posible identificar en la estructura de criomicroscopía electrónica los dominios que, en el volumen obtenido para el complejo DNA-PKcs:ADN, sufren cambios conformacionales como consecuencia de la interacción con el ADN. La unión al extremo del ADN provoca que las regiones cabeza y palma interaccionen entre sí. En esta interacción participarían la garra distal de la región palma y el dominio asignado a la región C-terminal del dominio FAT (véase la Figura 4.4.5). Como consecuencia de la unión al extremo de ADN el dominio FAT se dobla hacia la cabeza desde su posición inicial en la proteína libre. Como se mencionó anteriormente el cambio conformacional que conlleva la interacción con los extremos de ADN estaría estabilizado por dos puntos de contacto. En la estructura de criomicroscopía electrónica el primer punto de contacto se localiza cercano a la conexión entre la región brazo y la garra proximal. El segundo punto de contacto más externo está formado por la garra proximal de la región palma y el dominio FATC (véase la Figura 4.4.5). Los cambios conformacionales descritos pueden favorecer la unión y el mantenimiento de la interacción entre DNA-PKcs y los extremos de ADN. Además la interacción de DNA-PKcs con extremos de ADN provoca su activación como quinasa. Según esto, es muy posible que los cambios conformacionales que se observan en el complejo DNA-PKcs:ADN correlacionen con la activación de DNA-PKcs.

La estructura de criomicroscopía electrónica resuelta para DNA-PKcs permitió localizar el dominio catalítico en el dominio central de la región cabeza. Esta localización, junto con la información obtenida a partir del complejo DNA-PKcs:ADN, sugiere que el centro activo queda accesible incluso cuando la proteína DNA-PKcs se encuentre unida a extremos de ADN. En el modelo topológico propuesto para DNA-PKcs el centro activo queda comprendido entre los dominios FAT y FATC (Figuras 5.2.1A y 5.2.1B). El dominio FAT conecta directamente con el lóbulo N-terminal del dominio catalítico mientras que el dominio FATC lo hace con el lóbulo C-terminal. Como se ha visto anteriormente los cambios conformacionales descritos para el complejo DNA-PKcs:ADN afectan principalmente a los dominios FAT y FATC. Nuestros datos permiten

proponer un modelo de regulación de la actividad de DNA-PKcs dependiente de los dominios FAT y FATC. Según este modelo ambos dominios actuarían como sensores que sufrirían cambios conformacionales inducidos por la unión al ADN o incluso a otras proteínas. Esos cambios conformacionales podrían alterar la orientación relativa de los residuos que configuran el centro activo y que unen ATP y, de esta manera, provocar la activación de DNA-PKcs (*Figuras 5.2.1C y 5.2.1D*). Este sistema de regulación es un mecanismo clásico por el que se controla la actividad de las quinasas (Huse y Kuriyan, 2002). El dominio



**Figura 5.2.1. Modelo de regulación de la actividad quinasa de DNA-PKcs dependiente de la unión a ligando.** (A) Vista lateral de la estructura resuelta para DNA-PKcs utilizando criomicroscopía electrónica. Se indica la localización propuesta para los dominios FATC (C), FAT (F) y catalítico (PIKK). (B) Representación esquemática en la que se muestra el dominio catalítico de la PI3Kγ ajustado en el dominio central de la región cabeza (archivo PDB 1E8X). El lóbulo N-terminal del dominio catalítico se representa de color azul y el C-terminal de color rojo. Estos lóbulos conectan directamente con los dominios FAT y FATC respectivamente y entre ellos se une el ATP (de color morado en la figura). El dominio helicoidal se representa de color verde. (C) y (D) Modelo de activación de DNA-PKcs dependiente de la interacción con ligandos. La unión a extremos de ADN conlleva cambios conformacionales en los dominios FAT y FATC que se transmiten a los lóbulos del dominio catalítico (N y C en la figura). Estos cambios alteran la orientación relativa de los residuos que forman el centro activo y alojan la molécula de ATP (A en la figura).

catalítico de las distintas quinasas posee una arquitectura común basada en dos lóbulos, N-terminal y C-terminal, separados por una cavidad donde se aloja el ATP. Variaciones en la orientación relativa y separación de estos lóbulos permiten alternar entre las conformaciones activa e inactiva de estas proteínas. Datos que apoyan nuestro modelo vienen de experimentos en ratones en los que se ha visto que la pérdida de los 70 residuos más C-terminales de DNA-PKcs provoca la aparición de un fenotipo SCID. En estos ratones la actividad quinasa de DNA-PKcs dependiente de ADN se reduce a menos del 2% de la actividad normal (Bosma et al., 1983; Beamish et al., 2000). Estas observaciones son compatibles por tanto con un posible papel del dominio FATC como sensor de la interacción con ADN.

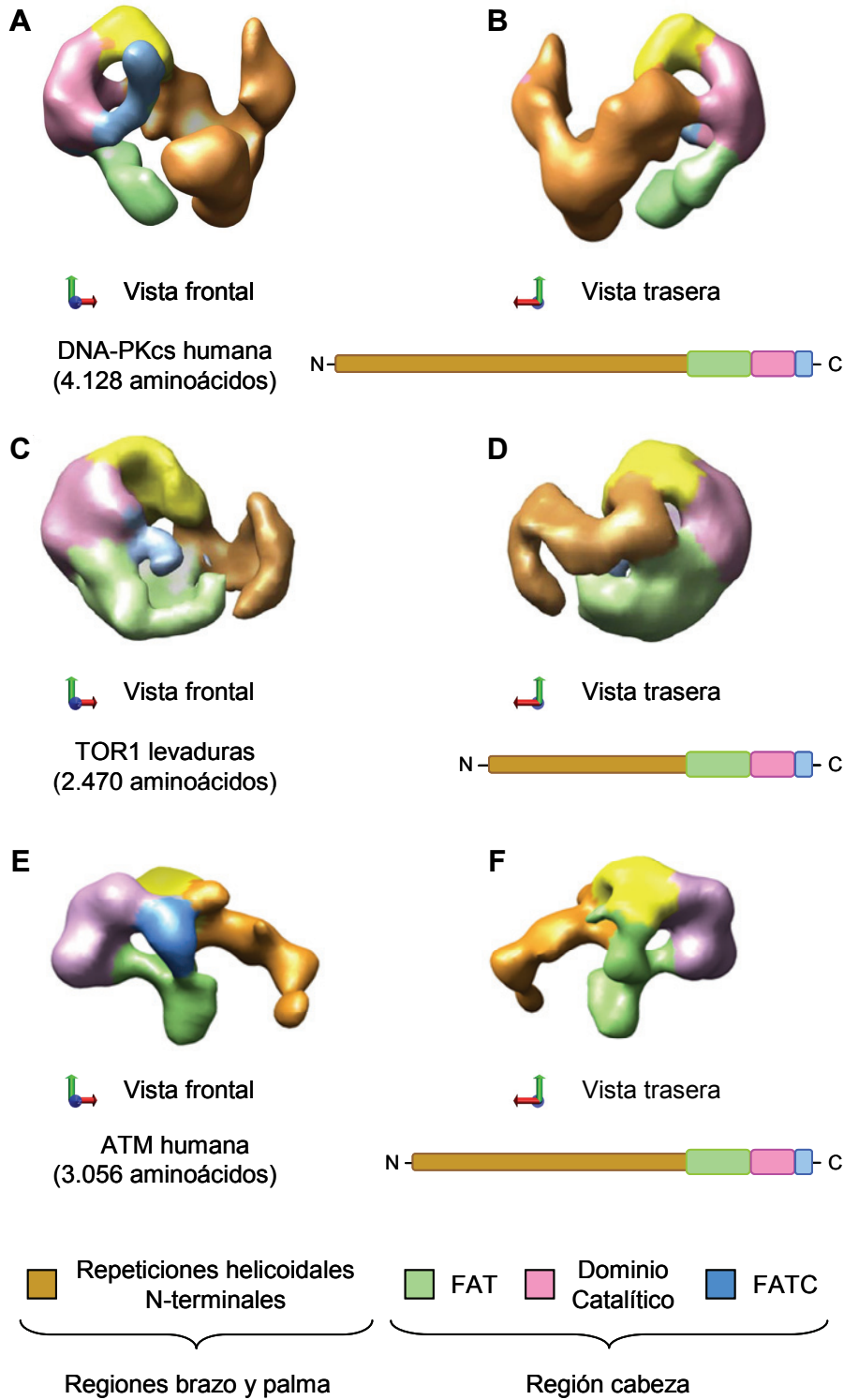
### 5.3 Existe un modelo estructural común para los miembros de la familia de las PIKKs

Las estructuras tridimensionales de ATM (350 kDa) y TOR (281 kDa) también han sido resueltas utilizando microscopía electrónica y análisis de partículas individuales teñidas negativamente (Llorca et al., 2003; Adami et al., 2007). Al comparar estas estructuras con la de DNA-PKcs se aprecia que los tres volúmenes comparten características estructurales significativas (*Figura 5.3.1*). TOR, ATM y DNA-PKcs se organizan según un patrón común que consiste en una región cabeza voluminosa de la que emerge la región brazo (*Figura 5.3.1*). En todas ellas estas dos regiones presentan dimensiones y orientaciones relativas similares. Además la organización estructural de la región cabeza es parecida en las tres proteínas.

La estructura de DNA-PKcs obtenida por criomicroscopía electrónica permite proponer un modelo estructural de la organización tridimensional de su cadena polipeptídica. Según este modelo la región palma corresponde a la región N-terminal de DNA-PKcs y se organiza en dos dominios garra. Las regiones brazo y palma forman estructuras tubulares discontinuas. Estas regiones son compatibles con experimentos de análisis de secuencia que predicen que la región N-terminal de DNA-PKcs no es un continuo de

repeticiones helicoidales (Perry y Kleckner, 2003; Brewerton et al., 2004). Al comparar las estructuras de ATM, TOR y DNA-PKcs la diferencia principal se da en la región palma (*Figura 5.3.1*). En TOR y ATM esta región consiste en una prolongación del brazo terminada en una garra distal de menores dimensiones que la garra equivalente de DNA-PKcs. Por el contrario la garra proximal está ausente en ambas proteínas. Las diferencias de tamaño entre ATM, TOR y DNA-PKcs corresponden a variaciones en la extensión de las regiones helicoidales N-terminales (*Figura 5.3.1*) que se reflejan, a nivel tridimensional, en la estructura de la región palma. Curiosamente, al eliminar la garra proximal de la estructura de criomicroscopía electrónica se obtiene un brazo cuya forma es muy similar a la que presenta esta región en la estructura de ATM (véase la *Figura 4.3.3*). El aspecto de tubo aplanado ligeramente curvado de las regiones brazo es característico de secuencias ricas en repeticiones helicoidales (véase el apartado 1.4.2 de Introducción). La garra distal podría ser una adición evolutiva específica de DNA-PKcs que le permitiría llevar a cabo actividades concretas durante el proceso de reparación por NHEJ. Por el contrario y como he mencionado anteriormente (véase el apartado 4.3.3 de Resultados), el dominio garra proximal podría aparecer como consecuencia de una inserción en la estructura canónica de la región brazo (Malek et al., 2003). En cualquier caso la baja homología de secuencia entre ATM, TOR y DNA-PKcs fuera de la región C-terminal impide determinar por análisis de secuencia si, respecto a ATM y TOR, la garra proximal de DNA-PKcs es una inserción o bien una extensión del extremo N-terminal.

**Figura 5.3.1. Arquitectura común de las PIKKs DNA-PKcs, ATM y TOR (página siguiente).** (A) y (B) Vista frontal y trasera respectivamente de la estructura resuelta para DNA-PKcs utilizando criomicroscopía electrónica. La estructura se muestra filtrada a una resolución aproximada de 25 Å. (C) y (D) Vista frontal y trasera respectivamente de la estructura resuelta para la proteína TOR1 de *Saccharomyces cerevisiae* a partir de especímenes teñidos negativamente. Estas vistas se han extraído de Adami et al., 2007. (E) y (F) Vista frontal y trasera respectivamente de la estructura resuelta para la proteína ATM humana utilizando tinción negativa (Llorca et al., 2003). Las diferentes estructuras se muestran a la misma escala y coloreadas de manera equivalente según los diferentes dominios característicos de la familia de las PIKKs. Estos dominios se representan de manera esquemática para cada una de las proteínas. Los dominios catalítico, FAT y FATC dan lugar a la región cabeza. La mitad N-terminal rica en repeticiones de motivos helicoidales genera la región brazo que puede terminar en una región palma más o menos voluminosa.



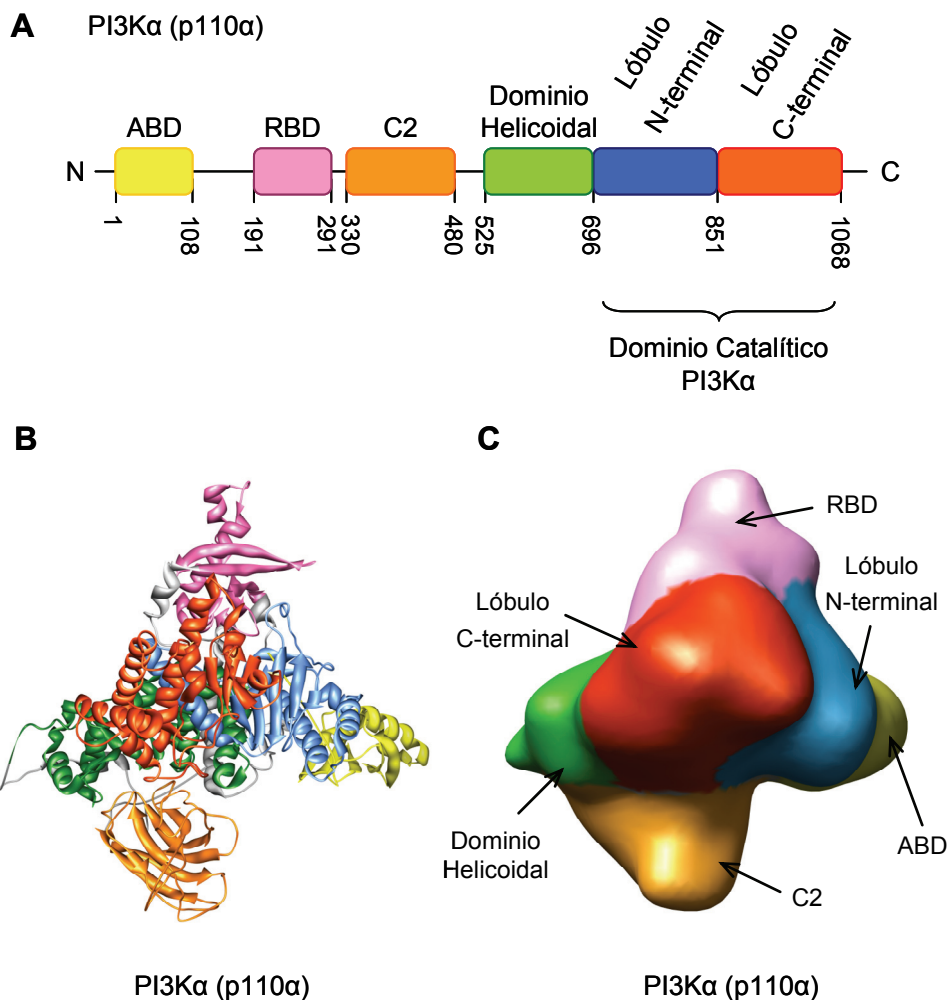
Las estructuras resueltas para TOR, ATM y DNA-PKcs permiten proponer un modelo de organización espacial común para los miembros que componen la familia de las PIKKs. Este modelo se compone de una región cabeza principal de la que emerge la región brazo. La región cabeza es la región en torno a la cual se organiza el resto de la estructura y consistiría en un dominio

catalítico central del que surgen los dominios FAT y FATC. Esta configuración es compatible además con la organización tridimensional que presentan las PI3Ks (*Figura 5.3.2*). (Walker et al., 1999; Miled et al., 2007). La estructura atómica completa de la subunidad catalítica de la PI3K $\alpha$  se organiza de manera similar a como se ha propuesto para la región cabeza de DNA-PKcs (Huang et al., 2007). Concretamente a partir del dominio catalítico central emergen distintos dominios en diferentes orientaciones. Estos dominios a pesar de estar distantes en la secuencia se disponen alrededor del dominio catalítico (*Figura 5.3.2*). En el modelo propuesto para las PIKKs el centro activo quedaría orientado hacia el exterior y su actividad quinasa, al igual que se ha descrito para DNA-PKcs, podría estar modulada por contactos intermoleculares e intramoleculares mediados por los dominios FAT y FATC. Además la orientación del centro activo condicionaría que los procesos de regulación dependientes de autofosforilación fueran en *trans*.

Una característica importante de las PIKKs es que funcionan a modo de andamio reclutando proteínas adicionales que participan en procesos de señalización o reparación (Keith y Schreiber, 1995). Esta característica se debe, principalmente, a la presencia de dominios formados por repeticiones de motivos helicoidales que constituyen módulos estructurales implicadas en interacciones proteína:proteína. Estos módulos jugarán por tanto un papel clave en el ensamblaje y regulación de los complejos proteicos en los que participan las PIKKs. El número de módulos que componen los distintos miembros de la familia de las PIKKs podrá variar debido a factores evolutivos. Estas variaciones se reflejarán a nivel tridimensional y podrían justificar también las diferencias funcionales entre las distintas PIKKs. El análisis de secuencia sugiere que estos módulos están formados por repeticiones de motivos helicoidales de tipo HEAT, ARM y PFT y se localizan en la región N-terminal de las PIKKs (Perry y Kleckner, 2003; Brewerton et al., 2004). Estas repeticiones tienen secuencias poco conservadas, sin embargo a nivel tridimensional forman estructuras tubulares curvadas características (véase el apartado 1.4.2 de Introducción). En el modelo propuesto de organización espacial de las PIKKs la región brazo estaría formada por segmentos ricos en repeticiones de motivos helicoidales. Las diferencias de tamaño entre las distintas PIKKs se dan principalmente en la región N-terminal. A nivel



tridimensional estas diferencias podrían corresponder a extensiones, eliminaciones o inserciones en el extremo de la región brazo que serían responsables de la aparición de una región palma más o menos voluminosa. Además las regiones palma y brazo también podrían jugar un papel clave en aquellas PIKKs que, al igual que DNA-PKcs y ATM, interaccionen con el ADN. Estas PIKKs podrían presentar un mecanismo común basado en movimientos tipo abrazo que permitiesen internar la molécula de ADN u otros ligandos y que, en último término, indujesen cambios conformacionales que afectarían a los



**Figura 5.3.2. Organización tridimensional de la subunidad catalítica de la PI3K $\alpha$ .** (A) Representación esquemática de los dominios que componen la subunidad catalítica de la PI3K $\alpha$  (p110 $\alpha$ ). (B) Estructura atómica resuelta para p110 $\alpha$  (archivo PDB 2RD0) (Huang et al., 2007). (C) Volumen obtenido al filtrar la estructura atómica que se muestra en (B) a una resolución aproximada de 25 Å. Ambas estructuras se muestran en la misma orientación. El código de colores utilizado para representar los distintos dominios en (B) y (C) equivale al que se muestra en (A).

dominios FAT y FATC. En cualquier caso entender por completo el mecanismo molecular de DNA-PKcs así como el de otras PIKKs requiere un conocimiento detallado de la estructura de estas proteínas.

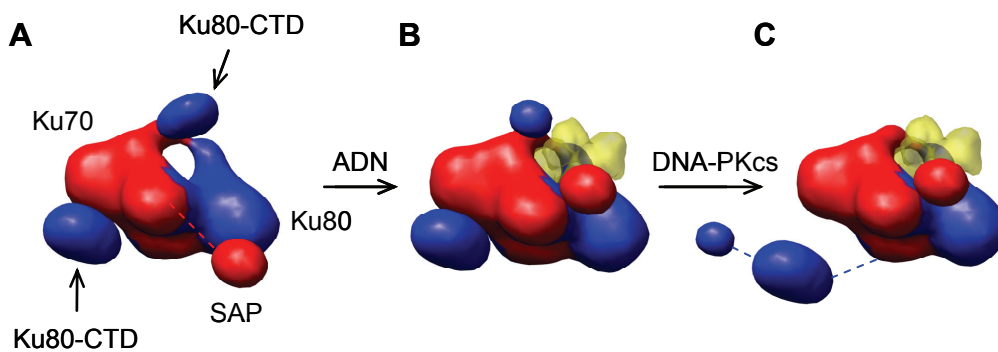
## 5.4 Modelo de las etapas iniciales del proceso de NHEJ

A lo largo de esta tesis se han resuelto las estructuras de diversos complejos biológicos que aparecen en los vertebrados durante la reparación de DSB mediante NHEJ. Es posible por tanto integrar toda esta información estructural en un modelo que explique cómo tienen lugar las etapas iniciales de este mecanismo de reparación.

### 5.4.1 La unión del heterodímero Ku a extremos de ADN conlleva cambios conformacionales en los dominios C-terminales de Ku70 y Ku80

En el heterodímero Ku las regiones C-terminales de Ku80 y Ku70 median funciones clave para la correcta reparación de DSB. Existe información estructural a resolución atómica para la mayor parte de la proteína Ku, sin embargo se desconocen los cambios conformacionales globales asociados a la interacción con los extremos de ADN y al posterior reclutamiento de DNA-PKcs (véase el apartado 1.4.1 de Introducción). Las estructuras presentadas en esta tesis para la proteína Ku humana completa, libre y unida a ADN, evidenciaron que los dominios C-terminales de Ku70 y Ku80 sufren reorganizaciones como consecuencia de la unión a los extremos de ADN (véase el apartado 4.1 de Resultados). Estos volúmenes, junto con la información estructural a resolución atómica existente, permiten proponer un modelo funcional para el heterodímero Ku asociado a las etapas iniciales del proceso de NHEJ (*Figura 5.4.1*). Ku consiste en un anillo preformado que enhebra el extremo de ADN generado tras la aparición de DSB. El reconocimiento inicial de los extremos conlleva el desplazamiento del dominio SAP de Ku70 desde su ubicación original en la base del dominio  $\alpha/\beta$  de Ku80. A partir de la estructura atómica correspondiente al núcleo del heterodímero Ku unido a ADN (fichero PDB 1JEY) se ha propuesto que, tras la unión al extremo de ADN, el dominio SAP podría

localizarse cercano al agujero central, en el lado de la proteína Ku más próximo a la parte continua de la molécula de ADN (Walker et al., 2001). Esta localización es compatible con la función propuesta para este dominio consistente en estabilizar al heterodímero Ku en el extremo de la rotura. El dominio SAP podría favorecer la interacción con DNA-PKcs al contactar con la molécula de ADN y aumentar así la afinidad de Ku por el ADN (*Figura 5.4.1B*). Los volúmenes resueltos para Ku y el complejo Ku:ADN son compatibles con un desplazamiento del dominio SAP desde la base de Ku80 a regiones próximas al agujero central (*véanse las Figuras 4.1.8 y 4.1.9*). Además el modelo propuesto predice que, antes del reconocimiento del extremo de ADN, el CTD de Ku80 se encontraría en un estado laxo y organizado en dos zonas (*Figura 5.4.1A*). Una de estas zonas se ubica debajo del dominio  $\alpha/\beta$  de Ku70 mientras que la otra zona se localiza detrás del agujero central (*véase la Figura 4.1.8*). Sin embargo, a diferencia de lo que sucede con el dominio SAP de Ku70, la unión al extremo del ADN no induce un cambio conformacional significativo en la región CTD de Ku80. Tras la unión el dominio CTD de Ku80 se organiza en las mismas zonas descritas anteriormente aunque de manera más compacta (*compárense las Figuras 4.1.8 y 4.1.9*). Es posible por tanto que la



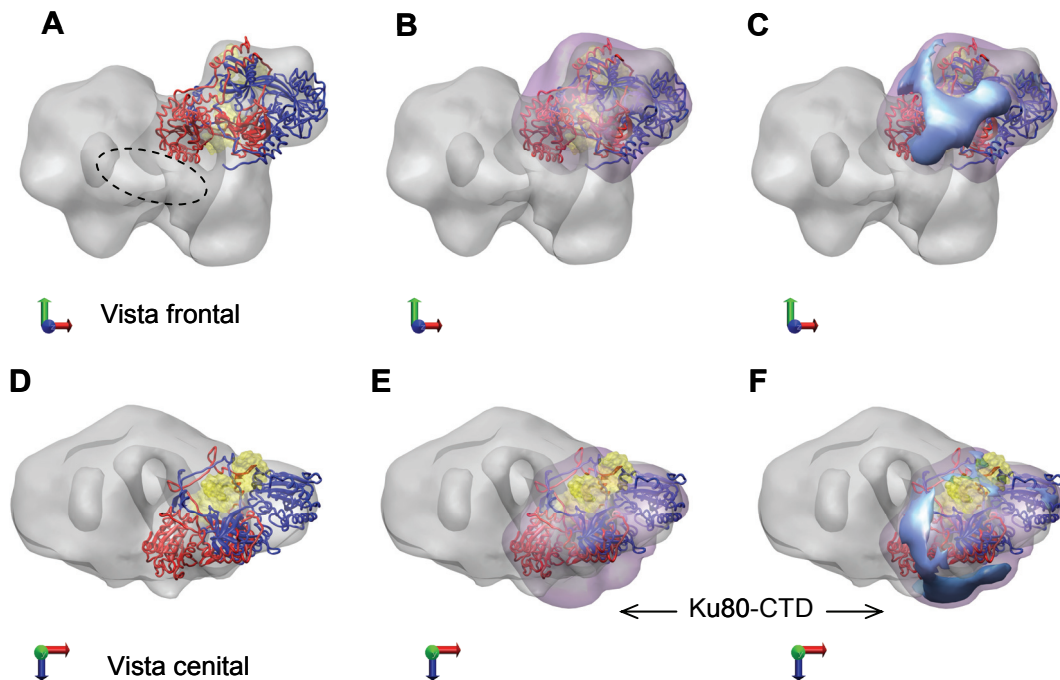
**Figura 5.4.1. Modelo esquemático de los cambios conformacionales que tienen lugar en el heterodímero Ku completo.** **(A)** Cuando Ku se encuentra libre el dominio CTD de Ku80 (de color azul en la figura) se organiza en dos regiones principales. El dominio SAP de Ku70 (de color rojo en la figura) se encuentra en la base del dominio  $\alpha/\beta$  de Ku80. **(B)** La unión al ADN (de color amarillo en la figura) conlleva cambios conformacionales que afectan principalmente a los dominios C-terminales de Ku70 y Ku80. Como resultado de la unión el CTD de Ku80 adquiere una conformación más compacta necesaria para el posterior reclutamiento de DNA-PKcs. El dominio SAP de Ku70 se desplaza a una posición cercana al agujero. **(C)** El reclutamiento de DNA-PKcs a los extremos de la rotura provoca un cambio conformacional importante en el CTD de Ku80. Este dominio se despegará del núcleo del heterodímero Ku para contactar con DNA-PKcs.

unión al extremo del ADN induzca algún cambio conformacional en la porción más C-terminal del CTD de Ku80 (*Figura 5.4.1B*). Esta región del CTD de Ku80 (residuos 720-732) es clave en el reclutamiento de DNA-PKcs a los sitios de rotura (Gell y Jackson, 1999). El análisis de secuencia predice que estaría desordenada en solución, aunque en determinadas circunstancias podría plegarse en una  $\alpha$ -hélice favoreciendo la interacción con DNA-PKcs (Zhang et al., 2004). Esta observación podría estar relacionada con los cambios conformacionales descritos a partir de la estructura resuelta para el complejo Ku:ADN. En cualquier caso la formación del complejo Ku:ADN conlleva alteraciones en la organización del heterodímero Ku responsables del reclutamiento posterior de DNA-PKcs (*Figura 5.4.1B*). De hecho la interacción entre Ku y DNA-PKcs únicamente es estable en presencia de extremos de ADN. Quizá un mecanismo de reclutamiento como el descrito para el heterodímero Ku y DNA-PKcs también aparezca en otras PIKKs (Falck et al., 2005). Experimentos recientes realizados por otros grupos sugieren también un modelo funcional en el que el Ku, debido a la unión a extremos de ADN, sufre cambios conformacionales en los dominios C-terminales de Ku70 y Ku80 (Lehman et al., 2008).

#### **5.4.2 El reclutamiento de DNA-PKcs por el complejo Ku:ADN conlleva cambios conformacionales en el dominio CTD de Ku80**

Es posible comparar las estructuras de tinción negativa correspondientes a los complejos Ku:ADN y DNA-PKcs:Ku:ADN y, de esta manera, explorar los cambios conformacionales que tienen lugar en el complejo Ku:ADN al reclutar a DNA-PKcs a los sitios donde se produjo la rotura (*Figura 5.4.2*). La comparación de ambas estructuras evidencia que el CTD de Ku80 se desplaza como consecuencia de la interacción con DNA-PKcs (*Figura 5.4.2*). Este dominio juega un papel clave en el reclutamiento de DNA-PKcs a los extremos del ADN. Es posible por tanto que tras la interacción con DNA-PKcs, el dominio CTD de Ku80 se desplace de su posición original e interactúe con regiones distantes de DNA-PKcs en el complejo DNA-PKcs:Ku:ADN (*Figura 5.4.1C*). La estructura resuelta para este complejo tiene una resolución limitada que dificulta la localización del dominio CTD de Ku80. Sin embargo, en la estructura

del complejo DNA-PKcs:Ku:ADN se distingue una zona de densidad ausente en estructuras resueltas previamente para DNA-PKcs (*Figura 5.4.2A y véase también la Figura 4.5.3*). Esta región podría asignarse siendo extremadamente prudentes al dominio CTD de Ku80. Los datos anteriores permiten proponer un modelo en el cual el dominio CTD de Ku80 sufriría un desplazamiento sustancial sólo cuando DNA-PKcs es reclutada (*Figura 5.4.1C*). En ese momento el CTD de Ku80 se despegaría del núcleo de Ku para interactuar con DNA-PKcs. El fragmento que une el núcleo de Ku80 con el dominio helicoidal del CTD tiene una longitud de 28 aminoácidos compatible con un desplazamiento de esa envergadura. La eliminación del CTD de Ku80 es



**Figura 5.4.2. Comparación de la estructura de Ku en los volúmenes obtenidos para los complejos DNA-PKcs:Ku:ADN y Ku:ADN.** (A) Vista frontal de la estructura resuelta por tinción negativa para el complejo DNA-PKcs:Ku:ADN en la que se ha ajustado la estructura atómica del heterodímero Ku (archivo PDB 1JEY). Ku70, Ku80 y la molécula de ADN se representan de color rojo, azul y amarillo respectivamente. La zona que se indica mediante una elipse discontinua corresponde a densidad adicional no observada en estructuras previas de DNA-PKcs. Esta densidad podría asignarse al CTD de Ku80. (B) Alineamiento de las estructuras de tinción negativa correspondientes al complejo DNA-PKcs:Ku:ADN y Ku:ADN. Para realizar el alineamiento se utilizó como patrón la estructura atómica del núcleo de Ku (archivo PDB 1JEY) que previamente se había ajustado en ambos volúmenes. (C) Cálculo de la diferencia entre los volúmenes que se indican en (B). El volumen diferencia se representa de color azul. (D), (E) y (F) Vistas cenitales de (A), (B) y (C) respectivamente. En (E) y (F) se indica la densidad presente en el volumen resuelto para el complejo Ku:ADN asignada previamente al CTD de Ku:80.

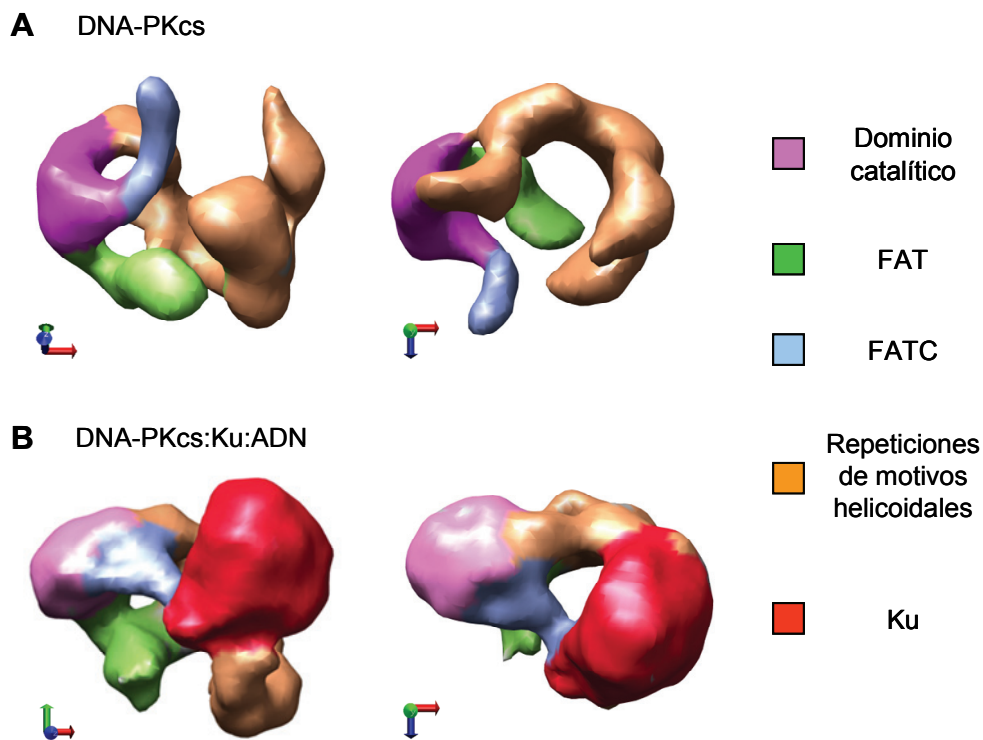
suficiente para impedir el reclutamiento de DNA-PKcs a los extremos del ADN (Falck et al., 2005), lo cual apoya que este dominio sea el responsable del reconocimiento inicial de DNA-PKcs. Su desplazamiento posterior permitiría además establecer contactos extensos entre ambas proteínas como los que se aprecian en las estructuras del complejo DNA-PKcs:Ku:ADN. Se ha descrito que el CTD de Ku80 puede romper el complejo DNA-PKcs:Artemis (Drouet et al., 2006). El modelo propuesto podría justificar esta observación. El reclutamiento de DNA-PKcs a los extremos del ADN implica el desplazamiento del CTD de Ku80 que competiría con Artemis por la unión a DNA-PKcs. Un mecanismo de ese tipo aseguraría que el complejo formado por DNA-PKcs y Ku sólo se ensamble en los sitios donde se produce la lesión.

#### **5.4.3 La formación del complejo holoenzima DNA-PK conlleva cambios conformacionales en DNA-PKcs**

El modelo topológico propuesto para DNA-PKcs puede extrapolarse a la estructura de criomicroscopía electrónica del complejo DNA-PKcs:Ku:ADN (*Figura 5.4.3*). Según esto, en la molécula de DNA-PKcs presente en el complejo se observan cambios conformacionales significativos en la región cabeza. Esos cambios afectan al dominio asignado a la porción más C-terminal del dominio FAT (*de color verde en la Figura 5.4.3*), si bien el cambio conformacional más llamativo se da en el dominio FATC (*de color azul en la Figura 5.4.3*). Este dominio está rotado respecto a su posición original en la estructura de criomicroscopía electrónica de DNA-PKcs. En esta nueva orientación el dominio FATC interacciona con la región palma y también con la densidad adicional asignada al heterodímero Ku (*Figura 5.4.3*). Al comparar las estructuras de criomicroscopía electrónica de DNA-PKcs y del complejo DNA-PKcs:Ku:ADN se aprecia que la región catalítica también modifica su conformación (*Figura 5.4.3*). Esta región presenta un aspecto más compacto como consecuencia de la interacción con el heterodímero Ku y el extremo de ADN. Como se comentó anteriormente los dominios FAT y FATC participarían en la regulación de la actividad quinasa de DNA-PKcs en respuesta a la interacción con otras macromoléculas. Por tanto los cambios conformacionales que se aprecian en la región catalítica podrían ser consecuencia de

movimientos que afecten a los dominios FAT y FATC. El dominio de DNA-PKcs que establece una interacción más íntima y extensa con la masa adicional asignada al heterodímero Ku corresponde a la garra proximal (*de color marrón en la Figura 5.4.3*). Curiosamente a partir de la comparación de estructuras resueltas para las PIKKs DNA-PKcs, TOR y ATM (véase el apartado 5.3) se ha propuesto que la garra proximal podría ser una característica específica de la proteína DNA-PKcs que estaría ausente en ATM y TOR.

La estructura resuelta para el complejo DNA-PKcs:Ku:ADN es totalmente compatible con los datos que actualmente se tienen del proceso de NHEJ (véase al apartado 1.4 de Introducción). Tras la aparición de DSB Ku recluta a DNA-PKcs a los extremos del ADN. A continuación Ku se desplaza hacia el interior de la molécula de ADN de manera que DNA-PKcs queda unida al extremo libre. Esta unión probablemente esté mediada por cambios



**Figura 5.4.3. Cambios conformacionales en DNA-PKcs asociados a la formación del ensamblado DNA-PKcs:Ku:ADN.** (A) y (B) Vistas equivalentes de las estructuras resueltas por criomicroscopía electrónica para la proteína DNA-PKcs y para el complejo DNA-PKcs:Ku:ADN respectivamente. Se han coloreado los dominios principales que componen la proteína DNA-PKcs. Comparando ambas estructuras es posible asignar los cambios conformacionales presentes en el complejo DNA-PKcs:Ku:ADN a dominios específicos.

conformacionales como los descritos anteriormente. El extremo de ADN debe tener una longitud mínima que permita acomodar ambas proteínas simultáneamente y que permita, además, inducir y estabilizar los cambios conformacionales que se observan en las estructuras correspondientes a los complejos DNA-PKcs:ADN y DNA-PKcs:Ku:ADN. El resultado final es la formación de un holoenzima DNA-PK funcional en el extremo de la rotura. A continuación DNA-PK participa en etapas subsiguientes del proceso de NHEJ fosforilando diversos sustratos y actuando, además, a modo de andamio donde se ensamblen componentes adicionales del proceso de reparación.

#### **5.4.4 Las características del complejo dimérico de DNA-PKcs:Ku:ADN son compatibles con las del complejo sináptico que interviene en el proceso de NHEJ**

La formación de dímeros de DNA-PKcs está descrita extensamente en la literatura. Esta capacidad podría ser una característica común de las PIKKs ya que se ha visto que ATM también forma dímeros (Bakkenist y Kastan, 2003). Estudios previos han sugerido la implicación directa de DNA-PKcs en favorecer la sinapsis de los extremos de ADN generados tras la rotura (DeFazio et al., 2002; Weterings et al., 2003; Weterings y van Gent, 2004). Esta sinapsis podría darse probablemente por asociación de las moléculas de DNA-PKcs unidas a los extremos de ADN.

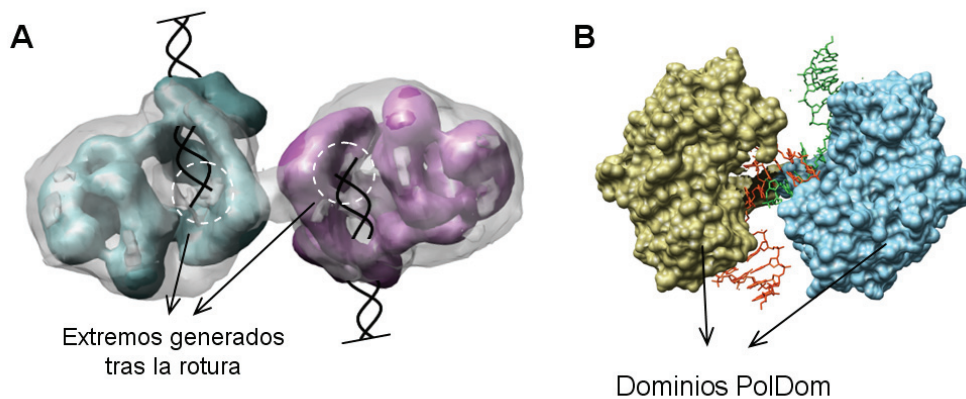
El análisis por microscopía electrónica evidencia la capacidad de DNA-PKcs de asociarse consigo misma y formar dímeros de manera independiente de Ku. El número de estos dímeros aumenta significativamente en presencia de Ku y ADN (véase el apartado 4.5.7 de Resultados). Esto sugiere una implicación directa de ambos componentes en el proceso de dimerización y por tanto en favorecer la sinapsis. La manera en que Ku ejerce esta función no es obvia. Ku se localiza hacia el exterior en cada uno de los complejos DNA-PKcs:Ku:ADN que forman el complejo dimérico, por tanto no forma parte de las superficies de contacto que se establecen entre los dos complejos DNA-PKcs:Ku:ADN. Quizá la unión de DNA-PKcs al complejo Ku:ADN podría inducir un cambio conformacional en la quinasa que se transmitiese a una región más distal.



DNA-PKcs contiene un dominio cremallera de leucina denominado LRR (del inglés “Leucine rich region”) en su región N-terminal (aminoácidos 1.503-1.602) (Gupta y Meek, 2005). Este tipo de dominios aparecen comúnmente en factores de transcripción que se unen al ADN como homodímeros. Mutantes puntuales en la región LRR de DNA-PKcs compensan parcialmente el fenotipo SCID característico de las células deficientes en DNA-PKcs (Gupta y Meek, 2005). Además alteraciones en la región LRR afectan a la capacidad de DNA-PKcs de unirse a ADN, sin embargo no afectan a su interacción con el heterodímero Ku ni a su actividad catalítica. El modelo topológico propuesto para DNA-PKcs predice que los primeros 2.000 aminoácidos de esta proteína se distribuyen entre la región palma y parte de la región brazo. Según este modelo el dominio LRR se localizaría en alguna zona de la región palma. De hecho los complejos DNA-PKcs:Ku:ADN que constituyen el ensamblado dímérico interaccionan por la cara externa de sus respectivas regiones palma. Quizá la interacción con Ku induzca un cambio conformacional en DNA-PKcs que, en último término, provoque el plegamiento correcto de la región LRR en un dominio cremallera de leucina funcional. La aparición de este dominio permitiría que el holoenzima DNA-PK interaccionase con el holoenzima presente en el extremo opuesto de la rotura de doble cadena en el ADN. Además la presencia de un dominio cremallera de leucina formado por dos holoenzimas DNA-PK podría ser clave para el correcto alineamiento de los extremos del ADN. Como se describe a continuación, la localización predicha para el extremo del fragmento de ADN presente en el complejo DNA-PKcs:Ku:ADN coincide con la ubicación propuesta para la región LRR. Un mecanismo de este tipo asegura que la formación del complejo sináptico únicamente se dé como consecuencia del ensamblaje del holoenzima DNA-PK en los extremos de la rotura.

En la estructura del complejo DNA-PKcs:Ku:ADN el heterodímero Ku se asienta sobre DNA-PKcs contactando principalmente con la región palma y también con la región cabeza. Como se mencionó anteriormente, el ajuste de la estructura atómica resuelta para Ku permite definir la trayectoria de la molécula de ADN dentro del complejo DNA-PKcs:Ku:ADN. A partir de este ajuste se aprecia que la molécula de ADN recorre el canal comprendido entre las

regiones cabeza y palma de DNA-PKcs (véase la Figura 4.5.5). Según esto, el extremo bloqueado de la molécula de ADN se dispone a continuación de Ku mientras que el extremo opuesto se localizaría próximo a la garra distal. Esta localización es totalmente compatible con los resultados obtenidos a partir del análisis estructural del complejo DNA-PKcs:ADN (véase la Figura 4.4.4). Al representar el complejo dimérico a un umbral correspondiente al 100% de la masa del ensamblado los dos complejos DNA-PKcs:Ku:ADN contactan entre sí (véase la Figura 4.5.9). Es posible por tanto que la densidad que conecta los dos complejos DNA-PKcs:Ku:ADN refleje la hibridación de las moléculas de ADN presentes en el complejo dimérico. Quizá la región palma de DNA-PKcs favorezca el establecimiento de apareamientos o microhomologías entre las moléculas de ADN presentes en el complejo dimérico. En cualquier caso en la estructura del ensamblado dimérico de DNA-PKcs:Ku:ADN los extremos de ADN se sitúan próximos en el espacio y, simultáneamente, quedan accesibles a otras enzimas debido al hueco comprendido entre las regiones palma enfrentadas (Figura 5.4.4A). Durante la reparación de DSB en el ADN los extremos generados han de mantenerse juntos evitándose así su difusión, sin embargo también han de estar accesibles a multitud de enzimas. Por tanto la



**Figura 5.4.4. Localización de las moléculas de ADN presentes en el complejo dimérico de DNA-PKcs:Ku:ADN.** (A) Representación de la trayectoria que podrían seguir las moléculas de ADN presentes en el complejo dimérico. Los extremos generados tras la rotura de doble cadena en el ADN quedarían accesibles y próximos en el espacio. (B) Estructura atómica del complejo sináptico formado los dominios polimerasa de LigD (archivo PDB 2R9L) (Brissett et al., 2007). Las dos moléculas de ADN equivalentes a los extremos de ADN que se generan tras la aparición de DSB hibridan en el centro de la estructura. Estas moléculas se muestran de color rojo y verde utilizando la representación de “bolas y palos”. Las superficies de los dominios PolDom que intervienen en la sinapsis se representan de color marrón y azul.

estructura resuelta para el complejo dimérico de DNA-PKcs:ADN es compatible con las características del complejo sináptico que aparece durante la reparación por NHEJ.

Recientemente se ha resuelto la estructura atómica del complejo sináptico entre dos extremos de ADN no complementarios mediado por dos dominios polimerasa de LigD (fichero PDB 2R9L) (*Figura 5.4.4B*) (Brissett et al., 2007). Como se comentó en el apartado 1.4 de Introducción, algunos procariontes cuentan con un sistema de NHEJ mínimo que consta de una proteína homóloga a Ku y de una ADN Ligasa multifunción (LigD). Es frecuente que esta ligasa contenga también dominios auxiliares polimerasa (PolDom) y nucleasa. En la estructura resuelta los dos dominios PolDom establecen contactos mínimos que permiten definir una cavidad central grande y accesible (*Figura 5.4.4B*). Los extremos de ADN se alojan en esta cavidad e hibridan entre sí estableciendo apareamientos que dirigen el alineamiento y la formación del complejo sináptico. Los dominios PolDom y los extremos de ADN interactúan de forma limitada, de manera que no se restringe la libertad conformacional de los extremos y se favorecen los apareamientos entre ellos. La estructura atómica resuelta para el complejo sináptico mediado por dominios PolDom es compatible con la estructura resuelta para el complejo dimérico de DNA-PKcs:Ku:ADN. En ésta los dos complejos DNA-PKcs:Ku:ADN se mantienen en una orientación definida gracias a contactos mínimos. Estos contactos se establecen entre las caras externas de las garras distales y permiten que la región comprendida entre los dos complejos DNA-PKcs:Ku:ADN quede totalmente accesible. Cerca de esa región se localizan los extremos no bloqueados de las moléculas de ADN presentes en el complejo dimérico. La estructura resuelta permite mantener los extremos de ADN próximos a la vez que se favorece el establecimiento de apareamientos entre ellos.

El complejo dimérico de DNA-PKcs:Ku:ADN podría actuar como plataforma de unión flexible sobre la cual se ensamblarían diversos complejos enzimáticos que, en último término, completarían la reparación de la rotura de doble cadena en el ADN. La organización del complejo dimérico podría variar en etapas posteriores del proceso de NHEJ en función de determinados eventos de

autofosforilación. En la estructura resuelta los dominios catalíticos de las dos moléculas de DNA-PKcs presentes en el complejo quedan orientados hacia el exterior. Esa configuración no parece ser la más idónea para que tenga lugar la autofosforilación en *trans* en las regiones ABCDE y PQR mencionadas anteriormente. Es posible por tanto que durante la reparación por NHEJ los complejos DNA-PKcs:Ku:ADN permanezcan asociados formando un complejo sináptico dinámico. Este complejo podría sufrir reorganizaciones estructurales dependientes de autofosforilación que favoreciesen el acceso a los extremos de ADN de otros componentes implicados en la reparación de NHEJ. Aún no está claro en qué etapa del proceso de reparación por NHEJ tiene lugar la autofosforilación de DNA-PKcs, aunque se sabe que disminuye su afinidad por los extremos de ADN (Merkle et al., 2002). La autofosforilación podría ser un requisito necesario para que tuviera lugar la disociación del complejo de los extremos del ADN. La estructura del complejo dimérico resuelta podría corresponder a una de estas configuraciones intermedias que aparecerían durante la reparación por NHEJ.

## 5.5 Reparación de DSB en el contexto de la cromatina

Los complejos proteína:ADN analizados durante el desarrollo de esta tesis se han obtenido utilizando pequeños fragmentos de ADN generados artificialmente. En las células sin embargo el ADN se encuentra asociado a nucleosomas, histonas externas y otras proteínas adicionales dando lugar a una estructura compacta denominada cromatina. Además las histonas juegan un papel importante en el contexto de la respuesta celular a DSB (van Attikum y Gasser, 2005). Concretamente la fosforilación de la histona H2AX es uno de los primeros eventos que tienen lugar en esta respuesta. La fosforilación es llevada a cabo por miembros de la familia de las PIKKs, y favorecería la acumulación y el mantenimiento de las distintas actividades enzimáticas implicadas en la reparación en el sitio donde se produjo la rotura (Celeste et al., 2003). Por tanto, en condiciones fisiológicas, es posible que los extremos generados tras la aparición de DSB presenten algún tipo de obstrucción que dificulte el acceso al ADN. Sin embargo *in vitro* el ensamblaje de complejos que

intervienen en el proceso de NHEJ requiere de extremos de ADN accesibles (Kysela et al., 2003).

El heterodímero Ku es la proteína encargada de reconocer los extremos generados tras la aparición de DSB (véase el apartado 1.4.1 de Introducción). Por tanto la capacidad de Ku de unirse a extremos que presenten algún tipo de bloqueo será determinante en la eficiencia del mecanismo de NHEJ. La estructura de la proteína Ku parece estar especialmente diseñada para reconocer extremos de ADN que no presenten obstrucciones (véase la Figura 1.4.2). Sin embargo se sabe que Ku interacciona con fragmentos de ADN unidos a la histona externa H1 provocando su desplazamiento (Roberts y Ramsden, 2007). La presencia de esta histona es suficiente para bloquear la actividad de la ADN Ligasa IV *in vitro* (Kysela et al., 2005). La capacidad de Ku de reconocer extremos próximos a la histona H1 permitiría a la maquinaria de NHEJ actuar en aquellos DSB que tengan lugar en regiones internucleosomales de la cromatina. Ku además reconoce extremos de ADN que aparezcan dentro del nucleosoma y que por tanto se disponen sobre la superficie de las histonas (Roberts y Ramsden, 2007). El mecanismo por el que el heterodímero Ku reconoce estos extremos es equivalente al descrito para extremos de ADN libres: enhebrando el ADN a través de su canal central. La unión de Ku a extremos de ADN asociados a los nucleosomas provoca el desprendimiento de la molécula de ADN de la superficie de las histonas (Roberts y Ramsden, 2007). Este desprendimiento puede afectar hasta un tercio de la longitud de la molécula de ADN que constituye el nucleosoma (50 pb aproximadamente). Este proceso, a diferencia de lo que sucede con otros procesos de remodelación de la cromatina, no conlleva consumo de energía ni tampoco la alteración de la estructura del nucleosoma. Por tanto el heterodímero Ku por sí solo tiene la capacidad de reconocer extremos de ADN en el contexto de la cromatina sin que tenga lugar una remodelación importante.

Las dimensiones de los fragmentos de ADN utilizados en el ensamblaje de los complejos Ku:ADN, DNA-PKcs:ADN y DNA-PKcs:Ku:ADN son compatibles con la longitud de la porción de ADN que el heterodímero Ku desprende de los nucleosomas. Según esto, los complejos observados así como los diferentes

cambios conformacionales y los modelos propuestos a partir de ellos podrían darse en condiciones fisiológicas en el contexto de la cromatina. Las etapas iniciales del proceso de NHEJ podrían tener lugar por tanto sin necesidad de remodelar la cromatina próxima al sitio donde se produjo la rotura. Esto contrasta con lo que sucede durante la reparación de DSB por recombinación homóloga (véase el apartado 1.2 de Introducción), que conlleva la retirada de nucleosomas en un área adyacente al sitio de rotura que abarca varias kb (Tsukuda et al., 2005). Esta diferencia podría explicar en parte por qué el proceso de NHEJ es el mecanismo preferido en los eucariotas superiores para reparar DSB. En células diferenciadas una reorganización significativa de la cromatina podría dar lugar a activaciones inapropiadas de determinados genes. El proceso de NHEJ es por tanto un mecanismo de urgencia que permite reparar DSB de forma sencilla y eficaz.

## 5.6 Otras proteínas adicionales son requeridas para completar la reparación del ADN

Como se ha mencionado anteriormente, las estructuras resueltas corresponden a complejos biológicos que participan en las etapas iniciales del proceso de NHEJ. Las características de los extremos generados tras la rotura requieren de la participación de actividades enzimáticas adicionales anteriores a la etapa final de ligación. Concretamente Artemis sería clave en el procesamiento de extremos de ADN que presenten configuraciones que impidan su ligación (Riballo et al., 2004; Kurosawa et al., 2008). El análisis bioquímico llevado a cabo con la muestra de Artemis purificada ha evidenciado la capacidad de oligomerizar de esta proteína (véase el apartado 4.6 de Resultados). En la estructura del complejo dimérico de DNA-PKcs:Ku:ADN los extremos del ADN se hallan próximos. Es posible que, durante el proceso de NHEJ, homodímeros de Artemis interaccionen con el complejo dimérico de DNA-PKcs:Ku:ADN y lleven a cabo el procesamiento simultáneo de los dos extremos generados tras la rotura. Existen otras nucleasas de la familia de las metalo- $\beta$ -lactamasas que realizan su función como homodímeros (Li de la Sierra-Gallay et al., 2006). Artemis puede actuar como endonucleasa y

exonucleasa (Moshous et al., 2001; Ma et al., 2002). Su actividad supone por tanto un riesgo intrínseco y por ello debe estar estrictamente regulada. Quizá la actividad de esta nucleasa esté regulada doblemente. Un primer punto de control podría requerir de la interacción con DNA-PKcs. Además la activación completa de Artemis dependería de que la proteína DNA-PKcs con la que interacciona se encuentre formando parte del complejo sináptico. Un mecanismo de regulación con estas características asegura que Artemis sólo esté activa durante un período concreto del proceso de NHEJ. En cualquier caso la caracterización bioquímica realizada ha demostrado la capacidad de DNA-PKcs y Artemis de formar un complejo estable en ausencia de extremos de ADN. Esta caracterización no permitió discriminar si el complejo DNA-PKcs:Artemis contiene una o dos moléculas de Artemis. Seguramente el análisis futuro de esta muestra por microscopía electrónica permitirá aclarar esta ambigüedad.

Junto con Artemis otras proteínas adicionales participan en el procesamiento de los extremos (véase el apartado 1.4.6 de Introducción). Concretamente la proteína PNK permitiría restaurar los extremos a la configuración 3' OH y 5' P necesaria para que tenga lugar la etapa final de ligación. Además intervienen polimerasas de la familia pol X (pol  $\lambda$  y pol  $\mu$ ) rellenando los huecos que quedan tras establecerse microhomologías entre los extremos de la rotura. Finalmente tendría lugar la etapa de ligación de los extremos. El complejo XRCC4:ADN Ligasa IV es el encargado de llevar a cabo esta etapa. Se sabe que la actividad enzimática de este complejo se estimula gracias a interacciones directas proteína-proteína con el holoenzima DNA-PKcs (Chen et al., 2000; Nick McElhinny et al., 2000).

La reparación por NHEJ conlleva por tanto la formación de ensamblados proteicos de complejidad variable. Estos ensamblados son dinámicos desde un punto de vista funcional y también en lo que respecta a su composición proteica, y llevan a cabo su actividad de manera coordinada. Caracterizar a nivel tridimensional los complejos implicados así como las interacciones que establecen serán claves para entender el proceso de NHEJ. Estos complejos sin embargo son escasos en la célula puesto que aparecen únicamente

durante la reparación de DSB, por ello reconstituirlos *in vitro* parece ser la mejor aproximación para poder llevar a cabo su caracterización estructural futura. Además existen también interrogantes de carácter bioquímico que han de resolverse en los próximos años para poder precisar cómo se da la reparación de DSB mediante NHEJ. Concretamente, será muy interesante definir el efecto de las fosforilaciones de DNA-PKcs sobre sí misma y también en el resto de componentes del proceso de NHEJ. Quizá DNA-PKcs sea el factor encargado de llevar a cabo la regulación local, en el sitio donde se produce la rotura, del proceso de NHEJ. Este proceso de reparación debe estar integrado en los mecanismos de respuesta globales que las células inician como consecuencia de la aparición de DSB (Boehme et al., 2008). Comprender cómo se da esa integración así como las relaciones entre los procesos de reparación NHEJ y HR durante las diversas etapas del ciclo celular se plantean como interrogantes interesantes a resolver en el futuro.

## 5.7 Conclusión

Entender los procesos biológicos que definen a los seres vivos requiere de su caracterización tridimensional. Durante el desarrollo de esta tesis se ha llevado a cabo el estudio estructural de ensamblados proteicos biológicos que participan en el proceso de reparación de DSB mediante NHEJ. Este mecanismo de reparación es clave en el mantenimiento de la estabilidad genómica en organismos superiores. Alteraciones en los componentes que intervienen en este proceso generan fenotipos SCID y provocan un aumento de la inestabilidad genómica favoreciendo, en último término, el desarrollo de fenotipos tumorales. Por ello la caracterización del proceso de NHEJ es, además, tremendamente interesante desde un punto de vista terapéutico. Para realizar el análisis estructural se ha utilizado la técnica consistente en la combinación de microscopía electrónica y procesamiento digital de imágenes individuales. Este análisis se ha llevado a cabo aumentando la complejidad de forma progresiva; primero se han analizado los especímenes libres y luego, una vez caracterizados los componentes individuales, se han estudiado los complejos en los que aparecen. El análisis de cada espécimen también se ha



realizado aumentando la complejidad progresivamente; primero mediante tinción negativa y luego por criomicroscopía electrónica. Los diversos especímenes estudiados han supuesto un reto importante que ha implicado el desarrollo de nuevas estrategias específicas. Concretamente se han caracterizado especímenes que, a excepción del complejo dimérico de DNA-PKcs:Ku:ADN, en ningún caso presentaban simetría. Además se han estudiado especímenes de bajo peso molecular (Ku) y mezclas heterogéneas de diversos especímenes (DNA-PKcs y DNA-PKcs:ADN). También, con el objetivo de evitar el sesgo introducido por el volumen inicial, se han desarrollado estrategias de procesamiento que actualmente se emplean de manera rutinaria en nuestro laboratorio durante los distintos procesos de refinamiento. La técnica estructural empleada ha demostrado ser una herramienta útil para avanzar en el conocimiento estructural del sistema de NHEJ permitiendo, en último término, proponer modelos que expliquen aspectos concretos de algunas de las etapas iniciales de este mecanismo de reparación. Sin embargo las estructuras resueltas presentan una resolución baja que impide hacer descripciones precisas de los diversos especímenes analizados. Por ello la combinación de información obtenida a partir de diferentes técnicas estructurales ha sido y será clave para avanzar en el conocimiento estructural de los distintos procesos biológicos.





## 6. Conclusiones



1. Se ha propuesto un modelo de organización tridimensional para el heterodímero Ku. Este modelo detalla los cambios conformacionales que tienen lugar durante la unión a los extremos del ADN y el posterior reclutamiento de DNA-PKcs.
2. La estructura de DNA-PKcs consiste en tres regiones claramente diferenciadas denominadas cabeza, brazo y palma. La región cabeza está constituida por los dominios conservados FAT, catalítico y FATC, mientras que las regiones brazo y palma comprenden las zonas ricas en repeticiones de motivos helicoidales N-terminales. Este esquema de organización tridimensional es común para el resto de miembros que componen la familia de las quinasas PIKKs.
3. La unión de DNA-PKcs al ADN conlleva cambios conformacionales importantes que regulan la actividad catalítica de DNA-PKcs.
4. Condiciones que favorecen la autofosforilación del complejo DNA-PKcs:Ku:ADN provocan su desensamblaje. Las estructuras resueltas sugieren que los distintos eventos de fosforilación tendrán lugar principalmente en *trans*.
5. Se ha descubierto la existencia de un complejo dimérico de DNA-PKcs:Ku:ADN que podría corresponder al complejo sináptico que interviene en el proceso de NHEJ.
6. El análisis estructural de DNA-PKcs y Ku así como de sus distintos complejos, ha permitido proponer un modelo que explica a nivel tridimensional las etapas iniciales del proceso de reparación por NHEJ.
7. Artemis tiene capacidad de formar oligómeros *in vitro*. En las condiciones ensayadas es posible ensamblar un complejo DNA-PKcs:Artemis mayoritario y homogéneo.
8. Desde un punto de vista metodológico, utilizar como volúmenes referencia ruido o elipsoides artificiales es una estrategia válida para disminuir el sesgo durante los procesos de refinamiento angular iterativo.





## 7. Bibliografía





- ❖ Abbott, A. (2002). The society of proteins. *Nature* **417**, 894-896.
- ❖ Abraham, R. T. (2004). PI 3-kinase related kinases: 'big' players in stress-induced signaling pathways. *DNA Repair (Amst)* **3**, 883-887.
- ❖ Adami, A., Garcia-Alvarez, B., Arias-Palomo, E., Barford, D., y Llorca, O. (2007). Structure of TOR and its complex with KOG1. *Mol Cell* **27**, 509-516.
- ❖ Adrian, M., Dubochet, J., Lepault, J., y McDowell, A. W. (1984). Cryo-electron microscopy of viruses. *Nature* **308**, 32-36.
- ❖ Ahnesorg, P., Smith, P., y Jackson, S. P. (2006). XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell* **124**, 301-313.
- ❖ Andres, S. N., Modesti, M., Tsai, C. J., Chu, G., y Junop, M. S. (2007). Crystal structure of human XLF: a twist in nonhomologous DNA end-joining. *Mol Cell* **28**, 1093-1101.
- ❖ Aravind, L., y Koonin, E. V. (2000). SAP - a putative DNA-binding motif involved in chromosomal organization. *Trends Biochem Sci* **25**, 112-114.
- ❖ Aravind, L., y Koonin, E. V. (2001). Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res* **11**, 1365-1374.
- ❖ Arosio, D., Cui, S., Ortega, C., Chovanec, M., Di Marco, S., Baldini, G., Falaschi, A., y Vindigni, A. (2002). Studies on the mode of Ku interaction with DNA. *J Biol Chem* **277**, 9741-9748.
- ❖ Bailey, S. M., Meyne, J., Chen, D. J., Kurimasa, A., Li, G. C., Lehnert, B. E., y Goodwin, E. H. (1999). DNA double-strand break repair proteins are required to cap the ends of mammalian chromosomes. *Proc Natl Acad Sci U S A* **96**, 14899-14904.
- ❖ Bakkenist, C. J., y Kastan, M. B. (2003). DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature* **421**, 499-506.
- ❖ Batista, D. A., Hawkins, A., Murphy, K. M., y Griffin, C. A. (2005). BCR/ABL rearrangement in two cases of Philadelphia chromosome negative chronic myeloid leukemia: deletion on the derivative chromosome 9 may or not be present. *Cancer Genet Cytogenet* **163**, 164-167.
- ❖ Beamish, H. J., Jessberger, R., Riballo, E., Priestley, A., Blunt, T., Kysela, B., y Jeggo, P. A. (2000). The C-terminal conserved domain of DNA-PKcs, missing in the SCID mouse, is required for kinase activity. *Nucleic Acids Res* **28**, 1506-1513.
- ❖ Bennett-Lovsey, R. M., Herbert, A. D., Sternberg, M. J., y Kelley, L. A. (2008). Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins* **70**, 611-625.
- ❖ Bernstein, N. K., Williams, R. S., Rakovszky, M. L., Cui, D., Green, R., Karimi-Busheri, F., Mani, R. S., Galicia, S., Koch, C. A., Cass, C. E., *et al.* (2005). The molecular architecture of the mammalian DNA repair enzyme, polynucleotide kinase. *Mol Cell* **17**, 657-670.

- ❖ Block, W. D., Yu, Y., Merkle, D., Gifford, J. L., Ding, Q., Meek, K., y Lees-Miller, S. P. (2004). Autophosphorylation-dependent remodeling of the DNA-dependent protein kinase catalytic subunit regulates ligation of DNA ends. *Nucleic Acids Res* **32**, 4351-4357.
- ❖ Boehme, K. A., Kulikov, R., y Blattner, C. (2008). p53 stabilization in response to DNA damage requires Akt/PKB and DNA-PK. *Proc Natl Acad Sci U S A*.
- ❖ Boguski, M. S., Murray, A. W., y Powers, S. (1992). Novel repetitive sequence motifs in the alpha and beta subunits of prenyl-protein transferases and homology of the alpha subunit to the MAD2 gene product of yeast. *New Biol* **4**, 408-411.
- ❖ Boisset, N., Penczek, P. A., Taveau, J. C., You, V., de Haas, F., y Lamy, J. (1998). Overabundant single-particle electron microscope views induce a three-dimensional reconstruction artifact. *Ultramicroscopy* **74** 201-207.
- ❖ Boskovic, J., Rivera-Calzada, A., Maman, J. D., Chacon, P., Willison, K. R., Pearl, L. H., y Llorca, O. (2003). Visualization of DNA-induced conformational changes in the DNA repair kinase DNA-PKcs. *Embo J* **22**, 5875-5882.
- ❖ Bosma, G. C., Custer, R. P., y Bosma, M. J. (1983). A severe combined immunodeficiency mutation in the mouse. *Nature* **301**, 527-530.
- ❖ Bosotti, R., Isacchi, A., y Sonnhammer, E. L. (2000). FAT: a novel domain in PIK-related kinases. *Trends Biochem Sci* **25**, 225-227.
- ❖ Bottcher, B., Wynne, S. A., y Crowther, R. A. (1997). Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature* **386**, 88-91.
- ❖ Bozzola, J. J. R., D. L. (1992). *Electron Microscopy: Principles and Techniques for Biologists*.
- ❖ Brewerton, S. C., Dore, A. S., Drake, A. C., Leuther, K. K., y Blundell, T. L. (2004). Structural analysis of DNA-PKcs: modelling of the repeat units and insights into the detailed molecular architecture. *J Struct Biol* **145**, 295-306.
- ❖ Brissett, N. C., Pitcher, R. S., Juarez, R., Picher, A. J., Green, A. J., Dafforn, T. R., Fox, G. C., Blanco, L., y Doherty, A. J. (2007). Structure of a NHEJ polymerase-mediated DNA synaptic complex. *Science* **318**, 456-459.
- ❖ Buck, D., Malivert, L., de Chasseval, R., Barraud, A., Fondaneche, M. C., Sanal, O., Plebani, A., Stephan, J. L., Hufnagel, M., le Deist, F., *et al.* (2006). Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell* **124**, 287-299.
- ❖ Caldecott, K. W. (2007). Mammalian single-strand break repair: mechanisms and links with chromatin. *DNA Repair (Amst)* **6**, 443-453.
- ❖ Callebaut, I., Moshous, D., Mornon, J. P., y de Villartay, J. P. (2002). Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. *Nucleic Acids Res* **30**, 3592-3601.
- ❖ Cary, R. B., Peterson, S. R., Wang, J., Bear, D. G., Bradbury, E. M., y Chen, D. J. (1997). DNA looping by Ku and the DNA-dependent protein kinase. *Proc Natl Acad Sci U S A* **94**, 4267-4272.

- ❖ Celeste, A., Fernandez-Capetillo, O., Kruhlak, M. J., Pilch, D. R., Staudt, D. W., Lee, A., Bonner, R. F., Bonner, W. M., y Nussenzweig, A. (2003). Histone H2AX phosphorylation is dispensable for the initial recognition of DNA breaks. *Nat Cell Biol* **5**, 675-679.
- ❖ Chacon, P., y Wriggers, W. (2002). Multi-resolution contour-based fitting of macromolecular structures. *J Mol Biol* **317**, 375-384.
- ❖ Chacon, P., Tama, F., y Wriggers, W. (2003). Mega-Dalton biomolecular motion captured from electron microscopy reconstructions. *J Mol Biol* **326**, 485-492.
- ❖ Chai, W., Ford, L. P., Lenertz, L., Wright, W. E., y Shay, J. W. (2002). Human Ku70/80 associates physically with telomerase through interaction with hTERT. *J Biol Chem* **277**, 47242-47247.
- ❖ Chappell, C., Hanakahi, L. A., Karimi-Busheri, F., Weinfeld, M., y West, S. C. (2002). Involvement of human polynucleotide kinase in double-strand break repair by non-homologous end joining. *Embo J* **21**, 2827-2832.
- ❖ Chen, L., Trujillo, K., Sung, P., y Tomkinson, A. E. (2000). Interactions of the DNA ligase IV-XRCC4 complex with DNA ends and the DNA-dependent protein kinase. *J Biol Chem* **275**, 26196-26205.
- ❖ Chen, D. H., Song, J. L., Chuang, D. T., Chiu, W., y Ludtke, S. J. (2006). An expanded conformation of single-ring GroEL-GroES complex encapsulates an 86 kDa substrate. *Structure* **14**, 1711-1722.
- ❖ Chiu, C. Y., Cary, R. B., Chen, D. J., Peterson, S. R., y Stewart, P. L. (1998). Cryo-EM imaging of the catalytic subunit of the DNA-dependent protein kinase. *J Mol Biol* **284**, 1075-1081.
- ❖ Chiu, W., Baker, M. L., Jiang, W., Dougherty, M., y Schmid, M. F. (2005). Electron cryomicroscopy of biological machines at subnanometer resolution. *Structure* **13**, 363-372.
- ❖ Chu, G. (1997). Double strand break repair. *J Biol Chem* **272**, 24097-24100.
- ❖ Cline, S. D., y Hanawalt, P. C. (2003). Who's on first in the cellular response to DNA damage? *Nat Rev Mol Cell Biol* **4**, 361-372.
- ❖ Collis, S. J., DeWeese, T. L., Jeggo, P. A., y Parker, A. R. (2005). The life and death of DNA-PK. *Oncogene* **24**, 949-961.
- ❖ Cromie, G. A., Connelly, J. C., y Leach, D. R. (2001). Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. *Mol Cell* **8**, 1163-1174.
- ❖ Crowther, R. A., Amos, L. A., Finch, J. T., De Rosier, D. J., y Klug, A. (1970). Three dimensional reconstructions of spherical viruses by fourier synthesis from electron micrographs. *Nature* **226**, 421-425.
- ❖ Cui, X., Yu, Y., Gupta, S., Cho, Y. M., Lees-Miller, S. P., y Meek, K. (2005). Autophosphorylation of DNA-dependent protein kinase regulates DNA end processing and may also alter double-strand break repair pathway choice. *Mol Cell Biol* **25**, 10842-10852.

- ❖ Dames, S. A., Mulet, J. M., Rathgeb-Szabo, K., Hall, M. N., y Grzesiek, S. (2005). The solution structure of the FATC domain of the protein kinase target of rapamycin suggests a role for redox-dependent structural and cellular stability. *J Biol Chem* **280**, 20558-20564.
- ❖ de Vries, E., van Driel, W., Bergsma, W. G., Arnberg, A. C., y van der Vliet, P. C. (1989). HeLa nuclear protein recognizing DNA termini and translocating on DNA forming a regular DNA-multimeric protein complex. *J Mol Biol* **208**, 65-78.
- ❖ DeFazio, L. G., Stansel, R. M., Griffith, J. D., y Chu, G. (2002). Synapsis of DNA ends by DNA-dependent protein kinase. *Embo J* **21**, 3192-3200.
- ❖ Dempster, A. P., Laird, N. M., y Rubin, D. B. (1977). Maximum-likelihood from incomplete data via the EM algorithm. *J Roy Stat Soc ser B* **39**, 1-38.
- ❖ Difilippantonio, M. J., Zhu, J., Chen, H. T., Meffre, E., Nussenzweig, M. C., Max, E. E., Ried, T., y Nussenzweig, A. (2000). DNA repair protein Ku80 suppresses chromosomal aberrations and malignant transformation. *Nature* **404**, 510-514.
- ❖ Djordjevic, S., y Driscoll, P. C. (2002). Structural insight into substrate specificity and regulatory mechanisms of phosphoinositide 3-kinases. *Trends Biochem Sci* **27**, 426-432.
- ❖ Doherty, A. J., y Jackson, S. P. (2001). DNA repair: how Ku makes ends meet. *Curr Biol* **11**, R920-924.
- ❖ Dore, A. S., Furnham, N., Davies, O. R., Sibanda, B. L., Chirgadze, D. Y., Jackson, S. P., Pellegrini, L., y Blundell, T. L. (2006). Structure of an Xrcc4-DNA ligase IV yeast ortholog complex reveals a novel BRCT interaction mode. *DNA Repair (Amst)* **5**, 362-368.
- ❖ Douglas, P., Moorhead, G. B., Ye, R., y Lees-Miller, S. P. (2001). Protein phosphatases regulate DNA-dependent protein kinase activity. *J Biol Chem* **276**, 18992-18998.
- ❖ Douglas, P., Cui, X., Block, W. D., Yu, Y., Gupta, S., Ding, Q., Ye, R., Morrice, N., Lees-Miller, S. P., y Meek, K. (2007). The DNA-dependent protein kinase catalytic subunit is phosphorylated in vivo on threonine 3950, a highly conserved amino acid in the protein kinase domain. *Mol Cell Biol* **27**, 1581-1591.
- ❖ Downs, J. A., y Jackson, S. P. (2004). A means to a DNA end: the many roles of Ku. *Nat Rev Mol Cell Biol* **5**, 367-378.
- ❖ Dresser, M. E. (2000). Meiotic chromosome behavior in *Saccharomyces cerevisiae* and (mostly) mammals. *Mutat Res* **451**, 107-127.
- ❖ Drouet, J., Frit, P., Delteil, C., de Villartay, J. P., Salles, B., y Calsou, P. (2006). Interplay between Ku, Artemis, and the DNA-dependent protein kinase catalytic subunit at DNA ends. *J Biol Chem* **281**, 27784-27793.
- ❖ Dubochet, J., Adrian, M., Chang, J. J., Homo, J. C., Lepault, J., McDowell, A. W., y Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys* **21**, 129-228.

- ❖ Dvir, A., Peterson, S. R., Knuth, M. W., Lu, H., y Dynan, W. S. (1992). Ku autoantigen is the regulatory component of a template-associated protein kinase that phosphorylates RNA polymerase II. *Proc Natl Acad Sci U S A* **89**, 11920-11924.
- ❖ Espejel, S., Franco, S., Rodriguez-Perales, S., Bouffler, S. D., Cigudosa, J. C., y Blasco, M. A. (2002a). Mammalian Ku86 mediates chromosomal fusions and apoptosis caused by critically short telomeres. *Embo J* **21**, 2207-2219.
- ❖ Espejel, S., Franco, S., Sgura, A., Gae, D., Bailey, S. M., Taccioli, G. E., y Blasco, M. A. (2002b). Functional interaction between DNA-PKcs and telomerase in telomere length maintenance. *Embo J* **21**, 6275-6287.
- ❖ Espejel, S., Martin, M., Klatt, P., Martin-Caballero, J., Flores, J. M., y Blasco, M. A. (2004). Shorter telomeres, accelerated ageing and increased lymphoma in DNA-PKcs-deficient mice. *EMBO Rep* **5**, 503-509.
- ❖ Falck, J., Coates, J., y Jackson, S. P. (2005). Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. *Nature* **434**, 605-611.
- ❖ Featherstone, C., y Jackson, S. P. (1999). Ku, a DNA repair protein with multiple cellular functions? *Mutat Res* **434**, 3-15.
- ❖ Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., y Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* **116**, 190-199.
- ❖ Frank, J. (2006). Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State, 2 edn (USA).
- ❖ Frank, K. M., Sekiguchi, J. M., Seidl, K. J., Swat, W., Rathbun, G. A., Cheng, H. L., Davidson, L., Kangaloo, L., y Alt, F. W. (1998). Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV. *Nature* **396**, 173-177.
- ❖ Gao, Y., Chaudhuri, J., Zhu, C., Davidson, L., Weaver, D. T., y Alt, F. W. (1998a). A targeted DNA-PKcs-null mutation reveals DNA-PK-independent functions for KU in V(D)J recombination. *Immunity* **9**, 367-376.
- ❖ Gao, Y., Sun, Y., Frank, K. M., Dikkes, P., Fujiwara, Y., Seidl, K. J., Sekiguchi, J. M., Rathbun, G. A., Swat, W., Wang, J., *et al.* (1998b). A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell* **95**, 891-902.
- ❖ Gao, Y., Ferguson, D. O., Xie, W., Manis, J. P., Sekiguchi, J., Frank, K. M., Chaudhuri, J., Horner, J., DePinho, R. A., y Alt, F. W. (2000). Interplay of p53 and DNA-repair protein XRCC4 in tumorigenesis, genomic stability and development. *Nature* **404**, 897-900.
- ❖ Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., *et al.* (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141-147.
- ❖ Gell, D., y Jackson, S. P. (1999). Mapping of protein-protein interactions within the DNA-dependent protein kinase complex. *Nucleic Acids Res* **27**, 3494-3502.

- ❖ Gellert, M. (2002). V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem* **71**, 101-132.
- ❖ Goodarzi, A. A., Yu, Y., Riballo, E., Douglas, P., Walker, S. A., Ye, R., Harer, C., Marchetti, C., Morrice, N., Jeggo, P. A., y Lees-Miller, S. P. (2006). DNA-PK autophosphorylation facilitates Artemis endonuclease activity. *Embo J* **25**, 3880-3889.
- ❖ Gottlieb, T. M., y Jackson, S. P. (1993). The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell* **72**, 131-142.
- ❖ Grant, P. A., Schieltz, D., Pray-Grant, M. G., Yates, J. R., 3rd, y Workman, J. L. (1998). The ATM-related cofactor Tra1 is a component of the purified SAGA complex. *Mol Cell* **2**, 863-867.
- ❖ Grawunder, U., Wilm, M., Wu, X., Kulesza, P., Wilson, T. E., Mann, M., y Lieber, M. R. (1997). Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature* **388**, 492-495.
- ❖ Green, M. R. (2005). Eukaryotic transcription activation: right on target. *Mol Cell* **18**, 399-402.
- ❖ Grigorieff, N. (2000). Resolution measurement in structures derived from single particles. *Acta Crystallogr D Biol Crystallogr* **56**, 1270-1277.
- ❖ Grob, P., Cruse, M. J., Inouye, C., Peris, M., Penczek, P. A., Tjian, R., y Nogales, E. (2006). Cryo-electron microscopy studies of human TFIIID: conformational breathing in the integration of gene regulatory cues. *Structure* **14**, 511-520.
- ❖ Groves, M. R., y Barford, D. (1999). Topological characteristics of helical repeat proteins. *Curr Opin Struct Biol* **9**, 383-389.
- ❖ Groves, M. R., Hanlon, N., Turowski, P., Hemmings, B. A., y Barford, D. (1999). The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs. *Cell* **96**, 99-110.
- ❖ Gu, J., Lu, H., Tippin, B., Shimazaki, N., Goodman, M. F., y Lieber, M. R. (2007). XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. *Embo J* **26**, 3506-3507.
- ❖ Gupta, S., y Meek, K. (2005). The leucine rich region of DNA-PKcs contributes to its innate DNA affinity. *Nucleic Acids Res* **33**, 6972-6981.
- ❖ Hammarsten, O., y Chu, G. (1998). DNA-dependent protein kinase: DNA binding and activation in the absence of Ku. *Proc Natl Acad Sci U S A* **95**, 525-530.
- ❖ Hammarsten, O., DeFazio, L. G., y Chu, G. (2000). Activation of DNA-dependent protein kinase by single-stranded DNA ends. *J Biol Chem* **275**, 1541-1550.
- ❖ Harauz, G., y Ottensmeyer, F. P. (1984). Nucleosome reconstruction via phosphorus mapping. *Science* **226**, 936-940.
- ❖ Harauz, G., y van Heel, M. (1986). Exact filters for general geometry three-dimensional reconstruction. *Optik* **73**, 146-156.

- ❖ Hardy, E., y Castellanos-Serra, L. R. (2004). "Reverse-staining" of biomolecules in electrophoresis gels: analytical and micropreparative applications. *Anal Biochem* **328**, 1-13.
- ❖ Harris, R., Esposito, D., Sankar, A., Maman, J. D., Hinks, J. A., Pearl, L. H., y Driscoll, P. C. (2004). The 3D solution structure of the C-terminal region of Ku86 (Ku86CTR). *J Mol Biol* **335**, 573-582.
- ❖ Hartley, K. O., Gell, D., Smith, G. C., Zhang, H., Divecha, N., Connelly, M. A., Admon, A., Lees-Miller, S. P., Anderson, C. W., y Jackson, S. P. (1995). DNA-dependent protein kinase catalytic subunit: a relative of phosphatidylinositol 3-kinase and the ataxia telangiectasia gene product. *Cell* **82**, 849-856.
- ❖ Helleday, T., Lo, J., van Gent, D. C., y Engelward, B. P. (2007). DNA double-strand break repair: from mechanistic understanding to cancer treatment. *DNA Repair (Amst)* **6**, 923-935.
- ❖ Henderson, R. (1995). The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q Rev Biophys* **28**, 171-193.
- ❖ Hentges, P., Ahnesorg, P., Pitcher, R. S., Bruce, C. K., Kysela, B., Green, A. J., Bianchi, J., Wilson, T. E., Jackson, S. P., y Doherty, A. J. (2006). Evolutionary and functional conservation of the DNA non-homologous end-joining protein, XLF/Cernunnos. *J Biol Chem* **281**, 37517-37526.
- ❖ Heymann, J. B. (2001). Bsoft: image and molecular processing in electron microscopy. *J Struct Biol* **133**, 156-169.
- ❖ Heymann, J. B., y Belnap, D. M. (2007). Bsoft: image processing and molecular modeling for electron microscopy. *J Struct Biol* **157**, 3-18.
- ❖ Hickson, I. D. (2003). RecQ helicases: caretakers of the genome. *Nat Rev Cancer* **3**, 169-178.
- ❖ Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., *et al.* (2002). Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180-183.
- ❖ Hoeijmakers, J. H. (2001). Genome maintenance mechanisms for preventing cancer. *Nature* **411**, 366-374.
- ❖ Hsieh, P. (2001). Molecular mechanisms of DNA mismatch repair. *Mutat Res* **486**, 71-87.
- ❖ Huang, C. H., Mandelker, D., Schmidt-Kittler, O., Samuels, Y., Velculescu, V. E., Kinzler, K. W., Vogelstein, B., Gabeli, S. B., y Amzel, L. M. (2007). The structure of a human p110alpha/p85alpha complex elucidates the effects of oncogenic PI3Kalpha mutations. *Science* **318**, 1744-1748.
- ❖ Humphrey, W., Dalke, A., y Schulten, K. (1996). VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-38, 27-38.

- ❖ Huse, M., y Kuriyan, J. (2002). The conformational plasticity of protein kinases. *Cell* **109**, 275-282.
- ❖ Ishikawa, H., Nakagawa, N., Kuramitsu, S., y Masui, R. (2006). Crystal structure of TTHA0252 from *Thermus thermophilus* HB8, a RNA degradation protein of the metallo-beta-lactamase superfamily. *J Biochem* **140**, 535-542.
- ❖ Itoh, T., y Horio, T. (2001). DNA-dependent protein kinase catalytic subunit is cleaved during UV-induced apoptosis. *J Dermatol Sci* **25**, 72-77.
- ❖ Jackson, S. P. (2002). Sensing and repairing DNA double-strand breaks. *Carcinogenesis* **23**, 687-696.
- ❖ Jeggo, P. A. (1990). Studies on mammalian mutants defective in rejoining double-strand breaks in DNA. *Mutat Res* **239**, 1-16.
- ❖ Jiang, W., y Ludtke, S. J. (2005). Electron cryomicroscopy of single particles at subnanometer resolution. *Curr Opin Struct Biol* **15**, 571-577.
- ❖ Jin, S., Kharbanda, S., Mayer, B., Kufe, D., y Weaver, D. T. (1997). Binding of Ku and c-Abl at the kinase homology region of DNA-dependent protein kinase catalytic subunit. *J Biol Chem* **272**, 24763-24766.
- ❖ Jones, T. A., Zou, J. Y., Cowan, S. W., y Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* **47 ( Pt 2)**, 110-119.
- ❖ Junop, M. S., Modesti, M., Guarne, A., Ghirlando, R., Gellert, M., y Yang, W. (2000). Crystal structure of the Xrcc4 DNA repair protein and implications for end joining. *Embo J* **19**, 5962-5970.
- ❖ Keith, C. T., y Schreiber, S. L. (1995). PIK-related kinases: DNA repair, recombination, and cell cycle checkpoints. *Science* **270**, 50-51.
- ❖ Kelley, L. A., MacCallum, R. M., y Sternberg, M. J. (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J Mol Biol* **299**, 499-520.
- ❖ Khanna, K. K., y Jackson, S. P. (2001). DNA double-strand breaks: signaling, repair and the cancer connection. *Nat Genet* **27**, 247-254.
- ❖ Kim, C. H., Park, S. J., y Lee, S. H. (2002). A targeted inhibition of DNA-dependent protein kinase sensitizes breast cancer cells following ionizing radiation. *J Pharmacol Exp Ther* **303**, 753-759.
- ❖ Kohonen, T. (1990). The self-organizing map. *Proc IEEE* **78**, 1464-1480.
- ❖ Kurosawa, A., Koyama, H., Takayama, S., Miki, K., Ayusawa, D., Fujii, M., Iizumi, S., y Adachi, N. (2007). The Requirement of Artemis in Double-Strand Break Repair Depends on the Type of DNA Damage. *DNA Cell Biol*.
- ❖ Kuzminov, A. (1995). Collapse and repair of replication forks in *Escherichia coli*. *Mol Microbiol* **16**, 373-384.



- ❖ Kysela, B., Doherty, A. J., Chovanec, M., Stiff, T., Ameer-Beg, S. M., Vojnovic, B., Girard, P. M., y Jeggo, P. A. (2003). Ku stimulation of DNA ligase IV-dependent ligation requires inward movement along the DNA molecule. *J Biol Chem* **278**, 22466-22474.
- ❖ Kysela, B., Chovanec, M., y Jeggo, P. A. (2005). Phosphorylation of linker histones by DNA-dependent protein kinase is required for DNA ligase IV-dependent ligation in the presence of histone H1. *Proc Natl Acad Sci U S A* **102**, 1877-1882.
- ❖ Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.
- ❖ Lavin, M. F., Khanna, K. K., Beamish, H., Spring, K., Watters, D., y Shiloh, Y. (1995). Relationship of the ataxia-telangiectasia protein ATM to phosphoinositide 3-kinase. *Trends Biochem Sci* **20**, 382-383.
- ❖ Lee, G. E., Yu, E. Y., Cho, C. H., Lee, J., Muller, M. T., y Chung, I. K. (2004). DNA-protein kinase catalytic subunit-interacting protein KIP binds telomerase by interacting with human telomerase reverse transcriptase. *J Biol Chem* **279**, 34750-34755.
- ❖ Lee, S. J., Sekimoto, T., Yamashita, E., Nagoshi, E., Nakagawa, A., Imamoto, N., Yoshimura, M., Sakai, H., Chong, K. T., Tsukihara, T., y Yoneda, Y. (2003). The structure of importin-beta bound to SREBP-2: nuclear import of a transcription factor. *Science* **302**, 1571-1575.
- ❖ Lehman, J. A., Hoelz, D. J., y Turchi, J. J. (2008). DNA-Dependent Conformational Changes in the Ku Heterodimer. *Biochemistry* **47**, 4359-4368.
- ❖ Lengauer, C., Kinzler, K. W., y Vogelstein, B. (1998). Genetic instabilities in human cancers. *Nature* **396**, 643-649.
- ❖ Leuther, K. K., Hammarsten, O., Kornberg, R. D., y Chu, G. (1999). Structure of DNA-dependent protein kinase: implications for its regulation by DNA. *Embo J* **18**, 1114-1123.
- ❖ Li de la Sierra-Gallay, I., Mathy, N., Pellegrini, O., y Condon, C. (2006). Structure of the ubiquitous 3' processing enzyme RNase Z bound to transfer RNA. *Nat Struct Mol Biol* **13**, 376-377.
- ❖ Li, Y., Chirgadze, D. Y., Bolanos-Garcia, V. M., Sibanda, B. L., Davies, O. R., Ahnesorg, P., Jackson, S. P., y Blundell, T. L. (2008). Crystal structure of human XLF/Cernunnos reveals unexpected differences from XRCC4 with implications for NHEJ. *Embo J* **27**, 290-300.
- ❖ Li, Z., Otevrel, T., Gao, Y., Cheng, H. L., Seed, B., Stamato, T. D., Taccioli, G. E., y Alt, F. W. (1995). The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. *Cell* **83**, 1079-1089.
- ❖ Lieber, M. R., Ma, Y., Pannicke, U., y Schwarz, K. (2003). Mechanism and regulation of human non-homologous DNA end-joining. *Nat Rev Mol Cell Biol* **4**, 712-720.
- ❖ Lieber, M. R. (2008). The mechanism of human nonhomologous DNA end joining. *J Biol Chem* **283**, 1-5.

- ❖ Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715.
- ❖ Lu, H., Pannicke, U., Schwarz, K., y Lieber, M. R. (2007). Length-dependent binding of human XLF to DNA and stimulation of XRCC4.DNA ligase IV activity. *J Biol Chem* **282**, 11155-11162.
- ❖ Ludtke, S. J., Baldwin, P. R., y Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* **128**, 82-97.
- ❖ Llorca, O., Rivera-Calzada, A., Grantham, J., y Willison, K. R. (2003). Electron microscopy and 3D reconstructions reveal that human ATM kinase uses an arm-like domain to clamp around double-stranded DNA. *Oncogene* **22**, 3867-3874.
- ❖ Ma, Y., Pannicke, U., Schwarz, K., y Lieber, M. R. (2002). Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **108**, 781-794.
- ❖ Ma, Y., Lu, H., Tippin, B., Goodman, M. F., Shimazaki, N., Koiwai, O., Hsieh, C. L., Schwarz, K., y Lieber, M. R. (2004). A biochemically defined system for mammalian nonhomologous DNA end joining. *Mol Cell* **16**, 701-713.
- ❖ Ma, Y., Lu, H., Schwarz, K., y Lieber, M. R. (2005a). Repair of double-strand DNA breaks by the human nonhomologous DNA end joining pathway: the iterative processing model. *Cell Cycle* **4**, 1193-1200.
- ❖ Ma, Y., Pannicke, U., Lu, H., Niewolik, D., Schwarz, K., y Lieber, M. R. (2005b). The DNA-dependent protein kinase catalytic subunit phosphorylation sites in human Artemis. *J Biol Chem* **280**, 33839-33846.
- ❖ Ma, Y., Schwarz, K., y Lieber, M. R. (2005c). The Artemis:DNA-PKcs endonuclease cleaves DNA loops, flaps, and gaps. *DNA Repair (Amst)* **4**, 845-851.
- ❖ Mahajan, K. N., Nick McElhinny, S. A., Mitchell, B. S., y Ramsden, D. A. (2002). Association of DNA polymerase mu (pol mu) with Ku and ligase IV: role for pol mu in end-joining double-strand break repair. *Mol Cell Biol* **22**, 5194-5202.
- ❖ Malek, S., Huang, D. B., Huxford, T., Ghosh, S., y Ghosh, G. (2003). X-ray crystal structure of an IkappaBbeta x NF-kappaB p65 homodimer complex. *J Biol Chem* **278**, 23094-23100.
- ❖ Mandel, C. R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J. L., y Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444**, 953-956.
- ❖ Marabini, R., Masegosa, I. M., San Martin, M. C., Marco, S., Fernandez, J. J., de la Fraga, L. G., Vaquerizo, C., y Carazo, J. M. (1996). Xmipp: An Image Processing Package for Electron Microscopy. *J Struct Biol* **116**, 237-240.
- ❖ Martensson, S., y Hammarsten, O. (2002). DNA-dependent protein kinase catalytic subunit. Structural requirements for kinase activation by DNA ends. *J Biol Chem* **277**, 3020-3029.
- ❖ McGuffin, L. J., Bryson, K., y Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404-405.

- ❖ Meek, K., Gupta, S., Ramsden, D. A., y Lees-Miller, S. P. (2004). The DNA-dependent protein kinase: the director at the end. *Immunol Rev* **200**, 132-141.
- ❖ Meek, K., Douglas, P., Cui, X., Ding, Q., y Lees-Miller, S. P. (2007). trans Autophosphorylation at DNA-dependent protein kinase's two major autophosphorylation site clusters facilitates end processing but not end joining. *Mol Cell Biol* **27**, 3881-3890.
- ❖ Memisoglu, A., y Samson, L. (2000). Base excision repair in yeast and mammals. *Mutat Res* **451**, 39-51.
- ❖ Merkle, D., Douglas, P., Moorhead, G. B., Leonenko, Z., Yu, Y., Cramb, D., Bazett-Jones, D. P., y Lees-Miller, S. P. (2002). The DNA-dependent protein kinase interacts with DNA to form a protein-DNA complex that is disrupted by phosphorylation. *Biochemistry* **41**, 12706-12714.
- ❖ Miled, N., Yan, Y., Hon, W. C., Perisic, O., Zvelebil, M., Inbar, Y., Schneidman-Duhovny, D., Wolfson, H. J., Backer, J. M., y Williams, R. L. (2007). Mechanism of two classes of cancer mutations in the phosphoinositide 3-kinase catalytic subunit. *Science* **317**, 239-242.
- ❖ Mimori, T., Akizuki, M., Yamagata, H., Inada, S., Yoshida, S., y Homma, M. (1981). Characterization of a high molecular weight acidic nuclear protein recognized by autoantibodies in sera from patients with polymyositis-scleroderma overlap. *J Clin Invest* **68**, 611-620.
- ❖ Mimori, T., y Hardin, J. A. (1986). Mechanism of interaction between Ku protein and DNA. *J Biol Chem* **261**, 10375-10379.
- ❖ Mindell, J. A., y Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* **142**, 334-347.
- ❖ Ming, D., Kong, Y., Wakil, S. J., Brink, J., y Ma, J. (2002). Domain movements in human fatty acid synthase by quantized elastic deformational model. *Proc Natl Acad Sci U S A* **99**, 7895-7899.
- ❖ Miyata, T., Oyama, T., Mayanagi, K., Ishino, S., Ishino, Y., y Morikawa, K. (2004). The clamp-loading complex for processive DNA replication. *Nat Struct Mol Biol* **11**, 632-636.
- ❖ Mizuta, R., Cheng, H. L., Gao, Y., y Alt, F. W. (1997). Molecular genetic characterization of XRCC4 function. *Int Immunol* **9**, 1607-1613.
- ❖ Modesti, M., Hesse, J. E., y Gellert, M. (1999). DNA binding of Xrcc4 protein is associated with V(D)J recombination but not with stimulation of DNA ligase IV activity. *Embo J* **18**, 2008-2018.
- ❖ Moon, A. F., Garcia-Diaz, M., Batra, V. K., Beard, W. A., Bebenek, K., Kunkel, T. A., Wilson, S. H., y Pedersen, L. C. (2007). The X family portrait: structural insights into biological functions of X family polymerases. *DNA Repair (Amst)* **6**, 1709-1725.
- ❖ Moore, J. K., y Haber, J. E. (1996). Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. *Mol Cell Biol* **16**, 2164-2173.

- ❖ Moshous, D., Callebaut, I., de Chasseval, R., Corneo, B., Cavazzana-Calvo, M., Le Deist, F., Tezcan, I., Sanal, O., Bertrand, Y., Philippe, N., *et al.* (2001). Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell* **105**, 177-186.
- ❖ Moshous, D., Pannetier, C., Chasseval Rd, R., Deist FI, F., Cavazzana-Calvo, M., Romana, S., Macintyre, E., Canioni, D., Brousse, N., Fischer, A., *et al.* (2003). Partial T and B lymphocyte immunodeficiency and predisposition to lymphoma in patients with hypomorphic mutations in Artemis. *J Clin Invest* **111**, 381-387.
- ❖ Mouche, F., Boisset, N., Lamy, J., Zal, F., y Lamy, J. N. (1999). Structural comparison of cephalopod hemocyanins: phylogenetic significance. *J Struct Biol* **127**, 199-212.
- ❖ Neale, M. J., Pan, J., y Keeney, S. (2005). Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* **436**, 1053-1057.
- ❖ Nick McElhinny, S. A., Snowden, C. M., McCarville, J., y Ramsden, D. A. (2000). Ku recruits the XRCC4-ligase IV complex to DNA ends. *Mol Cell Biol* **20**, 2996-3003.
- ❖ Nick McElhinny, S. A., Havener, J. M., Garcia-Diaz, M., Juarez, R., Bebenek, K., Kee, B. L., Blanco, L., Kunkel, T. A., y Ramsden, D. A. (2005). A gradient of template dependence defines distinct biological roles for family X polymerases in nonhomologous end joining. *Mol Cell* **19**, 357-366.
- ❖ O'Driscoll, M., Cerosaletti, K. M., Girard, P. M., Dai, Y., Stumm, M., Kysela, B., Hirsch, B., Gennery, A., Palmer, S. E., Seidel, J., *et al.* (2001). DNA ligase IV mutations identified in patients exhibiting developmental delay and immunodeficiency. *Mol Cell* **8**, 1175-1185.
- ❖ Orlova, E. V., Dube, P., Harris, J. R., Beckman, E., Zemlin, F., Markl, J., y van Heel, M. (1997). Structure of keyhole limpet hemocyanin type 1 (KLH1) at 15 Å resolution by electron cryomicroscopy and angular reconstitution. *J Mol Biol* **271**, 417-437.
- ❖ Ortiz, M. L., Calero, M., Fernandez Patron, C., Patron, C. F., Castellanos, L., y Mendez, E. (1992). Imidazole-SDS-Zn reverse staining of proteins in gels containing or not SDS and microsequence of individual unmodified electroblotted proteins. *FEBS Lett* **296**, 300-304.
- ❖ Pang, D., Yoo, S., Dynan, W. S., Jung, M., y Dritschilo, A. (1997). Ku proteins join DNA fragments as shown by atomic force microscopy. *Cancer Res* **57**, 1412-1415.
- ❖ Park, E. J., Chan, D. W., Park, J. H., Oettinger, M. A., y Kwon, J. (2003). DNA-PK is activated by nucleosomes and phosphorylates H2AX within the nucleosomes in an acetylation-dependent manner. *Nucleic Acids Res* **31**, 6819-6827.
- ❖ Pascal, J. M., O'Brien, P. J., Tomkinson, A. E., y Ellenberger, T. (2004). Human DNA ligase I completely encircles and partially unwinds nicked DNA. *Nature* **432**, 473-478.
- ❖ Pascual-Montano, A., Donate, L. E., Valle, M., Barcena, M., Pascual-Marqui, R. D., y Carazo, J. M. (2001). A novel neural network technique for analysis and classification of EM single-particle images. *J Struct Biol* **133**, 233-245.

- ❖ Pastwa, E., y Malinowski, M. (2007). Non-homologous DNA end joining in anticancer therapy. *Curr Cancer Drug Targets* **7**, 243-250.
- ❖ Penczek, P. A., Grassucci, R. A., y Frank, J. (1994). The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy* **53**, 251-270.
- ❖ Penczek, P. A., Frank, J., y Spahn, C. M. (2006a). A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation. *J Struct Biol* **154**, 184-194.
- ❖ Penczek, P. A., Yang, C., Frank, J., y Spahn, C. M. (2006b). Estimation of variance in single-particle reconstruction using the bootstrap technique. *J Struct Biol* **154**, 168-183.
- ❖ Perry, J., y Kleckner, N. (2003). The ATRs, ATMs, and TORs are giant HEAT repeat proteins. *Cell* **112**, 151-155.
- ❖ Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., y Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612.
- ❖ Pitcher, R. S., Brissett, N. C., y Doherty, A. J. (2007). Nonhomologous end-joining in bacteria: a microbial perspective. *Annu Rev Microbiol* **61**, 259-282.
- ❖ Radermacher, M., Wagenknecht, T., Verschoor, A., y Frank, J. (1986). A new 3-D reconstruction scheme applied to the 50S ribosomal subunit of *E. coli*. *J Microsc* **141**, RP1-2.
- ❖ Renau Piqueras, J. M. M., L. (1998). Manual de técnicas de microscopía electrónica (M.E.T): Aplicaciones biológicas).
- ❖ Resnick, M. A., y Moore, P. D. (1979). Molecular recombination and the repair of DNA double-strand breaks in CHO cells. *Nucleic Acids Res* **6**, 3145-3160.
- ❖ Revy, P., Buck, D., le Deist, F., y de Villartay, J. P. (2005). The repair of DNA damages/modifications during the maturation of the immune system: lessons from human primary immunodeficiency disorders and animal models. *Adv Immunol* **87**, 237-295.
- ❖ Riballo, E., Kuhne, M., Rief, N., Doherty, A., Smith, G. C., Recio, M. J., Reis, C., Dahm, K., Fricke, A., Krempler, A., *et al.* (2004). A pathway of double-strand break rejoining dependent upon ATM, Artemis, and proteins locating to gamma-H2AX foci. *Mol Cell* **16**, 715-724.
- ❖ Rich, T., Allen, R. L., y Wyllie, A. H. (2000). Defying death after DNA damage. *Nature* **407**, 777-783.
- ❖ Richardson, C., y Jasin, M. (2000). Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature* **405**, 697-700.
- ❖ Riha, K., Heacock, M. L., y Shippen, D. E. (2006). The role of the nonhomologous end-joining DNA double-strand break repair pathway in telomere biology. *Annu Rev Genet* **40**, 237-277.

- ❖ Rivera-Calzada, A., Maman, J. D., Spagnolo, L., Pearl, L. H., y Llorca, O. (2005). Three-dimensional structure and regulation of the DNA-dependent protein kinase catalytic subunit (DNA-PKcs). *Structure* **13**, 243-255.
- ❖ Roberts, S. A., y Ramsden, D. A. (2007). Loading of the nonhomologous end joining factor, Ku, on protein-occluded DNA ends. *J Biol Chem* **282**, 10605-10613.
- ❖ Robins, P., y Lindahl, T. (1996). DNA ligase IV from HeLa cell nuclei. *J Biol Chem* **271**, 24257-24261.
- ❖ Rooney, S., Sekiguchi, J., Zhu, C., Cheng, H. L., Manis, J., Whitlow, S., DeVido, J., Foy, D., Chaudhuri, J., Lombard, D., y Alt, F. W. (2002). Leaky Scid phenotype associated with defective V(D)J coding end processing in Artemis-deficient mice. *Mol Cell* **10**, 1379-1390.
- ❖ Rooney, S., Alt, F. W., Lombard, D., Whitlow, S., Eckersdorff, M., Fleming, J., Fugmann, S., Ferguson, D. O., Schatz, D. G., y Sekiguchi, J. (2003). Defective DNA repair and increased genomic instability in Artemis-deficient murine cells. *J Exp Med* **197**, 553-565.
- ❖ Rosenthal, P. B., y Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol* **333**, 721-745.
- ❖ Scheres, S. H., Valle, M., y Carazo, J. M. (2005a). Fast maximum-likelihood refinement of electron microscopy images. *Bioinformatics* **21 Suppl 2**, ii243-244.
- ❖ Scheres, S. H., Valle, M., Nunez, R., Sorzano, C. O., Marabini, R., Herman, G. T., y Carazo, J. M. (2005b). Maximum-likelihood multi-reference refinement for electron microscopy images. *J Mol Biol* **348**, 139-149.
- ❖ Schwarz, K., Ma, Y., Pannicke, U., y Lieber, M. R. (2003). Human severe combined immune deficiency and DNA repair. *Bioessays* **25**, 1061-1070.
- ❖ Sibanda, B. L., Critchlow, S. E., Begun, J., Pei, X. Y., Jackson, S. P., Blundell, T. L., y Pellegrini, L. (2001). Crystal structure of an Xrcc4-DNA ligase IV complex. *Nat Struct Biol* **8**, 1015-1019.
- ❖ Sonoda, E., Hohegger, H., Saberi, A., Taniguchi, Y., y Takeda, S. (2006). Differential usage of non-homologous end-joining and homologous recombination in double strand break repair. *DNA Repair (Amst)* **5**, 1021-1029.
- ❖ Sorzano, C. O., Marabini, R., Velazquez-Muriel, J., Bilbao-Castro, J. R., Scheres, S. H., Carazo, J. M., y Pascual-Montano, A. (2004). XMIPP: a new generation of an open-source image processing package for electron microscopy. *J Struct Biol* **148**, 194-204.
- ❖ Spagnolo, L., Rivera-Calzada, A., Pearl, L. H., y Llorca, O. (2006). Three-dimensional structure of the human DNA-PKcs/Ku70/Ku80 complex assembled on DNA and its implications for DNA DSB repair. *Mol Cell* **22**, 511-519.
- ❖ Takai, H., Wang, R. C., Takai, K. K., Yang, H., y de Lange, T. (2007). Tel2 regulates the stability of PI3K-related protein kinases. *Cell* **131**, 1248-1259.

- ❖ Takata, M., Sasaki, M. S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., Shinohara, A., y Takeda, S. (1998). Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *Embo J* **17**, 5497-5508.
- ❖ Tidow, H., Melero, R., Mylonas, E., Freund, S. M., Grossmann, J. G., Carazo, J. M., Svergun, D. I., Valle, M., y Fersht, A. R. (2007). Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex. *Proc Natl Acad Sci U S A* **104**, 12324-12329.
- ❖ Tomkinson, A. E., Vijayakumar, S., Pascal, J. M., y Ellenberger, T. (2006). DNA ligases: structure, reaction mechanism, and function. *Chem Rev* **106**, 687-699.
- ❖ Tsai, C. J., Kim, S. A., y Chu, G. (2007). Cernunnos/XLF promotes the ligation of mismatched and noncohesive DNA ends. *Proc Natl Acad Sci U S A* **104**, 7851-7856.
- ❖ Tsukuda, T., Fleming, A. B., Nickoloff, J. A., y Osley, M. A. (2005). Chromatin remodelling at a DNA double-strand break site in *Saccharomyces cerevisiae*. *Nature* **438**, 379-383.
- ❖ Uematsu, N., Weterings, E., Yano, K., Morotomi-Yano, K., Jakob, B., Taucher-Scholz, G., Mari, P. O., van Gent, D. C., Chen, B. P., y Chen, D. J. (2007). Autophosphorylation of DNA-PKCS regulates its dynamics at DNA double-strand breaks. *J Cell Biol* **177**, 219-229.
- ❖ Ullah, J. H., Walsh, T. R., Taylor, I. A., Emery, D. C., Verma, C. S., Gamblin, S. J., y Spencer, J. (1998). The crystal structure of the L1 metallo-beta-lactamase from *Stenotrophomonas maltophilia* at 1.7 Å resolution. *J Mol Biol* **284**, 125-136.
- ❖ van Attikum, H., y Gasser, S. M. (2005). The histone code at DNA breaks: a guide to repair? *Nat Rev Mol Cell Biol* **6**, 757-765.
- ❖ van Gent, D. C., Hoeijmakers, J. H., y Kanaar, R. (2001). Chromosomal stability and the DNA double-stranded break connection. *Nat Rev Genet* **2**, 196-206.
- ❖ van Heel, M., y Keegstram, W. (1981). IMAGIC: a fast, flexible and friendly image analysis software system. *Ultramicroscopy* **7**, 113-130.
- ❖ van Heel, M. (1984). Multivariate statistical classification of noisy images (randomly oriented biological macromolecules). *Ultramicroscopy* **13**, 165-183.
- ❖ van Heel, M. (1987). Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy* **21**, 111-123.
- ❖ van Heel, M., Gowen, B., Matadeen, R., Orlova, E. V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M., y Patwardhan, A. (2000). Single-particle electron cryo-microscopy: towards atomic resolution. *Q Rev Biophys* **33**, 307-369.
- ❖ Walker, E. H., Perisic, O., Ried, C., Stephens, L., y Williams, R. L. (1999). Structural insights into phosphoinositide 3-kinase catalysis and signalling. *Nature* **402**, 313-320.

- ❖ Walker, E. H., Pacold, M. E., Perisic, O., Stephens, L., Hawkins, P. T., Wymann, M. P., y Williams, R. L. (2000). Structural determinants of phosphoinositide 3-kinase inhibition by wortmannin, LY294002, quercetin, myricetin, and staurosporine. *Mol Cell* **6**, 909-919.
- ❖ Walker, J. R., Corpina, R. A., y Goldberg, J. (2001). Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* **412**, 607-614.
- ❖ Wei, Y. F., Robins, P., Carter, K., Caldecott, K., Pappin, D. J., Yu, G. L., Wang, R. P., Shell, B. K., Nash, R. A., Schar, P., y et al. (1995). Molecular cloning and expression of human cDNAs encoding a novel DNA ligase IV and DNA ligase III, an enzyme active in DNA repair and recombination. *Mol Cell Biol* **15**, 3206-3216.
- ❖ West, R. B., Yaneva, M., y Lieber, M. R. (1998). Productive and nonproductive complexes of Ku and DNA-dependent protein kinase at DNA termini. *Mol Cell Biol* **18**, 5908-5920.
- ❖ West, S. C. (2003). Molecular views of recombination proteins and their control. *Nat Rev Mol Cell Biol* **4**, 435-445.
- ❖ Weterings, E., Verkaik, N. S., Bruggenwirth, H. T., Hoeijmakers, J. H., y van Gent, D. C. (2003). The role of DNA dependent protein kinase in synapsis of DNA ends. *Nucleic Acids Res* **31**, 7238-7246.
- ❖ Weterings, E., y van Gent, D. C. (2004). The mechanism of non-homologous end-joining: a synopsis of synapsis. *DNA Repair (Amst)* **3**, 1425-1435.
- ❖ Weterings, E., y Chen, D. J. (2008). The endless tale of non-homologous end-joining. *Cell Res* **18**, 114-124.
- ❖ Williams, D. R., Lee, K. J., Shi, J., Chen, D. J., y Stewart, P. L. (2008). Cryo-EM Structure of the DNA-Dependent Protein Kinase Catalytic Subunit at Subnanometer Resolution Reveals alpha Helices and Insight into DNA Binding. *Structure* **16**, 468-477.
- ❖ Williams, R. C., y Fisher, H. W. (1970). Electron microscopy of tobacco mosaic virus under conditions of minimal beam exposure. *J Mol Biol* **52**, 121-123.
- ❖ Wiseman, H., y Halliwell, B. (1996). Damage to DNA by reactive oxygen and nitrogen species: role in inflammatory disease and progression to cancer. *Biochem J* **313 ( Pt 1)**, 17-29.
- ❖ Woo, Y., Wright, S. M., Maas, S. A., Alley, T. L., Caddle, L. B., Kamdar, S., Affourtit, J., Foreman, O., Akeson, E. C., Shaffer, D., et al. (2007). The nonhomologous end joining factor Artemis suppresses multi-tissue tumor formation and prevents loss of heterozygosity. *Oncogene* **26**, 6010-6020.
- ❖ Wriggers, W., Milligan, R. A., y McCammon, J. A. (1999). Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J Struct Biol* **125**, 185-195.
- ❖ Wu, P. Y., Ruhlmann, C., Winston, F., y Schultz, P. (2004). Molecular architecture of the *S. cerevisiae* SAGA complex. *Mol Cell* **15**, 199-208.



- ❖ Wu, X., y Lieber, M. R. (1997). Interaction between DNA-dependent protein kinase and a novel protein, KIP. *Mutat Res* **385**, 13-20.
- ❖ Yaneva, M., Kowalewski, T., y Lieber, M. R. (1997). Interaction of DNA-dependent protein kinase with DNA and with Ku: biochemical and atomic-force microscopy studies. *Embo J* **16**, 5098-5112.
- ❖ Yang, S., Yu, X., Galkin, V. E., y Egelman, E. H. (2003). Issues of resolution and polymorphism in single-particle reconstruction. *J Struct Biol* **144**, 162-171.
- ❖ Yannone, S. M., Roy, S., Chan, D. W., Murphy, M. B., Huang, S., Campisi, J., y Chen, D. J. (2001). Werner syndrome protein is regulated and phosphorylated by DNA-dependent protein kinase. *J Biol Chem* **276**, 38242-38248.
- ❖ Yano, K., Morotomi-Yano, K., Wang, S. Y., Uematsu, N., Lee, K. J., Asaithamby, A., Weterings, E., y Chen, D. J. (2008). Ku recruits XLF to DNA double-strand breaks. *EMBO Rep* **9**, 91-96.
- ❖ Yoo, S., y Dynan, W. S. (1999). Geometry of a complex formed by double strand break repair proteins at a single DNA end: recruitment of DNA-PKcs induces inward translocation of Ku protein. *Nucleic Acids Res* **27**, 4679-4686.
- ❖ Yu, X., Hingorani, M. M., Patel, S. S., y Egelman, E. H. (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**, 740-743.
- ❖ Zhang, X., Settembre, E., Xu, C., Dormitzer, P. R., Bellamy, R., Harrison, S. C., y Grigorieff, N. (2008). Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proc Natl Acad Sci U S A* **105**, 1867-1872.
- ❖ Zhang, Z., Zhu, L., Lin, D., Chen, F., Chen, D. J., y Chen, Y. (2001). The three-dimensional structure of the C-terminal DNA-binding domain of human Ku70. *J Biol Chem* **276**, 38231-38236.
- ❖ Zhang, Z., Hu, W., Cano, L., Lee, T. D., Chen, D. J., y Chen, Y. (2004). Solution structure of the C-terminal domain of Ku80 suggests important sites for protein-protein interactions. *Structure* **12**, 495-502.





## 8. Publicaciones



# Structural model of full-length human Ku70–Ku80 heterodimer and its recognition of DNA and DNA-PKcs

Angel Rivera-Calzada<sup>1</sup>, Laura Spagnolo<sup>2</sup>, Laurence H. Pearl<sup>2+</sup> & Oscar Llorca<sup>1++</sup>

<sup>1</sup>Centro de Investigaciones Biológicas (CIB), Spanish National Research Council (CSIC), Campus Complutense University, Madrid, Spain, and <sup>2</sup>Section of Structural Biology and Cancer Research UK DNA Repair Enzyme Research Group, Institute of Cancer Research, Chester Beatty Laboratories, London, UK

**Recognition of DNA double-strand breaks during non-homologous end joining is carried out by the Ku70–Ku80 protein, a 150 kDa heterodimer that recruits the DNA repair kinase DNA-dependent protein kinase catalytic subunit (DNA-PKcs) to the lesion. The atomic structure of a truncated Ku70–Ku80 was determined; however, the subunit-specific carboxy-terminal domain of Ku80—essential for binding to DNA-PKcs—was determined only in isolation, and the C-terminal domain of Ku70 was not resolved in its DNA-bound conformation. Both regions are conserved and mediate protein–protein interactions specific to mammals. Here, we reconstruct the three-dimensional structure of the human full-length Ku70–Ku80 dimer at 25 Å resolution, alone and in complex with DNA, by using single-particle electron microscopy. We map the C-terminal regions of both subunits, and their conformational changes after DNA and DNA-PKcs binding to define a molecular model of the functions of these domains during DNA repair in the context of full-length Ku70–Ku80 protein.**

Keywords: cryo-EM; DNA-PK; Ku70; Ku80; DNA repair

EMBO reports (2007) 8, 56–62. doi:10.1038/sj.embor.7400847

## INTRODUCTION

Non-homologous end joining (NHEJ) directly ligates DNA double-strand breaks (DSBs) independently of the template, and is the most prevalent DSB repair mechanism in mammalian cells (Jackson, 2002; Lees-Miller & Meek, 2003; Lieber *et al*, 2003).

<sup>1</sup>Centro de Investigaciones Biológicas (CIB), Spanish National Research Council (CSIC), Ramiro de Maeztu 9, Campus Complutense University, Madrid 28040, Spain

<sup>2</sup>Section of Structural Biology and Cancer Research UK DNA Repair Enzyme Research Group, Institute of Cancer Research, Chester Beatty Laboratories, 237 Fulham Road, London SW3 6JB, UK

\*Corresponding author. Tel: +44 20 7153 5422; Fax: +44 20 7153 5457;

E-mail: laurence.pearl@icr.ac.uk

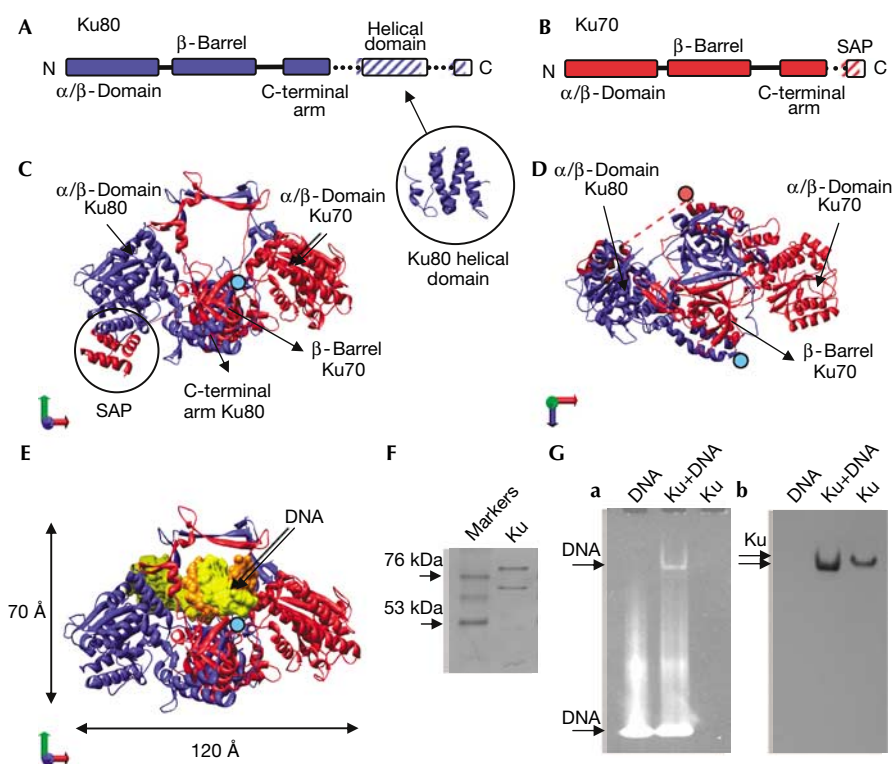
++Corresponding author. Tel: +34 91 837 3112 ext 4446; Fax: +34 91 5360432;

E-mail: ollorca@cib.csic.es

Received 6 June 2006; revised 11 September 2006; accepted 2 October 2006; published online 8 December 2006

The first step during NHEJ is the specific recognition of the DNA ends at the site of the lesion. This is carried out by the Ku70–Ku80 protein, a heterodimer consisting of the two highly related subunits Ku70 (70 kDa) and Ku80 (86 kDa), which build a preformed ring that can encircle DNA (Fig 1; Doherty & Jackson, 2001; Walker *et al*, 2001; Downs & Jackson, 2004). Ku70–Ku80 then recruits the catalytic subunit of the DNA-dependent protein kinase catalytic subunit (DNA-PKcs) to the DNA end (Gottlieb & Jackson, 1993; West *et al*, 1998; Meek *et al*, 2004; Falck *et al*, 2005). The medium resolution structures of the 460 kDa DNA-PKcs kinase and its complex with Ku have been solved by cryo-electron microscopy (cryo-EM) and show that the molecule is divided into a large head that comprises the conserved carboxy-terminal domains and an extended arm assigned to the amino terminus (Rivera-Calzada *et al*, 2005; Spagnolo *et al*, 2006).

The atomic structure of a truncated form of the Ku70–Ku80 dimer has been determined by using X-ray crystallography (Walker *et al*, 2001). Both subunits show a common topology consisting of an N-terminal  $\alpha/\beta$  domain, a central  $\beta$ -barrel domain and a helical C-terminal arm, which together form a pseudo-symmetrical structure with a preformed ring extending from a broad base (Fig 1A–D). DNA is embraced by the narrow bridge through sequence-independent interactions (Fig 1E). The ends of the C-terminal arms are located centrally in this atomic structure (Fig 1D, shown by blue and red filled circles), from which the subunit-specific C-terminal regions of each monomer emerge (Fig 1A). The 19 kDa C-terminal region of Ku80 (Ku80-CTD; residues 545–732) is implicated in the recruitment of DNA-PKcs by Ku to sites of damage, and has also been shown to interact with the Werner's syndrome protein (Gell & Jackson, 1999; Falck *et al*, 2005). Nuclear magnetic resonance studies of Ku80-CTD in isolation show a flat helical structure for the fragment comprising residues 592–709 (Harris *et al*, 2004; Zhang *et al*, 2004; Fig 1A), whereas the most C-terminal portion of Ku80-CTD (residues 720–732) is disordered. The Ku70-specific C-terminal region consists of a 5 kDa SAP domain (Ku70-SAP), the structure of which was determined both in isolation (Zhang *et al*, 2001) and as part of the Ku70–Ku80 crystal structure (Walker *et al*, 2001; Fig 1C).



**Fig 1** | Purification of the Ku70–Ku80 dimer. (A) Schematic representation of the domain organization of Ku80. Ku80-CTD contains a helical domain (PDB 1Q2Z; Harris *et al*, 2004). (B) Schematic representation of Ku70. (C,D) Two views of the Ku70–Ku80 atomic structure (PDB 1JEQ; Walker *et al*, 2001). Ku80 and Ku70 subunits are coloured blue and red, respectively. In (D), a dotted line represents the connection between the core of Ku70 and the SAP domain, not solved in the structure. (E) Front view of the Ku–DNA atomic structure (PDB 1JEY). In (C–E), the last amino acid of each carboxy-terminal arm, from which the subunit-specific C-terminal domains emerge, is indicated with either a blue (Ku80) or a red (Ku70) filled circle. (F) SDS gel of purified Ku70–Ku80. (G) DNA interaction assay. (a) 6% native polyacrylamide gel electrophoresis gel developed with UV to show the presence of DNA. (b) Western blot of the native gel using a monoclonal antibody against Ku70. SDS, sodium dodecyl sulphate.

Importantly, in the DNA-bound Ku70–Ku80 crystal structure, the Ku70-SAP domain was displaced from its original position and was not resolved. SAP domains have been defined as putative DNA-binding motifs (Aravind & Koonin, 2000); therefore, this region could also have a role in DNA binding during the NHEJ reaction.

The subunit-specific C-terminal domains of Ku mediate DNA recognition and DNA-PKcs recruitment specific to mammals, and the structural basis of these processes has remained unclear in the absence of a three-dimensional (3D) structure of the full-length Ku70–Ku80. Here, a 3D structural model of the full-length Ku protein is proposed on the basis of single-particle EM experiments.

## RESULTS

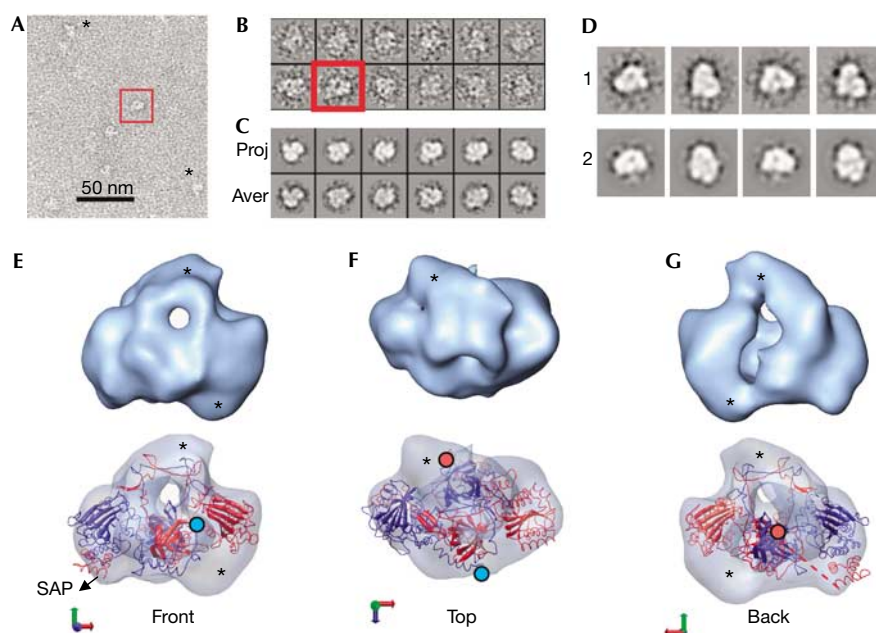
### Ku purification and its interaction with DNA

Ku70–Ku80 was purified from HeLa cells (Fig 1F). The ability of our preparation to form a stable complex with DNA was tested by incubation with a 54 bp oligonucleotide. We used concentrations and conditions that were identical to those used later in the EM. The complex was resolved by using native polyacrylamide gel electrophoresis, which was first incubated with ethidium bromide to show the presence of DNA (Fig 1Ga), followed by

immunoblotting with a monoclonal antibody specific for the Ku70 subunit (Fig 1Gb). A faster migrating band is detected both by the antibody and the ethidium bromide, and indicates that under these conditions we can assemble a homogenous DNA-bound Ku complex as there is only one band after incubation with DNA.

### EM and 3D structure of full-length Ku70–Ku80

Ku molecules were clearly seen in negatively stained electron micrographs (Fig 2A, indicated by asterisks) with some single-molecule images being strongly evocative of the typical view of Ku (Fig 2A,B, within a red square). A total of 3,419 images were subjected to iterative angular refinement using a strategy defined to avoid any influence of our previous knowledge of the atomic structures (supplementary information online), which resulted in a 3D reconstruction at a resolution of approximately 25 Å (at a Fourier shell correlation of 0.5; Fig 2E–G, top row; supplementary Movie S1 online). Projections and averages at the final step of refinement (Fig 2C) were consistent with averages derived from the initial data set using reference-free classification and 2D averaging methods implemented in EMAN (Electron Micrograph Analysis; Ludtke *et al*, 1999) and XMIPP (X-windows-based microscopy image processing package; Sorzano *et al*, 2004; Fig 2D).



**Fig 2** | Three-dimensional structure of full-length Ku70–Ku80. (A) Electron microscopy field of negatively stained Ku70–Ku80 dimers. Some molecules are indicated with asterisks. (B) Gallery containing a representative collection of single molecules. (A,B) A view evocative of the atomic structure of Ku is shown within a red square. (C) Panel containing projections (upper row) and their corresponding class averages (lower row) obtained after image refinement. (D) Galley of reference-free two-dimensional averages obtained using EMAN and XMIPP (1) compared with projections derived by angular refinement (2). (E–G) Views of the three-dimensional reconstruction of Ku (upper row) and their corresponding fitting of the atomic structure of Ku (PDB file 1JEQ; lower row). Ku80 and Ku70 are coloured blue and red, respectively. The last amino acid of each carboxyl-terminal arm, from which the subunit-specific C-terminal domains emerge, is indicated with either a blue (Ku80) or a red (Ku70) filled circle. Densities in the reconstruction that are not accounted for by the atomic structure are highlighted with asterisks.

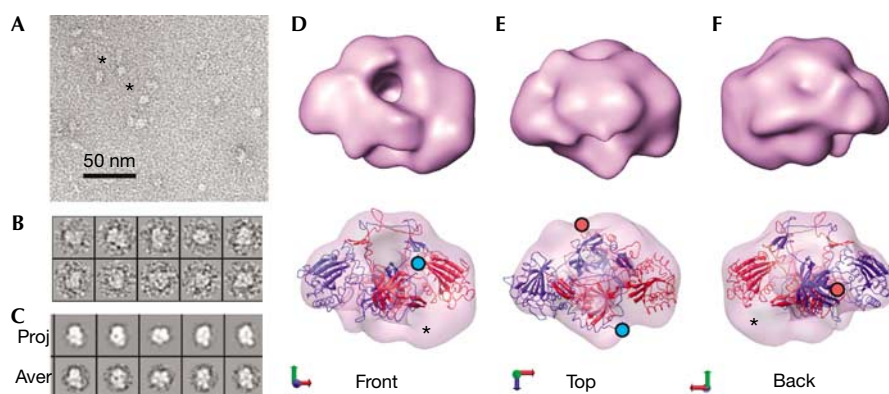
The homogeneity and consistency of the EM reconstructions in this work was further evaluated by using several approaches: (i) a number of volumes were used as starting references for angular refinement and all of these converged to the same solution (ii) regions of high variance within the 3D reconstructions calculated using a bootstrap technique of sampling (Penczek *et al*, 2006) were found to be very limited (supplementary Fig S1 online); and (iii) we employed the random conical tilt (RCT) method (Radermacher, 1988) on typical views. Internal consistency of the reconstructions defined by angular refinement was supported by their similarity to the RCT volumes, which were constructed from tilted data not used in the reconstruction using EMAN (supplementary Fig S2 online). These results support that the reconstructed volume is coherent with the input data and that the methodology used has probably minimized the effects of distortions caused by the staining agent at the particles in the 3D reconstruction.

The reconstructed volume contains the general features found in the atomic structure of the core dimer, with an obvious hole underneath a narrow bridge of density (Fig 2E). However, there are other masses evident at the main axis and also at the back of the hole (see asterisks in Fig 2E–G), which break the symmetry of the structure. As information on atomic resolution is available for much of the Ku molecule, although in several pieces (Fig 1), we fitted the structure of the crystallized dimer in the absence of DNA but containing the SAP domain (Walker *et al*, 2001) into the EM density maps using SITUS (Chacon & Wriggers, 2002). The top solution shows an excellent agreement between the low- and

high-resolution structures (cross-correlation value of 0.66), substantiating the quality of our map and defining the correct handedness (Fig 2E–G, bottom row; supplementary Movie S2 and supplementary Fig S3 online). The characteristic ring and elongated shape of the Ku molecule, plus the presence of the SAP domain limit the number of acceptable solutions during the fitting. The density of the SAP domain was clearly present in our reconstruction and, significantly, a region in our map was not accounted for by the atomic structure and was therefore considered as the likely position of Ku80-CTD (Fig 2E–G, indicated by asterisks).

### 3D structure of Ku70–Ku80 complexed with DNA

Ku70–Ku80 was incubated with a 54 bp oligonucleotide under conditions in which all the protein was bound to DNA (Fig 1G), and then observed using an electron microscope (Fig 3A). A total of 8,285 particles were extracted from the micrographs (Fig 3B) and the images were processed, as stated above into a structure, with a resolution of approximately 25 Å, which shows substantial conformational changes (Fig 3D–F, top row; supplementary Movie S3 online). To confirm these rearrangements, the reconstructions of Ku and DNA-Ku were used as input references of the DNA-bound and DNA free data sets respectively, which nonetheless refined to the correct final structure. We fitted the crystal structure of DNA-bound Ku into the EM map of the full-length DNA-Ku molecule but without the DNA (Fig 3D–F, bottom row; supplementary Fig S4 and supplementary Movie S4 online). In the atomic structure of DNA-bound Ku (Walker *et al*, 2001), the



**Fig 3** | Three-dimensional structure of full-length DNA-bound Ku70–Ku80 complex. (A) Electron microscopy field of negatively stained DNA-Ku complexes. Some particles are indicated with asterisks. (B) Gallery of representative single molecules. (C) Panel containing projections (upper row) and their corresponding class averages (lower row) obtained after refinement. (D–F) Views of the DNA-Ku reconstruction (upper row) and the fitting of the atomic structure of DNA-bound Ku (PDB file 1JEY; lower row). Ku80 and Ku70 are colored blue and red, respectively. The last amino acid of each carboxyl-terminal arm, from which the subunit-specific C-terminal domain emerge, is indicated with either a blue (Ku80) or a red (Ku70) filled circle. Densities in the reconstruction that are not accounted for by the atomic structure are indicated with asterisks.

SAP domain moves from its position in the apo-structure but is not resolved in the DNA-bound complex. Our EM reconstruction fits this atomic structure well (cross-correlation value of 0.75) and confirms this displacement of the SAP domain (Fig 3D). However, in the EM reconstruction, there is some additional density that is not accounted for by the atomic structure (Fig 3D–F, asterisks). Furthermore, the hole beneath the bridge, which is responsible for encircling DNA, now shows some density.

### Mapping of Ku70-SAP and Ku80-CTD

To locate the subunit-specific Ku70-SAP and Ku80-CTD in the full-length Ku70–Ku80, we calculated difference maps between the reconstructions and their corresponding fitted atomic structures (Fig 4, differences are shown in yellow). This approach seems to be supported by the consistency of our reconstructions (supplementary Figs S1,S2 online) and their close resemblance with the atomic information of Ku despite this not being used during refinement. In addition, the observed displacement of the SAP domain matches that found by X-ray crystallography, further supporting our reconstructions. Consequently, difference mapping showed only a few regions that possibly correspond to the missing C-terminal domains.

In the DNA-free protein, Ku70-SAP maps to one side, as seen in the atomic structure, and the additional densities that are not occupied by the atomic structure probably enclose the linker segments and the missing Ku80-CTD domain (Fig 4B–D, arrows 1 and 2). Putative density for Ku80-CTD is present underneath the base beside the  $\alpha/\beta$  domain of the Ku70 subunit (Fig 4B–D, arrow 1) and also in the back of the ring and slightly protruding from the top (Fig 4B–D, arrow 2). Ku80-CTD contains a helical domain connected by a flexible linker to the core domain of Ku, and followed by a C-terminal region, which is implicated in DNA-PKcs recruitment (Gell & Jackson, 1999; Fig 1A). Either of the regions of unoccupied density (Fig 4B–D, arrows 1 and 2) could accommodate this helical domain, and the combination of both regions is sufficient to account for the whole Ku80-CTD. Nevertheless, the NMR structures of the helical domain in Ku80-CTD

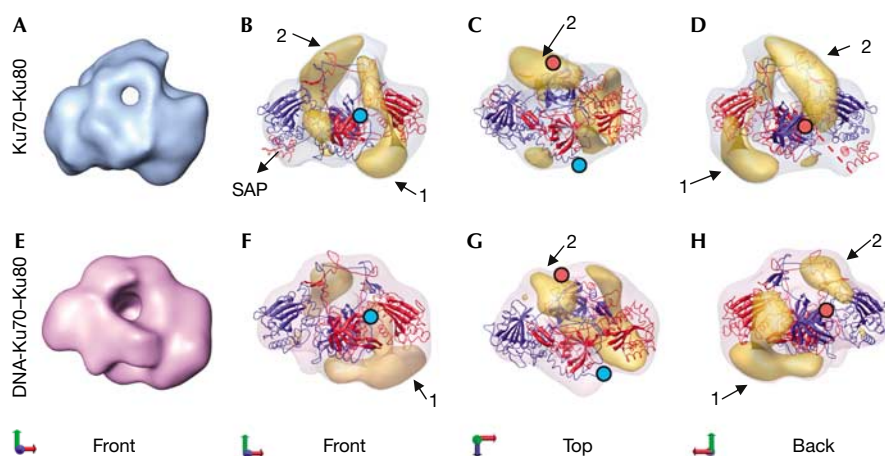
were better fitted in the region under the  $\alpha/\beta$  domain of the Ku70 subunit (Fig 4B,D, arrow 1), which, together with its proximity to the C-terminal end of Ku80 at the core domains (Fig 4B,C, blue filled circle), allows us to assign this region speculatively to the helical domain. The second additional region placed proximal to the ring might correspond to the most C-terminal portion to Ku80-CTD implicated in DNA-PKcs recruitment (Fig 4B–D, arrow 2).

Similar difference mapping analyses were carried out with the DNA-bound conformation (Fig 4E–H), showing densities under the  $\alpha/\beta$  domain of the Ku70 subunit in a similar orientation to that described previously in the DNA-free density (Fig 4B–D, arrow 1). Additional material was evident at the level of the hole, but here it was more compact and packed against the core domains (Fig 4G,H, arrow 2). These observations indicate that Ku80-CTD remains at a similar position after DNA binding but is probably affected by some changes in conformation. The largest change occurs to the SAP domain, which is in contact with the  $\alpha/\beta$  domain of Ku80 in the absence of DNA, a location found unoccupied after DNA binding. The atomic structure of DNA-bound Ku shows a similar change and Walker *et al* indicate that the domain could have moved to regions in contact with the DNA present in the structure (Walker *et al*, 2001). Those regions are located at the side of the Ku ring opposed to the free end of the DNA, facing the continuous DNA molecule; therefore, Ku70-SAP could be implicated in the maintenance of interactions between DNA and Ku. In our difference map, densities that might accommodate the displaced SAP domain could not be mapped owing to its small size.

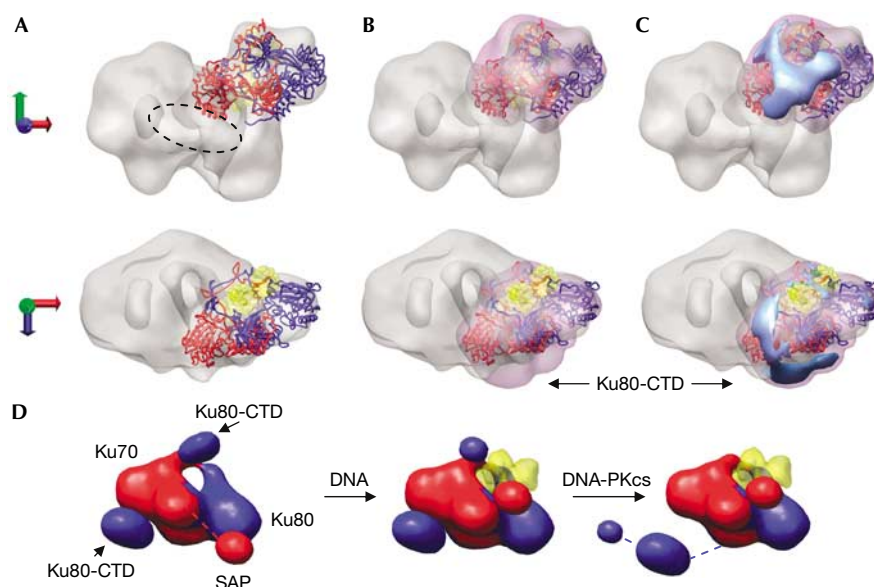
### Model of recruitment of DNA-PKcs by DNA-Ku

In the 3D structure of the DNA-PKcs–Ku complex (Spagnolo *et al*, 2006), Ku sits on the top of the palm domain of DNA-PKcs in a defined orientation that allows the DNA to follow a continuous channel formed by Ku and the cavity in DNA-PKcs (Fig 5A). We compared the structure of DNA-bound Ku with that after the recruitment of DNA-PKcs to explore the conformational changes that might occur in full-length Ku70–Ku80 on forming a complex





**Fig 4** | Mapping of Ku80-CTD and Ku70-CTD. (A) Front view of the Ku volume. (B–D) Reconstruction of Ku70–Ku80, in which densities shown by difference mapping are in yellow. (E) Front view of the DNA–Ku reconstruction. (F–H) DNA-bound Ku70–Ku80 with difference map in yellow. Ku80 and Ku70 are coloured blue and red, respectively. The last amino acid of each carboxyl-terminal arm, from which the subunit-specific C-terminal domains emerge, is indicated with either a blue (Ku80) or a red (Ku70) filled circle.



**Fig 5** | Model of domain rearrangements in full-length Ku70–Ku80 on DNA and DNA-dependent protein kinase catalytic subunit recognition. (A) Two views of the DNA–PKcs–Ku complex (grey density), in which the atomic structure of Ku has been fitted (Spagnolo *et al*, 2006). (B) Three-dimensional reconstruction of DNA–Ku (purple density) fitted with the DNA–PKcs–Ku complex and positioned in an orientation equivalent to that found for Ku in the complex. (C) Difference map between the reconstructions of DNA–Ku and the DNA–PKcs–Ku complex shown in blue. (D) Schematic model of the domain movements described in full-length Ku70–Ku80 after DNA interaction and DNA–PKcs recruitment. DNA is in yellow. Ku80 and Ku70 are coloured blue and red, respectively. DNA–PKcs, DNA-dependent protein kinase catalytic subunit.

with DNA–PKcs. We fitted the reconstruction of DNA-bound Ku into the DNA–PKcs–Ku complex, by matching the orientation of the core domains of DNA–Ku shown after fitting (Fig 3), with the position of the core domains of Ku in the DNA–PKcs–Ku complex (Spagnolo *et al*, 2006; Fig 5B). This fitting shows directly (Fig 5B) and after calculating a difference map (Fig 5C), that the Ku80-CTD present in DNA–Ku underneath the base is missing in the corresponding position after formation of the complex with

DNA–PKcs. As this region is essential for the interaction with DNA–PKcs, the Ku80-CTD domain has probably dislocated and it is likely to be now situated in the density of the complex. Tentatively, we searched for possible locations; visual inspection showed an area of protein density in the DNA–PKcs–Ku complex that was not found in isolated DNA–PKcs (Rivera-Calzada *et al*, 2005) and that might therefore be very cautiously attributed to this domain (Fig 5A, dotted circle).

## DISCUSSION

We propose a structural model of the functions of the subunit-specific C-terminal domains in the context of the full-length Ku70–Ku80 protein (Fig 5D). The Ku70-SAP domain changes its position on DNA binding from the base of the  $\alpha/\beta$  domain of Ku80 to a position that might be proximal to the ring in the side of Ku that faces the continuous DNA. At this site, the SAP domain may be in contact with the DNA, possibly acting to stabilize Ku binding at the end of the break and favouring its interaction with DNA-PKcs. By contrast, Ku80-CTD undergoes a substantial displacement only when DNA-PKcs is recruited, detaching from the Ku core to interact with DNA-PKcs. The length of the linker between the Ku80 core and the helical domain of Ku80-CTD would certainly allow such a displacement. As Ku and DNA-PKcs do not associate in the absence of DNA—whereas Ku80-CTD is sufficient to interact with DNA-PKcs—it is likely that the displacement of Ku80-CTD contributes to stabilizing the interaction. Such a mechanism would ensure that the DNA-PKcs–Ku repair complex assembles only at the site of the lesion. It has also been recently reported that Ku80-CTD can dissociate a DNA-PKcs–Artemis complex (Drouet *et al*, 2006). Our structure would suggest that the displacement of Ku80-CTD on DNA-PKcs binding could potentially remove Artemis by competition.

## METHODS

**Purification of the Ku70/Ku80 dimer.** Ku70 and Ku80 were purified from the HeLa nuclear extract (CilBiotech, Mons, Belgium) by using a combination of heparin, double-stranded DNA cellulose, phenyl sepharose high resolution and monoQ chromatographies.

**DNA interaction assay.** A 54bp blunt-ended double-stranded DNA was constructed by annealing two FPLC-purified complementary 54-base oligonucleotides (5'-biotGGCCGCACGCGTCC ACCATGGGGTACAACCTACGATCTAGCTTCATGCACCGGAC-3'; Oswell, Southampton, UK). Then, 2  $\mu$ l of purified Ku (0.45  $\mu$ M), an approximately tenfold molar excess of 3.7  $\mu$ M DNA (4  $\mu$ l) and 1  $\mu$ l of buffer (50 mM Tris pH 7.5, 20% glycerol, 1 mM MgCl<sub>2</sub>, 50 mM KCl, 10 mM EDTA) were incubated at 25 °C for 20 min. A 6% native polyacrylamide gel electrophoresis gel was run at 4 °C in Tris-glycine buffer, incubated with ethidium bromide and developed with UV. After washing, the gel was subjected to immunoblotting with a polyvinyl difluoride membrane (Bio-Rad Laboratories, Hercules, CA, USA) with a monoclonal anti-Ku70 (2C3.11, ab87; Abcam, Cambridge, UK) followed by a horseradish peroxidase-conjugated secondary antibody (ab6728; Abcam).

**Electron microscopy and image processing.** A few microlitres of the purified Ku and the DNA-bound complexes were adsorbed to glow discharged carbon-coated grids and negatively stained using 1% uranyl acetate. The grids were observed in a JEOL 1230 operated at 100 kV. Micrographs were taken at  $\times 50,000$  magnification under low-dose conditions and digitized at a final 10  $\mu$ m step size. Particles from the micrographs were extracted, normalized, centred, filtered and refined using different commands found in the EMAN (Ludtke *et al*, 1999). A total of 3,419 particles for Ku and 8,285 for the DNA-containing protein were collected. The resolution was estimated at approximately 25 Å for both reconstructions by Fourier shell correlation using the *eotest* command and a correlation of 0.5. The density maps and the atomic structures were visualized with University of California,

San Francisco Chimera (Pettersen *et al*, 2004). In addition, pairs of micrographs were obtained and processed according to the RCT method and using the XMIPP platform (Sorzano *et al*, 2004). 3D variance was estimated using a bootstrap technique of sampling as proposed by Penczek *et al* (2006) with SPIDER (Frank *et al*, 1996; supplementary information online).

**Fitting atomic structures into the EM volumes and calculation of difference maps.** Fitting was carried out using files from the Protein Data Bank (1JEY and 1JEQ) by a six-dimensional search using *colores* (Chacon & Wriggers, 2002) from the software SITUS. Fittings were performed using both possible hands in the volumes and it was found that one hand allowed the most adequate fitting. To calculate the difference maps between the fitted atomic structures and the EM density, the atomic structures were converted to density maps at comparable resolution. The EM maps of the Ku70–Ku80 dimer have been deposited in the 3D EM database (<http://www.ebi.ac.uk/msd/>) under accession codes EMD-1270 (Ku70–Ku80) and EMD-1271 (DNA-bound Ku70–Ku80).

**Supplementary information** is available at *EMBO reports* online (<http://www.emboreports.org>).

## ACKNOWLEDGEMENTS

We thank P. Penczek (The University of Texas, Houston) for providing batches and help with the estimation of 3D variance. This work was supported by projects SAF2002-01715, SAF2005-00775 and GEN2003-20239-C06-06 from the Spanish Ministry of Education and Science (O.L.) and a Programme grant from Cancer Research UK (L.H.P.). A.R.C. is supported by an Itinerario Integrado De Insercion Profesional PhD fellowship from the Spanish National Research Council (CSIC) and the European Social Fund.

## REFERENCES

- Aravind L, Koonin EV (2000) SAP—a putative DNA-binding motif involved in chromosomal organization. *Trends Biochem Sci* **25**: 112–114
- Chacon P, Wriggers W (2002) Multi-resolution contour-based fitting of macromolecular structures. *J Mol Biol* **317**: 375–384
- Doherty AJ, Jackson SP (2001) DNA repair: how Ku makes ends meet. *Curr Biol* **11**: R920–R924
- Downs JA, Jackson SP (2004) A means to a DNA end: the many roles of Ku. *Nat Rev Mol Cell Biol* **5**: 367–378
- Drouet J, Frit P, Delteil C, de Villartay JP, Salles B, Calsou P (2006) Interplay between Ku, artemis and DNA-PKcs at DNA ends. *J Biol Chem* **281**: 27784–27793
- Falck J, Coates J, Jackson SP (2005) Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. *Nature* **434**: 605–611
- Frank J, Radermacher M, Penczek P, Zhu J, Li Y, Ladjadj M, Leith A (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* **116**: 190–199
- Gell D, Jackson SP (1999) Mapping of protein–protein interactions within the DNA-dependent protein kinase complex. *Nucleic Acids Res* **27**: 3494–3502
- Gottlieb TM, Jackson SP (1993) The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell* **72**: 131–142
- Harris R, Esposito D, Sankar A, Maman JD, Hinks JA, Pearl LH, Driscoll PC (2004) The 3D solution structure of the C-terminal region of Ku86 (Ku86CTR). *J Mol Biol* **335**: 573–582
- Jackson SP (2002) Sensing and repairing DNA double-strand breaks. *Carcinogenesis* **23**: 687–696
- Lees-Miller SP, Meek K (2003) Repair of DNA double strand breaks by non-homologous end joining. *Biochimie* **85**: 1161–1173
- Lieber MR, Ma Y, Pannicke U, Schwarz K (2003) Mechanism and regulation of human non-homologous DNA end-joining. *Nat Rev Mol Cell Biol* **4**: 712–720

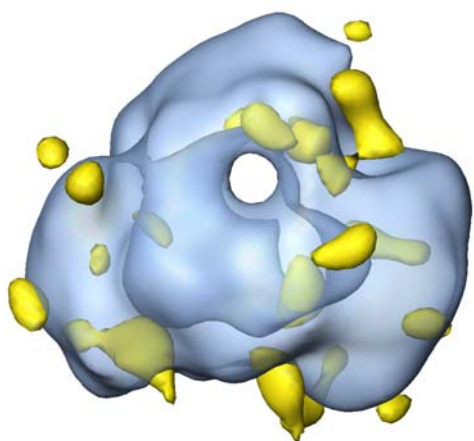
- Ludtke SJ, Baldwin PR, Chiu W (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* **128**: 82–97
- Meek K, Gupta S, Ramsden DA, Lees-Miller SP (2004) The DNA-dependent protein kinase: the director at the end. *Immunol Rev* **200**: 132–141
- Penczek PA, Yang C, Frank J, Spahn CM (2006) Estimation of variance in single-particle reconstruction using the bootstrap technique. *J Struct Biol* **154**: 168–183
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* **25**: 1605–1612
- Radermacher M (1988) Three-dimensional reconstruction of single particles from random and nonrandom tilt series. *J Electron Microscop Tech* **9**: 359–394
- Rivera-Calzada A, Maman JD, Spagnolo L, Pearl LH, Llorca O (2005) Three-dimensional structure and regulation of the DNA-dependent protein kinase catalytic subunit (DNA-PKcs). *Structure* **13**: 243–255
- Sorzano CO, Marabini R, Velazquez-Muriel J, Bilbao-Castro JR, Scheres SH, Carazo JM, Pascual-Montano A (2004) XMIPP: a new generation of an open-source image processing package for electron microscopy. *J Struct Biol* **148**: 194–204
- Spagnolo L, Rivera-Calzada A, Pearl LH, Llorca O (2006) Three-dimensional structure of the human DNA-PKcs/Ku70/Ku80 complex assembled on DNA and its implications for DNA DSB repair. *Mol Cell* **22**: 511–519
- Walker JR, Corpina RA, Goldberg J (2001) Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* **412**: 607–614
- West RB, Yaneva M, Lieber MR (1998) Productive and nonproductive complexes of Ku and DNA-dependent protein kinase at DNA termini. *Mol Cell Biol* **18**: 5908–5920
- Zhang Z, Zhu L, Lin D, Chen F, Chen DJ, Chen Y (2001) The three-dimensional structure of the C-terminal DNA-binding domain of human Ku70. *J Biol Chem* **276**: 38231–38236
- Zhang Z, Hu W, Cano L, Lee TD, Chen DJ, Chen Y (2004) Solution structure of the C-terminal domain of Ku80 suggests important sites for protein–protein interactions. *Structure* **12**: 495–502

**SUPPLEMENTARY INFORMATION FOR**

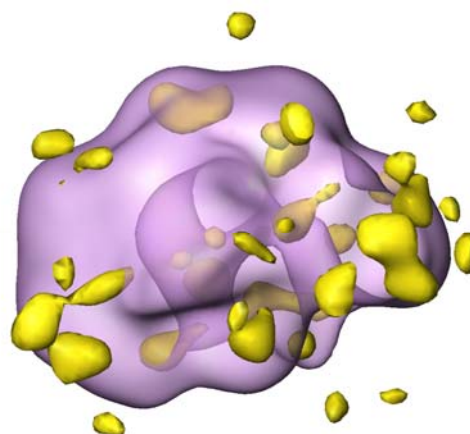
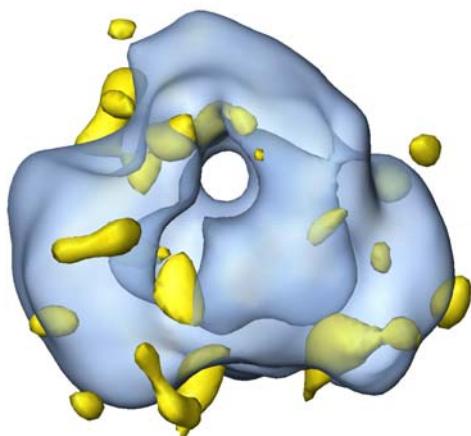
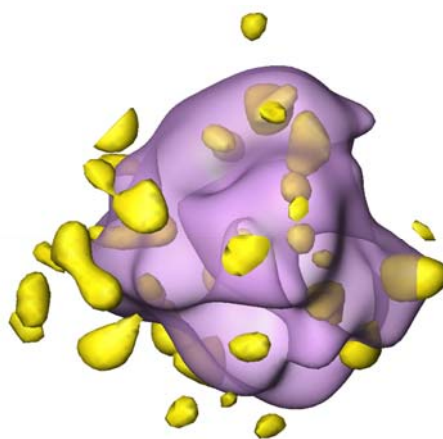
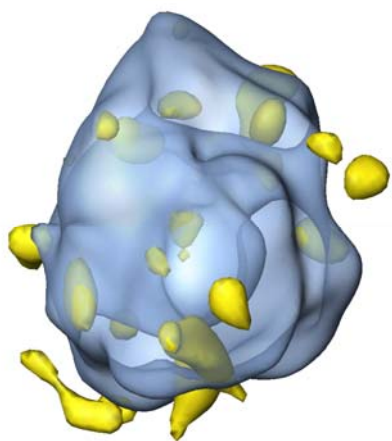
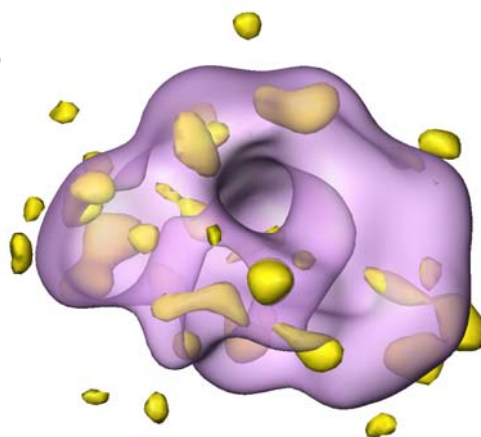
**Structural model of full-length human Ku70/Ku80 heterodimer and its  
recognition of DNA and DNA-PKcs**

**Angel Rivera-Calzada, Laura Spagnolo, Laurence H. Pearl and Oscar  
Llorca**

A



B

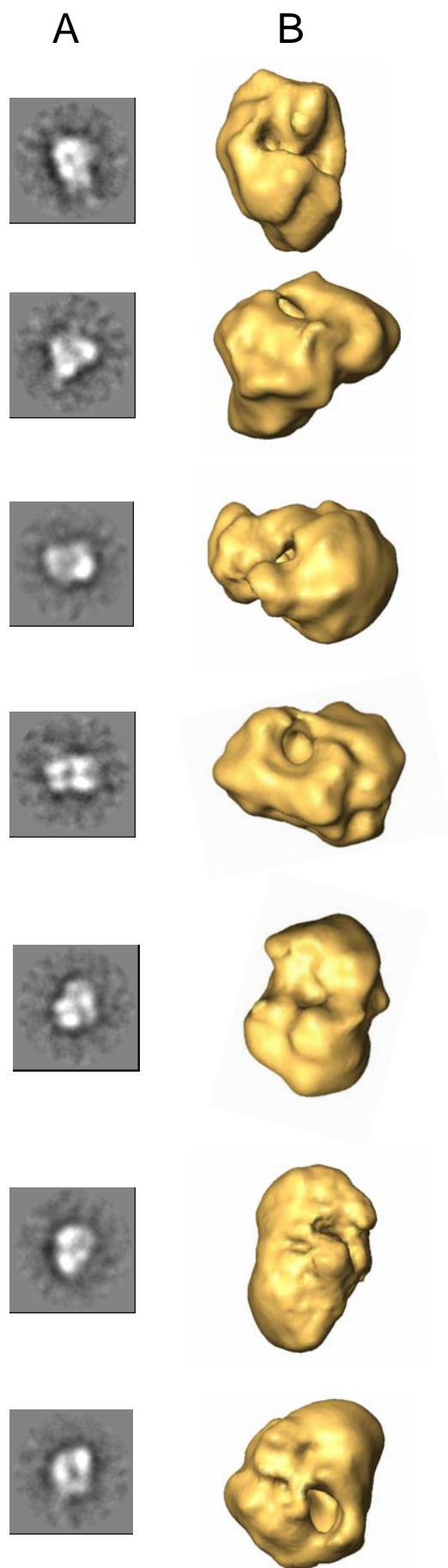


Rivera-Calzada et al.

Figure S1

**Figure S1. Variance of the 3D reconstruction of Ku (A) and DNA-bound Ku (B).** The 3D variance of the two reconstructions was calculated following the method described by Penczek et al. (Penczek et al., 2006) and using the SPIDER platform (Frank et al., 1996). The high densities of the 3D variance map are represented in yellow colour and superposed onto the 3D reconstruction of Ku (A) and DNA-Ku (B) shown as a transparent density. Several views are shown for each case. The regions of high variance within the actual density of the reconstructions are minor and they do not localize in specific domains. Even at the threshold used in this figure, most of the regions of higher variance locate around the reconstruction (be aware that some variance might look to map inside the volume at one view but it is revealed to be outside when other views of the reconstructions are inspected). Taken together, these results support the overall consistency of the reconstructions.

SPIDER batches used for estimation of the 3D variance were kindly provided by P.A. Penczek (The University of Texas, Houston, USA). We transformed our data into SPIDER format using XMIPP (Sorzano et al., 2004). A run of refinement with the particles and the final reconstruction derived using EMAN was performed with the “projection matching” command in XMIPP so that a spider file could be created that contained the Euler angles for each particle. 500 bootstrap versions of the data set were picked randomly and 3D reconstructions were performed using backprojection in SPIDER and using the Euler angles defined from the projection matching with XMIPP. The bootstrap estimate of the 3D variance was calculated using 500 resampled volumes filtered at 30 Å. The same procedure was used to calculate the bootstrap variance of the background noise. “Noise particles” were extracted directly from the micrographs. The real space 3D variance was then calculated from the bootstrap variances and the number of particles using the program VA 3R (SPIDER) as described (Penczek et al., 2006).



Rivera-Calzada et al.

Figure S2

**Figure S2. Random Conical Tilt Reconstruction (RCT) of Ku.** In order to validate the structural homogeneity of the sample and further support the 3D reconstruction derived by angular refinement, we analyzed the DNA-Ku sample by the Random Conical Tilt method (RCT) (Radermacher, 1988). This allows deriving 3D structures from defined views of the specimen.

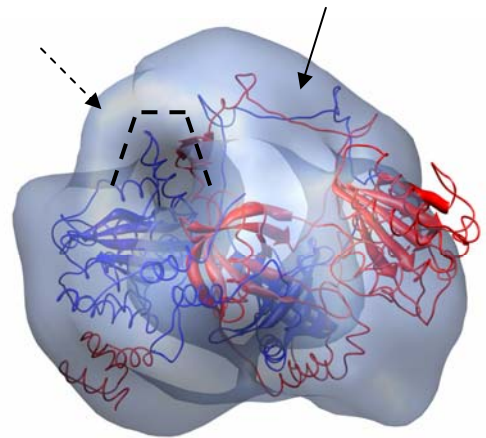
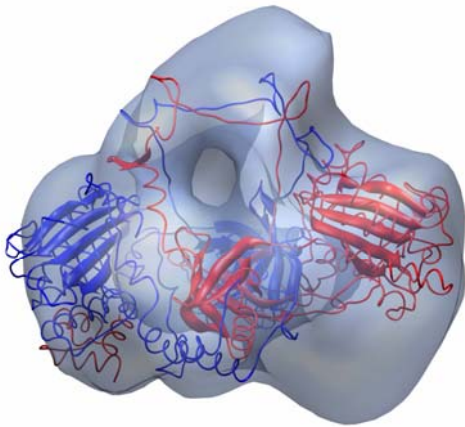
Pairs of micrographs were taken for each area, with and without tilting the specimen holder. Around 1400 pairs of particles were selected and all the processed was performed using XMIPP (Sorzano et al., 2004). Untitled data was classified and aligned in 2D. Seven distinct 2D averages of Ku were selected (A) and their tilted data used to build 3D reconstructions (B). These RCT 3D structures have been derived from defined views and therefore they are not affected by the averaging of several views taking place in angular refinement methods. The RCT structures were found to preserve the overall features of the model derived by angular refinement. The RCT reconstructions varied mainly in their relative orientation indicating they most likely represented the same structure bound to the support film in the grid in different orientations. Figure (B) only displays the view of the RCT reconstruction matching the 2D average selected for the corresponding RCT (A).



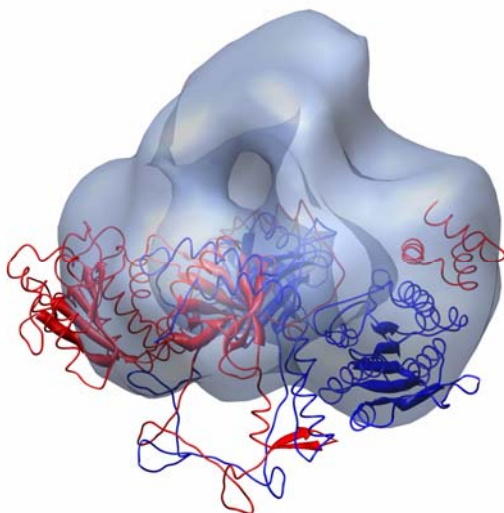
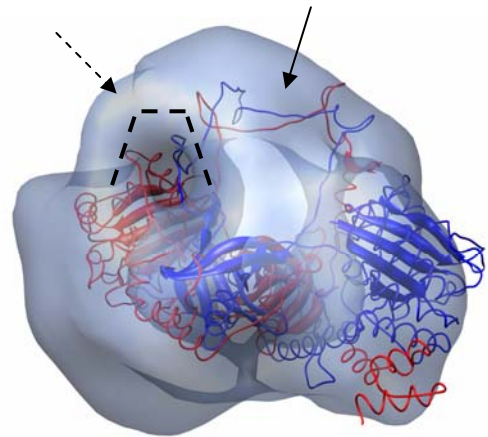
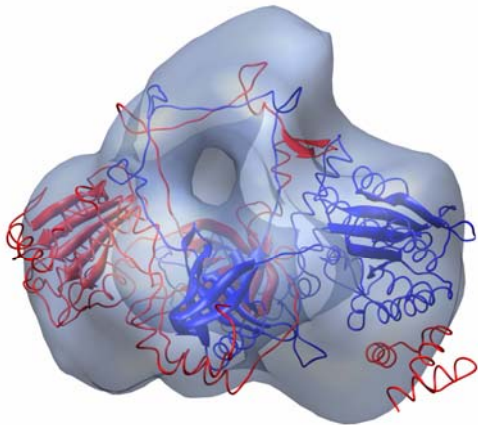
Selected handedness

Non selected handedness

1. Top solution



2. Lower rank solutions



The arrows point out that the arch of Ku, which is clearly identified in the EM reconstruction, is not fitted by the atomic structure

Rivera-Calzada et al.

Figure S3

**Figure S3. Fitting of the atomic structures into the EM volume of Ku70/Ku80.**

Atomic structures deposited in the Protein Data Bank and used in the fitting experiments were as follows: 1JEY (Ku core bound to DNA), 1JEQ (Ku core and the Ku70-SAP domain) and 1Q2Z (helical domain of the Ku80-CTD). Fitting was carried out by a 6-dimensional search using *colores* (Chacon and Wriggers, 2002) from the software SITUS (Wriggers et al., 1999). Fittings were performed using both possible handednesses in the volumes and it was found that only one allowed an adequate fitting, and this was chosen to represent the reconstructions. The figure shows results obtained for each volume in the two possible hands.

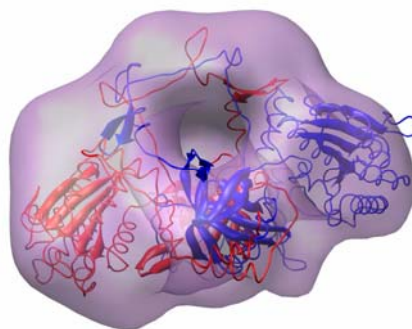
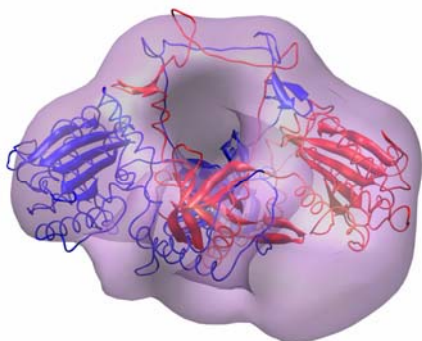
In the case of the selected handedness (left panels) the top solution placed the atomic structure within the EM density without leaving much of the atomic structure out of the volume and there was a nice agreement between the arch domain visible in the EM reconstruction and its fitting with the atomic structure. We believe that the particular elongated shape of Ku and the presence of the ring limit the options for an acceptable fitting. Also, the SAP domain introduces some asymmetry in the molecule that could contribute to define the orientation of the atomic structure within the EM map. Lower rank solutions leaved bits of the structure out of the density.

The fitting experiments performed with the non-selected handedness did not provide any satisfactory solution. In all outputs the arch domain in the atomic structure (arrow) was fitted far away from its obvious position in the reconstruction (labeled in the figure with a line and arrow).

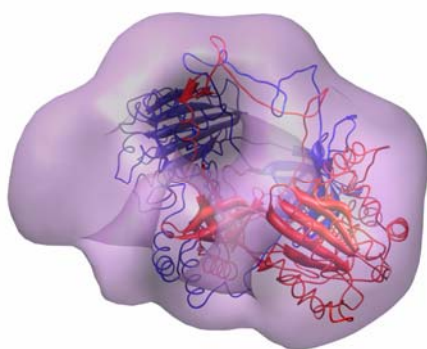
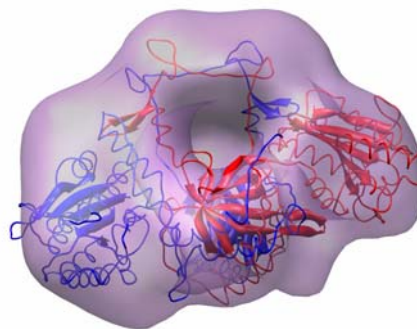
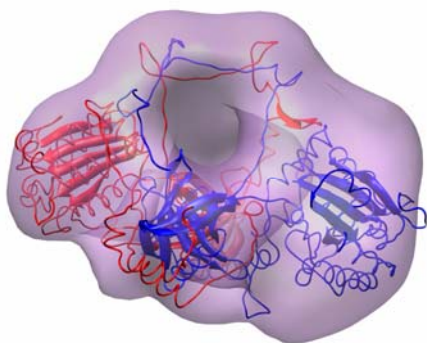
Selected handedness

Non selected handedness

1. Top solution



2. Lower rank solutions



Rivera-Calzada et al.  
Figure S4

**Figure S4. Fitting of the atomic structures into the EM volume of DNA-bound Ku70/Ku80.** Fitting experiments and representations are as in Figure S4. As shown in Fig. S4 for Ku70/Ku80, the fitting experiments performed with the non-selected handedness did not provide satisfactory solutions because the ring domain was not fitted adequately.

In the case of the selected handedness (left panels) the top solution placed the atomic structure within the EM density without leaving much of the atomic structure out of the volume and there was a correspondence between the arch domain in the EM reconstruction and the atomic structure. Importantly, the pseudo 2-fold symmetry of the atomic structure of DNA-bound Ku70/Ku80 (which lacks the Ku70-CTD and Ku80-CTD, in contrast to the atomic structure of free Ku that contains the SAP domain) implied that a 180 degree rotation of the fitting obtained as top solution provided essentially the same model. Because the density assigned to Ku80-CTD is present in both DNA-free and DNA-bound EM structures, we chose the orientation of the Ku70 and Ku80 subunits in the atomic model that matched between the two reconstructions.

**Movie S1.** 3D reconstruction of Ku70/Ku80.

**Movie S2.** 3D reconstruction of Ku70/Ku80 after the fitting of the atomic structure of Ku (Walker et al., 2001) containing Ku70-SAP domain, but lacking Ku80-CTD.

**Movie S3.** 3D reconstruction of Ku70/Ku80 DNA-bound complex.

**Movie S4.** 3D reconstruction of DNA-bound Ku70/Ku80 after the fitting of the atomic structure of DNA-Ku (Walker et al., 2001) where DNA has been removed, Ku70-SAP is not resolved and Ku80-CTD is missing.

### **General strategy of electron microscopy and image processing**

We designed a strategy for the study of the Ku70/Ku80 dimer in the electron microscope focusing in avoiding model bias so as to have a complete confidence that the resulting reconstructions were not influenced by our previous knowledge from the atomic structures (Harris et al., 2004; Walker et al., 2001; Zhang et al., 2004; Zhang et al., 2001). Images were processed using angular refinement methods in EMAN (Ludtke et al., 1999) and our tactic was not to use the atomic structures at any point of the reconstructions. Image processing was performed without assuming any symmetry, so that given the small size of the molecule we decided to use negative stain. To minimize the bias introduced by the initial model, several starting volumes were used. Initially, images were classified into homogeneous groups using reference-free methods and the best averages used to build several reference volumes using common lines (*startAny* command in EMAN). We also used artificial noisy ellipses and Gaussian blobs as starting models, and in all cases and after several iterations, the different strategies used converged to similar solutions. Then, the strategy coming from a noisy blob was further refined to obtain the final volume. Also, several experiments using the *multirefine* command in EMAN were performed to explore if there were more than one possible solution for each data set. All these experiments derived similar reconstructions for each dataset. As a further control, we employed the random conical tilt method on a few of the typical views obtained for the sample (**Fig S2**). The resulting volumes supported that all these views corresponded to different orientations of the same conformation of the Ku70/Ku80 dimer and confirmed the structure of the protein defined by angular refinement methods.

The use of a staining agent was required due to the small size of the protein, but this could potentially cause distortions in the images. During each step of the refinement, particles have been discarded during averaging within each class following procedures implemented in EMAN by playing with the values of the *classkeep* parameter. The objective was to average only those images more highly coherent and iteratively remove possible distorted images as refinement progresses. Discarded particles at each step do not contribute to the 3D output for that refinement stage but they are not removed from the initial data set to allow them to find a possible match in the next round. Consequently, approximately 30% of the single images from the initial data set were removed in the final step and did not contribute to the final reconstructions. These procedures probably help to reduce the level of distortion of the final 3D reconstruction that could be caused by the use of negative staining. Several authors have demonstrated that the 3D averaging of negatively stained images by the use of angular refinement or

angular reconstitution methods, when combined with procedures to control the quality of the particles that contribute to the 3D average, usually minimize the distortions caused by the staining agent, as measured by comparison with cryo-EM or atomic structures (Clarey et al., 2006; Llorca et al., 2005; Okorokov et al., 2004; Rodal et al., 2005; Stroupe et al., 2006). In the case of the present structures, this conclusion is further supported by:

- a) The similarity between the EM maps and the atomic structure of Ku;
- b) The agreement between the DNA-induced displacements of the SAP-domain observed in our structure and in the X-ray map;
- c) Also, reference-free averages were calculated using the *startAny* and *refine2d.py* commands in EMAN (Ludtke et al., 1999) and 2D maximum-likelihood methods (*MLalign2d* command) implemented in XMIPP (Sorzano et al., 2004). These averages are therefore independent of a 3D refinement and the fact that they are similar to those averages obtained at the final stages of angular refinement using EMAN indicates that the reconstructed volume is consistent with the original dataset.

The handedness chosen to represent the volumes was that providing the best fit with the atomic structure of Ku (**Fig. S3 and S4**). The density maps and the atomic structures were visualized with UCSF Chimera from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (Pettersen et al., 2004). The threshold used for the reconstruction was that required to approximately represent 100% of the protein mass calculated with the *volume* command in EMAN.

## Bibliography

- Clarey, M.G., Erzberger, J.P., Grob, P., Leschziner, A.E., Berger, J.M., Nogales, E. and Botchan, M. (2006) Nucleotide-dependent conformational changes in the DnaA-like core of the origin recognition complex. *Nat Struct Mol Biol*, **13**, 684-690.
- Chacon, P. and Wriggers, W. (2002) Multi-resolution contour-based fitting of macromolecular structures. *J Mol Biol*, **317**, 375-384.
- Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M. and Leith, A. (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol*, **116**, 190-199.
- Harris, R., Esposito, D., Sankar, A., Maman, J.D., Hinks, J.A., Pearl, L.H. and Driscoll, P.C. (2004) The 3D solution structure of the C-terminal region of Ku86 (Ku86CTR). *J Mol Biol*, **335**, 573-582.
- Ludtke, S.J., Baldwin, P.R. and Chiu, W. (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol*, **128**, 82-97.
- Llorca, O., Arias-Palomo, E., Zugaza, J.L. and Bustelo, X.R. (2005) Global conformational rearrangements during the activation of the GDP/GTP exchange factor Vav3. *Embo J*, **24**, 1330-1340.
- Okorokov, A.L., Orlova, E.V., Kingsbury, S.R., Bagneris, C., Gohlke, U., Williams, G.H. and Stoeber, K. (2004) Molecular structure of human geminin. *Nat Struct Mol Biol*, **11**, 1021-1022.
- Penczek, P.A., Yang, C., Frank, J. and Spahn, C.M. (2006) Estimation of variance in single-particle reconstruction using the bootstrap technique. *J Struct Biol*, **154**, 168-183.

- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*, **25**, 1605-1612.
- Radermacher, M. (1988) Three-dimensional reconstruction of single particles from random and nonrandom tilt series. *J Electron Microsc Tech*, **9**, 359-394.
- Rodal, A.A., Sokolova, O., Robins, D.B., Daugherty, K.M., Hippenmeyer, S., Riezman, H., Grigorieff, N. and Goode, B.L. (2005) Conformational changes in the Arp2/3 complex leading to actin nucleation. *Nat Struct Mol Biol*, **12**, 26-31.
- Sorzano, C.O., Marabini, R., Velazquez-Muriel, J., Bilbao-Castro, J.R., Scheres, S.H., Carazo, J.M. and Pascual-Montano, A. (2004) XMIPP: a new generation of an open-source image processing package for electron microscopy. *J Struct Biol*, **148**, 194-204.
- Stroupe, M.E., Tange, T.O., Thomas, D.R., Moore, M.J. and Grigorieff, N. (2006) The three-dimensional architecture of the EJC core. *J Mol Biol*, **360**, 743-749.
- Walker, J.R., Corpina, R.A. and Goldberg, J. (2001) Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature*, **412**, 607-614.
- Wriggers, W., Milligan, R.A. and McCammon, J.A. (1999) Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J Struct Biol*, **125**, 185-195.
- Zhang, Z., Hu, W., Cano, L., Lee, T.D., Chen, D.J. and Chen, Y. (2004) Solution structure of the C-terminal domain of Ku80 suggests important sites for protein-protein interactions. *Structure*, **12**, 495-502.
- Zhang, Z., Zhu, L., Lin, D., Chen, F., Chen, D.J. and Chen, Y. (2001) The three-dimensional structure of the C-terminal DNA-binding domain of human Ku70. *J Biol Chem*, **276**, 38231-38236.

# Visualization of DNA-induced conformational changes in the DNA repair kinase DNA-PKcs

Jasminka Boskovic, Angel Rivera-Calzada, Joseph D.Maman<sup>1</sup>, Pablo Chacón, Keith R.Willison<sup>2</sup>, Laurence H.Pearl<sup>1</sup> and Oscar Llorca<sup>3</sup>

Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas (CSIC), Ramiro de Maeztu, 9, Campus Universidad Complutense, 28040 Madrid, Spain. <sup>1</sup>Section of Structural Biology and Cancer Research UK DNA Repair Enzyme Research Group and <sup>2</sup>Cancer Research UK Centre for Cell and Molecular Biology, Institute of Cancer Research, Cancer Research UK, Chester Beatty Laboratories, 237 Fulham Road, London SW3 6JB, UK

<sup>3</sup>Corresponding author  
e-mail: ollorca@cib.csic.es

J.Boskovic, A.Rivera-Calzada and J.D.Maman contributed equally to this work

**The catalytic subunit of the DNA-dependent protein kinase (DNA-PKcs) is essential for the repair of double-stranded DNA breaks (DSBs) in non-homologous end joining (NHEJ) and during V(D)J recombination. DNA-PKcs binds single- and double-stranded DNA *in vitro*, and *in vivo* the Ku heterodimer probably helps recruit it to DSBs with high affinity. Once loaded onto DNA, DNA-PKcs acts as a scaffold for other repair factors to generate a multiprotein complex that brings the two DNA ends together. Human DNA-PKcs has been analysed by electron microscopy in the absence and presence of double-stranded DNA, and the three-dimensional reconstruction of DNA-bound DNA-PKcs displays large conformational changes when compared with the unbound protein. DNA-PKcs seems to use a palm-like domain to clip onto the DNA, and this new conformation correlates with the activation of the kinase. We suggest that the observed domain movements might help the binding and maintenance of DNA-PKcs' interaction with DNA at the sites of damage, and that these conformational changes activate the kinase.**

**Keywords:** DNA-PK/DNA repair/electron microscopy/NHEJ

## Introduction

DNA is constantly being exposed to insults that compromise its integrity and its information content. Ionizing radiation and some chemotherapeutic drugs generate double-stranded DNA breaks (DSBs) that must be repaired to avoid genomic instability (Jackson, 2002). DSBs also appear as intermediates in normal nuclear functions, such as V(D)J recombination in developing mammalian lymphocytes. The main DSB repair mechanism in mammals, known as non-homologous end joining (NHEJ), requires a multiprotein complex built around the DNA-PK kinase, and the same basic processes are also

used for V(D)J recombination. Deficiency in components of this pathway has been linked to tumour development (Pierce and Jasin, 2001).

DNA-PK comprises a 469 kDa catalytic subunit (DNA-PKcs) and a DNA targeting partner consisting of the 70 and 86 kDa heterodimeric protein Ku. DNA-PKcs and Ku associate very weakly in the absence of DNA. *In vivo*, Ku appears to recognize DSBs with high affinity as a pre-formed circular assembly, which then helps load DNA-PKcs onto the DNA (DeFazio *et al.*, 2002). It is presumed that the DNA-bound Ku–DNA-PKcs complex facilitates the loading and activation of other NHEJ components, such as the ligase IV–XRCC4 complex (Jackson, 2002). Also, two DNA-PKcs molecules can probably interact to bring the two DNA ends together (DeFazio *et al.*, 2002). Purified DNA-PKcs binds to single- and double-stranded DNA (dsDNA) ends. This interaction has been intensively demonstrated and characterized using biochemical techniques (Leuther *et al.*, 1999; Hammarsten *et al.*, 2000; Merkle *et al.*, 2002; Martensson and Hammarsten, 2002), surface plasmon resonance (West *et al.*, 1998), atomic force microscopy (Yaneva *et al.*, 1997) and electron microscopy (EM) (DeFazio *et al.*, 2002). Upon DNA binding, DNA-PKcs is activated as a serine/threonine kinase (Leuther *et al.*, 1999; Martensson and Hammarsten, 2002). This kinase activity is essential for the DNA-PKcs-dependent repair response but not for the actual recognition of DSBs. Two medium resolution structures of DNA-PKcs obtained by electron microscopy of either single particles (Chiu *et al.*, 1998) or 2D crystals (Leuther *et al.*, 1999) have been published.

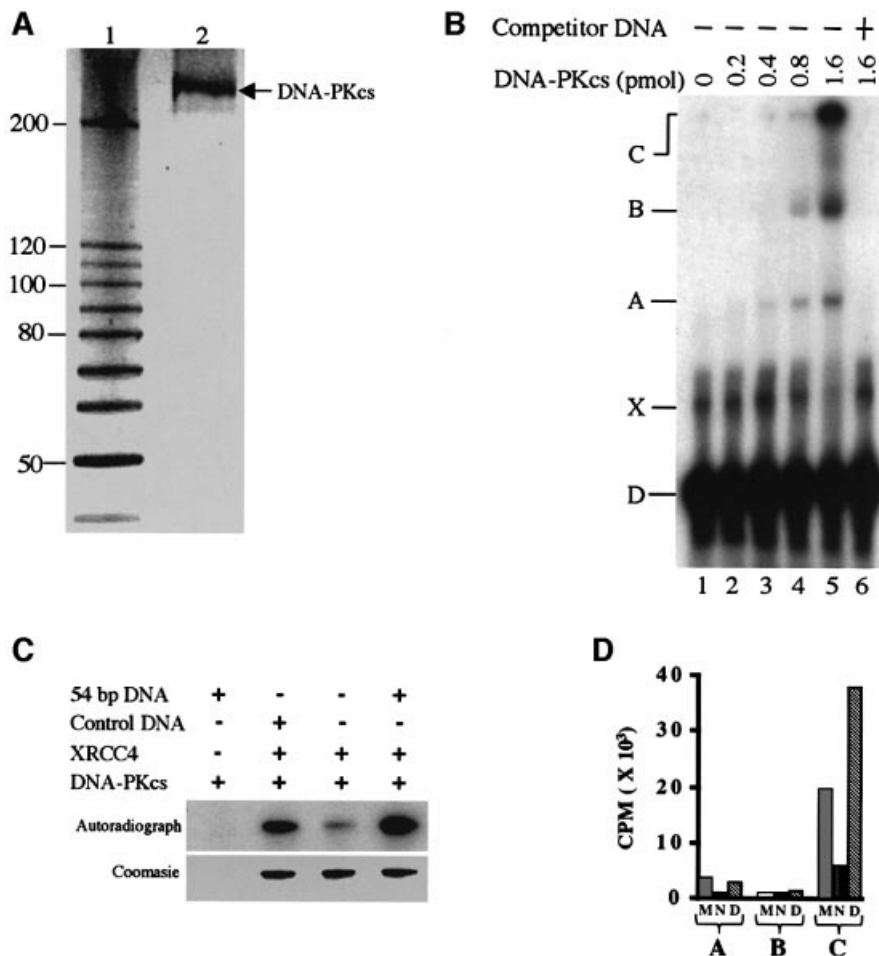
DNA-PKcs belongs to the family of the phosphatidylinositol-3-OH kinase-related kinases (PIKKs) (Jackson, 2002) that in mammals includes two other DNA repair proteins, the ATM and ATR kinases. They are all high molecular weight polypeptides sharing a conserved C-terminal region with homology to the catalytic domain of the phosphatidylinositol-3-OH kinase (Rouse and Jackson, 2002; Shiloh, 2003). ATM, ATR and DNA-PKcs are all implicated in the cell response to DSBs, although each family member has specialized functions (Jackson, 2002; Shiloh, 2003). In order to obtain further insight into the structural basis for DNA recognition by DNA-PKcs and the PIKK family of kinases, we have studied purified protein and DNA-bound complexes by single-particle EM to obtain 3D reconstructions and compare them with those of ATM (Llorca *et al.*, 2003).

## Results and discussion

### **DNA-PKcs purification, DNA binding and kinase activation**

The purified human DNA-PKcs from HeLa nuclear extract appears as a single silver-stained band in a 4–20% SDS–

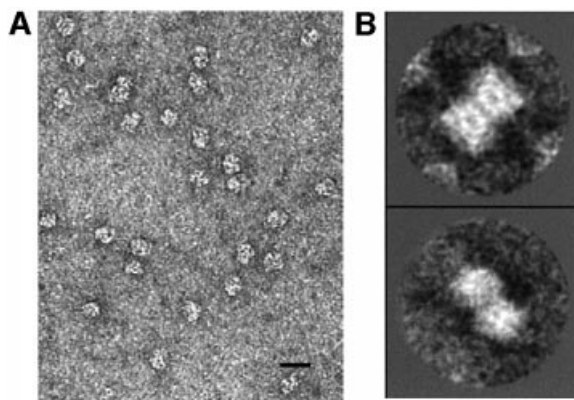




**Fig. 1.** Purification and functional assays of DNA-PKcs. (A) Silver-stained SDS-gel of purified human DNA-PKcs. Lane 1, 10 kDa ladder protein marker (Gibco); lane 2, purified DNA-PKcs. (B) EMSA of DNA-PKcs bound to 54 bp DNA. 5'-<sup>32</sup>P-labelled dsDNA (20 fmol) was titrated with an increasing amount of DNA-PKcs. In lane 6, 500 ng of unlabelled 54 bp DNA were added as competitor. (C) Autoradiography of DNA-dependent kinase activity over XRCC4. Protein bands correspond to the XRCC4 in the experiment. Control DNA, 100 bp ladder (New England BioLabs). (D) Radioactivity measurement of DNA-dependent kinase activity on DNA-PKcs (autophosphorylation) (lanes A and B) and XRCC4 (lane C). Protein bands obtained in experiments similar to those shown in 'C' were cut and the radioactivity was measured by Cerenkov counting. CPM, counts per minute; A, autophosphorylation of DNA-PKcs in the absence of XRCC4; B, autophosphorylation of DNA-PKcs in the presence of XRCC4; C, phosphorylation of XRCC4; M, phosphorylation in the presence of control DNA; N, phosphorylation in the absence of DNA; D, phosphorylation in the presence of 54 bp DNA.

polyacrylamide gel (Figure 1A). Western blot of this protein, probed with an anti-DNA-PKcs antibody (Santa Cruz Biotechnology), also resulted in a single band at the same position (unpublished data). We tested the ability of the purified protein to bind to duplex DNA by electrophoretic mobility shift assay (EMSA) (Figure 1B). We find that DNA-PKcs forms three distinct complexes (lane 5) with 54 bp DNA, and these complexes disappear in the presence of excess unlabelled DNA (lane 6). We did not investigate further the nature of these three complexes, but they have been observed previously (Leuther *et al.*, 1999) and it was concluded that complex A contains a single DNA-PKcs molecule, complex B contains two DNA-PKcs molecules and complex C (well position) contains aggregates that arise from protein-protein interactions. Band X, specific to this study, is not a protein-DNA complex because it is present in free DNA (lane 1) and its intensity decreases in the presence of excess DNA-PKcs. We suggest that this band is probably the result of DNA secondary structure as the antisense oligo is capable of forming dimers.

To gain further understanding of the functional relevance of this interaction, we investigated the DNA dependency of DNA-PKcs kinase activity (Figure 1C and D). XRCC4 has been reported to be an excellent substrate for DNA-PKcs, in a DNA-dependent manner (Critchlow *et al.*, 1997). The presence of the 54 bp DNA in the kinase reaction increases the phosphorylation of XRCC4 by DNA-PKcs by ~6-fold (Figure 1C, and D lane C), compared with the absence of DNA. Phosphorylation of XRCC4 in the presence of the 54 bp duplex was ~2-fold higher than in the presence of the control DNA (100 bp ladder DNA marker; New England BioLabs), probably because of the higher molar concentration of the 54 bp DNA. As well as phosphorylating several proteins *in vitro*, DNA-PKcs was shown to possess autophosphorylation activity that was DNA dependent (Chan and Lees-Miller, 1996), and we tested this possibility. We show that this autophosphorylation activity is present in our DNA-PKcs preparation (Figure 1D, lanes A and B).



**Fig. 2.** Electron microscopy. (A) EM field from purified DNA-PKcs. Scale bar corresponds to 100 nm. (B) 2D averaging of selected dimeric DNA-PKcs.

### Electron microscopy of human DNA-PKcs

When the purified protein was adsorbed to carbon-coated grids, negatively stained and observed in the electron microscope, we found a clean distribution of particles in different orientations (Figure 2A). Apparently, and as later confirmed, most of the particles in the EM fields corresponded to DNA-PKcs monomers, but a small percentage of images (<5%) were found to form dimers. Though the number of ‘dimer’ images collected was not very high, we managed to obtain some 2D averages (Figure 2B) that suggested that they could consist of two molecules joined side by side. The significance of our DNA-PKcs dimers is unclear, but they may reflect an intrinsic ability of the DNA-PKcs monomers to interact with each other. Such a property might be relevant for its regulation through autophosphorylation, since several reports indicate that a DNA-PKcs molecule can phosphorylate another molecule *in trans* (Chan and Lees-Miller, 1996; Soubeyrand *et al.*, 2003). Alternatively, and as proposed for ATM (Bakkenist and Kastan, 2003), the dimers might reflect a regulatory mechanism or an intermediate in the NHEJ repair process (DeFazio *et al.*, 2002).

### 3D structure of DNA-PKcs

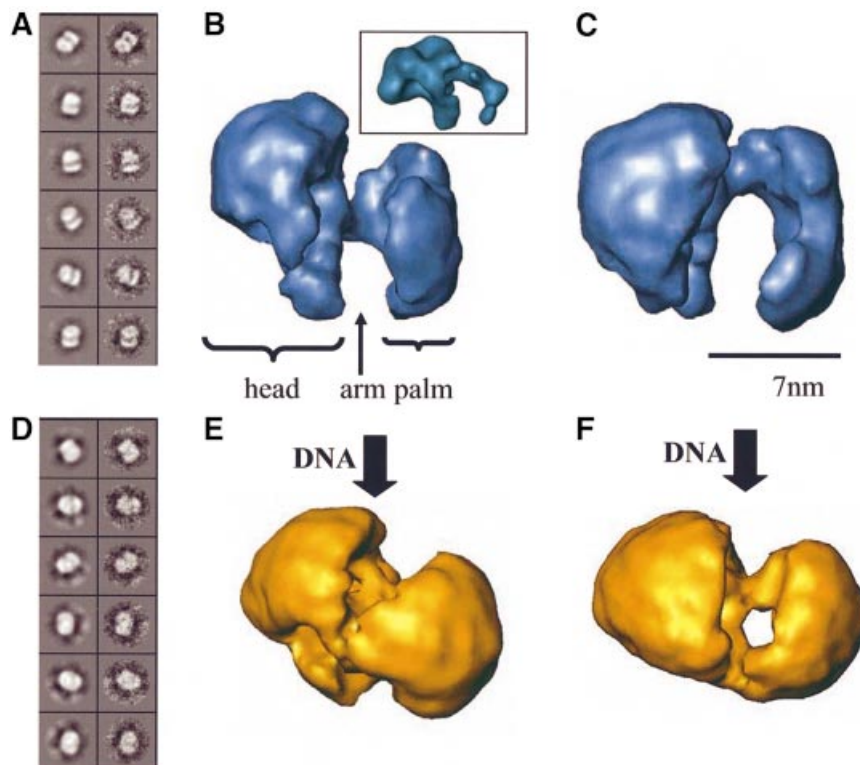
We designed a strategy to analyse the structure of DNA-PKcs and its DNA-bound complexes. As a first control, 2746 single molecule images of DNA-PKcs were extracted from the micrographs and used in a reference-free iterative refinement process (Ludtke *et al.*, 1999; Llorca *et al.*, 2001), resulting in a 3D volume (unpublished data). DNA-PKcs was also incubated with the 54 bp DNA using a 1:6 protein:DNA molar ratio to ensure that a high proportion of DNA-PKcs molecules would contain DNA. Under these conditions, 1:1 protein:DNA complexes should be favoured. EM was performed and images from single molecules were separated into different sets congruent with unique volumes. Hence, all particles observed in the EM fields were selected and extracted without any *a priori* criteria, accounting for a total of 10 468 individual images. All of them were used in a 3D classification and refinement procedure independent of any previous volume (Ludtke *et al.*, 1999). An initial volume was generated after model-free image classifica-

tion and using common lines, which was then used to generate a set of projections with a uniform distribution of orientations. Each single image was classified using these projections as references. Two criteria were used to select particles congruent with a unique volume. First, within each class, only the more similar particles were used to generate an average. Secondly, this collection of new class averages was used to reconstruct a new volume, but only the classes that best matched this new structure were actually included in the reconstruction. During the first steps of the refinement process, volumes were built from a small percentage of the total number of initial particles but, after a few iterations, and keeping these two selection criteria, a congruent volume could be generated containing a few thousands images. The particles that were not included in this first volume were processed independently following similar procedures, and a new structure was obtained.

Therefore, we found that the initial set of particles could be classified into two sets of averages (Figure 3A and D), each one compatible with just one structure. A total of 3049 particles generated class averages (Figure 3A) congruent with one volume (Figure 3B and C) which correlated with the 3D structure obtained previously from images of DNA-PKcs without the addition of any ligand. On the other hand, 3280 particles not included in the first set (Figure 3D) built a volume different from any of the two previous structures (Figure 3E and F). The remaining particles could not be classified unambiguously into any of the volumes, and were discarded. The quality of the particle assignment to each volume was confirmed because, when the two resulting volumes were used as starting models for the other set of images (to force a bias in the refinement), the intermediate volumes during the processing evolved towards the previous structures. Consequently, we had resolved two complexes from our sets of images: DNA-PKcs (Figure 3A and C) and DNA-bound DNA-PKcs (Figure 3D and F). To reinforce this interpretation, DNA within the latter volume has also been identified (see below).

The DNA-PKcs structure at 30 Å resolution (Figure 3B and 3C) contains three distinct regions we have labelled as ‘head’, ‘palm’ and an ‘arm’ that connects the head to the palm. The head encloses a cavity surrounded by protein density and two openings in the front of the molecule. DNA-PKcs has a longitudinal axis of ~14 nm and a thickness ranging from 7 to 10 nm. This reconstruction closely resembles the previously published structure of the protein obtained by electron crystallography (Leuther *et al.*, 1999), though some clear differences are also visible and mainly located at the level of the ‘palm’ domain. In the crystal structure, this domain forms a compact block that contains a flexible arm, whereas in our 3D reconstruction this domain creates a palm-shaped feature. Upon inspection of the two volumes, these differences seem to be due to an ~90° rotation of the palm domain in our reconstruction compared with the crystallographic structure.

The 3D structure of ATM, at a similar resolution, has been solved recently (Llorca *et al.*, 2003; see inset in Figure 3B) using a single-particle EM approach, thus allowing direct comparison of the two structures. DNA-PKcs and ATM display strong structural similarities, including head and arm domains. Recently, it has been



**Fig. 3.** 3D structure of DNA-PKcs and its DNA-bound complexes. (A) Classes from the DNA-PKcs volume. Each pair of images represents a projection from the final volume and the corresponding average from the original data. (B and C) Two views of the DNA-PKcs volume. The inset in (B) shows a comparable view of ATM. Scale bar represents 7 nm. (D) Classes from the DNA-bound structure. (E and F) Two views of the DNA-bound complex at similar orientations to those in (B and C).

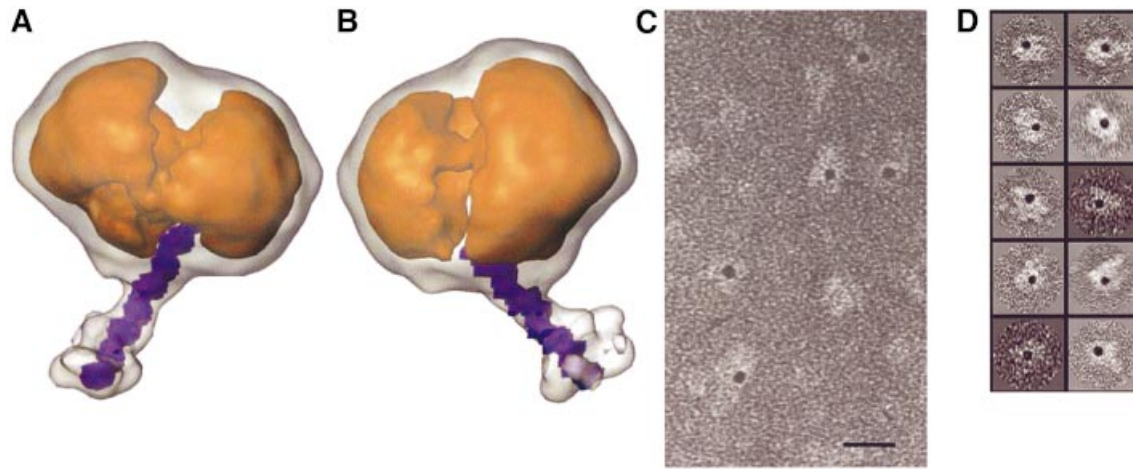
reported that the PIKK family of kinases may share a common structural architecture based on a low sequence conservation helical unit known as HEAT repeats (Perry and Kleckner, 2003), and the 3D structures of DNA-PKcs and ATM support these hypotheses.

#### **Conformational changes in DNA-PKcs upon DNA binding**

DNA-bound DNA-PKcs, three-dimensionally reconstructed at 30 Å resolution, shows discernible head, arm and palm domains, but substantial conformational changes have taken place compared with the unbound protein (Figure 3E and F). Views of the DNA-free and DNA-bound complexes at compatible orientations are shown (Figure 3). The palm domain in the DNA-bound DNA-PKcs is bent and comes into contact with the head. In top views of the two reconstructions (Figure 3C versus F), these domain movements are most clearly visible. As a consequence of DNA binding, not only is the palm bending but also the arm moves inwards so that the palm actually achieves contact with the head (Figure 3F). As a result of all these modifications, a small channel is left between head and palm, which is sufficiently large to accommodate dsDNA. It is tempting to speculate that these DNA-driven conformational changes could be responsible for the activation of the kinase activity of DNA-PKcs that we observed in the presence of the 54 bp DNA.

#### **Localization of the dsDNA in DNA-bound DNA-PKcs complexes**

A common theme solving DNA–protein complexes by single-particle EM is the difficulty in the detection of the DNA component, perhaps due to DNA’s flexibility or the way it interacts with the staining agent. The DNA-bound DNA-PKcs complex (Figure 3E and F) displays large conformational changes, but no DNA is clearly observed. However, when this volume is examined at a slightly lower threshold to visualize other structural features clearly still above the background noise, a long density comes out of the reconstruction (Figure 4A and B, white transparency superimposed to the same volume at the usual threshold). This density has the shape and dimensions that would be expected for a short segment of dsDNA. When the reconstruction from unbound DNA-PKcs, generated out of images from the same micrographs as the DNA-bound complex, was analysed using the same lower threshold, this extra density was never observed (unpublished data). This observation supports our view that we are in fact visualizing the DNA in the complex and, additionally, reinforces our claim that the original single images have been correctly assigned to either of the two volumes. The DNA comes out of the complex from just one side and in a reasonably well defined position. The lower threshold needed to observe the DNA density might reflect that we are averaging a small range of positions within the protein complex.



**Fig. 4.** Localization of DNA in the DNA-bound DNA-PKcs complexes. (A and B) Views of the DNA–DNA-PKcs volume at two different thresholds. The image at the lower threshold is shown as a white transparency superimposed on the volume at the usual threshold as an orange solid object. A 36 bp DNA was scaled to 4.9 Å/voxel, docked into the extra density and displayed in purple. (C) An area from micrographs where gold–DNA complexes were incubated with DNA-PKcs. (D) A gallery of selected individual images from (C).

As an independent method to detect the presence of DNA bound to DNA-PKcs in the images obtained by EM, we labelled the 54 bp oligo with a gold conjugate. A 5'-biotinylated derivative version of the DNA was incubated with a streptavidin-coupled gold conjugate with a gold particle of 5 nm in diameter to generate gold–DNA complexes. Gold–DNA was incubated with purified DNA-PKcs at a 1:1 molar ratio, and the preparation was added to grids and observed in the microscope. The presence of streptavidin–gold in one end of the DNA should preclude the interaction of DNA-PKcs, leaving just one free end at the opposite side. As controls for the specificity of the results, we also observed streptavidin–gold on its own, DNA-PKcs plus streptavidin–gold without DNA, and DNA-PKcs plus streptavidin–gold plus a non-biotinylated form of dsDNA (unpublished data). Only in the presence of the gold–DNA complex were a large number of DNA-PKcs molecules labelled with one gold particle observed (Figure 4C and D), and this confirms the presence of DNA-bound DNA-PKcs in our sample. Some gold densities were also found that were not associated with the protein, corresponding to free gold–DNA.

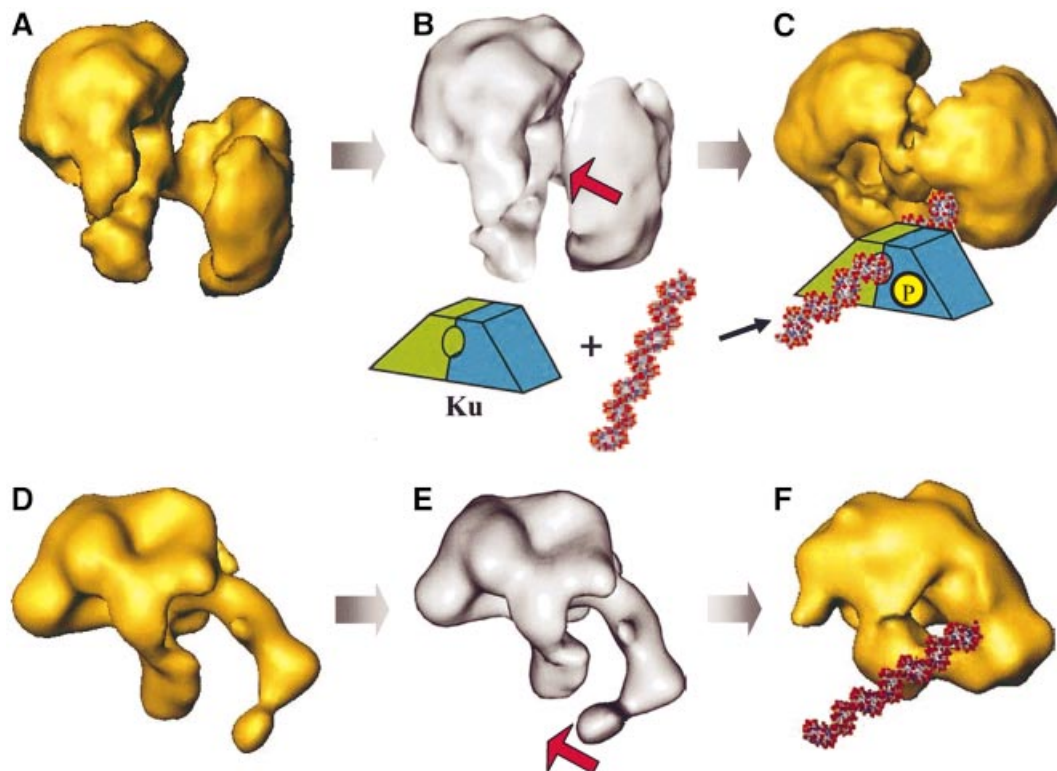
#### **Model for kinase activation upon DNA binding**

As shown in Figure 1, binding of the 54 bp DNA to DNA-PKcs strongly activated its kinase activity on its natural substrate, the XRCC4 protein. The DNA density coming out from DNA-bound DNA-PKcs could be docked very well to a 36 bp dsDNA (Figure 4A and B; DNA in purple within the white transparency). Therefore, it seems that ~18 bp, out of the 54 bp DNA, are in contact with DNA-PKcs. Previous studies indicate that a minimum length of ~15 bp of dsDNA was required to activate DNA-PKcs (Leuther *et al.*, 1999; Hammarsten *et al.*, 2000; Martensson and Hammarsten, 2002) and, although smaller dsDNA molecules bind with comparable affinity, they do not activate the kinase. Furthermore, increasing the length of the dsDNA beyond 20 bp does not elicit a further significant increase in kinase activation. When considered in light of our EM reconstruction, these data are most

consistent with a model in which DNA-PKcs kinase activation on DNA binding may be the result of closure of the cavity between the head and palm. The stabilization of this conformational change could require a minimum of two points of contact between DNA-PKcs and dsDNA: (i) a binding site for the broken end of the DNA in the base of the palm; and (ii) an interaction between the edge of the head domain and the sugar–phosphate backbone of the DNA 15–20 bp out from the end, as the DNA emerges from the cavity between the head and palm (Figure 4A and B). An activating DNA molecule would thereby bridge the palm and head domains and stabilize the closed conformation, while a short dsDNA molecule would only occupy the end-binding site and would fail to activate the kinase. Extending the length of dsDNA over and above that required for making the bridging contact with the head would add no further stability to the closed conformation and would not further activate the kinase, consistent with the observed data.

#### **Model for dsDNA recognition by DNA-PKcs and ATM**

Previous reports have already provided information and some hypotheses on how DNA-PKcs could recognize DNA (Leuther *et al.*, 1999; DeFazio *et al.*, 2002). Some models suggest that both single-stranded and dsDNA are bound by DNA-PKcs through its different holes and cavities, and that a flexible subdomain could be used to mediate this interaction (Leuther *et al.*, 1999; Martensson and Hammarsten, 2002). We show, for the first time, a comparison between the 3D structures of DNA-PKcs and a DNA-bound complex at medium resolution. We observe that upon DNA incubation, DNA-PKcs displays conformational changes that encircle the DNA through a clamp closure motion (Figure 5A and C). The gross mechanistic aspects of this model can be generalized to ATM (Figure 5D and F). The palm domain of DNA-PKcs can grab and encircle DNA, and this interaction takes place through DNA ends, which are somehow specifically recognized. During NHEJ, it is likely that the Ku protein



**Fig. 5.** Model for DNA recognition by DNA-PKcs (A–C) and ATM (D–F). (A and D) The apo-EM maps; (C) and (F) correspond to the DNA-bound reconstructions. (B and E) Theoretical intermediate states calculated using vibrational analysis and corresponding to the lowest-frequency normal mode captured from the DNA-PKcs (A) and ATM (D) DNA-free electron density maps. Red arrows indicate the direction of the conformational changes in the arm domains. P stands for inorganic phosphate.

acts as a recruiting factor to allow high affinity binding of DNA-PKcs to broken DNA. However, after Ku–DNA-PKcs interaction, Ku translocates into the DNA leaving DNA-PKcs at the DNA free end, bound probably through the conformational changes observed here. A Ku-like factor may also help ATM to bind to DNA, although it has not been included in the model. At this stage, DNA-PK could act as a scaffolding protein for the assembly of other components in the repair machinery. In addition, DNA-induced conformational changes probably activate the kinase to phosphorylate substrates implicated in DNA repair, including Ku itself. A minimum length of DNA is required to induce and stabilize these conformational changes and the subsequent kinase activation.

#### **Prediction of domain motions using vibrational analysis**

Vibrational analysis has been shown to be a powerful tool to predict potential motions of macromolecular machines where the low frequency modes efficiently describe the experimentally observed global domain movements (Ming *et al.*, 2002; Chacón *et al.*, 2003). We have carried out this type of analysis of both DNA-PKcs and ATM using their DNA-free EM density maps (Figure 5A and D) and without any other additional input information. The normal mode analysis was performed with an elastic model built from the EM structures with the help of vector quantization techniques to represent the density maps in real space with a number of landmark points. A total of 1000 landmarks with a connecting spring distance cut-off

of 14 Å were used. Normal mode analysis decomposed the possible motions of the elastic model in an orthonormal basis of modes (i.e. displacements) sorted by frequency. The first mode displacements were applied to the density maps by interpolation with the 3D thin-plate spline method (Bookstein, 1991). The amplitudes of the motions were scaled to better visualize the mode motion.

The intrinsic global flexibility of both proteins predicted large motions where the lowest frequency mode describes the closure of the palm domain towards the head domain in DNA-PKcs (Figure 5B and Supplementary movie1 available at *The EMBO Journal Online*) and the closure clamp motion of the arm-like domain in ATM (Figure 5E and Supplementary movie2). These predictions are in full agreement with the conformational changes we observed experimentally in our EM reconstructions (Figure 5C and F). These analyses support the likelihood of the conformational changes we have observed by EM and provide a model of how the transition between the DNA-free and DNA-bound conformations might take place. Note that even though the lowest frequency mode described the basic motions, other low frequency modes or a lineal combination of modes can complement the description of the movements (unpublished data).

In conclusion, we have found that both ATM (Llorca *et al.*, 2003) and DNA-PKcs are built with a similar general architecture where an equivalent structural region, the arm domain, is at the centre of the conformational changes upon DNA recognition. At least in the case of DNA-PKcs, these movements seem to be responsible for

the activation of its kinase activity, but we can hypothesize that other PIKK kinases might share these common properties during DNA binding.

## Materials and methods

### Sample preparation

DNA-PKcs was purified as previously described (Gell and Jackson, 1999) with a few modifications. Briefly, HeLa nuclear extract (Computer Cell Culture, Mons, Belgium) was fractionated on a Q-Sepharose column (Pharmacia) followed by a heparin-agarose column (Sigma). Fractions containing DNA-PKcs were separated by a phenyl-Sepharose column (Pharmacia) in 0.75 M ammonium sulfate, developed by a linear gradient of 0.75–0 M ammonium sulfate. DNA-PKcs-containing fractions were purified further by a Mono-S column (Pharmacia). Pure DNA-PKcs-containing fractions were pooled, dialysed against buffer A [25 mM HEPES pH 7.6, 50 mM KCl, 1 mM MgCl<sub>2</sub>, 1 mM dithiothreitol (DTT), 0.5 mM EDTA and 20% glycerol] and stored at –80°C.

The 54 bp blunt-ended dsDNA was made by annealing two, HPLC- or PAGE-purified, complementary 54 base oligonucleotides (5'-GGCCG-CACGCGTCCACCATGGGGTACAACACTACGATCTAGCTTCATGC-ACCGGAC-3'). The quality of the annealing was tested by measuring the hyperchromicity following digestion by snake venom phosphodiesterase 1 (Amersham).

Human XRCC4 was cloned from human kidney cDNA (Clontech) into pRSET-B (Invitrogen) and expressed in BL21-CodonPlus(DE3) cells (Stratagene). The His<sub>6</sub>-tagged recombinant protein was purified by an immobilized metal-affinity column (Clontech) followed by ion exchange chromatography on a HiTrap-Q column (Pharmacia) and, finally, by gel filtration on a Superdex 200 column (Pharmacia). The purified XRCC4 was stored at –80°C in 20 mM Tris-HCl pH 8.0, 10% glycerol, 1 M NaCl, 1 mM EDTA, 2 mM DTT, 0.1 mM phenylmethylsulfonyl fluoride (PMSF).

### Electrophoretic mobility shift assay (EMSA)

EMSA was carried out essentially as described (Hammarsten and Chu, 1998) with minor modifications. Binding reactions were performed in buffer B (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 1 mM EDTA, 1 mM DTT and 5% glycerol) using 0.5 ng of <sup>32</sup>P-labelled dsDNA (54 bp). The protein was always added last, followed by 30 min incubation at room temperature. Protein-DNA complexes were resolved on a 4% polyacrylamide gel in ×0.25 TGE buffer at 4°C.

### Kinase assays

The DNA-dependent kinase activity was tested by the phosphorylation of XRCC4 in the presence or absence of DNA. A 1 µg aliquot of XRCC4 was phosphorylated in the presence of 10 ng/µl DNA, 10 mM Tris-HCl pH 7.5, 50 mM KCl, 10 mM MgCl<sub>2</sub>, 1 mM DTT, 2 µl of [<sup>32</sup>P]ATP (6000 Ci/mmol, 10 mCi/ml) and 200 ng DNA-PKcs, in a volume of 20 µl, for 30 min, at 30°C. The reaction products were analysed by SDS-PAGE, and the gel was stained with Coomassie blue, dried onto 3MM paper and developed by autoradiography. Protein bands corresponding to DNA-PKcs and XRCC4 were cut and the radioactivity was measured by Cerenkov counting.

### Electron microscopy

A few microlitres of purified DNA-PKcs and its DNA-bound complexes were diluted to 0.2 mg/ml in buffer (50 mM Tris-HCl pH 7.5, 50 mM KCl, 1 mM MgCl<sub>2</sub> and 20% glycerol) and applied to carbon-coated grids. The sample was negatively stained with 2% uranyl acetate and observed in Philips CM100 and a JEOL 1230 electron microscopes. The microscopes were operated at 100 kV and the micrographs recorded at 0° tilt and a magnification of 39 000× under low dose conditions. A Leafscan45 scanner (Leaf Systems Inc.) was used to scan the pictures with a 10 µm step size.

### Image processing and three-dimensional reconstructions

A total of 2746 particles from purified DNA-PKcs and 10 468 particles from the DNA incubated sample were selected from the micrographs at a final 4.9 Å/pixel using XMIPP (Marabini *et al.*, 1996). Particles initially were centred and averaged in two dimensions using XMIPP. Image classification and 3D refinement were carried out without assuming any starting model using EMAN (Ludtke *et al.*, 1999). Resolution was estimated by Fourier shell correlation between two independent volumes. Using the criteria of a 0.5 correlation coefficient, the volumes showed a

resolution of 30 Å. The handedness of the structure was determined by comparison with the absolute handedness calculated from 2D crystals (Leuther *et al.*, 1999). Docking of the 36 bp DNA into the DNA-bound volume was done after scaling to 4.9 Å/voxel.

### Gold labelling

Streptavidin-gold with a gold particle of 5 nm in diameter (BB International) was incubated with a 5'-biotinylated 54 bp DNA to generate gold-DNA complexes. Afterwards, DNA-PKcs was incubated with the pre-formed gold-DNA complex for 15 min at a 1:1 molar ratio, which is lower than that used in the 3D reconstructions, in order to avoid a strong background of free gold-DNA in the EM fields. The reaction was applied to carbon-coated grids, negatively stained and observed in the electron microscope.

### Vibrational analysis

The full procedure to carry out this analysis has already been described in detail (Chacón *et al.*, 2003). Briefly, a normal mode analysis is performed with an elastic model built from a given EM structure, DNA-free ATM (Llorca *et al.*, 2003) and DNA-PKcs in our case. Normal mode analysis decomposes the possible motions of the elastic model in an orthonormal basis of modes (i.e. displacements) sorted by frequency. The low frequency modes have been found to capture efficiently the functional global motions exhibited by large macromolecules (Chacón *et al.*, 2003).

### Supplementary data

Supplementary data are available at *The EMBO Journal* Online.

## Acknowledgements

We thank Marin van Heel at Imperial College London for the use of their microscopes, and John Barber and John Nield for the use of his Leafscan45 scanner (Leaf Systems Inc.). We are grateful to Aidan Doherty for contributions at an early stage of the project and to Steve Jackson for advice and assistance with DNA-PKcs purification. This work has been supported by a project SAF2002-01715 from the 'Ministerio de Ciencia y Tecnología' of Spain (O.L.), and Programme Grants from Cancer Research UK (K.R.W. and L.H.P.).

## References

- Bakkenist,C.J. and Kastan,M.B. (2003) DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature*, **30**, 421, 499–506.
- Bookstein,F.L. (1991) *Morphometric Tools for Landmark Data*. Cambridge University Press, Cambridge.
- Chacón,P., Tama,F. and Wriggers,W. (2003) Mega-Dalton biomolecular motion captured from electron microscopy reconstructions. *J. Mol. Biol.*, **326**, 485–492.
- Chan,D.W. and Lees-Miller,S.P. (1996) The DNA-dependent protein kinase is inactivated by autophosphorylation of the catalytic subunit. *J. Biol. Chem.*, **271**, 8936–8941.
- Chiu,C.Y., Cary,R.B., Chen,D.J., Peterson,S.R. and Stewart,P.L. (1998) Cryo-EM imaging of the catalytic subunit of the DNA-dependent protein kinase. *J. Mol. Biol.* **284**, 1075–1081.
- Critchlow,S.E., Bowater,R.P. and Jackson,S.P. (1997) Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. *Curr. Biol.*, **7**, 588–598.
- DeFazio,L.G., Stansel,R.M., Griffith,J.D. and Gilbert,C. (2002) Synapsis of DNA ends by DNA-dependent protein kinase. *EMBO J.*, **12**, 3192–3200.
- Gell,D. and Jackson,S.P. (1999) Mapping of protein-protein interactions within the DNA-dependent protein kinase complex. *Nucleic Acids Res.*, **27**, 3494–3502.
- Hammarsten,O. and Chu,G. (1998) DNA-dependent protein kinase: DNA binding and activation in the absence of Ku. *Proc. Natl Acad. Sci. USA*, **95**, 525–530.
- Hammarsten,O., DeFazio,L.G. and Chu,G. (2000) Activation of DNA-dependent protein kinase by single-stranded DNA ends. *J. Biol. Chem.*, **275**, 1541–1550.
- Jackson,S.P. (2002) Sensing and repairing DNA double-strand breaks. *Carcinogenesis*, **23**, 687–696.
- Leuther,K.K., Hammarsten,O., Kornberg,R.D. and Gilbert,C. (1999) Structure of DNA-dependent protein kinase: implications for its regulation by DNA. *EMBO J.*, **18**, 1114–1123.

- Llorca,O. *et al.* (2001) The 'sequential allosteric ring' mechanism in the eukaryotic chaperonin-assisted folding of actin and tubulin. *EMBO J.*, **20**, 4065–75.
- Llorca,O., Rivera-Calzada,A., Grantham,J. and Willison,K.R. (2003) Electron microscopy and 3D reconstructions reveal that human ATM kinase uses an arm-like domain to clamp around double-stranded DNA. *Oncogene*, **22**, 3867–3874.
- Ludtke,S.J., Baldwin,P.R. and Chiu,W. (1999) EMAN: semi automated software for high-resolution single-particle reconstructions. *J. Struct. Biol.*, **128**, 82–97.
- Marabini,R. *et al.* (1996) Xmipp: an image processing package for electron microscopy. *J. Struct. Biol.*, **116**, 237–240.
- Martensson,S. and Hammarsten,O. (2002) DNA-dependent protein kinase catalytic subunit. *J. Biol. Chem.*, **277**, 3020–3029.
- Merkle,D. *et al.* (2002) The DNA-dependent protein kinase interacts with DNA to form a protein–DNA complex that is disrupted by phosphorylation. *Biochemistry*, **41**, 12706–12714.
- Ming,D., Kong,Y., Wakil,S.J., Brink,J. and Ma,J. (2002) Domain movements in human fatty acid synthase by quantized elastic deformational model. *Proc. Natl Acad. Sci. USA*, **99**, 7895–9.
- Perry,J. and Kleckner,N. (2003) The ATRs, ATMs and TORs are giant HEAT repeat proteins. *Cell*, **112**, 151–155.
- Pierce,A.J. and Jasin,M. (2001) NHEJ deficiency and disease. *Mol. Cell*, **8**, 1160–1161.
- Rouse,J. and Jackson,S.P. (2002) Interfaces between the detection, signalling and repair of DNA damage. *Science*, **297**, 547–551.
- Shiloh,Y. (2003) ATM and related protein kinases: safeguarding genome integrity. *Nature Rev. Cancer*, **3**, 155–168.
- Soubeyrand,S., Pope,L., Pakuts, B and Hache,R.J. (2003) Threonines 2638/2647 in DNA-PK are essential for cellular resistance to ionizing radiation. *Cancer Res.*, **15**, 1198–1201.
- West,R.B., Yaneva,M. and Lieber,M.R. (1998) Productive and non-productive complexes of Ku and DNA-dependent protein kinase at DNA termini. *Mol. Cell Biol.*, **18**, 5908–5920.
- Yaneva,M., Kowalewski,T. and Lieber,M.R. (1997) Interaction of DNA-dependent protein kinase with DNA and with Ku: biochemical and atomic-force microscopy studies. *EMBO J.*, **16**, 5098–5112.

*Received July 18, 2003; revised September 4, 2003;  
accepted September 9, 2003*

# Three-Dimensional Structure and Regulation of the DNA-Dependent Protein Kinase Catalytic Subunit (DNA-PKcs)

Angel Rivera-Calzada,<sup>1</sup> Joseph P. Maman,<sup>2,3</sup>  
Laura Spagnolo,<sup>2</sup> Laurence H. Pearl,<sup>2,\*</sup>  
and Oscar Llorca<sup>1,\*\*</sup>

<sup>1</sup>Centro de Investigaciones Biológicas  
Consejo Superior de Investigaciones Científicas  
(CSIC)

Ramiro de Maeztu 9  
Campus Universidad Complutense  
28040 Madrid  
Spain

<sup>2</sup>Cancer Research UK DNA Repair Enzyme  
Research Group

Section of Structural Biology  
Institute of Cancer Research  
Chester Beatty Laboratories  
237 Fulham Road  
London SW3 6JB  
United Kingdom

## Summary

DNA-PKcs is a large PI3-kinase-related protein kinase (PIKK) that plays a central role in DNA double-strand break (DSB) repair via nonhomologous end joining. Using cryo-electron microscopy we have now generated a  $\sim 13$  Å three-dimensional map of DNA-PKcs, revealing the overall architecture and topology of the 4128 residue polypeptide chain and allowing location of domains. The highly conserved C-terminal PIKK catalytic domain forms a central structure from which FAT and FATC domains protrude. Conformational changes observed in these domains on DNA binding suggest that they transduce DNA-induced conformational changes to the catalytic core and regulate kinase activity. The N-terminal segments form long curved tubular-shaped domains based on helical repeats to create interacting surfaces required for macromolecular assembly. Comparison of DNA-PKcs with another PIKK DNA repair factor, ATM, defines a common architecture for this important protein family

## Introduction

Cells are exposed to DNA-damaging agents that can affect their genome integrity and pose a huge challenge for their survival (Khanna and Jackson, 2001). The most cytotoxic form of DNA damage, double-stranded breaks (DSBs), are generated by ionizing radiation and radiomimetic drugs used in chemotherapy. DSBs are repaired by one of two pathways: homologous recombination (HR), which is functional only during late S and G2 phases of the cell cycle, and nonhomologous end joining (NHEJ), which is the predominant mechanism

during G<sub>0</sub>, G<sub>1</sub>, and early S phase (Jackson, 2002; Lees-Miller and Meek, 2003; Lieber et al., 2003). As well as dealing with adventitious DSBs, the NHEJ pathway is essential for repair of programmed DSBs generated by V(D)J recombination during development of immune diversity, so that genetic deficiencies in NHEJ components are typified by radiosensitivity and immunodeficiency (Taccioli et al., 1993).

In the current view of NHEJ, primary recognition of DSBs is the function of the DNA-dependent protein kinase (DNA-PK) holoenzyme, consisting of a heterodimer of the homologous Ku70 and Ku86 proteins, and the 460 kDa catalytic subunit (DNA-PKcs) (Jackson, 2002). The Ku70:86 heterodimer forms a clamp structure (Walker et al., 2001) which slides onto broken DNA ends (Khanna and Jackson, 2001; Jackson, 2002). Recruitment of DNA-PKcs to DNA bound Ku is facilitated by interaction with a small helical domain at the C terminus of Ku86 (Singleton et al., 1999; Gell and Jackson, 1999; Harris et al., 2004). DNA-PKcs itself interacts directly with DNA ends, which substantially stimulates its kinase activity (Chan et al., 1996; Critchlow et al., 1997; Leuther et al., 1999; Merkle et al., 2002), resulting in phosphorylation of DNA-PKcs itself and other proteins involved in NHEJ. The precise role of DNA-PKcs phosphorylation is not fully understood, but it is required for regulation of the NHEJ pathway (Lieber et al., 2003), and its pharmacological inhibition renders cells radiosensitive (Ismail et al., 2004). In addition to its catalytic activity, DNA-PKcs acts as a scaffold, facilitating synthesis of broken ends by homodimerization (DeFazio et al., 2002) and provides binding sites for downstream components of NHEJ, such as Artemis, a nuclease required for processing of broken ends (Ma et al., 2002), and the XRCC4-DNA ligase IV complex that performs the final repair step of ligation (Critchlow et al., 1997).

DNA-PKcs is a serine/threonine protein kinase belonging to the phosphatidylinositol-3-OH kinase (PI3K)-related kinase family (PIKKs), which includes other large proteins involved in DNA repair signaling, such as ATM and ATR (Shiloh, 2003) and the nutrient responsive regulator of translation mTOR (Raught et al., 2001). A common feature of these proteins is their size (2549, 2644, 3056, and 4128 residues, respectively, for human mTOR, ATR, ATM, and DNA-PKcs) and lack of homology apart from the C-terminal PI3K-related catalytic domain (residues 3649–4011 in DNA-PKcs), and a weakly conserved  $\sim 500$ -residue region (the FAT domain) (Bosotti et al., 2000) immediately N-terminal of this (Figure 1A). The catalytic domain in DNA-PKcs is homologous to the kinase domain of the PI3K $\gamma$  of the classical PI3-kinases, the only PI3-kinase of any class for which atomic structural data are available (Djordjevic and Driscoll, 2002; Walker et al., 1999). Homology is restricted to  $\sim 360$  residues, which comprise the two lobes of the catalytic region, and to a lesser extent to the adjacent helical-repeat segment that forms the backbone of the PI3K $\gamma$  structure. Apart from the significantly larger stretch of polypeptide at the N terminus, the catalytic domains in the PIKK enzymes have two additional seg-

\*Correspondence: ollorca@cib.csic.es (O.L.); laurence.pearl@icr.ac.uk (L.H.P.)

<sup>3</sup>Present address: Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, United Kingdom.



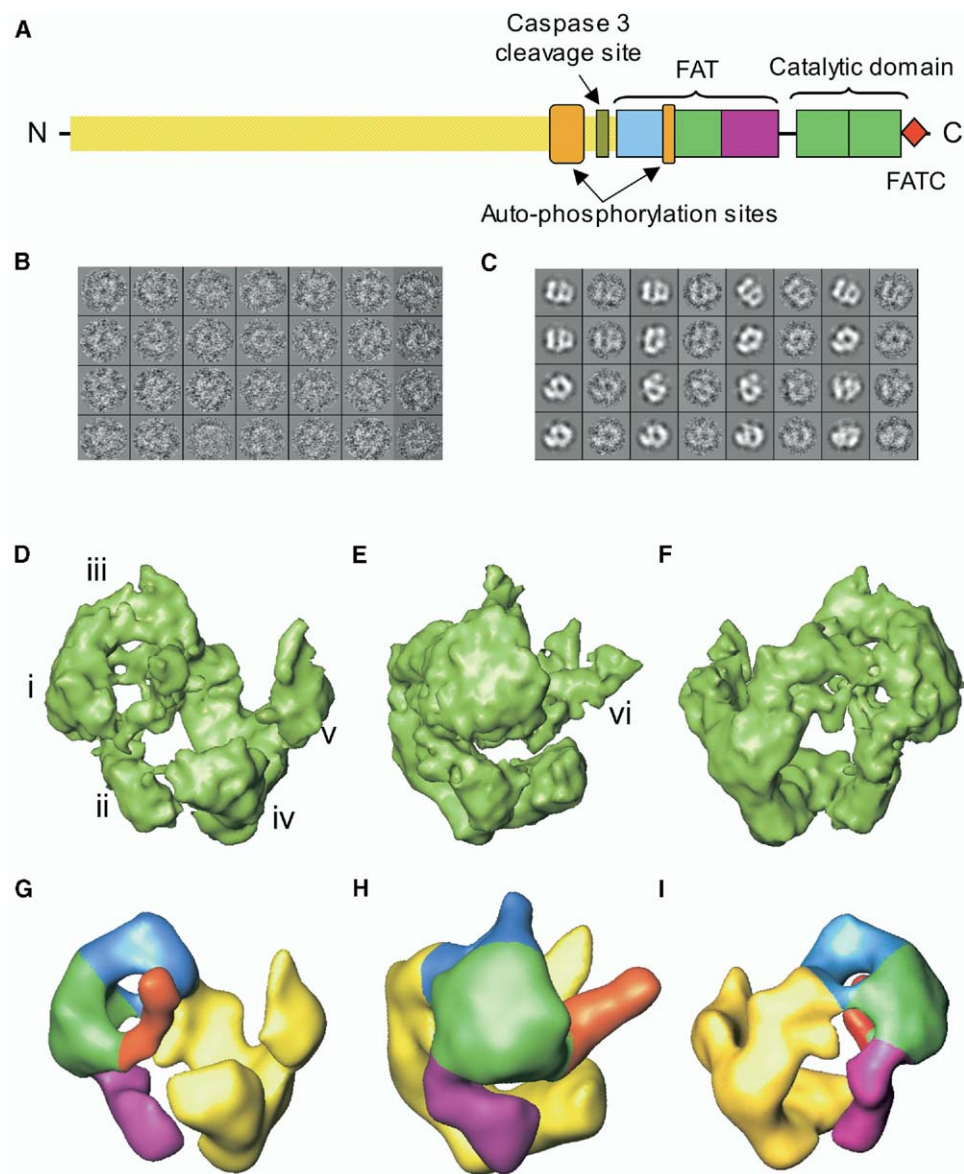


Figure 1. Three-Dimensional Cryo-EM Structure of DNA-PKcs

(A) Linear sequence and conserved domains in DNA-PKcs. Color codes match those of Figure 1G through 1I according to the domain assignment to the structure (see text).

(B) Representative collection of single particles obtained for DNA-PKcs using cryo-electron microscopy.

(C) Each pair of images represents a theoretical projection and its corresponding class average obtained after image refinement.

(D–F) Different views of the reconstructed volume. In (D) and (E), different domains are labeled “i” through “vi.”

(G–I) Simplified views of (D)–(F) generated by low-pass filtering of the cryo-reconstruction at 30 Å. Each assigned structural region (see text) has been highlighted with a different color.

ments specific to that family and absent in PI3K $\gamma$ : an  $\sim 50$  amino acid insertion between the two lobes of the catalytic domain and an  $\sim 100$ -residue extension at the end of the catalytic domain which forms the extreme C terminus of the protein (FATC domain).

In contrast to the C-terminal catalytic region, which is of similar size in all the large PIKK enzymes, the N-terminal regions show substantial variation in size and no apparent homology. However, recent studies (Perry and Kleckner, 2003; Brewerton et al., 2004) have suggested

that the N-terminal regions in all these proteins consists of multiple repeats of a degenerate helical motif. These HEAT repeats, originally identified in mTOR (Andrade and Bork, 1995), are  $\sim 40$  residue long and are arranged as a pair of antiparallel  $\alpha$  helices separated by a tight turn. Multiple copies of these helix pairs stack in parallel to create flattened tubular structures composed of double layers of  $\alpha$  helices that typically curve to form a convex and a concave face (Groves and Barford, 1999).

A complete understanding of the molecular mecha-

nisms of DNA-PKcs and other PIKKs requires detailed study of their structure. Though no atomic data is available, three medium-resolution structures of DNA-PKcs (Chiu et al., 1998; Boskovic et al., 2003; Leuther et al., 1999) and one of ATM (Llorca et al., 2003) have been obtained using electron microscopy (EM). Two studies of DNA-PKcs utilized negatively stained samples, one using two-dimensional crystals of the protein (Leuther et al., 1999) and the other isolated molecules (Boskovic et al., 2003), to generate structures at 22 Å and 30 Å resolution, respectively. Both studies showed that DNA-PKcs consists of three main regions: a “head,” a “palm,” and an “arm” connecting them. Studies in the presence of dsDNA have shown that DNA-PKcs utilizes its palm domain to bind DNA and that DNA-induced conformational changes trigger an interaction between initially distant palm and head regions (Boskovic et al., 2003). Interestingly, the 30 Å resolution structure of the related ATM kinase showed a comparable organization into head and arm domains, which reflects the likely structural homology within the PI3K-like kinase family (Llorca et al., 2003). A third study of DNA-PKcs, which used single particle electron reconstruction on electron microscope images of unstained samples in vitrified ice (cryo-EM), generated a structure with many holes and cavities (Chiu et al., 1998), which is difficult to reconcile with the other reconstructions.

To gain further insight into the structure of DNA-PKcs, we have performed cryo-EM experiments on single DNA-PKcs molecules using a Field Emission Gun (FEG) electron microscope. The resulting ~13 Å resolution map now provides detailed insights into the molecular architecture of DNA-PKcs, allowing determination of the overall polypeptide topology and identification and fitting of distinct protein domains, and suggests a conformational mechanism for kinase activation by DNA binding.

## Results and Discussion

**Cryo-EM Three-Dimensional Structure of DNA-PKcs**  
Human DNA-PKcs was purified as described (Boskovic et al., 2003), vitrified, and observed on a Tecnai G<sup>2</sup> FEG electron microscope. Individual particles were clearly discernible (Figure 1B) and ~7000 images were extracted and corrected for the contrast transfer function (CTF) of the microscope prior to image processing. Refinement was performed using EMAN (Ludtke et al., 1999) (Figure 1C), and an ~13 Å resolution map was generated (Figures 1D–1F and Supplemental Data available with this article online) (see Experimental Procedures). The full dataset correlated very well with the output volume, suggesting the sample was structurally homogeneous. The map obtained is fully consistent with the lower-resolution volume obtained from negatively stained microscopy of single particles (Boskovic et al., 2003), but the higher resolution allows a significantly more detailed definition of structural domains and reveals several new features.

The head, palm, and arm regions, previously described from the low-resolution structure (Boskovic et al., 2003), are resolved in the cryo-EM structure into their constituent domains (labels “i” through “vi” in Fig-

ures 1D and 1E). For simplicity, similar views of the reconstruction are shown after filtration at 30 Å and coloring of the distinct structural areas (Figures 1G–1I and Supplemental Data). The largest domain within the head region is located at one pole of the DNA-PKcs structure and has the overall shape of a triangular wedge, ~90 Å × 60 Å, and ~20 Å thick (Figure 1D, label “i,”; Figures 1G–1I, green color). This region acts as a nucleus within the head, from which other protein domains emerge. Three distinct domains protrude in different directions from the vertices of the wedge (red, magenta, and blue in Figures 1G–1I). The smallest of these protrusions is a curved stalk ~80 Å long and ~10 Å across at its narrowest, projecting from the body of the protein and ending in a bulbous head (Figures 1D–1F, labeled as “vi” in Figure 1E, and Figures 1G–1I, red color). Although this protrusion was not evident as a distinct feature in previous studies, careful comparison with the cryo-EM structure shows that it was present but folded back onto the edge of the central head density in stained specimens, suggesting a considerable degree of flexibility at the point where it connects to the central domain (Leuther et al., 1999; Boskovic et al., 2003). The second protrusion also projects from the central core to a similar length but is much more substantial, ending in a large globular region (Figure 1D, labeled as “ii” and Figures 1G–1H, magenta color). The third protrusion, the “shoulder,” is a tubular structure extending from the acute vertex of the triangle (Figure 1D, labeled as “iii,” and Figures 1G and 1H, blue color) bending by ~90° to connect the head perpendicularly to one end of the arm (Figures 1G–1I, yellow color).

The arm itself is a curved and flattened tube of density of rounded rectangular cross-section, running from the end of the shoulder to merge into the palm. In negative stain, the palm was a single smooth structure. However in the cryo-EM it is resolved into two distinct segments or “claws” (Figure 1D, labeled “iv” and “v,” and Figures 1G–1I, colored yellow), which jointly maintain a similar orientation to that of the overall palm domain observed at low resolution (Boskovic et al., 2003). Both claws project perpendicularly from the narrow edge of the arm, which forms the base of the palm. One claw (henceforth the distal claw) (Figure 1D labeled as “iv”) extends directly from the end of the main arm segment opposite to the head, while the other (proximal claw) inserts into the primary arm midway along its length (Figure 1D, labeled as “v”). Interestingly, eliminating the inserted claw from the density leaves an arm segment whose shape is very reminiscent of the arm observed in the medium-resolution structure of ATM (Llorca et al., 2003; Figure 2A).

The cryo-EM volume we have obtained at 13 Å describes a continuous manifold surface, built from seven identifiable domains, that forms a cage around a central volume. This differs substantially from a previously published cryo-EM reconstruction at 21 Å (Chiu et al., 1998), in which DNA-PKcs was described as a highly fragmented structure heavily perforated by myriad holes and discontinuous cavities.

## Location of the Conserved C-Terminal Region

Bioinformatic analyses of DNA-PKcs and related PIKKs identify a conserved C-terminal portion con-

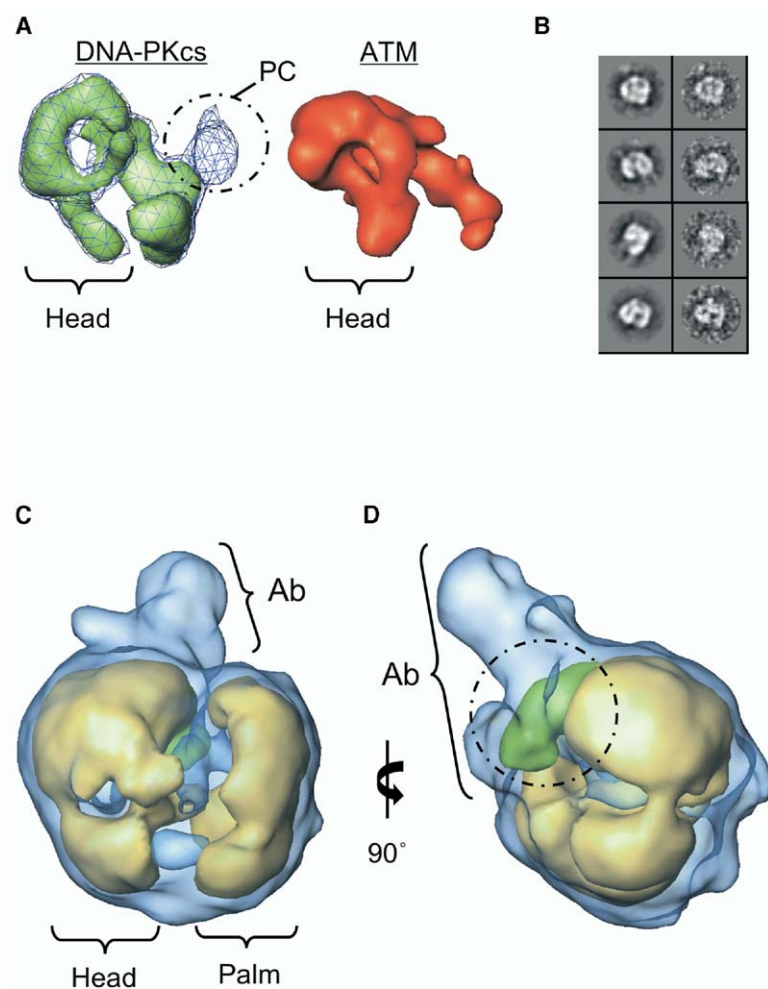


Figure 2. Location of the Kinase Domain in DNA-PKcs

(A) View of a 30 Å filtered version of DNA-PKcs structure (black grid) enclosing the same volume (green solid object) after removal of the “proximal claw” (PC circled). A comparable view from ATM reconstruction (Llorca et al., 2003) is shown alongside as an orange solid object. The major differences in sequence between ATM and DNA-PKcs occur in the N terminus while the differences in structure map to the arm and palm.

(B) Immunocomplexes of monoclonal sc-5282 (Santa Cruz) and DNA-PKcs were prepared in solution and observed under the microscope. After image processing, a collection of class averages were obtained. In the figure each pair of images shows one representative projection and its corresponding average image.

(C) View of the antibody-DNA-PKcs complex highlighting the head/arm location of the protein and the position of the antibody (Ab). The inside volume, contoured to contain 100% of the DNA-PKcs volume mass (solid object), shows a clear antibody label (green and circled) which extends very clearly from the head of DNA-PKcs at a lower contour level (blue transparent surface).

(D) Perpendicular view of (C).

sisting of the FAT, PIKK, and FATC domains, while the variable N-terminal portion is thought to be built primarily of helical repeats (Bosotti et al., 2000; Brewerton et al., 2004; Perry and Kleckner, 2003; Shiloh, 2003). At the current resolution, individual secondary structures cannot be resolved, but the overall architecture and arrangement of domains in DNA-PKcs can be defined, so that given suitable atomic models a quasi-atomic model can be constructed. However, no experimental high-resolution structure is currently available for any fragment of DNA-PKcs, so that interpretation of the present map has utilized a combination of comparative electron microscopy, bioinformatics, molecular modeling, and docking.

As a first step, we sought to locate the PIKK domain, which provides the catalytic activity in DNA-PKcs and related proteins such as ATM, ATR, and mTOR. Several lines of evidence suggest that this domain resides in the head part of the protein rather than the palm. The C-terminal halves of the DNA-PKcs and ATM amino acid sequences, which contain the conserved PIKK domain, can be aligned with each other (and those of ATR and TOR) with little relative insertion and deletion, so that the majority of the ~1000 residue difference in size between the two proteins occurs in the N-terminal half.

Comparison of negative stain structures of DNA-PKcs (Boskovic et al., 2003) and ATM (Llorca et al., 2003) shows a strong similarity between the head regions in shape and size, whereas the palm in ATM appears significantly smaller than in DNA-PKcs (Figure 2A). Taken together these two observations are most consistent with the variable N-terminal half of the protein sequence mapping to the palm, while the conserved C-terminal part, including the catalytic PIKK domain, maps to the head.

As this assignment contradicts a previous suggestion as to the location of the PIKK domain (Brewerton et al., 2004), we sought to test our sequence/structure mapping with direct experimental evidence. To this end, we used a commercial monoclonal antibody (sc-5282 from Santa Cruz Biotechnology) raised against a C-terminal fragment of DNA-PKcs (residues 2965–4127) containing the conserved FATC, PIKK, and FAT domains. This antibody was first verified as capable of recognizing our DNA-PKcs preparations (data not shown) and then used to form an immunocomplex in solution, which was observed under the electron microscope using negative stain, and single particles picked and refined (Figure 2B) into a three-dimensional (3D) average volume. The resulting structure, at ~30 Å reso-

lution, is similar to those previously obtained for free DNA-PKcs, but with a large additional mass evident, consistent with the bound monoclonal antibody protruding from the head and not the palm (Figures 2C and 2D). When the contour level is slightly reduced this feature becomes even more evident, and without significant surrounding noise (Figures 2C and 2D). Although the precise epitope recognized by this antibody has not been defined, it appears to reside at the beginning of the tubular shoulder structure at the junction between the globular core of the head and the arm. The combination of this experimental mapping with the structure comparison of DNA-PKcs and ATM provides strong support for locating the C terminus of DNA-PKcs in the head.

#### Identification and Modeling of Head Domains

While the structure of the catalytic domain from DNA-PKcs or any other large PIKK protein is as yet unknown, the crystal structure of an ~840-residue C-terminal catalytic fragment of a type I PI-kinase, the  $\gamma$ -isoform phosphatidylinositol-3-kinase (PI3K $\gamma$ ), has been determined at 2.2 Å resolution (Walker et al., 1999). We used optimal sequence threading techniques (Kelley et al., 2000; McGuffin et al., 2000) to map the DNA-PKcs C-terminal sequence onto the PI3K $\gamma$  structure and construct a working model. In addition to the core domain, there is evidence in DNA-PKcs for sequences, N-terminal to the kinase, that resemble the noncanonical helical repeats that pack against the back of the kinase domain and form the backbone of the PI3K $\gamma$  structure. As this segment in PI3K $\gamma$  appears to be structurally inseparable from the kinase domain, it is highly likely that a structure corresponding to the helical repeats will also be present in DNA-PKcs, although it may be embedded in a more extensive helical structure, and its precise mapping in the upstream sequence is less clear. We therefore constructed an extended model consisting of the minimal PIKK domain (residues 727–1092 of PI3K $\gamma$ ) and the helical repeats (residues 546–726 of PI3K $\gamma$ ) and used this in automated docking searches of the cryo-EM density map using SITUS (Wriggers et al., 1999). The highest ranked solution from a full six-dimensional (6D) search of the final refined map placed the PIKK domain model in the head region, where it fully occupied the triangular slab-shaped density at the center of the head region (Figures 3A and 3B). In different SITUS searches using maps at different stages of refinement and with models filtered to different resolutions, a second high-ranked solution was obtained at the same position in the map, but with the model rotated by 180° around the long axis of the triangular density, relative to the best solution. Visual inspection confirms that either orientation of the search model, whose filtered volume has a rough 2-fold symmetry, can be accommodated within this roughly symmetrical density equally well. While the limited resolution and pseudo-symmetry of the model generate some ambiguity in the orientation of the PIKK domain within the head, the determination of its location within the central triangular slab that forms the heart of the head region is unequivocal and fully consistent with the overall sequence/structure mapping and the antibody labeling.

Of the two orientations for the PIKK domain model within the head, defined by the docking procedure, only one provides a consistent explanation of the connectivity with the rest of the visible density. Furthermore, this orientation places the catalytic site facing outwards and therefore accessible for protein substrates in the DNA bound conformation when the internal cavity is occupied by DNA (Boskovic et al., 2003). In this orientation, the C terminus of the docked PIKK model comes close to the connection between the core density and the smaller of the two “peninsula” projections (Figure 1E, labeled “vi,” and Figures 1G–1I, red color). The narrow stalk of this projection could comfortably accommodate a single  $\alpha$  helix but not any substantially larger secondary structure, suggesting that it is a terminal feature. Together with its proximity to the C terminus of the docked PIKK domain, this identifies it as the C-terminal FATC domain (Figure 3C). The predicted secondary structure (McGuffin et al., 2000) for this ~100-residue extension beyond the C terminus of the PI3K $\gamma$  structure consists of an ~35-residue helix at the N-terminal end, running into two smaller helices interspersed by coil segments, and a rough model based on this was found to fit this protrusion remarkably well (Figure 3C).

With the small protrusion assigned to the FATC domain, the junction between the larger peninsula of density (Figure 1D, labeled “ii,” and Figures 1G–1I, magenta color) and the core density is close to the loop connecting the C terminus of the PIKK-helical segment in PI3K $\gamma$  to the N-terminal lobe of the kinase domain itself. This helical segment, conserved in all PI3-kinases (Djordjevic and Driscoll, 2002), consists of ~200 residues arranged as five pairs of antiparallel helices. The topology of these is reminiscent of a HEAT repeat, but with shorter helices, giving it a more circular average cross-section. Sequence analyses suggest that the weakly conserved region immediately N-terminal to the kinase in the large PIKK proteins also consists of a distinct class of helical repeats, but over a much larger stretch of sequence, ~500 residues, identified as the FAT domain (Bosotti et al., 2000). The large protrusion (Figures 1G–1H, magenta color) would then correspond to the C-terminal end of this FAT helical repeat domain, whose middle segment would be the equivalent of the shorter repeat found in PI3K $\gamma$ , while the N-terminal end would be directed toward the apex of the triangular core density connecting to and probably becoming the tubular shoulder segment that connects to the arm and N terminus of the protein (Figures 1G and 1H). Together these three segments form a continuous stretch of helical elements, more than 700 residues long, running alongside the core kinase domain in the head, down to the tip of the FAT protrusion, and with a flexible connection back to the N-terminal lobe of the core kinase domain (Figures 3D and 3E).

#### Modeling of Arm and Palm Domains

Immediately beyond the shoulder, at the opposite end to its connection to the head, the density of the arm takes on a flattened tubular appearance, which leads into the two claw domains of the palm. With the C-terminal domains assigned to the head, this and its con-

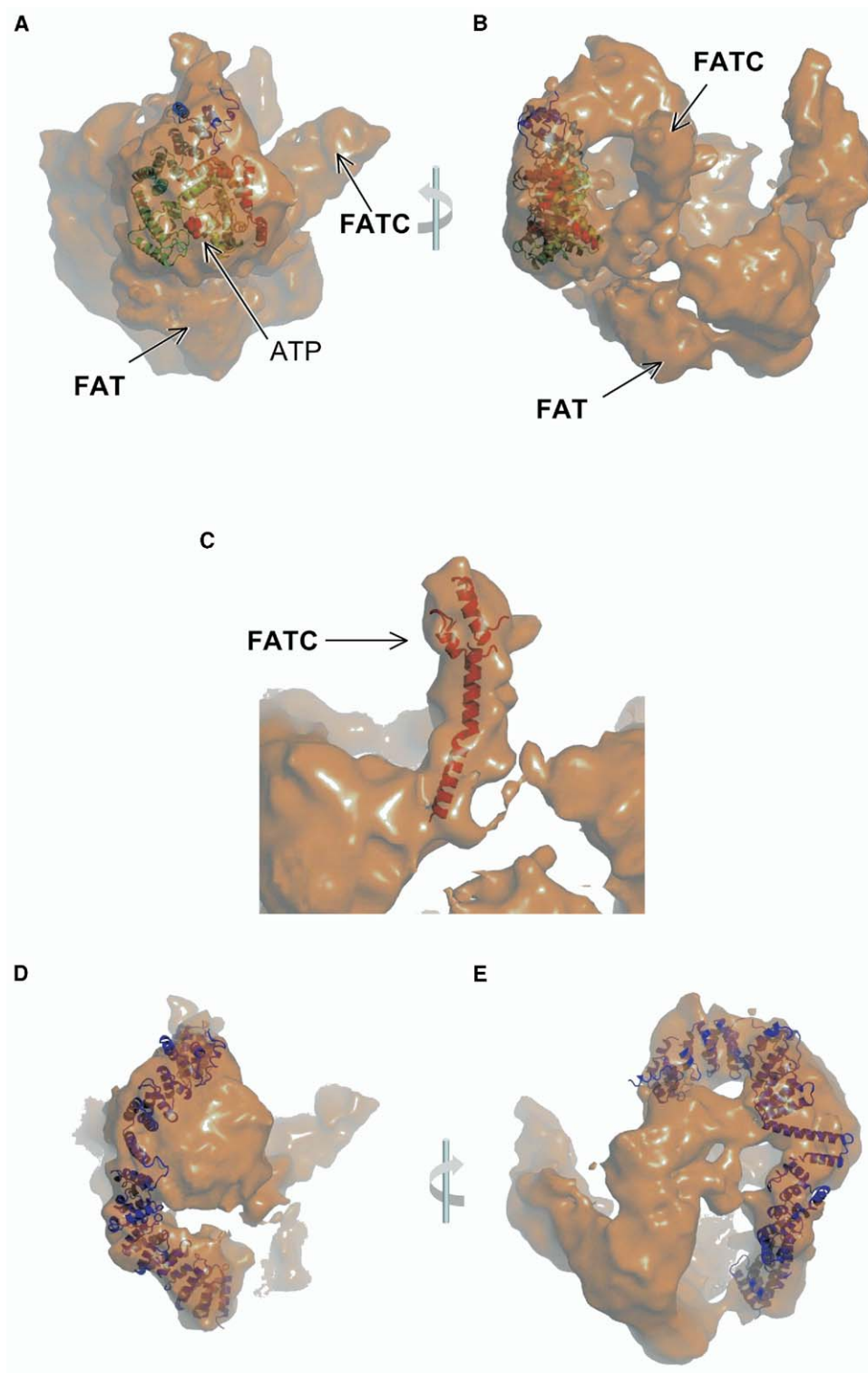


Figure 3. Assignment of Head Domains

(A) Automated 6D search of the cryo-EM map using SITUS (Wriggers et al., 1999) using a homologous model places the DNA-PKcs PIKK domain in the central triangular density of the head region. The docked model is shown as a secondary structure cartoon rainbow colored blue-red, from the N to C terminus. The position of ATP based on the PI3K $\gamma$  crystal structure (Walker et al., 1999) is shown as a red CPK model.

(B) As (A) but rotated 90° around the vertical.

(C) Location of the FATC domain—an ~100 residue helical extension beyond the C terminus of PI3K $\gamma$ , specific to the large PIKK enzymes.

(D) Location of the FAT domain. This ~800 residue segment immediately N-terminal to the catalytic domain is predicted to have a helical repeat structure distinct from HEAT repeats (Brewerton et al., 2004), encompassing the shorter helical repeat segment observed in the PI3K $\gamma$  crystal structure (Walker et al., 1999). The atomic model in this figure was constructed from multiple copies of the helical repeat segment found in PI3K $\gamma$ .

(E) As (D) but rotated 90° around the vertical.

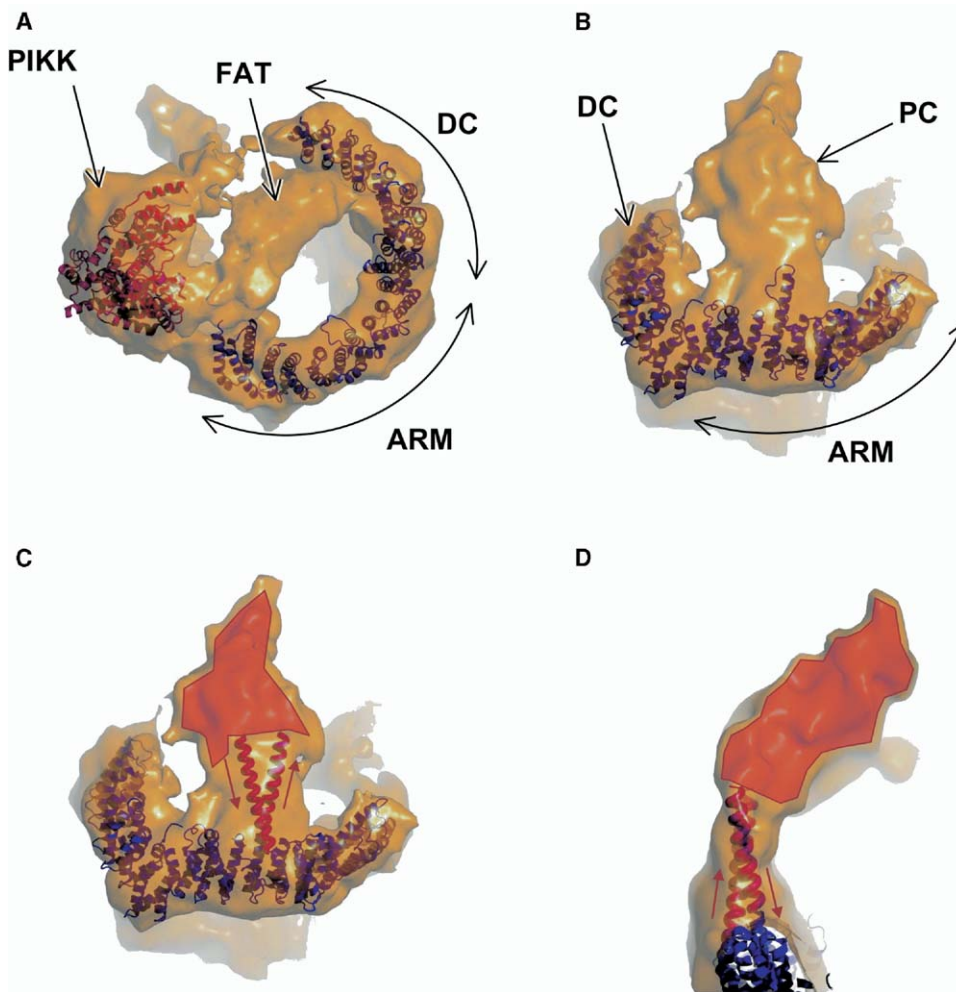


Figure 4. Assignment of Arm and Palm Domains

(A) Location of main HEAT repeat segments. The curved belt of density constituting the arm and distal claw (DC) of the palm closely match the dimensions of HEAT repeats, found in proteins such as importin  $\beta$ 1 (Lee et al., 2003) and the PR65 $\alpha$  subunit of protein phosphatase 2A (Groves et al., 1999).

(B) As (A) but rotated  $\sim 90^\circ$  around the horizontal. Variations in the width of the arm/distal claw density correspond to variations in the length of individual helices in the HEAT repeats. Comparable variations in helix length occur in the HEAT repeat crystal structures docked into the cryo-EM density.

(C) Hypothetical model for the insertion of a domain (the proximal claw) into the HEAT repeat array of the arm by extension of two consecutive helices and elaboration of the intervening loop.

(D) As (C) but rotated by  $90^\circ$  around the vertical. The thickness of the density connecting the edge of the arm to the proximal claw is consistent with the model for domain insertion into the HEAT repeat by helical extension.

ected density must correspond to the poorly conserved N-terminal part of the molecule. The shape and dimensions of the arm density along its full length are consistent with its being formed by a succession of HEAT repeat segments as suggested by sequence-based predictions (Andrade and Bork, 1995; Brewerton et al., 2004; Perry and Kleckner, 2003). The basic architecture of HEAT repeats defines a smooth curvature of the packed pairs of helices (Groves and Barford, 1999), and this is clearly evident in the arm density in the DNA-PKcs cryo-EM map, although there are several dislocations, suggesting that the HEAT repeat architecture is not monotonic and continuous. Nonetheless, much of the arm can be fitted satisfactorily by docking carefully

chosen segments of HEAT repeats extracted directly from crystal structures (Figures 4A and 4B).

The major departures from a canonical HEAT repeat structure occur at the junctions between the arm and the claws, which extend perpendicularly from the narrow edge of the arm, to generate a bifurcated structure (Figure 4B). Comparison with the negative stained structure for ATM (Llorca et al., 2003) shows the combined arm and distal claw structures to be a common feature of both proteins, whereas the proximal claw is an evolutionary addition specific to DNA-PKcs (Figure 2A). As homology between ATM and DNA-PKcs outside of the C-terminal region is extremely low, it has not been possible to determine from sequence analysis

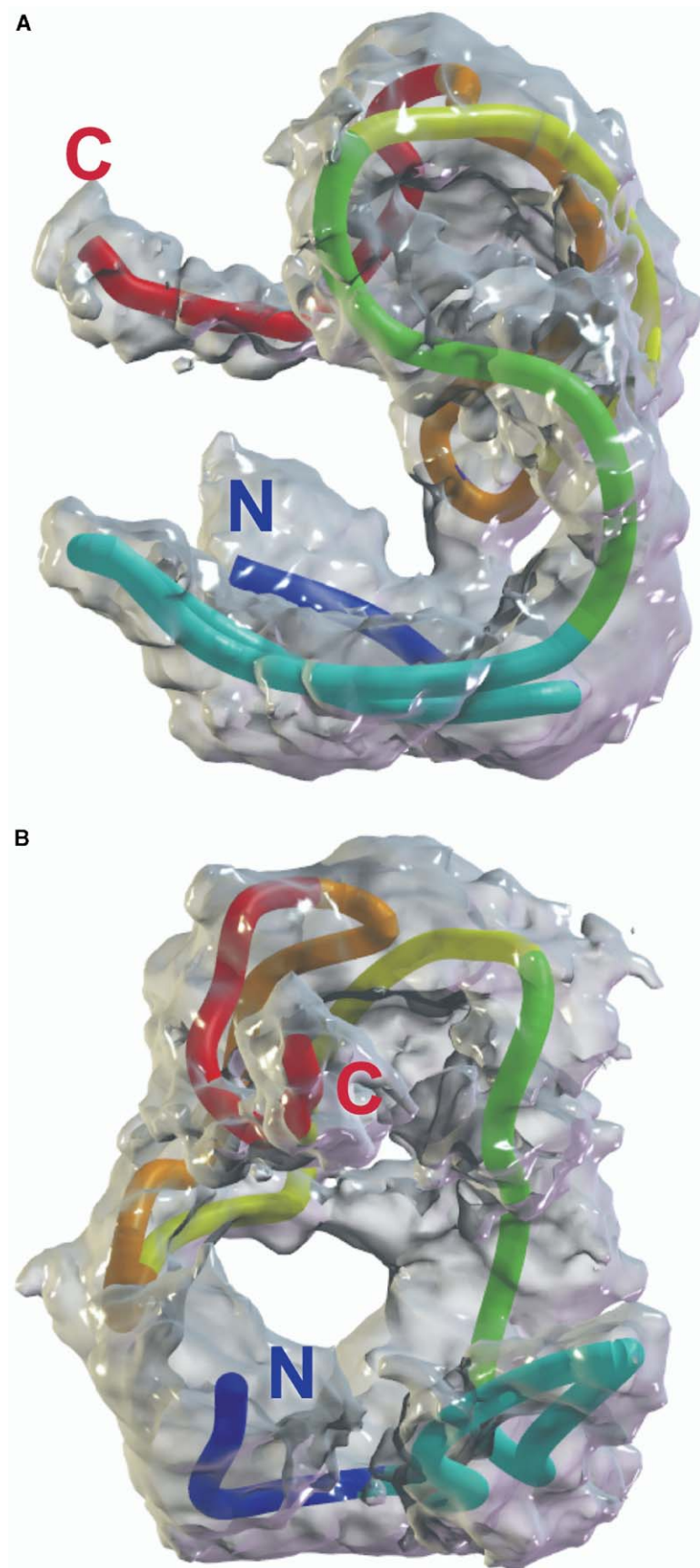


Figure 5. Chain Topology of DNA-PKcs  
(A) Transparent models of the cryo-EM volume, threaded with a tube (color-coded N terminus; blue to C terminus; red, which shows the overall topological "flow" of the 4128 residue polypeptide chain).  
(B) As (A) but rotated 90° anticlockwise around the vertical.

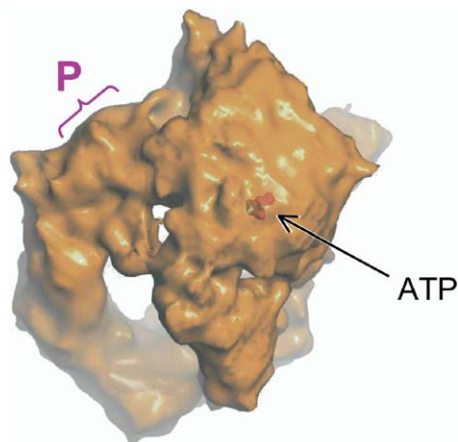


Figure 6. Relative Location of Autophosphorylation and Active Sites

The ATP binding site, based on docking of a PI3K $\gamma$ -based model, lies on the outer face of the main density in the head region of DNA-PKcs and would be accessible in the presence of bound DNA. The main autophosphorylation sites (P) which are predicted to lie within the shoulder region would be inaccessible to the ATP binding site in the same molecule and would require phosphorylation by a second DNA-PKcs molecule in trans.

alone whether relative to ATM, the DNA-PKcs proximal claw is an extension at the N terminus or an internal insertion. Evolutionary elongation of helical (and other) repeat structures is well described, so that a model for adding further repeats to the N terminus of an existing array could be proposed. However as helical repeats are inherently polar structures, with N and C termini at opposite ends, the bifurcated arrangement of the claws would require the array of HEAT repeats to fold back on itself. One claw would then contain an antiparallel pair of efferent and afferent HEAT repeat segments and be at least twice the diameter of the other. As both claws are of roughly similar girth and neither thick enough to contain a pair of HEAT repeat arrays, it is unlikely that the proximal claw is an N-terminal extension.

Alternatively, the proximal claw could be an insertion into the HEAT repeat structure of the arm. Insertions in other types of linearly repeated secondary structure are known although no example has yet been described for HEAT repeats. Nonetheless, insertion of a domain without disruption of the repeat architecture could be readily achieved by elongation of a consecutive pair of helices, with the inserted domain replacing the loop connecting them (Figures 4C and 4D). This would be fully consistent with protrusion of the proximal claw perpendicularly from the narrow edge of the arm and with the observed density at the junction between the claw and the arm, whose size and shape matches a pair of  $\alpha$  helices. It is not possible at the present time to define the structure of the proximal claw, although its estimated location in the sequence suggests that it is predominantly  $\alpha$  helical. Similarly, the detailed structure of the distal claw cannot be determined at this resolution; however, its shape is consistent with a segment of HEAT repeat, with progressively longer helices to-

ward the N terminus generating a more globular structure at the tip.

### Sequence-Structure Mapping

The present resolution of the cryo-EM map does not permit accurate placement of residues or secondary structures, and higher-resolution EM or crystallographic data will be required to confirm our interpretation. However, the current map has allowed us to predict the location of the main domains and supersecondary structural elements, making it possible to suggest the gross “flow” of the polypeptide chain through the structure and estimate approximate physical locations of “landmark” residues along the sequence (Figure 5). The first  $\sim 800$  residues constitute the distal claw of the palm and associated arm segment, with residues 800–1900 forming the major segment of the arm and the inserted domain that constitutes the proximal claw. Residues  $\sim 1900$ –2500 form the final segment of the arm and the beginning of the narrower tubular density that forms the shoulder. Beyond this, residues  $\sim 2500$ –3700 form the rest of the shoulder and the noncanonical helical repeat FAT segment running from the shoulder, along the side of the catalytic domain, and into the larger of the two head protrusions, connecting back to the N-terminal lobe of the catalytic domain. The kinase domain itself occupies residues  $\sim 3700$ –4000, with the final  $\sim 100$  residues forming the protruding C-terminal FATC domain.

Although approximate, this mapping does provide insight into the location of several functionally implicated regions (Figure 1A). The major patch of autophosphorylation sites (Douglas et al., 2002) (see below) and sites cleaved by caspase-3 (DEVD 2712, DWVD 2982) (Song et al., 1996) all map in or close to the narrow density of the shoulder connecting the end of the arm to the beginning of the head. The location of these sites in this accessible segment of the protein is consistent with the requirement for access by other proteins. A segment of DNA-PKcs C-terminal of this, and including parts of the FAT and catalytic domains, has been shown to coprecipitate the Ku70:86 heterodimer in an in vitro transcription/translation system (Jin et al., 1997). A substantial binding site for Ku70:86 on the head region of DNA-PKcs is topologically consistent with a model of the DNA-PK holoenzyme complex, reconstructed from crystallographic and negative-stain EM structural studies of separate Ku and DNA-PKcs DNA complexes (Llorca and Pearl, 2004).

### DNA Binding and Kinase Activation of DNA-PKcs

With the head region organized as described above, the kinase active site is located at the bottom edge of the core density, midway between the FAT and FATC domains, opening onto the external surface of the head. This orientation would still permit access to the active site by protein substrates in the presence of bound dsDNA, which occupies the central cavity (Boskovic et al., 2003). In addition to other components of the NHEJ system, DNA-PKcs efficiently phosphorylates itself, and seven sites of autophosphorylation have been mapped (Douglas et al., 2002). One (S3205) lies within the head region, in the central part of the FAT



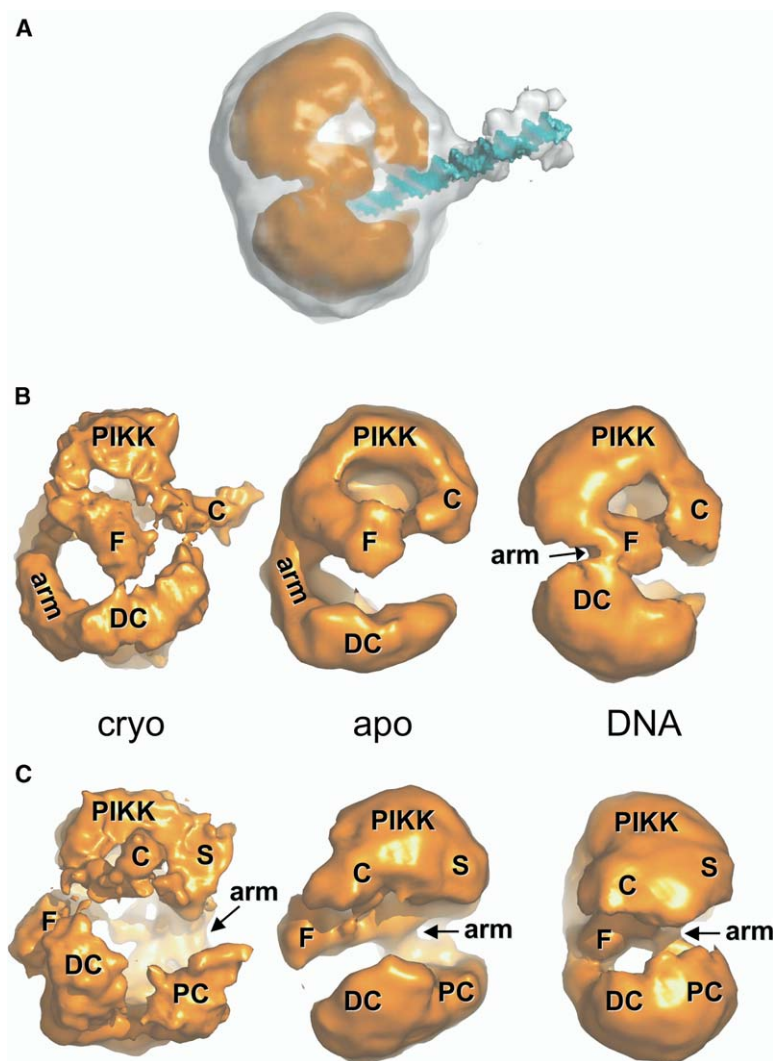


Figure 7. Conformational Changes of DNA-PKcs upon DNA Binding

(A) Negative-stain EM structure (gold) of DNA-PKcs bound to one end of a 54 bp DNA duplex (Boskovic et al., 2003). The DNA (blue) which is clearly visible at a lower contour level (white) passes between and is in contact with the edges of head and palm along its backbone, with the blunt end in contact with the inner face of the palm.

(B) Montage of the 13 Å cryo-EM structure of DNA-PKcs (this work), and the ~30 Å negative-stain structures for DNA-PKcs (Boskovic et al., 2003) in the absence and presence of DNA. Functional domains identified in the cryo-EM structure and their counterparts in the negative-stain structures are labeled as follows: PIKK—catalytic domain; arm—main array of HEAT repeats; F—protruding end of the FAT domain; C—FATC domain; S—shoulder connecting the PIKK domain to the arm; DC—distal claw of the palm; PC—proximal claw. The view is the same as (A).

(C) As (B) but rotated ~90° around the vertical. The change in position of the FATC domain, FAT protrusion, and claws on DNA binding is evident.

domain, while the rest (T2609, S2612, T2620, S2624, T2638, T2647) occur as a cluster, in a region that would map to the shoulder. Without very substantial remodeling, all of these sites are inaccessible to the active site of the same molecule, so that autophosphorylation by DNA-PKcs could only occur in trans via interaction with a second molecule (Figure 6). Dimers of DNA-PKcs have been observed in EM studies (DeFazio et al., 2002; Boskovic et al., 2003) and are functionally implicated in the process of synapsis whereby the association of two broken DNA ends is facilitated.

The kinase activity of DNA-PKcs is strongly stimulated by DNA binding (Chan et al., 1996; Critchlow et al., 1997; Leuther et al., 1999; Merkle et al., 2002). Using negative-stain reconstructions with and without bound dsDNA, we have been able to directly observe the location of a blunt-ended dsDNA molecule in DNA-PKcs and correlate the observed conformational change on DNA binding with kinase activation (Boskovic et al., 2003). Double-stranded DNA was found to bind in the open cavity between the head and the palm, with the blunt end buried in the inner curved face of the palm

and the edge of the head in contact with the DNA backbone (Figure 7A). DNA binding elicited a substantial change in overall conformation so that regions of the palm and head are brought into intimate contact. As the negative-stain and cryo reconstructions of DNA-PKcs are very similar in overall appearance, we have been able to relate regions observed in the negative-stain structure to the distinct domains evident in the cryo-EM map, with sufficient accuracy to identify those regions that undergo conformational changes on DNA binding and that are involved in DNA interaction (Figures 7B and 7C).

Protein-protein contacts induced on DNA binding involve the proximal claw in the palm and the largest of the head projections, ascribed to the FAT protrusion, which bends from its position in the free protein, in toward the head. Interaction with the blunt end of dsDNA appears to be mediated by a region close to the junction between the arm and the proximal claw of the palm. Interaction with the DNA backbone involves the distal claw from the palm, and the narrow projecting region from the head, ascribed to the FATC domain,

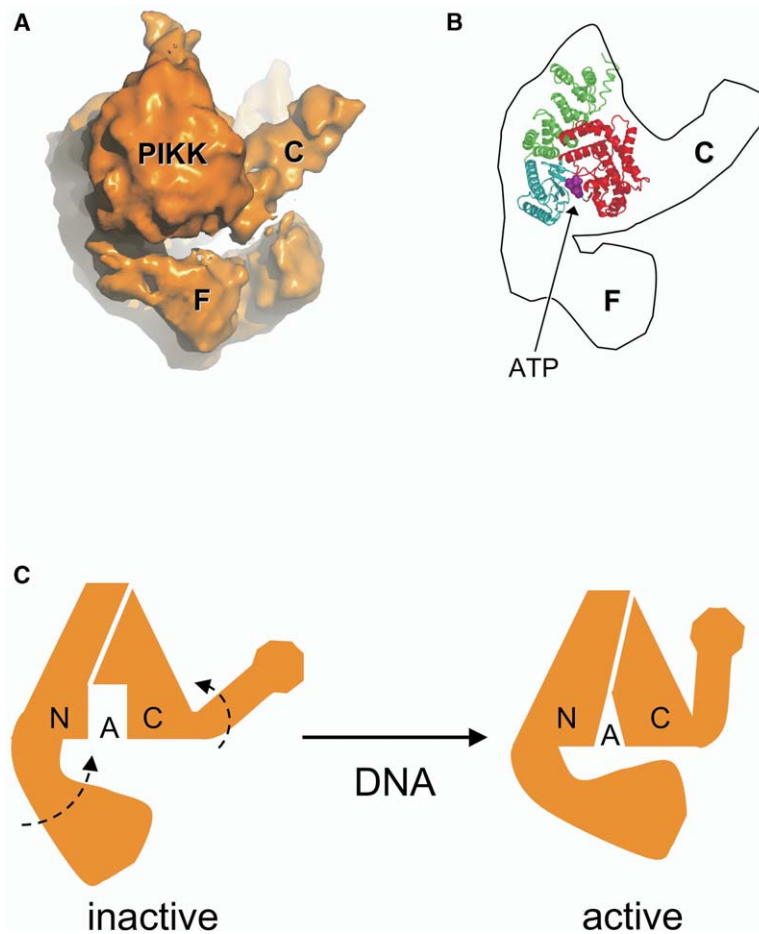


Figure 8. Mechanism of Kinase Regulation by DNA Binding

(A) View of the head region showing the FATC (C) and FAT protrusion (F) extending from the central density of the catalytic domain (PIKK).

(B) Schematic showing the location of the docked PI3K $\gamma$  catalytic domain within the head density. The N-terminal lobe (cyan) and C-terminal lobe (red) which entrap the ATP are directly connected to the FAT (F) and FATC (C) protrusions respectively.

(C) Model for ligand-dependent regulation of DNA-PKcs kinase activity. Conformational changes observed in the positions of the FAT and FATC protrusions on DNA binding are communicated to the linked N- and C-terminal lobes of the kinase domain, altering the juxtaposition of catalytic residues and bound ATP.

which is folded back against the main part of the head in the negative-stained structures. The FATC domain is unique to the large PIKK enzymes and shows the greatest conformational difference between the negative-stain and cryo-EM reconstructions, of any part of the structure (Figure 7).

Both the FAT and FATC domains are directly attached to the catalytic PIKK core domain of DNA-PKcs but are connected to opposite lobes. Relative motions of these two protrusions could directly influence catalytic activity by altering the position of the catalytic groups relative to each other and/or to ATP bound between them. Thus, the FAT and FATC domains together may act as a “sensor” that couples conformational changes induced upon DNA binding to directly activate the catalytic center (Figure 8). Such a role for the FATC domain is completely consistent with the observation that loss of the last ~70 residues of DNA-PKcs, the underlying genetic defect in the classic SCID mouse (Bosma et al., 1983), reduces its DNA-dependent kinase activity to <2% that of the wild-type enzyme (Beamish et al., 2000).

#### Structural Architecture of PI3K-Related Kinases

Using negative stain EM, we have previously shown that the two largest of the PIKK enzymes, DNA-PKcs and ATM, have grossly similar shapes (Llorca et al., 2003; Boskovic et al., 2003). With the ~13 Å cryo-EM

structure for DNA-PKcs described here, we are now able to understand the similarities and differences between these two proteins and potentially extend this analysis to related PI3K-related kinases (Figure 2A). Both proteins are constructed around a common pattern, consisting of a large head with a protruding arm. Taking into account differences in resolution and the effect of stain, the structural organization of the head regions nonetheless appears very similar in both proteins, and each domain in the more highly resolved DNA-PKcs has its counterpart in ATM. The similarity can also be extended to the arm regions, which both show the flattened tubular structure characteristic of HEAT repeats. The ATM arm domain appears to be a subset of the DNA-PKcs arm, matching the main arm and distal-claw segment of the larger protein. The additional proximal claw found in DNA-PKcs must have appeared as an evolutionary insertion to deal with DNA repair-linked tasks specific to metazoa, arthropods, and some insects.

The approximate mapping of the DNA-PKcs sequence to the cryo-EM structure suggests that several key structural features are likely to be common to the other members of the family, including ATR and mTOR. In particular, the organization of the head region around the kinase domain and attached FAT and FATC domains is likely to be conserved throughout, with the active site

on the outside face of the head sterically restricted to autophosphorylation in trans, and regulated via inter- and intramolecular contacts mediated by the FAT and FATC domains (Figure 7). Major differences amongst PI3K-related kinases are expected to occur in the arm, which forms the N-terminal part of the protein. Sequence analysis suggests that the arm regions in all members of this family are organized in the form of blocks of HEAT repeats. While the underlying architecture may be conserved, there is virtually no detectable homology in these regions between the different proteins, and based on the structural differences observed between ATM and DNA-PKcs, it is here that differences in function are likely to be manifest as relative extensions, deletions, and elaborations of the distal end of the arm.

One of the distinctive features of the large PIKK proteins additional to their kinase activity is their role as scaffolds, recruiting other proteins involved in downstream signaling (Jackson, 2002; Shiloh, 2003)—for the TRRAP protein, which lacks catalytic activity, this is its primary function (Vassilev et al., 1998). The helical repeat domains that form the N-terminal parts of these proteins are evolutionarily adaptable structural units, forming protein-protein interaction modules in many systems (Groves and Barford, 1999). Their presence throughout the large PIKK proteins suggests that they will also play key roles in assembling and regulating multiprotein complexes.

#### Experimental Procedures

##### Sample Preparation

DNA-PKcs was purified from HeLa nuclear extracts as described (Boskovic et al., 2003). Antibody-DNA-PKcs immunocomplexes were prepared by mixing a 3:1 molar ratio of monoclonal antibody sc-5282 from Santa Cruz and DNA-PKcs. Reactions were incubated for 30 min at room temperature prior to their analysis. Immunocomplex formation was tested by immunoprecipitation.

##### Cryo-Electron Microscopy and Three-Dimensional Reconstruction of DNA-PKcs

Four microliters of the DNA-PKcs sample were vitrified and observed in a Tecnai G<sup>2</sup> operated at 200 kV and equipped with a Gatan liquid nitrogen specimen holder. Images were recorded at a nominal magnification of 50,000 and different defocus on Kodak SO-163 film. Micrographs were digitized on a Zeiss SCAI scanner at 7  $\mu\text{m}/\text{pixel}$  and averaged to a final 2.8  $\text{\AA}/\text{pixel}$  at the specimen. Around 7000 particles were extracted using the “boxer” command from EMAN (Ludtke et al., 1999), and the contrast transfer function of the microscope (CTF) for each image corrected before image processing. CTF parameters were estimated using CTFIND3 (Mindell and Grigorieff, 2003) and further adjusted with the ctfit program from EMAN. CTF correction was performed by flipping phases with ctfit. Corrected particles were then subjected to 3D refinement as implemented in EMAN (Ludtke et al., 1999). The starting volume was generated by image classification of the whole data set into a few average images to build a reconstruction using common lines. The resolution of the 3D map was estimated by Fourier Shell Correlation (FSC) of two volumes generated after splitting the data set in two halves to 12.46  $\text{\AA}$ , 13.30  $\text{\AA}$ , and 19.78  $\text{\AA}$  according to the 0.14 (Rosenthal and Henderson, 2003), 3 $\sigma$ , and 0.5 criteria, respectively (see Supplemental Data). Recently it has been demonstrated that a cutoff of 0.14 at the FSC curve is a good estimate of the actual resolution of the EM maps built from the complete dataset (Rosenthal and Henderson, 2003). The high-frequency components of the final reconstruction were eliminated by low-pass filtering at 13  $\text{\AA}$ . The volume was rendered to show 100% of the protein mass assuming an average density of 1.35 g/ml and visualized using VMD (Humphrey et al., 1996), PyMol (DeLano Scientific, www.pymol.org),

and Blender (Blender Foundation, www.blender.org). Handedness was chosen to match the absolute hand defined previously by an electron diffraction study of 2D crystals (Leuther et al., 1999).

##### Negatively Stained Electron Microscopy and Three-Dimensional Reconstruction of Antibody-DNA-PKcs Complexes

A few microlitres of the immunocomplex were applied to carbon-coated grids, negatively stained with 1% uranyl acetate, and observed in a JEOL 1230 electron microscope operated at 100 kV. Micrographs were recorded at a magnification of 50,000 under low-dose conditions and processed without CTF correction using EMAN (Ludtke et al., 1999). Roughly 2000 out of a total of 3500 particles were used to build a volume where some additional density in the outskirts of the protein was apparent. Approximately 30% of the initial dataset was found to only be consistent with unlabeled protein after independent processing of the particles that did not match the global volume. The final resolution was estimated to be 30  $\text{\AA}$  with the FSC 0.5 correlation coefficient criteria.

##### Modeling of Atomic Data into the EM Volume

Segments of atomic structures representing identifiable domains were extracted from crystal structures deposited in the Protein Data Bank as follows: PIKK and helical segments of FAT—1E8X; HEAT repeats—1B3U and 1UKL. Initial unbiased positions for the core PIKK domain and segments of HEAT repeat were obtained by the 6D rigid-body fitting procedure of SITUS (Wriggers et al., 1999). Subsequent conformational adjustment was performed manually using PyMol (DeLano Scientific, www.pymol.org) and O (Jones et al., 1991).

##### Supplemental Data

Supplemental Data can be found with this article online at <http://www.structure.org/cgi/content/full/13/2/243/DC1/>.

##### Acknowledgments

This work has been supported by project SAF2002-01715 from the “Ministerio de Ciencia y Tecnología” of Spain (OL) and a Programme Grant from Cancer Research UK (L.H.P.). We thank Prof. José L. Carrascosa and Prof. José M. Valpuesta for the use of their FEG electron microscope facility. We are also very thankful to Jose-Jesús Fernandez for his important help with CTF correction. We are also very thankful to Prof. Keith R. Willison at the ICR (London) and Rocio Arranz at the CNB (Madrid).

Received: October 14, 2004

Revised: December 15, 2004

Accepted: December 17, 2004

Published: February 8, 2005

##### References

- Andrade, M.A., and Bork, P. (1995). HEAT repeats in the Huntington's disease protein. *Nat. Genet.* 17, 115–116.
- Beamish, H.J., Jessberger, R., Riballo, E., Priestley, A., Blunt, T., Kysela, B., and Jeggo, P.A. (2000). The C-terminal conserved domain of DNA-PKcs, missing in the SCID mouse, is required for kinase activity. *Nucleic Acids Res.* 28, 1506–1513.
- Boskovic, J., Rivera-Calzada, A., Maman, J.D., Chacon, P., Willison, K.R., Pearl, L.H., and Llorca, O. (2003). Visualization of DNA-induced conformational changes in the DNA repair kinase DNA-PKcs. *EMBO J.* 22, 5875–5882.
- Bosma, G.C., Custer, R.P., and Bosma, M.J. (1983). A severe combined immunodeficiency mutation in the mouse. *Nature* 307, 527–530.
- Bosotti, R., Isacchi, A., and Sonnhammer, E.L. (2000). FAT: a novel domain in PIK-related kinases. *Trends Biochem. Sci.* 25, 225–227.
- Brewerton, S.C., Dore, A.S., Drake, A.C., Leuther, K.K., and Mindell, T.L. (2004). Structural analysis of DNA-PKcs: modelling of the

- repeat units and insights into the detailed molecular architecture. *J. Struct. Biol.* **145**, 295–306.
- Chan, D.W., Mody, C.H., Ting, N.S., and Lees-Miller, S.P. (1996). Purification and characterization of the double-stranded DNA-activated protein kinase, DNA-PK, from human placenta. *Biochem. Cell Biol.* **74**, 67–73.
- Chiu, C.Y., Cary, R.B., Chen, D.J., Peterson, S.R., and Stewart, P.L. (1998). Cryo-EM imaging of the catalytic subunit of the DNA-dependent protein kinase. *J. Mol. Biol.* **284**, 1075–1081.
- Critchlow, S.E., Bowater, R.P., and Jackson, S.P. (1997). Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. *Curr. Biol.* **7**, 588–598.
- DeFazio, L.G., Stansel, R.M., Griffith, J.D., and Chu, G. (2002). Synapsis of DNA ends by DNA-dependent protein kinase. *EMBO J.* **21**, 3192–3200.
- Djordjevic, S., and Driscoll, P.C. (2002). Structural insight into substrate specificity and regulatory mechanisms of phosphoinositide 3-kinases. *Trends Biochem. Sci.* **27**, 426–432.
- Douglas, P., Sapkota, G.P., Morrice, N., Yu, Y., Goodarzi, A.A., Merkle, D., Meek, K., Alessi, D.R., and Lees-Miller, S.P. (2002). Identification of *in vitro* and *in vivo* phosphorylation sites in the catalytic subunit of the DNA-dependent protein kinase. *Biochem. J.* **368**, 243–251.
- Gell, D., and Jackson, S.P. (1999). Mapping of protein-protein interactions within the DNA-dependent protein kinase complex. *Nucleic Acids Res.* **27**, 3494–3502.
- Groves, M.R., and Barford, D. (1999). Topological characteristics of helical repeat proteins. *Curr. Opin. Struct. Biol.* **9**, 383–389.
- Groves, M.R., Hanlon, N., Turowski, P., Hemmings, B.A., and Barford, D. (1999). The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs. *Cell* **96**, 99–110.
- Harris, R., Esposito, D., Sankar, A., Maman, J.D., Hinks, J.A., Pearl, L.H., and Driscoll, P.C. (2004). The 3D solution structure of the C-terminal region of Ku86 (Ku86CTR). *J. Mol. Biol.* **335**, 573–582.
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD—visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38.
- Ismail, I.H., Martensson, S., Moshinsky, D., Rice, A., Tang, C., Howlett, A., McMahon, G., and Hammarsten, O. (2004). SU11752 inhibits the DNA-dependent protein kinase and DNA double-strand break repair resulting in ionizing radiation sensitization. *Oncogene* **23**, 873–882.
- Jackson, S.P. (2002). Sensing and repairing DNA double-strand breaks. *Carcinogenesis* **23**, 687–696.
- Jin, S., Kharbanda, S., Mayer, B., Kufe, D., and Weaver, D.T. (1997). Binding of Ku and c-Abl at the kinase homology region of DNA-dependent protein kinase catalytic subunit. *J. Biol. Chem.* **272**, 24763–24766.
- Jones, T.A., Zou, J.-Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.* **A47**, 110–119.
- Kelley, L.A., MacCallum, R.M., and Sternberg, M.J. (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.* **299**, 499–520.
- Khanna, K.K., and Jackson, S.P. (2001). DNA double-strand breaks: signalling, repair and the cancer connection. *Nat. Genet.* **27**, 247–254.
- Lee, S.J., Sekimoto, T., Yamashita, E., Nagoshi, E., Nakagawa, A., Imamoto, N., Yoshimura, M., Sakai, H., Chong, K.T., Tsukihara, T., et al. (2003). The structure of importin-beta bound to SREBP-2: nuclear import of a transcription factor. *Science* **302**, 1571–1575.
- Lees-Miller, S.P., and Meek, K. (2003). Repair of DNA double strand breaks by non-homologous end joining. *Biochimie* **85**, 1161–1173.
- Leuther, K.K., Hammarsten, O., Kornberg, R.D., and Chu, G. (1999). Structure of DNA-dependent protein kinase: implications for its regulation by DNA. *EMBO J.* **18**, 1114–1123.
- Lieber, M.R., Ma, Y., Pannicke, U., and Schwarz, K. (2003). Mechanism and regulation of human non-homologous DNA end-joining. *Nat. Rev. Mol. Cell Biol.* **4**, 712–720.
- Llorca, O., and Pearl, L.H. (2004). Electron microscopy studies on DNA recognition by DNA-PK. *Micron* **35**, 625–633.
- Llorca, O., Rivera-Calzada, A., Grantham, J., and Willison, K.R. (2003). Electron microscopy and 3D reconstructions reveal that human ATM kinase uses an arm-like domain to clamp around double-stranded DNA. *Oncogene* **22**, 3867–3874.
- Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97.
- Ma, Y., Pannicke, U., Schwarz, K., and Lieber, M.R. (2002). Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **108**, 781–794.
- McGuffin, L.J., Bryson, K., and Jones, D.T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405.
- Merkle, D., Douglas, P., Moorhead, G.B., Leonenko, Z., Yu, Y., Cramb, D., Bazett-Jones, D.P., and Lees-Miller, S.P. (2002). The DNA-dependent protein kinase interacts with DNA to form a protein-DNA complex that is disrupted by phosphorylation. *Biochemistry* **41**, 12706–12714.
- Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347.
- Perry, J., and Kleckner, N. (2003). The ATRs, ATMs, and TORs are giant HEAT repeat proteins. *Cell* **112**, 151–155.
- Raught, B., Gingras, A.C., and Sonenberg, N. (2001). The target of rapamycin (TOR) proteins. *Proc. Natl. Acad. Sci. USA* **98**, 7037–7044.
- Rosenthal, P.B., and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **31**, 721–745.
- Shiloh, Y. (2003). ATM and related protein kinases: safeguarding genome integrity. *Nat. Rev. Cancer* **3**, 155–168.
- Singleton, B.K., Torres-Arzayus, M.I., Rottinghaus, S.T., Taccioli, G.E., and Jeggo, P.A. (1999). The C terminus of Ku80 activates the DNA-dependent protein kinase catalytic subunit. *Mol. Cell Biol.* **19**, 3267–3277.
- Song, Q., Lees-Miller, S.P., Kumar, S., Zhang, Z., Chan, D.W., Smith, G.C., Jackson, S.P., Alnemri, E.S., Litwack, G., Khanna, K.K., et al. (1996). DNA-dependent protein kinase catalytic subunit: a target for an ICE-like protease in apoptosis. *EMBO J.* **15**, 3238–3246.
- Taccioli, G.E., Rathbun, G., Oltz, E., Stamato, T., Jeggo, P.A., and Alt, F.W. (1993). Impairment of V(D)J recombination in double-strand break repair mutants. *Science* **260**, 207–210.
- Vassilev, A., Yamauchi, J., Kotani, T., Prives, C., Avantaggiati, M.L., Qin, J., and Nakatani, Y. (1998). The 400 kDa subunit of the PCAF histone acetylase complex belongs to the ATM superfamily. *Mol. Cell* **2**, 869–875.
- Walker, E.H., Perisic, O., Ried, C., Stephens, L., and Williams, R.L. (1999). Structural insights into phosphoinositide 3-kinase catalysis and signalling. *Nature* **402**, 313–320.
- Walker, J.R., Corpina, R.A., and Goldberg, J. (2001). Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* **412**, 607–614.
- Wriggers, W., Milligan, R.A., and McCammon, J.A. (1999). Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.* **125**, 185–195.

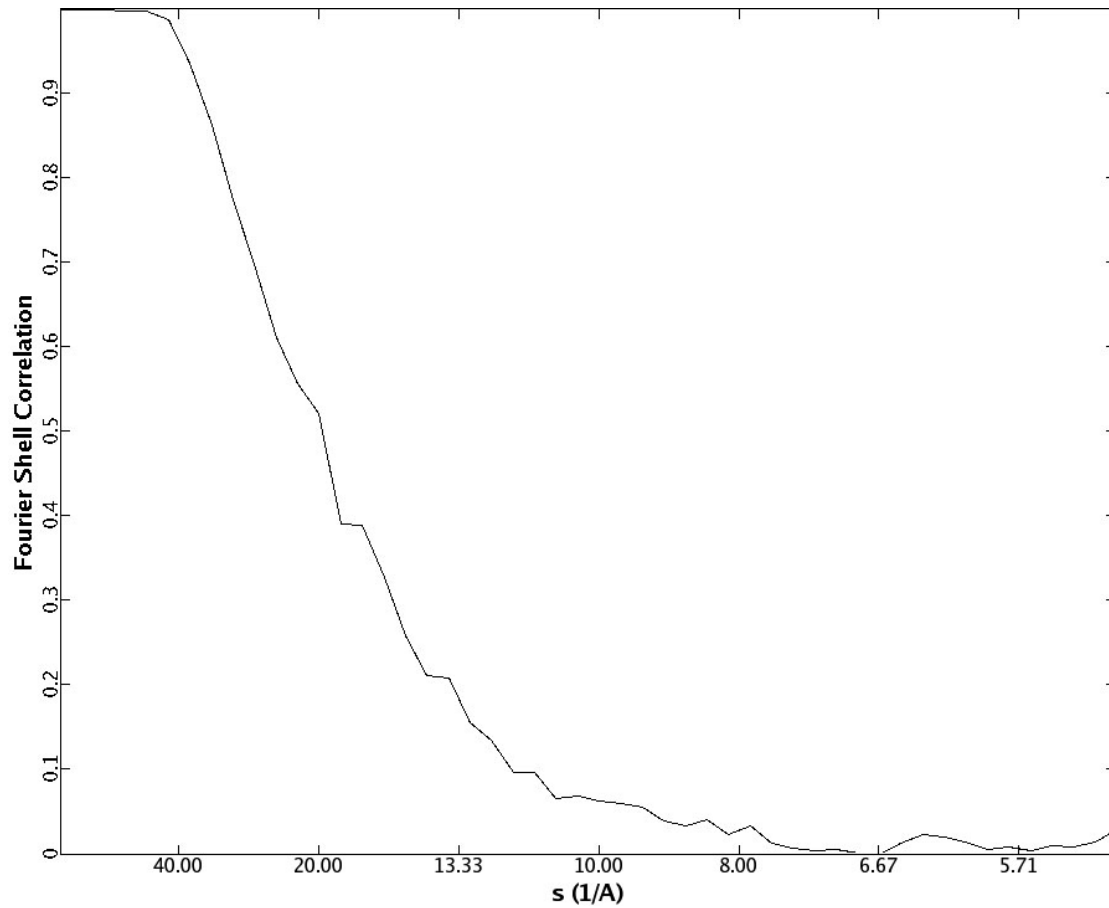
#### Accession Numbers

The cryo-EM map of DNA-PKcs has been deposited in the 3D EM database (<http://www.ebi.ac.uk/msd/>) under accession code EMD-1102.

Supplemental Data

**Three-Dimensional Structure and Regulation  
of the DNA-Dependent Protein Kinase  
Catalytic Subunit (DNA-PKcs)**

Angel Rivera-Calzada, Joseph P. Maman, Laura Spagnolo, Laurence H. Pearl, and Oscar Llorca



**Figure S1. Fourier Shell Correlation Plot for cryoEM Data**

# Three-Dimensional Structure of the Human DNA-PKcs/Ku70/Ku80 Complex Assembled on DNA and Its Implications for DNA DSB Repair

Laura Spagnolo,<sup>2</sup> Angel Rivera-Calzada,<sup>1</sup>  
Laurence H. Pearl,<sup>2,\*</sup> and Oscar Llorca<sup>1,\*</sup>

<sup>1</sup>Centro de Investigaciones Biológicas  
Consejo Superior de Investigaciones Científicas  
Ramiro de Maeztu, 9  
Campus Complutense University  
28040 Madrid

Spain

<sup>2</sup>Section of Structural Biology and  
Cancer Research UK DNA Repair Enzyme  
Research Group  
Institute of Cancer Research  
Chester Beatty Laboratories  
237 Fulham Road  
London SW3 6JB  
United Kingdom

## Summary

DNA-PKcs is a large (~470 kDa) kinase that plays an essential role in the repair of DNA double-strand breaks (DSBs) by nonhomologous end joining (NHEJ). DNA-PKcs is recruited to DSBs by the Ku70/Ku80 heterodimer, with which it forms the core of a multiprotein complex that promotes synapsis of the broken DNA ends. We have purified the human DNA-PKcs/Ku70/Ku80 holoenzyme assembled on a DNA molecule. Its three-dimensional (3D) structure at ~25 Å resolution was determined by single-particle electron microscopy. Binding of Ku and DNA elicits conformational changes in the FAT and FATC domains of DNA-PKcs. Dimeric particles are observed in which two DNA-PKcs/Ku70/Ku80 holoenzymes interact through the N-terminal HEAT repeats. The proximity of the dimer contacts to the likely positions of the DNA ends suggests that these represent synaptic complexes that maintain broken DNA ends in proximity and provide a platform for access of the various enzymes required for end processing and ligation.

## Introduction

DSBs are the most toxic form of DNA damage, with only one or two unrepaired breaks being sufficient to kill a dividing cell. In S and G2 phases of the cell cycle, where sister chromatids are available, DSBs are efficiently repaired in an essentially error-free process of homologous recombination dependent on the Rad51 pathway (West, 2003). In G0 and G1, however, DSBs are repaired by an NHEJ process in which the broken ends are resected and/or processed, with potential loss of genetic information, and the backbones ligated to restore strand continuity. In addition to its role in repair of adventitious DSBs, the NHEJ system is responsible for repair of programmed DSBs generated as intermediates in V(D)J recombination in lymphocytes during development of

immune diversity. Central to the NHEJ process is the formation of a synaptic complex that seems to be driven by the DNA-dependent protein kinase catalytic subunit (DNA-PKcs) and the Ku70 and Ku80 proteins, which maintain the broken ends in proximity and provide a platform for the recruitment of enzymes such as Artemis (Ma et al., 2002) and PNK (Koch et al., 2004), required to restore normal DNA structure at the broken ends, X family polymerases, which promote microhomology and cohesion between the broken ends (Ma et al., 2004), and DNA ligase IV-XRCC4 (Hsu et al., 2002), which closes the phosphodiester backbone on both strands.

Primary recognition of free DNA ends is a function of the Ku protein, a highly abundant heterodimer of the homologous Ku70 and Ku80 proteins. The crystal structure of the Ku70/Ku80 heterodimer (except the evolutionary distinct C-terminal domain of Ku80) shows an extended platform structure with a narrow bridge, creating a preformed ring that can sterically encircle DNA without establishing sequence-specific contacts (Walker et al., 2001) (Figure 1A). DNA bound Ku directs the recruitment of the catalytic subunit DNA-PKcs via a small helical domain at the C terminus of Ku80, whose structure has been determined in isolation (Harris et al., 2004; Zhang et al., 2004). DNA-PKcs itself is a large ~470 kDa serine/threonine protein kinase belonging to the phosphatidylinositol-3-OH kinase (PI3K)-related kinase family (PIKKs), which includes other DNA damage-sensing enzymes, such as ATM and ATR, as well as proteins involved in nutrient sensing (mTOR) and nonsense-mediated decay (hSMG1) (Abraham, 2004). DNA-PKcs binds DNA ends independently of Ku with concomitant stimulation of its kinase activity (Yaneva et al., 1997). Activated DNA-PKcs can phosphorylate itself and a variety of other proteins, including other NHEJ components. Although autophosphorylation appears to be essential to DNA-PK function in NHEJ (Ding et al., 2003; Block et al., 2004), apart from activating Artemis (Ma et al., 2005) and promoting dissociation of histone H1 from nucleosomes (Kysela et al., 2005), the physiological relevance and mechanistic function of other phosphorylation reactions is far from clear (for example Douglas et al. [2005]).

The size and complexity of this multicomponent repair system presents severe challenges for structural analysis. Several electron microscopy (EM) studies have provided low-resolution views of DNA-PKcs (Chiu et al., 1998; Leuther et al., 1999; Boskovic et al., 2003), but only recently have we attained sufficient resolution to provide a reasonable model of its molecular architecture (Rivera-Calzada et al., 2005). The cryo-EM structure of DNA-PKcs shows the organization of the protein into a bulky and largely globular “head” structure connected to a flat tubular “arm” segment terminating in two projecting “claws” (Rivera-Calzada et al., 2005). From a combination of antibody labeling, bioinformatics, domain docking, and comparison with a low-resolution structure of the related ATM (Llorca et al., 2003), the majority of the DNA-PKcs structure has been assigned to conserved regions of its 4128 residue amino acid

\*Correspondence: laurence.pearl@icr.ac.uk (L.H.P.); ollorca@cib.csic.es (O.L.)

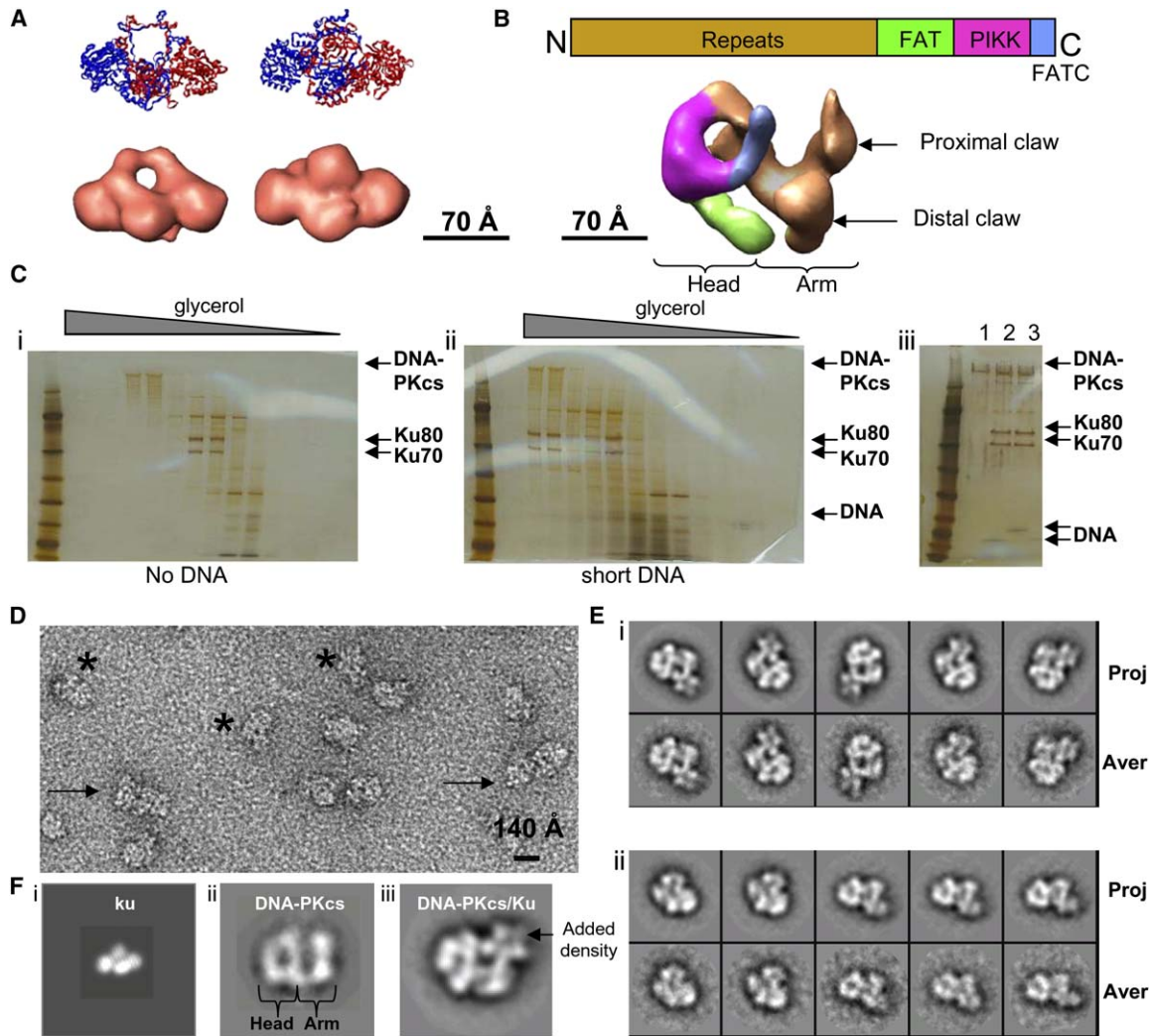


Figure 1. Purification and EM of DNA Bound DNA-PKcs/Ku70/Ku80 Complexes

(A) Atomic structure of the Ku70/Ku80 dimer (PDB entry 1JEQ), lacking the C-terminal evolutionary distinct domains, before (top panels) and after (bottom panels) filtration to 25 Å resolution. Each subunit in the atomic representation has been colored differently. The scale bar represents 70 Å. (B) 3D structure of DNA-PKcs taken from [Rivera-Calzada et al. \(2005\)](#) filtered at 30 Å. Coloring depicts the assignment of domains in the sequence of DNA-PKcs (top row) into the 3D structure, according to [Rivera-Calzada et al. \(2005\)](#). The scale bar represents 70 Å. (C) Purification of DNA-PKcs, Ku70, and Ku80 and isolation of their complex. Glycerol gradients performed with the purified proteins in the absence (i) and presence of DNA (ii). Only the “short” DNA is shown, but similar results were obtained with the “long” DNA (see [Experimental Procedures](#) and [Figure S1](#)). The purified complexes (iii) contain DNA-PKcs, Ku70, Ku80, and DNA. (1) DNA-PKcs alone, (2) complexes with short DNA, and (3) complexes with long DNA. (D) EM field of purified complexes. Asterisks (\*) point to some representative views. Arrows point to putative dimeric aggregates. The scale bar represents 140 Å. (E) A collection of selected pairs of projections and their class averages of the complex as provided by EMAN (i) from untilted micrographs and (ii) from tilted micrographs. (F) 2D projections of the Ku dimer crystal structure (i), DNA-PKcs cryo-EM structure (ii), and the DNA-PKcs/Ku70/Ku80 complex (iii).

sequence ([Figure 1B](#)). The long and distinctive N-terminal region of DNA-PKcs, probably formed by an extended series of HEAT and related helical repeats ([Brewerton et al., 2004](#)), maps into the long curved tubular-shaped domains within the arm region ([Figure 1B](#), orange). The C terminus containing the conserved PI3K-related catalytic domain (residues 3649–4011 in DNA-PKcs), a weakly conserved ~500 residue helical-repeat region immediately N-terminal of this (the FAT domain), and a narrow ~100 residue C-terminal extension at the end of the catalytic domain (FATC domain) ([Bosotti et al., 2000](#)) all locate in the head region

([Figure 1B](#), magenta, green, and blue). A recent solution structure for the last part of the FATC domain from the related mTOR is fully consistent with this assignment in the EM structure ([Dames et al., 2005](#)).

## Results and Discussion

We have now purified to homogeneity a human DNA-PKcs/Ku70/Ku80 complex assembled on DNA ([Figure 1C](#)). Full-length DNA-PKcs, Ku70, and Ku80 were purified from HeLa cell nuclear extracts by using column chromatography and glycerol gradient centrifugation

(see [Experimental Procedures](#)). Under these conditions, and in the absence of added DNA, the Ku70 and Ku80 proteins copurified as expected but did not copurify with DNA-PKcs ([Figure 1C, i](#)). After observations of negative regulation of DNA binding by DNA-PKcs autophosphorylation ([Merkle et al., 2002](#)), DNA-PKcs fractions were dephosphorylated with lambda phosphatase to improve the homogeneity of the preparation and restore the DNA-PKcs to a high-affinity DNA binding state more in accordance with an initial stage of the NHEJ reaction. To form a DNA complex, a DNA molecule containing a 35 base pair duplex segment with a 19-base Y structure at one end (short DNA) was incubated with the purified DNA-PKcs and Ku70/Ku80 proteins. A DNA molecule containing 50 bp plus the 19 base Y structure (long DNA) was also used and similar results obtained ([Figure S1](#) available in the [Supplemental Data](#) with this article online). The DNA was designed on the basis of previous structural studies of DNA-PKcs ([Boskovic et al., 2003](#)) and Ku ([Walker et al., 2001](#)) to be long enough to bind both but with a blocked end to prevent migration along the DNA and no protruding duplex DNA that might permit loading of multiple Ku heterodimers. The rationale behind this design was the result of a compromise between two independent requirements: assure simultaneous binding of both Ku70/Ku80 and DNA-PKcs and reduce the conformational flexibility and structural heterogeneity of the DNA-protein complex, also avoiding DNA protruding in several orientations. The analysis of macromolecules using single-particle EM has the requisite of averaging views of the same complex in the same conformation, and conformational heterogeneity can distort and complicate these types of studies. We added either 21 or 36 additional bp to the DNA length used by [Walker et al. \(2001\)](#) to allow for the accommodation of DNA-PKcs. Both DNA sizes were capable of assembling a DNA/Ku70/Ku80/DNA-PKcs complex, and we decided to focus on the short DNA, which should be more adequate for the EM studies.

When this DNA-protein mixture was resubjected to glycerol gradient centrifugation, a heavy species sedimenting distinctly from any of the individual components was isolated in good yield ([Figure 1C, ii](#)) and found to contain a stoichiometric complex of DNA-PKcs, Ku80, Ku70, and DNA as judged from silver-stained gels ([Figure 1C, iii](#)). This complex was loaded onto EM grids and observed under the electron microscope after negative staining with uranyl acetate ([Figure 1D](#) and [Figure S2](#)). Fields showed the characteristic views of DNA-PKcs particles seen in previous studies ([Chiu et al., 1998](#); [Leuther et al., 1999](#); [Boskovic et al., 2003](#); [Rivera-Calzada et al., 2005](#)) but with some additional density clearly evident ([Figure 1D](#), asterisk; and [Figure S2](#), squares). In addition, some particles seemed to correspond to larger assemblies ([Figure 1D](#), arrowed; and [Figure S2](#), circles), which were analyzed separately (see below). Initially, ~6000 particles of individual molecules were extracted and processed in 2D and 3D with the EMAN software package ([Ludtke et al., 1999](#)). This analysis revealed that the large majority of particles corresponded to side views ([Figure 1E, i](#)), indicating a preferred orientation in binding to the EM grid. To overcome this problem, the sample holder was tilted 40° and an-

other ~7000 images collected ([Figure 1E, ii](#)). Finally, 14,239 images were used and found to cover Euler angles adequately (data not shown). The 2D projections and averages obtained from the combined data displayed the distinctive morphology of DNA-PKcs in side view, with additional lobules of density at the level of the arm region, very evocative of a projection of the Ku dimer ([Figure 1F](#)).

A 3D reconstruction of DNA bound DNA-PKcs/Ku70/Ku80 at 25 Å resolution was built by angular refinement ([Figures 2B, i](#), displayed at a threshold showing ~55% of the calculated protein mass in order to highlight structural details; [iii](#), displayed at a threshold to show 100% of the protein mass). Minimization of model bias was one of the main goals during 3D refinement and was achieved by the use of several starting volumes, including randomly generated Gaussian blobs, all of which progressed toward similar solutions. The structure of the complex is fully compatible with the cryo-EM map of isolated DNA-PKcs ([Figures 2A](#) and [1B](#)) so that domains can be mapped by visual comparison. Ku is found to sit on top of DNA-PKcs with contact points that expand from the back of the head region to the tubular N-terminal arm. The location of Ku was scrutinized more precisely by calculating the difference between the cryo-EM map of DNA-PKcs and the 3D structure of the complex after proper filtering, scaling, and alignment ([Figure 2B, ii](#)). A substantial region of additional density is present in the complex compared to the isolated kinase, indicating the probable location of the bound Ku. The structure of Ku70/Ku80 (without DNA) was fitted into this map after removing the small C-terminal domains of both subunits, which are known to dislocate when bound to DNA, and whose individual position could not be reliably determined at this resolution ([Figure 2B, iv](#)). Although correlation-based density fitting ([Wriggers et al., 1999](#)) is less accurate in negative stain maps compared to cryo-EM maps, all top-ranked solutions obtained by using Situs ([Wriggers et al., 1999](#)) placed Ku such that it fully occupied the additional lobular density, specifically within the main volume revealed by the difference map. The best solution was readjusted manually to better account not only for the new densities but also for their connectivity ([Figure 2B, iv](#), DNA shown as yellow balls). The docked position of Ku within the complex suggests that Ku makes extensive interactions with several distinct regions of DNA-PKcs, including the HEAT repeats and their projecting claws, and contacts the head close to the expected position of the kinase domain. Consistent with this, a region immediately N-terminal to the kinase domain (residues 3002–3850) has previously been implicated in interactions with Ku ([Jin et al., 1997](#)). The C-terminal domain (CTD) of Ku80 comprises a small helical region ([Harris et al., 2004](#); [Zhang et al., 2004](#)), which has been shown to contain a module that drives the interaction with DNA-PKcs, in a mechanism that might be shared by other PIKK kinases ([Falck et al., 2005](#)). Our structure suggests that this CTD might be required for initial recruitment of DNA-PKcs, but more extensive contacts involving large parts of DNA-PKcs probably stabilize the interaction. Additionally, DNA-loaded DNA-PKcs/Ku70/Ku80 complexes were incubated with ATP to permit autophosphorylation, and the sample was then studied by



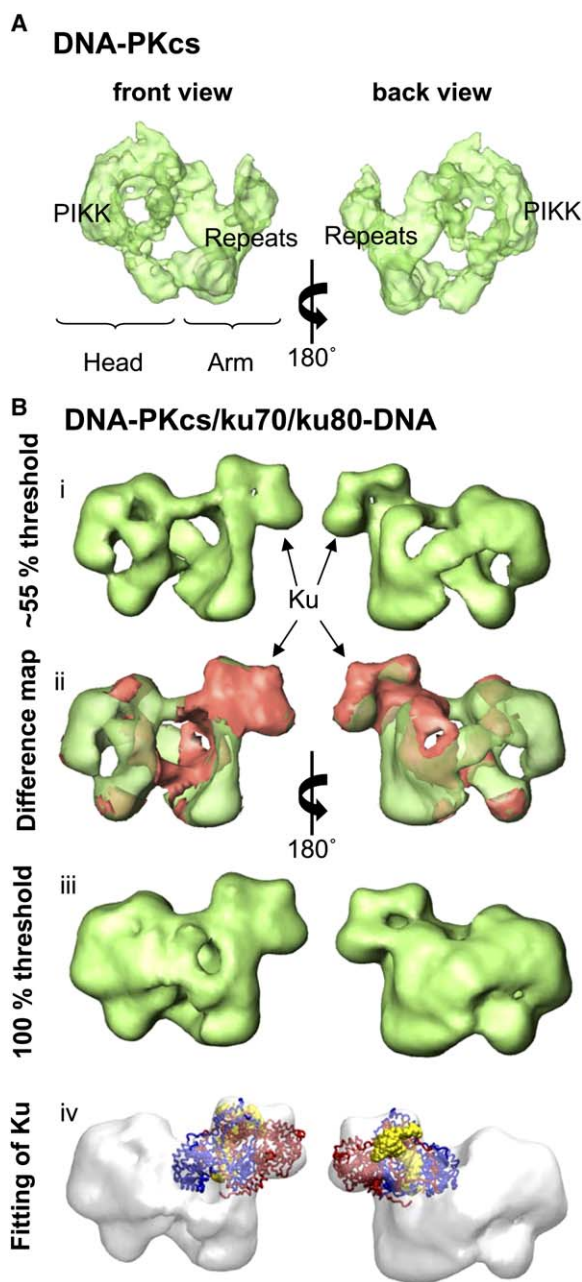


Figure 2. 3D Structure of DNA-PKcs/Ku70/Ku80 Using Negative Staining EM

(A) Front and back views of the 3D structure of DNA-PKcs taken from Rivera-Calzada et al. (2005). Regions approximately corresponding to the catalytic PIKK domain and the N-terminal repeats are labeled. (B) Front and back views of the 3D structure of DNA bound DNA-PKcs/Ku70/Ku80 complexes at a threshold displaying ~55% of the volume (i) and 100% (iii). Panel (ii) shows the difference map between the structures of free DNA-PKcs and the complex shown at a similar threshold to that in (i). Red areas denote the presence of additional density in the Ku-containing complex. (iv) Fitting of core Ku dimer into the 3D reconstruction of the complex, using the DNA bound Ku structure (PDB entry 1JEY). Ku70 and Ku80 subunits in Ku are colored as red and blue ribbons, whereas DNA, as taken from the DNA bound atomic structure of Ku, is shown as yellow spheres. Thresholds showing ~55% of the protein mass are used to visually highlight some of the structural features present in the map that can be masked when displaying a 100% of the density.

EM. We derived two main kinds of averages: one type resembling those previously found in isolated DNA-PKcs (Rivera-Calzada et al., 2005) and another that had a smaller size compatible with a Ku70/Ku80 dimer (data not shown). Therefore, this result supported that the density assigned to Ku70/Ku80 in the complex was probably so because this density was missing under conditions promoting a disassembling of DNA-PKcs/Ku70/Ku80.

To further support the results obtained from negative stain, we also performed a 3D reconstruction of the complex from specimens vitrified in liquid ethane (cryo-EM) so to better preserve the native conformation of the molecule. Limitations due to the buffers used in purification, the amount of purified material, and its concentration precluded the collection of very large numbers of images, but ~4000 images were obtained to generate a ~30 Å map (Figure 3). This map was reconstructed independently, without using information from the negative stain reconstruction, using either random Gaussian blobs or a model built after reference-free image classification and common lines as implemented in Eman (Ludtke et al., 1999). The problem of preferred orientation encountered in negative staining was greatly reduced in the cryo-EM data, allowing a 3D reconstruction without tilting the sample holder. Although the resolution of the cryo-EM structure is lower than the structure derived from negative stain, and the different domains in Ku are less well defined, the two structures are extremely similar and the location of Ku in the complex is unambiguous in both cases (Figure 3B). We performed a fitting of Ku70/Ku80 into this cryo-EM volume but constrained to the Ku density revealed after difference mapping with the cryo-EM structure of isolated DNA-PKcs. Under these conditions, SITUS provided a solution fully compatible with that obtained for the negatively stained structure (Figure 3B, ii).

The different regions in this volume could be readily assigned to domains in DNA-PKcs by comparison with its cryo-EM structure in isolation (Rivera-Calzada et al., 2005) using both the volume data and a partial pseudo-atomic model (data not shown). This comparison reveals substantial conformational changes upon DNA and Ku binding (Figures 3D and 3E) within the kinase itself, which are reminiscent of those previously observed on DNA binding by DNA-PKcs (Boskovic et al., 2003), and can be related to changes in the position of individual segments of the protein by reference to the much higher resolution cryo-EM structure of isolated DNA-PKcs. In particular, there are changes at the level of the head region that correspond to movements of the projecting segment of the FAT domain and changes in the N-terminal repeat region that correspond to movements of the distal claw. Most striking is a change in the conformation of the extreme C-terminal FATC domain, which forms a slender tubular projection from the head. In the DNA-PKcs/Ku70/Ku80 complex, this feature is rotated by ~90° from its position in the apo-DNA-PKcs to interact with the distal claw of the N-terminal HEAT-repeat segment and with the additional density corresponding to the Ku heterodimer itself. The direct connection between the FATC domain and the C-terminal lobe of the kinase domain of DNA-PKcs suggests a key role for the FATC domain in

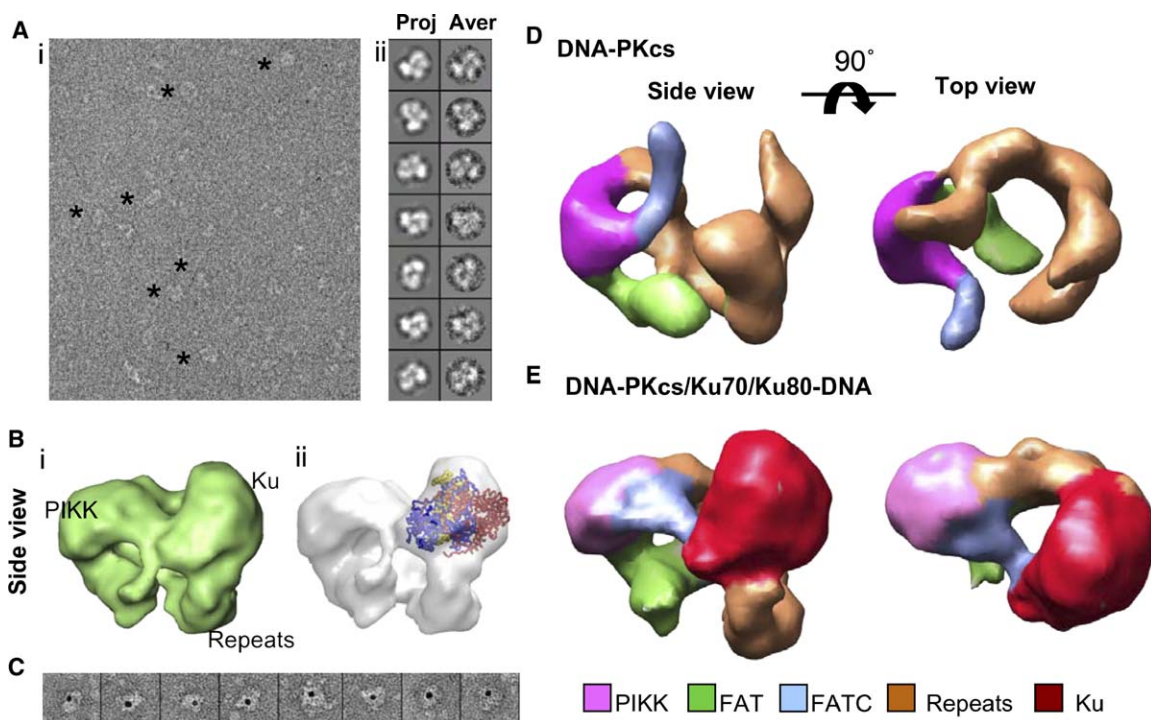


Figure 3. Cryo-EM Reconstruction of DNA Bound DNA-PKcs/Ku70/Ku80 Complexes

(A) (i) A typical cryo-EM field containing particles of the complex (\*). (ii) Selected projections (left column) and their class averages (right column) obtained after 3D refinement of the cryo-EM data.

(B) One view of the cryo-EM structure of DNA bound DNA-PKcs/Ku70/Ku80 complexes, before (i) and after (ii) the fitting of the atomic structure of DNA-Ku (PDB entry 1JEY). In (ii), the 3D volume is shown as white transparency containing the ribbon structure of Ku. DNA in the structure is shown as yellow spheres.

(C) Mapping of iminobiotin DNA within the complex using a gold-streptavidin cluster. Several selected single molecule images are shown.

(D) Side and top views of free DNA-PKcs (taken from [Rivera-Calzada et al. \[2005\]](#)) after filtration to 30 Å and coloring of domains according to the legend shown below.

(E) Comparable views to those of free DNA-PKcs but of the DNA bound DNA-PKcs/Ku70/Ku80 complex.

regulating the kinase activity of DNA-PKcs in response to binding of other macromolecules ([Rivera-Calzada et al., 2005](#)) and is fully consistent with the extreme sensitivity of DNA-PKcs to mutations and truncations of this domain ([Beamish et al., 2000](#)). The density for the proximal claw, which appears to be a specific feature of DNA-PKcs not found in ATM ([Llorca et al., 2003](#); [Rivera-Calzada et al., 2005](#)), is intimately connected to the additional density attributed to Ku in the complex maps and suggests a major role for this domain in mediating interactions. Nevertheless, a very precise description of the interactions between DNA-PKcs and Ku70/Ku80 cannot be supported at this level of resolution, and these results should be interpreted only in overall terms. A more detailed description of the complex could be potentially possible by labeling portions of the proteins. Nonetheless, these approaches can frequently only map a broad area within a reconstruction.

The presence of DNA in the particles used for the 3D reconstructions can be demonstrated by taking advantage of an iminobiotin tag incorporated in one end of the DNA molecule and labeling with streptavidin-gold conjugate as we have previously described ([Boskovic et al., 2003](#)) ([Figure 3C](#)). However, distinct density for the bound DNA is not clearly evident in either the negative-stained or cryo-EM maps, most likely because the DNA density is intimately associated with the protein.

The likely track of bound duplex DNA can be estimated by superimposing the crystal structure of the Ku-DNA complex ([Walker et al., 2001](#)) onto the docked position of the Ku heterodimer in the EM map. When docked in this position, Ku bound DNA would run approximately perpendicular to the longitudinal axis of DNA-PKcs, through the Ku “ring,” and directed along the inner surface of the two claws that form part of the N-terminal HEAT repeats. Insertion of the Ku bound DNA between the head and palm regions of DNA-PKcs is fully consistent with previous observations for DNA bound to DNA-PKcs alone, but the direction of the DNA is rotated by ~15–20° around the long axis of DNA-PKcs compared to its orientation in the absence of Ku.

The larger assemblies observed in the negative-stain fields were processed independently and found to comprise mainly dimers containing two DNA-PKcs molecules. Up to 2151 images of both tilted and untilted particles were extracted and analyzed without imposing any symmetry. These complexes represented approximately one-fifth of the DNA-PKcs molecules in the micrographs. Rotational analysis revealed that many of the single particles conform to a strong 2-fold symmetry ([Figure S3](#)). Reference-free 2D averaging of this data showed that the proximity of pairs of DNA-PKcs molecules is not merely accidental but constitutes a defined assembly with a consistent 2-fold symmetrical

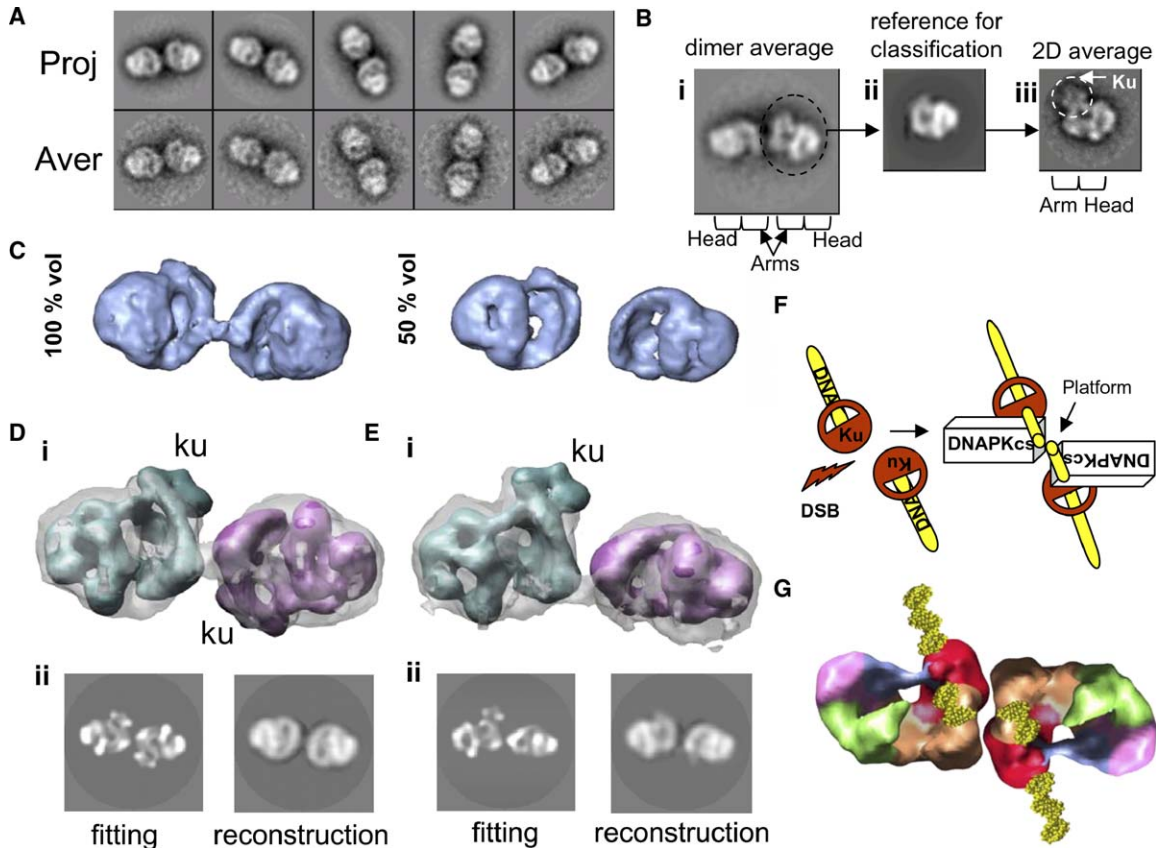


Figure 4. 3D Structure of DNA-PKcs/Ku70/Ku80 Dimeric Complex

(A) A collection of selected projections and averages of dimeric complexes obtained after 3D refinement using 2-fold symmetry. (B) (i) Projections shown in (A) were masked to display just one of the putative monomers and (ii) used as references to classify and align single images of monomers. (iii) Images within each class were then averaged. (C) One view of the 3D structure of the dimer complex at a threshold to display 100% (left) and ~50% (right) of the volume. (D and E) Two views of the 3D structure of the dimer complex, related by rotation along its longitudinal axis. (i) Dimer is shown as transparency and at a threshold to display 100% of the protein mass, whereas the fitted monomers are displayed showing ~50% of their mass to highlight their structural features. (ii) Projections of the dimer along the direction of the view shown above. To the left, projection of the model derived from the fitting. To the right, compatible projection of the 3D reconstruction obtained from the experimental data. (F) Cartoon model of the NHEJ reaction. (G) Model of the putative NHEJ synaptic complex based on the data presented. Color codes are as in Figure 3.

orientation of the paired DNA-PKcs/Ku70/Ku80 complexes (Figure S4A). Furthermore, when a Gaussian blob was used as starting model for a 3D refinement of these images, all the different views of these dimers observed in 2D projection were found to be compatible with a unique volume. Improved averages were obtained (Figure S4E), which were fully consistent with the reference-free averages, and a 3D model refined without imposing any symmetry (Figure S4F). Two-fold symmetry was then imposed in the very last steps of the refinement to yield improved averages (Figure 4A) and an improved volume at a resolution of 33Å (Figure 4C and Figure S4G). Projections of this volume were masked to display just one of the putative monomers (Figure 4B, i) and used as references to classify, align, and average single images of DNA-PKcs/Ku70/Ku80-DNA monomers (Figure 4B, ii). The averages demonstrate that each side of the dimeric projections can be accounted for by a single DNA-PKcs/Ku70/Ku80-DNA complex interacting with the other monomer through the arm region (Figure 4B, iii). These dimeric complexes constituted a significant fraction of the total molecules in

the sample, whereas they were only very rarely observed in similar DNA-PKcs preparations without Ku and DNA added. Furthermore, a comparable proportion of dimers was also observed in the cryo-EM fields (Figure S4B), showing that their formation was not dependent on the addition of stain. In addition to the appearance of dimers in the EM fields, native gel electrophoresis of the DNA-loaded DNA-PKcs/Ku70/Ku80 complexes shows a distinct band running at a significantly higher molecular mass than the bulk of the protein, consistent with the presence of DNA-PKcs/Ku70/Ku80-DNA dimers in solution (Figure S4C).

A 3D map of the complex reconstructed at a resolution of 33 Å shows two DNA-PKcs/Ku70/Ku80 complexes, essentially identical in structure to those reconstructed in isolation but interacting along a common long axis with an axis of 2-fold symmetry perpendicular to that. When the map is contoured to contain 100% of the predicted protein mass (Figure 4C, left, and Figure S4G), the volumes for the two DNA-PKcs/Ku70/Ku80 complexes are clearly connected by a bridge of density, consistent with a symmetric contact between the outer faces of the

distal claw from each DNA-PKcs molecule. The orientation of each monomer within the dimer was impartially examined by computationally fitting the reconstruction of the monomer within the 3D structure of the 2-fold symmetrized complex by using the latest version of SITUS. All top-ranked solutions placed a monomer within either side of the dimer in identical orientations and with the head domains facing outwards (Figure 4D, i, Figure 4E, i, and Figure S5). This orientation is entirely compatible with the structural features of the reconstruction (Figure S4H), and projections of the fitted model correlate well with the experimental projections of the reconstruction (Figures 4D and 4E, ii). In fields where the samples were labeled with streptavidin-gold, strong density corresponding to gold clusters was consistently observed near the center of dimeric particles, approximating to the position of the connecting bridge density observed in the 3D reconstruction (Figure S4D). The localization of gold clusters close to the bridging density shows that the iminobiotin-labeled end of at least one of the bound DNA molecules in the dimer complex is close to, or indeed forms part of, the bridging structure and in that position is accessible to the ~55 kDa streptavidin protein.

Previous studies have suggested a direct role for DNA-PKcs in promoting the synapsis of broken DNA ends (DeFazio et al., 2002; Ma et al., 2004), most probably by self-association of DNA end bound DNA-PKcs molecules. In this present study, we observe formation of dimeric DNA-loaded DNA-PKcs/Ku70/Ku80 complexes of a consistent architecture, suggesting strongly that we have reconstructed such an NHEJ synaptic complex. Although we and others have observed DNA-PKcs self-association in the absence of Ku, the considerably higher yield of dimeric species in its presence suggests an indirect role for Ku in promoting synapsis, possibly by conditioning the conformation of DNA-PKcs. However, Ku itself is not involved in the dimeric contacts but is positioned on the outside face of each DNA-PKcs in the dimer. The structure of this putative NHEJ synaptic complex is thus fully consistent with the current understanding of the early steps of NHEJ, in which the DNA ends are recognized by Ku dimers, which then slide further onto the DNA as DNA-PKcs interacts with the ends (Meek et al., 2004) (Figure 4F). Accordingly, the two Ku dimers in the structure are far apart, whereas the DNA-PKcs molecules contact each other through the claw domains in the N terminus of the kinase (Figure 4G). At the resolution of the present reconstruction, the precision with which the orientation of the Ku heterodimer in the complex can be defined is limited. So, of consequence is the precision with which the trajectory of an extended Ku bound DNA can be predicted. However, it is clear that the angle between the distal segments of the DNA molecules passing through the Ku heterodimers is relatively small and that no substantial distortion of the surrounding DNA (and chromatin) would be required for this putative NHEJ synaptic complex to assemble.

Completion of NHEJ involves ligation of the broken ends by XRCC4-DNA ligase IV complex, whose end-joining activity is substantially stimulated by direct protein-protein interactions with DNA-PKcs/Ku70/Ku80 (Hsu et al., 2002). Depending on the nature of the broken

DNA ends, substantial modification may be required before XRCC4-DNA ligase IV can act. This may involve endo- and/or exonuclease cleavage by Artemis (Ma et al., 2002) and addition and/or removal of terminal phosphates by PNK (which is also bound to XRCC4) (Koch et al., 2004) as well as microhomology-primed extension by polymerases  $\lambda$  and/or  $\mu$  (Ma et al., 2004). Thus, considerable protein traffic must be accommodated without loss of synapsis, so the DNA ends bound to the DNA-PKcs/Ku70/Ku80 complex must be tethered but accessible. Although it is still not clear at which stage of the NHEJ process autophosphorylation actually occurs, its effect on DNA-PKcs is to decrease its affinity for DNA and it may be required to release DNA ends and make them available for downstream enzymes. However, it is likely that DNA-PKcs/Ku70/Ku80 remains closely associated and undergoes a structural remodeling on autophosphorylation rather than simply dissociating (Block et al., 2004; Reddy et al., 2004). With the abutting broken ends of the bound DNA localized close to the interface of the claw domains and accessible to streptavidin-gold, the structure of the DNA bound DNA-PKcs/Ku70/Ku80 dimer we have determined appears well suited to maintaining end proximity and encouraging microannealing while providing a flexible binding platform for the assembly of the multienzyme complex that ultimately achieves the rejoining of the double-strand break.

#### Experimental Procedures

##### Isolation of DNA Bound DNA-PKcs/Ku70/Ku80 Complexes

DNA-PKcs, Ku70, and Ku80 were purified from HeLa nuclear extract (CilBiotech, Belgium) with a buffer system based on 20 mM HEPES (pH 7.5), 1 mM DTT, 0.5 mM EDTA, 0.001%  $\beta$ -octylglucoside, and 10% glycerol, with NaCl concentrations ranging from 50 to 1000 mM in different chromatographic steps. After fractionation on heparin and dsDNA cellulose (GE Healthcare), fractions containing DNA-PKcs, Ku70, and Ku80 were loaded onto a 1 ml ResourceQ column (GE Healthcare) and eluted with a linear NaCl gradient. Fractions containing all three protein components were incubated with  $\lambda$ -phosphatase for 1 hr at 4°C. The sample was divided into three parts, one of which was not treated further; the others were then incubated for an additional hour at 4°C with either “short” or “long” dsDNA molecules iminobiotinylated (IB) at the 5' end of the longer oligo.

Short dsDNA was obtained by annealing oligos 1 and 2. Long dsDNA was obtained by annealing oligos 3 and 4. All oligos were purchased from IBA GmbH, Goettingen, Germany. The sequences of the oligos (with the linear duplex segments in uppercase) were: 1, 5'-cgcgccagctttcccagctaataaactaaaacTATTATTATGGCCGCA CGCGT-3'; 2, 5'-ACGCGTGCGGCCATAATAATAGTTTTAGTTTTAT TGGGCGCG-3'; 3, 5'-cgcgccagctttcccagctaataaactaaaacTAT TATTATGGCCGACGCGTCCACCATGGGGTACAAC-3'; and 4, 5'-AGTTGTACCCCATGGTGACGCGTGCGGCCATAATAATAGTTT TTAGTTTATTGGGCGCG-3'.

The samples were then loaded on a 18%–60% glycerol gradient, with NaCl at the constant concentration of 200 mM, and further separation was achieved by using a Beckman SW28 rotor, spinning for 72 hr at 25,000 rpm. Fractions were collected from the bottom and analyzed by means of SDS-PAGE on 4%–12% Bis-Tris NOVEX gradient gels (Invitrogen). DNA-PKcs/Ku70/Ku80-DNA complexes were additionally analyzed under native PAGE conditions on 3%–9% Tris-Acetate NOVEX gradient gels to assess their oligomerisation state.

##### Electron Microscopy and 3D Reconstruction of Negatively Stained DNA-PKcs/Ku70/Ku80 Complexes

The complexes were applied to carbon-coated grids, negatively stained with 1% uranyl acetate, observed in a JEOL 1230 electron

microscope operating at 100 kV, and micrographs were recorded at a magnification of 50,000 under low dose conditions. Images were digitized by using a MINOLTA Dimage Scan Multi Pro scanner at 2400 dpi and averaged to a final 4.4 Å/pixel at the specimen. 14,239 images of individual molecules from both untilted and 40° tilted specimens were boxed, processed, and refined by using EMAN (Ludtke et al., 1999). Initial analysis of the data was performed in 2D by reference-free classification of the data set into a collection of 2D averages. For 3D analysis, the particles were processed by angular refinement techniques implemented in EMAN. To avoid model bias, several models were used as starting volumes for refinement but, importantly, never any information from any previous works. Some models were constructed by randomly generated noisy Gaussian blobs using EMAN, whereas others were built after reference-free image classification, averaging, and using common lines as implemented in EMAN. In all cases, volumes rapidly evolved toward a common 3D solution, fully consistent with previous reconstructions. Resolution was estimated with the Fourier Shell Correlation (FSC) of two independent reconstructions refined by using even and odd halves of the particle data, which was found to correspond to 25 Å and 21 Å by using the 0.5 correlation coefficient and 3  $\sigma$  criteria, respectively. Gold labeling was performed by incubating streptavidin-gold conjugates with the purified complex as previously described (Boskovic et al., 2003).

Around 2151 images of larger dimeric complexes from both tilted and untilted micrographs were similarly processed to generate a map at an estimated resolution of 33 Å (0.5 correlation coefficient) and 28 Å (3  $\sigma$ ). No symmetry was assumed except at the very final stages of image processing, where 2-fold rotational symmetry was applied. In all cases, handedness was chosen to match the absolute hand defined previously by an electron diffraction study of 2D crystals (Leuther et al., 1999).

#### Cryo-EM and 3D Reconstruction of DNA-PKcs/Ku70/Ku80

The DNA-PKcs/Ku70/Ku80 sample was dialyzed into 20 mM HEPES (pH 7.4), 200 mM KCl, 0.1 mM DTT, and 0.5 mM EDTA prior to vitrification to remove detergent and other components of the purification buffer that were found to strongly interfere with the vitrification. Five microliters of freshly dialyzed material was vitrified by plunging into liquid ethane and observed in a Tecnai G<sup>2</sup> FEG 200 operated at 200 kV and equipped with a Gatan liquid nitrogen specimen holder. Images were recorded at a magnification of 50,000 and by using different defocus on Kodak SO-163 film. Micrographs were digitized and averaged to a final 4.4 Å/pixel. Due to the limited amount and concentration of the material recover after dialysis, around 4000 particles were extracted. The contrast transfer function (CTF) of the microscope for each image was estimated by using CTFIND3 and corrected by flipping phases with ctfilt command from EMAN as described (Rivera-Calzada et al., 2005). Image processing was carried as described above and by using several starting volumes but never using the cryo-EM structure from DNA-PKcs. FSC of two volumes generated after splitting the initial data set in two halves provided an estimate of 32 Å and 28 Å, according to the 0.5 and 3  $\sigma$  criteria, respectively. The volume was rendered to show either 100% or 50% of the protein mass assuming an average density of 1.35 g/ml and visualized by using VMD (Humphrey et al., 1996) and UCSF Chimera (Pettersen et al., 2004).

#### Supplemental Data

Supplemental Data include Supplemental References and five figures and can be found with this article online at <http://www.molecule.org/cgi/content/full/22/4/511/DC1/>.

#### Acknowledgments

This work has been supported by projects SAF2002-01715, SAF2005-00775, and GEN2003-20239-C06-06 from the Spanish Ministry of Education and Science (O.L.) and a Programme Grant from Cancer Research UK (L.H.P.). A.R.-C. is supported by an I3P PhD fellowship from the Spanish National Research Council (CSIC) and the European Social Fund. We thank Prof. José L. Carrascosa, Prof. José M. Valpuesta, and Prof. José M. Carazo for the use of their FEG electron microscope facility and Drs. Edward Morris and Paula di Fonseca for helpful discussion.

Received: March 2, 2006

Revised: March 29, 2006

Accepted: April 11, 2006

Published: May 18, 2006

#### References

- Abraham, R.T. (2004). PI 3-kinase related kinases: 'big' players in stress-induced signaling pathways. *DNA Repair (Amst.)* 3, 883–887.
- Beamish, H.J., Jessberger, R., Riballo, E., Priestley, A., Blunt, T., Kysela, B., and Jeggo, P.A. (2000). The C-terminal conserved domain of DNA-PKcs, missing in the SCID mouse, is required for kinase activity. *Nucleic Acids Res.* 28, 1506–1513.
- Block, W.D., Yu, Y., Merkle, D., Gifford, J.L., Ding, Q., Meek, K., and Lees-Miller, S.P. (2004). Autophosphorylation-dependent remodeling of the DNA-dependent protein kinase catalytic subunit regulates ligation of DNA ends. *Nucleic Acids Res.* 32, 4351–4357.
- Boskovic, J., Rivera-Calzada, A., Maman, J.D., Chacon, P., Willison, K.R., Pearl, L.H., and Llorca, O. (2003). Visualization of DNA-induced conformational changes in the DNA repair kinase DNA-PKcs. *EMBO J.* 22, 5875–5882.
- Bosotti, R., Isacchi, A., and Sonnhammer, E.L. (2000). FAT: a novel domain in PIK-related kinases. *Trends Biochem. Sci.* 25, 225–227.
- Brewerton, S.C., Dore, A.S., Drake, A.C., Leuther, K.K., and Blundell, T.L. (2004). Structural analysis of DNA-PKcs: modelling of the repeat units and insights into the detailed molecular architecture. *J. Struct. Biol.* 145, 295–306.
- Chiu, C.Y., Cary, R.B., Chen, D.J., Peterson, S.R., and Stewart, P.L. (1998). Cryo-EM imaging of the catalytic subunit of the DNA-dependent protein kinase. *J. Mol. Biol.* 284, 1075–1081.
- Dames, S.A., Mulet, J.M., Rathgeb-Szabo, K., Hall, M.N., and Grzesiek, S. (2005). The solution structure of the FATC domain of the protein kinase target of rapamycin suggests a role for redox-dependent structural and cellular stability. *J. Biol. Chem.* 280, 20558–20564.
- DeFazio, L.G., Stansel, R.M., Griffith, J.D., and Chu, G. (2002). Synapsis of DNA ends by DNA-dependent protein kinase. *EMBO J.* 21, 3192–3200.
- Ding, Q., Reddy, Y.V., Wang, W., Woods, T., Douglas, P., Ramsden, D.A., Lees-Miller, S.P., and Meek, K. (2003). Autophosphorylation of the catalytic subunit of the DNA-dependent protein kinase is required for efficient end processing during DNA double-strand break repair. *Mol. Cell. Biol.* 23, 5836–5848.
- Douglas, P., Gupta, S., Morrice, N., Meek, K., and Lees-Miller, S.P. (2005). DNA-PK-dependent phosphorylation of Ku70/80 is not required for non-homologous end joining. *DNA Repair (Amst.)* 4, 1006–1018.
- Falck, J., Coates, J., and Jackson, S.P. (2005). Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. *Nature* 434, 605–611.
- Harris, R., Esposito, D., Sankar, A., Maman, J.D., Hinks, J.A., Pearl, L.H., and Driscoll, P.C. (2004). The 3D solution structure of the C-terminal region of Ku86 (Ku86CTR). *J. Mol. Biol.* 335, 573–582.
- Hsu, H.L., Yannone, S.M., and Chen, D.J. (2002). Defining interactions between DNA-PK and ligase IV/XRCC4. *DNA Repair (Amst.)* 1, 225–235.
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38.
- Jin, S., Kharbanda, S., Mayer, B., Kufe, D., and Weaver, D.T. (1997). Binding of Ku and c-Abl at the kinase homology region of DNA-dependent protein kinase catalytic subunit. *J. Biol. Chem.* 272, 24763–24766.
- Koch, C.A., Agyei, R., Galicia, S., Metalnikov, P., O'Donnell, P., Starostine, A., Weinfeld, M., and Durocher, D. (2004). Xrcc4 physically links DNA end processing by polynucleotide kinase to DNA ligation by DNA ligase IV. *EMBO J.* 23, 3874–3885.
- Kysela, B., Chovanec, M., and Jeggo, P.A. (2005). Phosphorylation of linker histones by DNA-dependent protein kinase is required for DNA ligase IV-dependent ligation in the presence of histone H1. *Proc. Natl. Acad. Sci. USA* 102, 1877–1882.

- Leuther, K.K., Hammarsten, O., Kornberg, R.D., and Chu, G. (1999). Structure of DNA-dependent protein kinase: implications for its regulation by DNA. *EMBO J.* *18*, 1114–1123.
- Llorca, O., Rivera-Calzada, A., Grantham, J., and Willison, K.R. (2003). Electron microscopy and 3D reconstructions reveal that human ATM kinase uses an arm-like domain to clamp around double-stranded DNA. *Oncogene* *22*, 3867–3874.
- Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* *128*, 82–97.
- Ma, Y., Pannicke, U., Schwarz, K., and Lieber, M.R. (2002). Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* *108*, 781–794.
- Ma, Y., Lu, H., Tippin, B., Goodman, M.F., Shimazaki, N., Koiwai, O., Hsieh, C.L., Schwarz, K., and Lieber, M.R. (2004). A biochemically defined system for mammalian nonhomologous DNA end joining. *Mol. Cell* *16*, 701–713.
- Ma, Y., Pannicke, U., Lu, H., Niewolik, D., Schwarz, K., and Lieber, M.R. (2005). The DNA-dependent protein kinase catalytic subunit phosphorylation sites in human artemis. *J. Biol. Chem.* *280*, 33839–33846.
- Meek, K., Gupta, S., Ramsden, D.A., and Lees-Miller, S.P. (2004). The DNA-dependent protein kinase: the director at the end. *Immunol. Rev.* *200*, 132–141.
- Merkle, D., Douglas, P., Moorhead, G.B., Leonenko, Z., Yu, Y., Cramb, D., Bazett-Jones, D.P., and Lees-Miller, S.P. (2002). The DNA-dependent protein kinase interacts with DNA to form a protein-DNA complex that is disrupted by phosphorylation. *Biochemistry* *41*, 12706–12714.
- Petersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* *25*, 1605–1612.
- Reddy, Y.V., Ding, Q., Lees-Miller, S.P., Meek, K., and Ramsden, D.A. (2004). Non-homologous end joining requires that the DNA-PK complex undergo an autophosphorylation-dependent rearrangement at DNA ends. *J. Biol. Chem.* *279*, 39408–39413.
- Rivera-Calzada, A., Maman, J.D., Spagnolo, L., Pearl, L.H., and Llorca, O. (2005). Three-dimensional structure and regulation of the DNA-dependent protein kinase catalytic subunit (DNA-PKcs). *Structure* *13*, 243–255.
- Walker, J.R., Corpina, R.A., and Goldberg, J. (2001). Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* *412*, 607–614.
- West, S.C. (2003). Molecular views of recombination proteins and their control. *Nat. Rev. Mol. Cell Biol.* *4*, 435–445.
- Wriggers, W., Milligan, R.A., and McCammon, J.A. (1999). Situs: a package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.* *125*, 185–195.
- Yaneva, M., Kowalewski, T., and Lieber, M.R. (1997). Interaction of DNA-dependent protein kinase with DNA and with Ku: biochemical and atomic-force microscopy studies. *EMBO J.* *16*, 5098–5112.
- Zhang, Z., Hu, W., Cano, L., Lee, T.D., Chen, D.J., and Chen, Y. (2004). Solution structure of the C-terminal domain of Ku80 suggests important sites for protein-protein interactions. *Structure* *12*, 495–502.

#### Accession Numbers

The EM maps of the DNA-PKcs-Ku complex have been deposited in the 3D EM database (<http://www.ebi.ac.uk/msd/>) under accession codes EMD-1209 (monomeric DNA bound DNA-PKcs/Ku70/Ku80 complex using negative staining), EMD-1210 (dimeric DNA bound DNA-PKcs/Ku70/Ku80 complex using negative staining), and EMD-1208 (DNA bound DNA-PKcs/Ku70/Ku80 complex using cryo-EM).

## Supplemental Data

### Three-Dimensional Structure of the Human

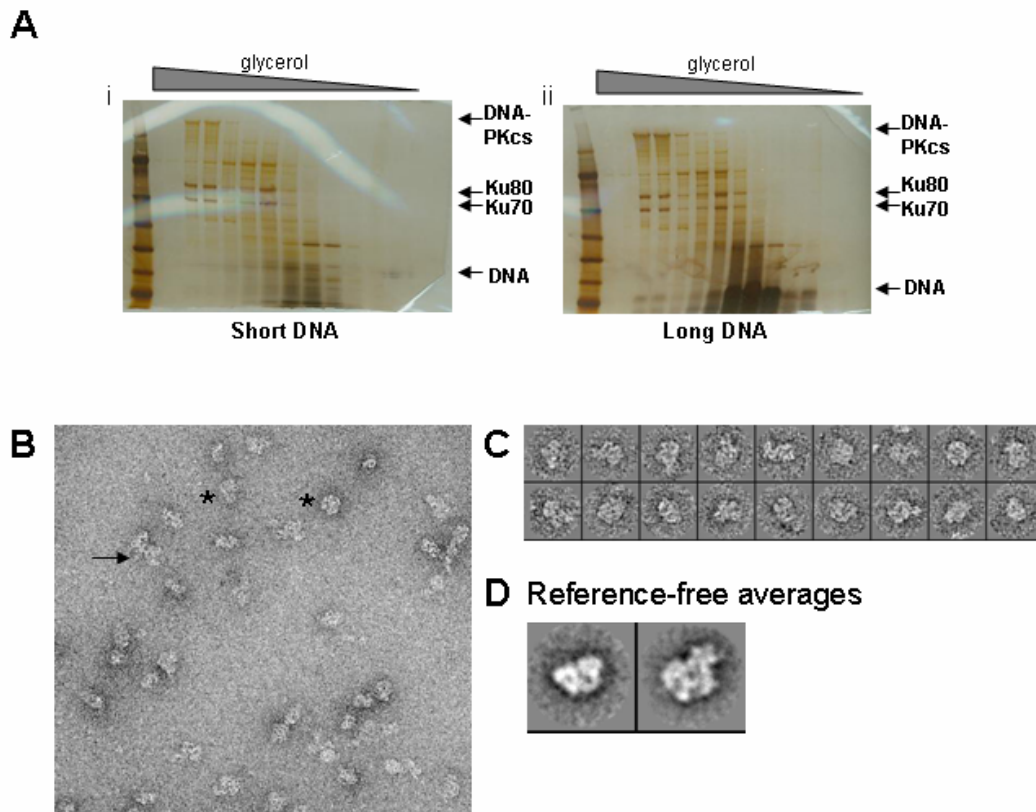
### DNA-PKcs/Ku70/Ku80 Complex Assembled

### on DNA and Its Implications for DNA DSB Repair

Laura Spagnolo, Angel Rivera-Calzada, Laurence H. Pearl, and Oscar Llorca

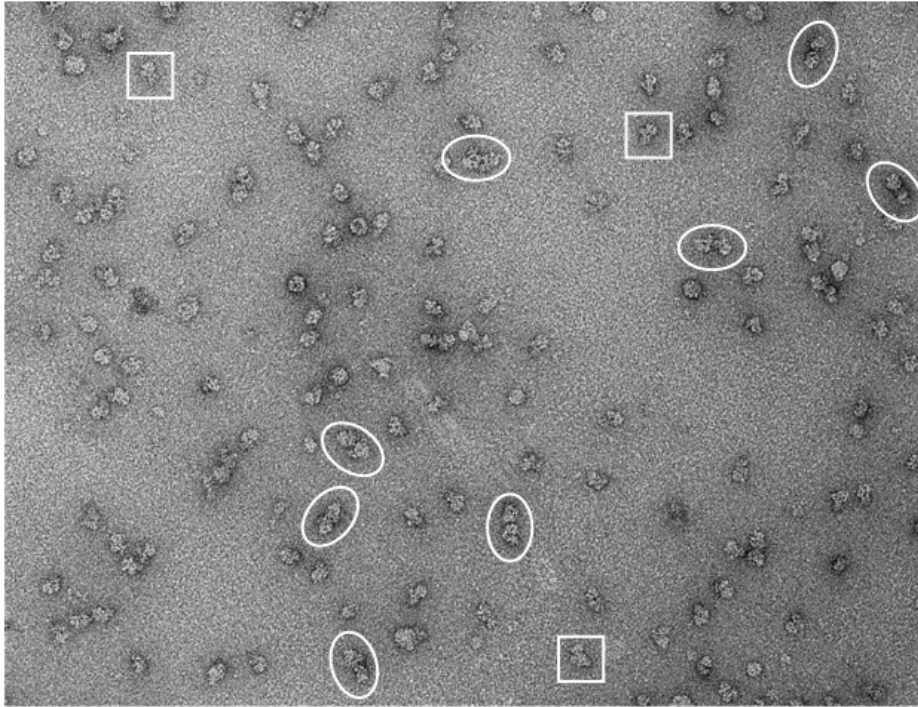
## Supplemental References

- S1. R.A. Crowther, L. Amos, *J. Mol. Biol.* **60**, 123 (1971).
- S2. A. Pascual-Montano, L. E. Donate, M. Valle, M. Bárcena, R. D. Pascual-Marqui, J. M. Carazo, *J Struct Biol.* **133**, 233 (2001).
- S3. C.O. Sorzano, R. Marabini, J. Velazquez-Muriel, J.R. Bilbao-Castro, S.H. Scheres, J.M. Carazo, A. Pascual-Montano, *J Struct Biol.* **148**, 194 (2004).
- S4. L.E. Donate, O. Llorca, M. Barcena, S.E. Brown, N.E. Dixon, J.M. Carazo, *J Mol Biol.* **303**, 383 (2000).

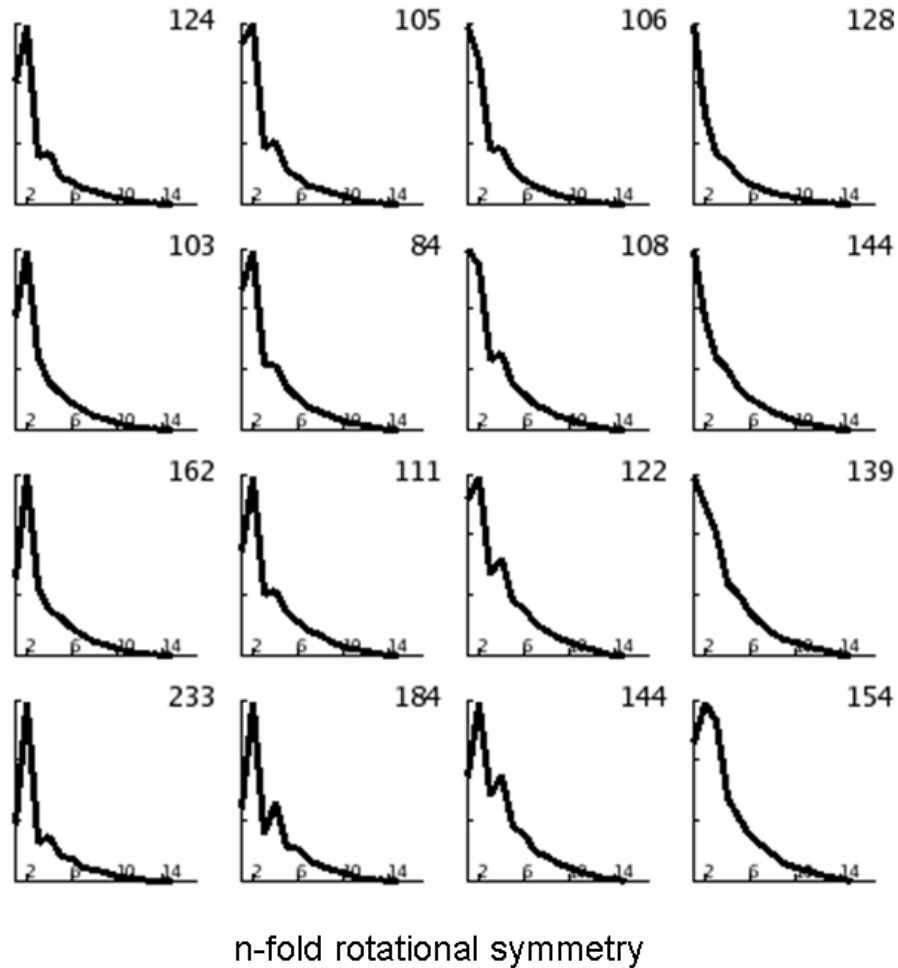


**Figure S1.** DNA-PKcs/Ku70/Ku80-DNA complexes using the “long DNA”. (A) Purification of DNA-PKcs, Ku70 and Ku80 and isolation of their complex. Glycerol gradients performed with the purified proteins in the presence of the “short DNA” (i) and the “long DNA” DNA (ii). (B) EM field of purified complexes. Monomers (\*). Dimeric aggregates. (C) Selected raw images of monomers. (D) A few reference-free class averages of the monomeric images.

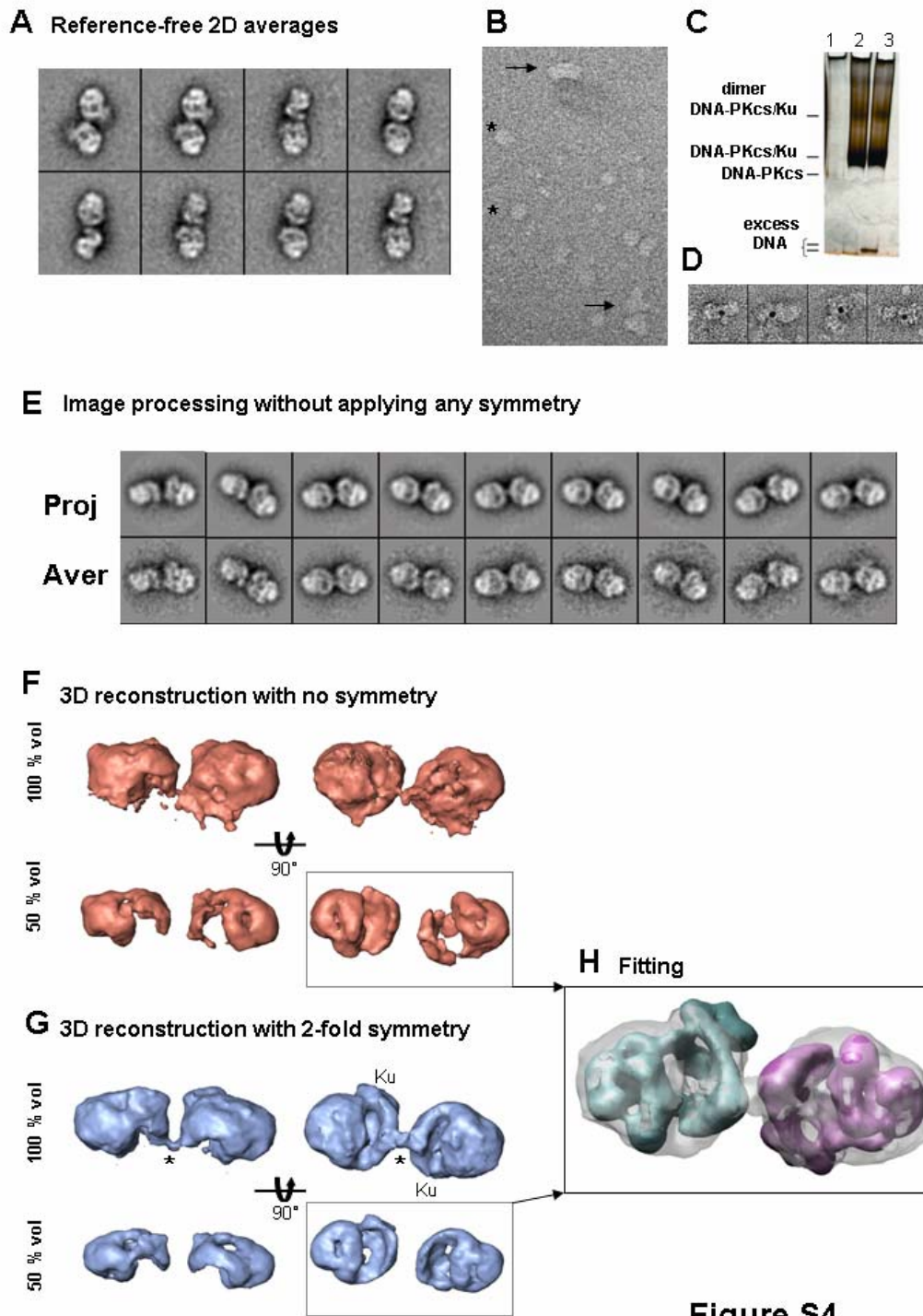




**Figure S2.** EM field from a micrograph containing DNA-PKcs/Ku70/Ku80-DNA complexes after negative staining. Some monomers are highlighted within white squares and some dimers within white elongated circles.



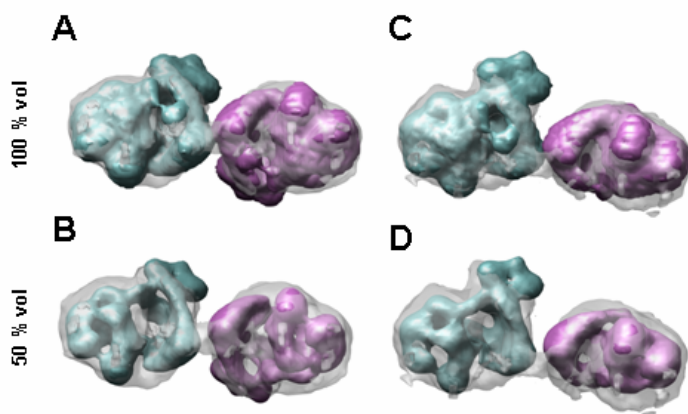
**Figure S3.** Rotational analysis of dimeric complexes. We calculated the rotational power spectra of the previously translationally centered dimeric images (*S1*). These spectra were then analyzed by means of the Kernel Probability Density Estimator Self-Organizing Map (KenDerSOM) algorithm (*S2*) implemented in the XMIPP software platform (<http://www.cnb.uam.es/~bioinfo/>) (*S3*) and as described before (*S4*). This is a totally unsupervised algorithm that makes no assumption at the outset of the expected trend of variability. Heterogeneities in the symmetry of the particles of the original population were sorted, therefore, based on the power spectrum calculated for each of the images. Figure depicts the output 4×4 two-dimensional map of code vectors obtained, where the upper right corner of each panel indicates the number of images assigned to each group. Within each panel, the X-axis indicates the n-foldness of rotational symmetry. Most groups reveal a very strong 2-fold component and hardly any other symmetry.



**Figure S4**

**Figure S4.** Image Processing and analysis of DNA-PKcs/Ku70/Ku80-DNA dimers. (A) A collection of selected averages of dimeric images obtained after reference-free 2D classification, alignment and averaging. (B) EM field from a cryo-EM micrograph

containing monomeric (\*) and dimeric (arrows) complexes. (C), native gel of the purified DNA-PKcs/Ku70/Ku80 complex, showing that ~20 % of the protein mass migrates as a larger aggregate compatible with the dimeric species observed by EM. (1) DNA-PKcs alone. (2 and 3) Complexes of DNA-PKcs/Ku70/Ku80 in the presence of short and long DNA respectively. (D) Mapping of imino-biotin DNA within the dimer complex using a gold-streptavidin cluster. The figure displays a few selected dimers and the gold density was always found to sit within the middle segment of the molecule. (E) Pairs of projections and their class averages generated during 3D refinement of the dimers data without imposing any symmetry. (F-G), Two views of the 3D structure of the dimer complex, related by a 90 degree rotation, at a threshold to display 100% (top row) and ~50% (bottom row) of the volume. (\*) contact area between the two monomers in the complex. (F) Without applying 2-fold symmetry. (G) After applying 2-fold symmetry. (H) The computational fitting of the reconstructed monomer within the dimer. The orientation of each monomer in model obtained after the fitting correlates well with the structural features of the reconstructed volumes from the experimental data.



**Figures S5.** (A and C) Two views of the 3D structure of the dimer complex, related by rotation along the longitudinal axis. The dimer is shown as transparency and at a threshold to display 100% of the protein mass. Fitted monomers displayed showing 100% of their volume. (B and D) Similar views to those above but with the fitted monomers displayed showing ~50% of their mass to highlight their structural features.



