



Universidad Autónoma de Madrid
Departamento de Biología Molecular
Facultad de Ciencias

**Estudio de la estructura cuaternaria
de p53 y el complejo p53/ADN
mediante microscopía electrónica**

TESIS DOCTORAL

Roberto Melero del Río
Madrid, 2010



En relación a la propuesta de defensa oral de Tesis Doctoral realizada por D. Roberto Melero del Río, los abajo firmantes

INFORMAN QUE

D. Roberto Melero del Río, licenciado en Químicas por la Universidad de Valladolid y licenciado en Bioquímica por la Universidad Complutense de Madrid, ha realizado bajo nuestra dirección el trabajo experimental titulado “Estudio de la estructura cuaternaria de p53 y el complejo p53/ADN mediante microscopía electrónica”, cuya memoria se presenta para optar al título de Doctor en Ciencias (Biología Molecular) por la Universidad Autónoma de Madrid.

La investigación se ha realizado en el Centro Nacional de Biotecnología (C.N.B.), en el marco de la Unidad de Biocomputación del Departamento de Estructura de Macromoléculas del CNB.

Dicha investigación se ha centrado en el estudio estructural de p53 libre y en interacción con ADN a partir de imágenes de microscopía electrónica de ambos complejos. Mediante técnicas de procesamiento de imagen y promediado de las imágenes individuales se han obtenido mapas tridimensionales que han servido para comprender la organización de p53 y su reconocimiento de secuencias específicas de ADN para la formación del complejo tetramérico. El trabajo desarrollado por el doctorando se encuadra en un marco multidisciplinar, pues se han empleado técnicas y metodologías procedentes de campos como la bioquímica, la bioinformática y, principalmente, la microscopía electrónica y el procesamiento digital de imágenes, para solucionar los problemas que surgen en el análisis estructural de estas macromoléculas biológicas. Dentro de las técnicas de microscopía electrónica tridimensional se ha hecho uso de la microscopía electrónica de transmisión convencional que permite observar las muestras en condiciones muy próximas a las fisiológicas.

El doctorando, que ha realizado personalmente todos los procedimientos experimentales recogidos en la memoria de tesis, se ha revelado durante este periodo como un investigador sumamente capaz y motivado, con lo que ha adquirido una excelente formación tanto por amplitud como por profundidad.

El trabajo desarrollado en esta propuesta de Tesis es novedoso, desde el punto de vista conceptual y de ejecución; los resultados obtenidos aportan nueva información que mejora el entendimiento del papel de p53 dentro de la célula como factor de transcripción. Sus Directores estamos por tanto firmemente convencidos de la idoneidad tanto del doctorando como de la propuesta de Tesis, y proponemos sin reservas su consideración para defensa oral.

Y para que conste a los efectos oportunos, lo firman los directores en

Madrid a 25 de Febrero de 2010.



Dr. José M. Carazo García
Profesor de Investigación del CSIC



Dr. Mikel Valle Rodríguez
Investigador Principal del CICBIOGUNE

Departamento de Estructura de Macromoléculas
Centro Nacional de Biotecnología
CSIC

Abstract

Tumor suppressor p53 is a tetrameric, multidomain transcription factor that plays central roles in the cell cycle and in maintaining genomic integrity. p53 binds to specific DNA response elements, and is integrated in various signalling networks by a multitude of protein-protein interactions. Each peptide chain consists of four different domains: (i) the N-terminal transactivation domain (residues 1-94), (ii) the core or DNA-binding domain (residues 94-294), (iii) the tetramerization domain (residues 323-360), and (iv) the C-terminal domain (CTD; residues 360-393). The CTD contains an arginine- and lysine-rich basic domain which was shown to bind DNA in a sequence non-specific manner. A sliding model has been postulated, where this non-specific binding of the CTD is employed to search for p53 response elements along the DNA, which are ultimately bound in a sequence-specific manner via the core domain.

We isolated full-length p53 tetramers bound to DNA. These complexes were chemically cross-linked using a combined glycerol-glutaraldehyde gradient (GRAFIX), and visualized by negative staining. To explore the conformational flexibility of this complex, we performed 3D maximum likelihood-based classification.

The results reveal dynamic and variable p53 tetramers that adopt different architectures. Some of the detected conformations seem to oscillate between specific and unspecific DNA recognition, a behaviour that might be relevant for the regulation of p53 transcriptional activity.

Índice General

| | |
|---|-----------|
| I. INTRODUCCIÓN..... | 1 |
| I.1 Sistemas de regulación del ciclo celular..... | 1 |
| I.2 El gen supresor tumoral p53..... | 2 |
| I.3 La proteína p53..... | 2 |
| I.3.1 Función de p53..... | 2 |
| I.3.2 Regulación de la actividad de p53..... | 5 |
| I.3.3 Estructura de p53..... | 7 |
| a. Dominio amino terminal..... | 7 |
| b. Dominio de unión de ADN..... | 8 |
| c. Dominio de tetramerización..... | 13 |
| d. Dominio carboxilo terminal..... | 14 |
| I.3.4 Regulación de unión a ADN..... | 14 |
| I.4 Microscopía electrónica..... | 19 |
| I.4.1 El microscopio electrónico de transmisión..... | 19 |
| I.4.2 Estudio de muestras biológicas en un microscopio electrónico..... | 20 |
| I.4.3 Procesamiento de imagen..... | 22 |
| I.4.4 Reconstrucción tridimensional..... | 22 |
| OBJETIVOS..... | 25 |
| M. MATERIALES Y MÉTODOS..... | 27 |
| M.1 Muestra de p53..... | 27 |
| M.2 Muestras de ADN..... | 27 |
| M.3 Formación del complejo p53-ADN..... | 28 |
| M.4 Preparación de muestras para microscopía electrónica..... | 29 |
| M.4.1 Análisis de la muestra total..... | 29 |
| M.4.2 GRAFIX..... | 30 |
| M.5 Obtención de imágenes en el microscopio electrónico..... | 32 |
| M.6 Procesamiento de imagen..... | 32 |
| M.6.1 Clasificación 2D..... | 33 |
| M.6.2 Reconstrucción 3D..... | 33 |
| M.6.3 Clasificación 3D..... | 34 |
| M.6.4 Ajuste de estructuras atómicas..... | 34 |

| | |
|---|-----------|
| R. RESULTADOS..... | 37 |
| R.1 p53 libre..... | 37 |
| a. Caracterización bioquímica..... | 37 |
| b. Reconstrucción tridimensional..... | 38 |
| R.2 Complejo p53-ADN..... | 41 |
| R.2.1 Análisis de la muestra total..... | 41 |
| a. Caracterización bioquímica..... | 41 |
| b. Reconstrucción tridimensional..... | 42 |
| R.2.2 GRAFIX: purificación en gradiente de densidad/fijación..... | 47 |
| R.2.2.1 Purificación de tetrámeros de p53 unidos a ADN de 60 pb..... | 48 |
| a. Caracterización bioquímica..... | 48 |
| b. Reconstrucción tridimensional..... | 49 |
| c. Polimorfismo cuaternario de p53 en su interacción con ADN de 60 pb.... | 52 |
| R.2.2.2 Purificación de tetrámeros de p53 unidos a ADN de 44pb..... | 57 |
| a. Caracterización bioquímica..... | 57 |
| b. Reconstrucción tridimensional..... | 58 |
| c. Polimorfismo cuaternario de p53 en su interacción con ADN de 44 pb.... | 60 |
| D. DISCUSIÓN..... | 65 |
| D.1 p53 libre, antes de su interacción con el ADN..... | 66 |
| D.2 p53 formando complejo con el ADN..... | 69 |
| D.3 Mecanismo de formación del complejo tetramérico p53-ADN..... | 71 |
| D.4 Papel regulador del dominio básico C-terminal en la formación del complejo p53-ADN..... | 74 |
| D.5 Estudio de la influencia del tamaño del ADN en la formación del complejo p53-ADN..... | 76 |
| D.6 Papel del dominio C-terminal cuando no hay deslizamiento sobre el ADN..... | 79 |
| D.7 Implicaciones biológicas..... | 81 |
| CONCLUSIONES..... | 85 |
| BIBLIOGRAFÍA..... | 87 |
| GLOSARIO..... | 97 |
| Direcciones Web..... | 98 |
| Apéndice..... | 99 |

Índice de Figuras

| | |
|--|----|
| 1. Esquema del mecanismo de regulación negativa de p53, por MDM2, Ciclina-G y Wip-1..... | 5 |
| 2. Modelo de activación de p53. | 6 |
| 3. Activación de p53..... | 7 |
| 4. Diagrama esquemático de los dominios de p53. | 7 |
| 5. Estructura atómica de la región (12-26) del dominio N-terminal de p53..... | 8 |
| 6. Estructura atómica del dominio central de p53..... | 9 |
| 7. Simetrías de la secuencia específica de ADN. | 10 |
| 8. Estudio de los efectos causados por las mutaciones en p53. | 11 |
| 9. Curvas de desnaturalización con urea del dominio central, tanto de p53 silvestre (WT) como de mutantes a pH = 7.2..... | 12 |
| 10. Comparación entre estructuras cristalinas del dominio de unión de ADN..... | 13 |
| 11. Estructura atómica del dominio de tetramerización..... | 13 |
| 12. Distribución del punto isoeléctrico de p53..... | 14 |
| 13. Modelos de activación de p53 en su unión a una secuencia específica de ADN. | 16 |
| 14. Modelo del mecanismo de unión p53-ADN..... | 17 |
| 15. Organización estructural de p53 de ratón. | 18 |
| 16. Rango de estructuras macromoleculares estudiadas por microscopía electrónica de transmisión..... | 20 |
| 17. Comparación de imágenes de microscopía electrónica mediante el método normal (análisis de la muestra total) y el método Gráfico..... | 30 |
| 18. Caracterización bioquímica de p53 libre..... | 37 |
| 19. Microscopía electrónica de p53 libre..... | 38 |
| 20. Galería de proyecciones y medias equivalentes de p53 libre. | 39 |
| 21. Estructura tridimensional del tetrámero p53 libre..... | 40 |
| 22. Resolución de mapa 3D de p53 libre. | 41 |
| 23. Caracterización bioquímica del complejo p53-ADN. | 42 |
| 24. Microscopía electrónica del complejo p53-ADN. | 43 |
| 25. Refinamiento angular de p53-ADN. | 44 |
| 26. Estructura tridimensional del complejo p53-ADN..... | 46 |
| 27. Resolución de mapa 3D del complejo p53-ADN..... | 46 |

| | |
|--|-----------|
| 28. Caracterización bioquímica de p53 y ADN de 60pb tras el gradiente de densidad/fijación (Método Grafix)..... | 49 |
| 29. Microscopía electrónica del complejo tetramérico p53 con ADN de 60 pb por el método Grafix. | 50 |
| 30. Galería de proyecciones y medias equivalentes del complejo p53-ADN (de 60pb) tras el método Grafix, obtenidas en el proceso de refinamiento angular iterativo. | 51 |
| 31. Mapa de microscopía electrónica de p53-ADN (de 60 pb) tras el método Grafix. | 52 |
| 32. Resolución del mapa 3D de p53-ADN (de 60 pb) tras el método Grafix..... | 52 |
| 33. Clasificación mediante <i>ML3D</i> del complejo p53-ADN (de 60 pb) tras el método Grafix..... | 54 |
| 34. Ajuste de estructuras atómicas en las clases I y III del complejo p53-ADN (de 60pb) | 55 |
| 35. Ajuste del dominio C-terminal en los mapas 3D de las clases I y III del complejo p53-ADN (de 60pb) | 56 |
| 36. Caracterización bioquímica de p53 y ADN de 44pb tras el gradiente de densidad/fijación..... | 57 |
| 37. Microscopía electrónica del complejo tetramérico p53con ADN de 44 pares de bases por el método Grafix. | 58 |
| 38. Galería de proyecciones y medias equivalentes del complejo p53-ADN (de 44pb) tras el método Grafix..... | 59 |
| 39. Mapa de microscopía electrónica de p53-ADN (de 44 pb) tras el método Grafix. | 60 |
| 40. Resolución del mapa 3D de p53-ADN (de 44 pb) tras el método Grafix..... | 60 |
| 41. Clasificación mediante <i>ML3D</i> del complejo p53-ADN (de 44 pb) tras el método Grafix..... | 61 |
| 42. Comparación de los mapas de las clases II y III con ADN de 44 pb..... | 63 |
| 43. Comparación de los mapas obtenidos por <i>ML3D</i> de p53 con dos muestras de ADN de distinta longitud (60 y 44pb). | 64 |
| 44. Comparación de mapas de p53 libre obtenidos mediante dos técnicas de análisis de estructura de macromoléculas distintos: SAXS y microscopía electrónica. | 67 |

| | |
|---|----|
| 45. Comparación de los resultados obtenidos por microscopía electrónica y SAXS para p53 humana respecto a los resultados obtenidos por crio-microscopía electrónica para p53 de ratón. | 68 |
| 46. Ejes de simetría de p53. | 69 |
| 47. Comparación de mapas p53-ADN obtenidos mediante SAXS y microscopía electrónica..... | 70 |
| 48. Mapas resultantes de la clasificación por <i>ML3D</i> del complejo p53-ADN (de 60 pb) tras el método Grafix..... | 71 |
| 49. Simetría H14..... | 73 |
| 50. Papel del dominio C-terminal en los mapas 3D de las clases I y III del complejo p53-ADN (de 60pb)..... | 74 |
| 51. Esquema comparativo de las distintas conformaciones posibles que puede adoptar p53 en su interacción con dos muestras de ADN de distinta longitud (60 y 44 pb). | 76 |
| 52. Mapa de microscopía electrónica de p53-ADN de 44 pares de bases tras el método Grafix..... | 77 |
| 53. Mapas resultantes de la clasificación por <i>ML3D</i> del complejo p53-ADN (de 44 pb) tras el método Grafix..... | 77 |
| 54. Unión directa de p53 a una secuencia específica de ADN. | 80 |
| 55. Esquema comparativo de las distintas conformaciones posibles que adopta p53 en su interacción con ADN de 60 y 44 pb. | 81 |

Índice de Tablas

| | |
|---|----|
| 1. Ejemplo de algunos de los genes que regula p53, positiva y negativamente. | 4 |
| 2. Oligos utilizados en la hibridación del ADN..... | 28 |
| 3. Estructuras atómicas encajadas en los mapas 3D obtenidos por microscopía electrónica de p53. | 34 |
| 4. Tabla de probabilidad del tipo de simetría para algunos de los genes que regula p53 y valores de estabilidad (K_D en nM) de la formación del complejo (p53-ADN). | 73 |
| 5. Unión específica e inespecífica entre p53 y ADN..... | 78 |

Capítulo 1

Introducción

I.1 Sistemas de regulación del ciclo celular

El ciclo celular está supervisado por determinados sistemas de control, que reciben un conjunto de señales reguladoras del crecimiento y de la división celular. Estos sistemas deciden si la célula debe o no continuar con su ciclo de vida, de tal forma que los fallos que tengan lugar en uno de estos sistemas de control, por acumulación de mutaciones en unos genes específicos, puede ocasionar la transformación maligna de la célula [Nurse et al., 1998]

La proliferación celular está regulada por varios tipos de genes:

Los **protooncogenes**, que codifican proteínas que regulan positivamente el ciclo celular activando la proliferación. Cuando éstos mutan se transforman en **oncogenes**, que activan la división celular descontrolada, de modo que algunos de ellos fuerzan la maquinaria celular para que sintetice de forma masiva determinados factores de crecimiento dando lugar a la aparición de tumores [Bishop et al., 1991].

Los **genes supresores de tumores**, que codifican proteínas que regulan negativamente el ciclo. En condiciones normales mantienen una correcta tasa de crecimiento y división. La falta de respuesta de estos genes ante un daño celular, por estar mutados o deleccionados, se traduce en una proliferación descontrolada y la consecuente aparición de tumores [Weinberg et al., 1991]. En la actualidad se han descrito más de 30 genes considerados supresores de tumores, de los cuales, la mitad aproximadamente, se han encontrado mutados o deleccionados en cánceres humanos.

I.2 El gen supresor tumoral p53

El gen *p53* se descubrió en 1979 [Linzer et al., 1979; Lane et al., 1979] y fue considerado durante una década como uno oncogén, capaz de inducir un crecimiento celular descontrolado, ya que el gen descubierto daba lugar a una forma mutada e inactiva de la proteína. Posteriormente, en 1989 se confirmó que era un gen supresor tumoral que se presenta mutado en una gran cantidad de células cancerígenas [Baker et al., 1989; Finlay et al., 1989].

Debido al importante papel que juega en la regulación del ciclo celular y en el mantenimiento e integridad del genoma, recibe el nombre de “el guardián del genoma” [Lane et al., 1992].

El gen *p53* se localiza en el brazo corto del cromosoma 17 (17p13) y codifica una proteína de 393 aminoácidos con un peso molecular de de 43.700 Dalton (43.7 KDa). El nombre de la proteína que codifica hace referencia a su masa molecular aparente, debido a que en un gel SDS-PAGE migra como una proteína de 53 KDa. La gran cantidad de residuos de prolina que contiene hacen que la proteína migre más lentamente, haciendo que parezca más pesada de lo que realmente es.

I.3 La proteína p53

Esta proteína se ha convertido en objeto de estudio intenso cuando se observó que, aproximadamente en un 50 % de los cánceres humanos, como consecuencia de mutaciones en el gen *p53*, permanece inactiva. En muchos otros casos está inactiva al encontrarse unida a proteínas virales [Levine et al., 1997].

La interacción de p53 con proteínas capaces de regular su actividad y el amplio número de genes que regula, definen la función de esta molécula.

I.3.1 Función de p53: factor de transcripción.

Los factores de transcripción son proteínas que, sin formar parte de la ARN polimerasa, intervienen en la regulación de la transcripción del ADN en la síntesis del ARN. Estas proteínas actúan reconociendo y uniéndose a secuencias específicas de ADN en la

región promotora de los genes. A estas secuencias específicas se las conoce también como elementos de respuesta (RE). Los factores de transcripción, para poder ejercer su función, necesitan ser activados por señales citoplasmáticas. Una vez activos, son capaces de activar o reprimir la expresión de determinados genes en el núcleo de la célula [Latchman et al., 1997].

Mutaciones en los genes que codifican a ciertos factores de transcripción hacen que estos se mantengan permanentemente activos, en ausencia de señales externas. Estas mutaciones generan la síntesis descontrolada de proteínas implicadas en el desarrollo del ciclo celular ocasionando la aparición de tumores [Libermann et al., 2006].

El supresor tumoral p53 es un **factor de transcripción** homotetramérico que juega un papel importante en el control de la proliferación celular, la integridad del ADN y la supervivencia de células expuestas a agentes que dañan el ADN [Levine et al., 1991; Hollstein et al., 1991]. Además p53 puede intervenir en el reconocimiento directo y reparación del ADN dañado [Tanaka et al., 2000; Albrechtsen et al., 1999].

La proteína p53 en condiciones normales está presente en su forma inactiva a muy baja concentración, mientras que en situaciones de estrés celular o de daño en el ADN, se produce su activación y estabilización. Si el daño es leve detiene la progresión del ciclo celular para permitir su reparación y posteriormente reiniciarlo. Ante un daño grave o irreparable, activa el proceso de apoptosis, reduciendo de esta manera la posibilidad de que células con mutaciones puedan sobrevivir y contribuir al proceso de carcinogénesis [Rotter et al., 1994; Kastan et al., 1991].

Para realizar su función como factor de transcripción, los complejos homotetraméricos de p53, en su forma activa, se unen a secuencias específicas de ADN dentro de la región promotora de algunos genes y regula su transcripción [Hainaut et al., 2005]. Los genes que regula, positiva o negativamente, están implicados en la parada del ciclo celular, reparación de ADN dañado, inhibición de la angiogénesis e inducción de apoptosis [Xiong et al., 1993; El-Deiry et al., 1993; Tanaka et al., 2000; Kluck et al., 1997; Oda et al., 2000; Robles et al., 2001] (*Tabla.1*).

| GEN | Función: Regulación positiva | GEN | Función: Regulación Negativa |
|----------------|----------------------------------|---------------|-----------------------------------|
| Waf1 | Paso de G1/S en el ciclo celular | PCNA | Replicación de ADN |
| Rb | Paso de G1/S en el ciclo celular | C-fos | Regulación de la transcripción |
| GADD45 | Reparación de ADN | C-jun | Regulación de la expresión génica |
| Mdm2 | Vía de autorregulación de p53 | Mdm2 | Transducción |
| Bax | Apoptosis | IL-2 | Regula el crecimiento |
| Fas | Apoptosis | Bcl | Inhibición de apoptosis |
| Thromospondina | Inhibidor de la angiogénesis | Genes virales | Diferentes funciones |
| IGF-BP3 | Paso de G1/M en el ciclo celular | | |

Tabla.1: Ejemplo de algunos de los genes que regula p53, positiva y negativamente.

Genes relacionados con la parada del ciclo celular:

Gen *p21 (Waf1)*: este gen codifica un inhibidor de las quinasas dependientes de ciclinas (Cdk) denominada proteína p21. p21 se une a los complejos ciclina/Cdk inhibiendo su actividad quinasa, permitiendo que la proteína Rb permanezca hipofosforilada y activa, reteniendo al factor de transcripción E2F (un factor de transcripción requerido para pasar el punto de restricción en G1) lo cuál determina una parada del ciclo celular en la fase G1.

Gen *Mdm2*: este gen codifica una proteína (mdm2) que regula en forma negativa la función de p53, mediante un mecanismo de feedback, promoviendo la rápida de gradación de la forma fosforilada de p53.

Gen *GADD45*: (Growth-Arrest and DNA Damage-inducible): el producto de este gen estimula los mecanismos de reparación/escisión del ADN e inhibe la entrada de la célula a fase S.

Genes relacionados con la inducción del proceso apoptótico:

Gen *Bax*; p53 activa la transcripción del gen *Bax* el cual codifica una proteína proapoptótica (bax) perteneciente a la familia de proteínas Bcl2. Bax forma parte de la maquinaria apoptótica y realiza sus funciones formando poros en la membrana mitocondrial, permitiendo la liberación del citocromo c que pasa al citoplasma celular. El citocromo c en conjunto con Apaf1 (factor activador de proteasas apoptóticas 1) activa a la caspasa 9 iniciando así una cascada de eventos que culminan con la muerte de la célula por apoptosis.

Gen *del receptor Fas (APO1)*: este gen activado por p53 convierte a las células que expresan el receptor fas en sensibles a su ligando natural, Fas-L, el cual causa la trimerización del receptor y posterior activación de la caspasa 8, dando lugar al proceso apoptótico.

Genes relacionados con la inhibición de la angiogénesis:

Gen *Tsp1 (thromospondina1)* y *BAIL*: p53 induce la expresión de los genes *Tsp1* y *BAIL* que codifican factores antiangiogénicos. El mecanismo de acción de estos factores sería la inhibición de la proliferación de células adyacentes.

I.3.2 Regulación de la actividad de p53

Teniendo en cuenta que p53 es una proteína con funciones críticas para la célula, no es de extrañar que su actividad esté regulada por múltiples mecanismos, relacionados con el aumento de su estabilidad y el paso de su forma inactiva a activa.

En condiciones normales, en células no expuestas a agresiones, la vida media de p53 es relativamente corta, de unos 6 a 30 minutos dependiendo de tipo de célula. En estas condiciones, los niveles de p53 están regulados por tres mecanismos de retroalimentación negativa, en los que están implicadas las proteínas Mdm2, Ciclina G y Wip-1. (Figura.1).

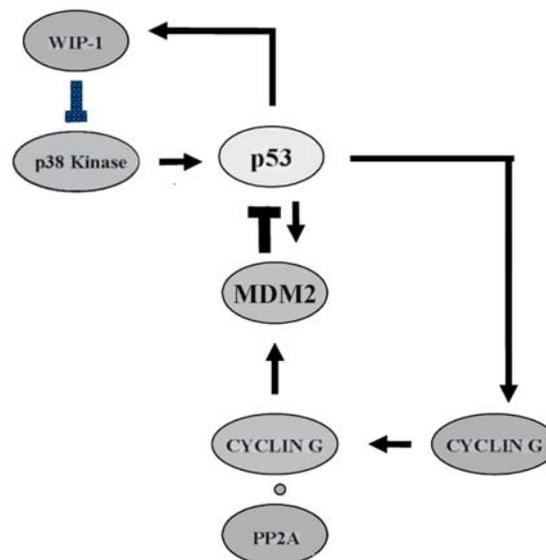


Figura.1. Esquema del mecanismo de regulación negativa de p53, por MDM2, Ciclina-G y Wip-1

Mdm2 es una ligasa E3 de ubiquitina que interacciona y marca a p53 en su extremo amino terminal con varias copias de un pequeño péptido de ubiquitina, pudiendo ser así identificada por la maquinaria de degradación del proteosoma. Los niveles de Mdm2 están a su vez regulados transcripcionalmente por p53, lo que genera un sistema autorregulable [Kussie et al., 1996, Barak et al., 1993].

p53 también regula el gen *Ciclina G*, este gen expresa la proteína Ciclina G que junto a la fosfatasa PP2A elimina las fosforilaciones de Mdm2, aumentando sus niveles y actividad, reduciendo así los de p53 [Okamoto et al., 2002].

Por último, p53 regula la expresión del gen *Wip-1*. Este codifica Wip-1, una fosfatasa que actúa inhibiendo la actividad de p38 kinasa (MAP kinasa). La desfosforilación de p38 kinasa hace que no fosforile a p53, reduciendo así su actividad [Takekawa et al., 2000].

Las señales o sucesos que conducen a la activación de p53 están principalmente asociados a daños en el ADN (por radiación ionizante o ultravioleta), hipoxia, estrés celular y acción de oncogenes. Estas señales son detectadas por una serie de sistemas de reconocimiento (ATM/ATR/DNA-PK/ChK1/ChK2/JNK) que avisan de la presencia de dicho daño modificando postraduccionalmente (fosforilaciones, ubiquitinaciones, metilaciones, acetilaciones y sumolizaciones) algunos de los dominios de p53, principalmente el amino terminal y el carboxilo terminal. Estas modificaciones provocan la activación de la proteína y el aumento de su estabilidad (acumulación de p53) [Hupp et al., 1994.a]. Una vez activa, p53 forma un complejo tetramérico con secuencias específicas de ADN de algunos genes implicados en la parada ciclo celular, reparación del daño en el ADN, inhibición de la angiogénesis e inducción de la apoptosis [Kastan et al., 1991] (*Figura.2*).

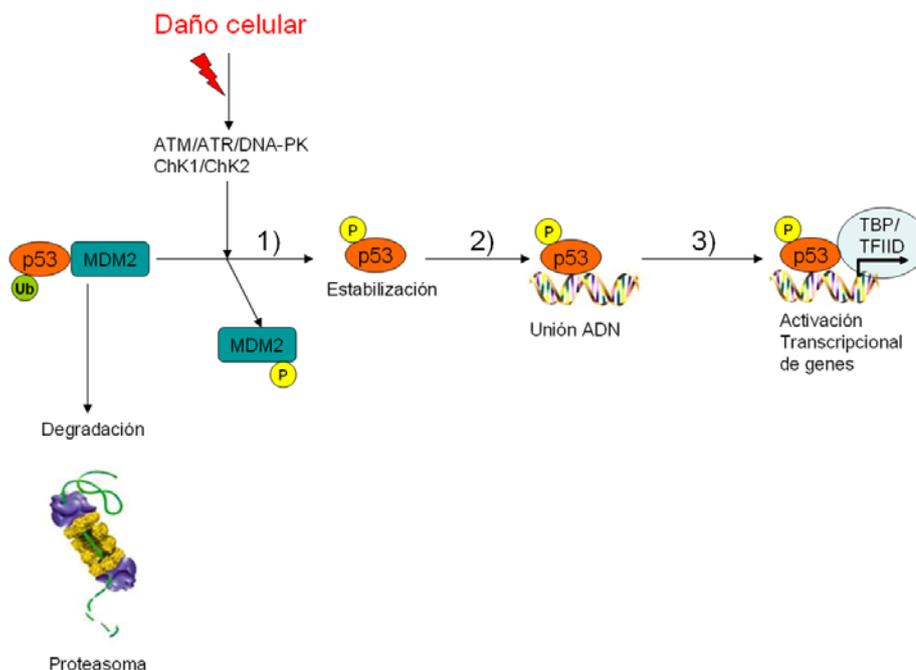


Figura. 2. Modelo de activación de p53. Consiste generalmente en tres pasos. **1)** Estabilización de p53 (Las modificaciones postraduccionales generadas por sistemas de reconociendo de daño celular rompen su interacción con mdm2, aumentando así su estabilidad y acumulación. **2)** Unión a secuencias específicas de genes. **3)** Regulación de la maquinaria transcripcional.

En el caso que p53 no sea funcional, bien por tratarse de una forma mutada o por estar formando complejos con proteínas virales (LTA del virus SV40, E1b de adenovirus, E6 del virus del papiloma humano), ante un daño celular permanece inactiva, lo que conlleva una proliferación celular descontrolada [Linzer et al., 1979; Sarnow et al., 1982; Werness et al., 1990] (*Figura.3*).

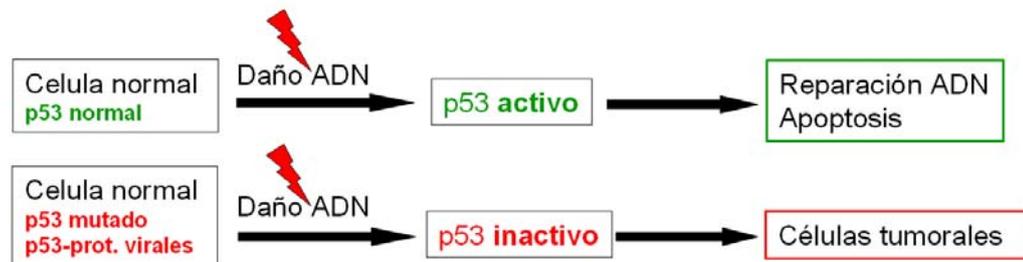


Figura.3: Activación de p53. El paso de p53 a su estado activo es necesario para que pueda generar algún tipo de respuesta (reparación de ADN, apoptosis) ante una señal de estrés (Ej. daño en el ADN)

Para poder entender los mecanismos de activación y regulación de p53, es necesario conocer su estructura, ya que cada uno de sus dominios juega un papel crucial en el paso a su estado activo.

I.3.3 Estructura de p53

p53 es una fosfoproteína homotetramérica, donde cada monómero de 393 aminoácidos, se divide en distintos dominios funcionales (*Figura.4*).



Figura.4. Diagrama esquemático de los dominios de p53.

a. Dominio amino terminal (N-terminal)

Los primeros 94 aminoácidos de p53 contienen dos dominios de transactivación (AD1, AD2), una región rica en prolinas, una señal de exportación nuclear (NES) y múltiples sitios que, ante una señal de estrés, sufren modificaciones postraduccionales

(fosforilaciones). Los dominios de transactivación AD1 (20-40) y AD2 (40-60) son imprescindibles para la activación transcripcional de genes diana [Chang et al., 1995; Candau et al., 1997]. La región rica en prolinas (60-90) contiene cinco copias de la secuencia PXXP, y esta implicada en la regulación de la apoptosis mediada por p53 [Baptiste et al., 2002; Sakamuro et al., 1997]. Este dominio juega un importante papel en la activación de la proteína, ya que la fosforilación del N-terminal hace que se libere de Mdm2, que es su regulador celular natural. Hay pocos trabajos sobre los efectos directos del dominio N-terminal sobre la unión al ADN [Jayaraman et al., 1999], se ha demostrado que anticuerpos contra epítopos del dominio N-terminal (Ej.: Ab1801) aumentan la estabilidad del complejo p53-ADN, protegiéndolo así de la desnaturalización [Friedlander et al., 1996].

Ha sido resuelta la estructura atómica de una región del dominio comprendida entre los aminoácidos 12-26 formando complejo con Mdm2 mediante Resonancia Magnética Nuclear (RMN) [Rosal et al., 2004] (*Figura.5*).

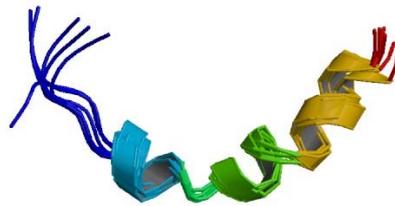


Figura.5 Estructura atómica de la región (12-26) del dominio N-terminal unida a Mdm2 (código PDB, 1Q2F) [Rosal et al., 2004].

b. Dominio de unión de ADN (Core)

Esta ubicado entre los aminoácidos 94-294. Recibe el nombre de Core debido a su resistencia a proteasas [Bargonetti et al., 1993] y contiene la secuencia evolutivamente más conservada entre las proteínas p53 de distintas especies y entre diferentes miembros de su misma familia (p63 y p73) [Yang et al., 2002].

Este dominio es crucial para el reconocimiento y unión de secuencias específicas de ADN localizadas en las regiones promotoras de los genes diana regulados por p53 [Kern et al., 1991; Wei et al., 2006; Cho et al., 1994, Ho et al., 2006; Kitayner et al., 2006]. Las secuencias específicas que reconoce están formadas por dos copias de la secuencia: 5'-R1R2R3C4W5W6G7Y8Y9Y10-3', donde W puede ser A o T, R una base

púrica (A, G) e Y una base pirimidínica (T, C). Entre ambas puede haber un espaciador de 1 a 13 pares de bases [El-Deiry et al., 1992].

La presencia de mutaciones o su interacción con otras proteínas, harán que p53 no pueda unir ADN, permaneciendo inactiva. La gran importancia de este dominio se debe a que el 80% de las mutaciones puntuales de p53 que se detectan en los cánceres humanos están aquí localizadas (puntos más frecuentemente mutados: R175, G245, R248, R249, R273 y R282) [Wong et al., 1999].

En 1994 fue resuelta la estructura atómica del dominio central unido a una secuencia específica de ADN. La estructura está formada por dos láminas β antiparalelas, dos grandes “loops” L2 (163-195) y L3 (236-251) y un motivo loop-lámina-hélice al final del dominio. La conformación de los grandes “loops” L2 y L3 está estabilizada por un ión Zinc (bola gris, *Figura.6.a*), coordinado tetraméricamente por tres cisteínas y una histidina (Cys176, His179, Cys238 y Cys242). El motivo “loop-lámina-hélice” y los “loops” L2 y L3, forman una superficie rica en aminoácidos básicos que unen el ADN por dos zonas, una por el surco mayor, por L1 (Lys120, Arg280) y otra al menor, mediante L3 (Arg248) [Cho et al., 1994].

Posteriormente, en 2006, se resolvió la estructura atómica de cuatro dominios centrales unidos a una secuencia específica de ADN formando un tetrámero (*Figura.6.b*). La estructura está formada por un dímero de dímeros, donde cada uno interacciona con 10 pares de bases de secuencia consenso de ADN [Kytainer et al., 2006; Malecka et al., 2009; Chen et al., 2010]. Estos resultados han permitido estudiar el papel de los distintos aminoácidos en la formación del complejo (*Figura.6*).

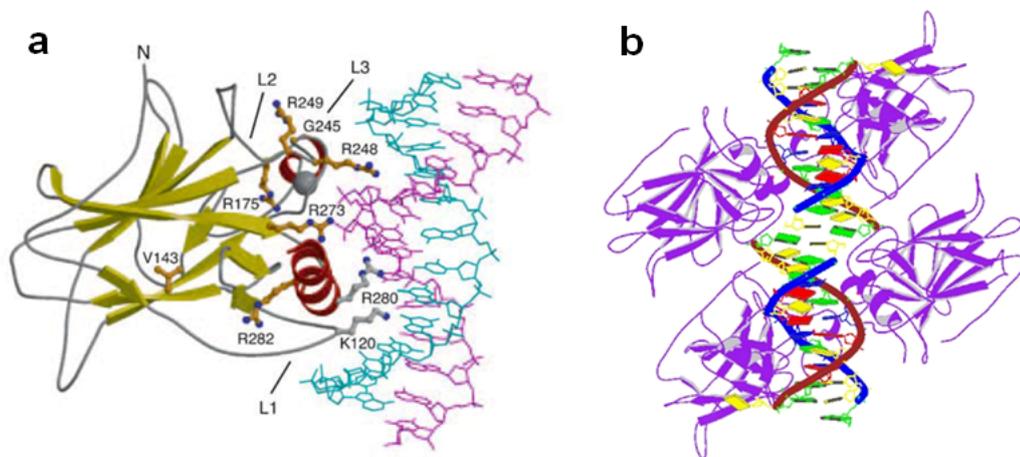


Figura.6. Estructura atómica del dominio central de p53. **(a)** Estructura atómica del dominio central unido a una molécula de ADN (código PDB: 1TSR [Cho et al., 1994]. **(b)** Estructura atómica de cuatro dominios centrales unidos a una molécula de ADN (código PDB: 2ADY [Kytainer et al., 2006].

En la unión de p53 con el ADN, hay tres aspectos importantes a tener en cuenta.

El primero es la importancia del puente de Zn, ya que la mutación en cualquiera de las cisteínas y/o histidina o el cambio del Zn por cualquier otro metal, inhibiría su unión con el ADN.

El segundo es la importancia de los aminoácidos, ya que la mutación puntual de alguno puede ocasionar la inestabilidad estructural y consiguiente desnaturalización (si se trata de aminoácidos de la lamina-β) o el no reconocimiento de la secuencia específica de ADN (si se trata de alguno de los aminoácidos que interactúan directamente con el ADN (Ej.: R273, R248).

El tercer aspecto importante está relacionado con la secuencia consenso o elemento de respuesta (RE) que reconoce p53. El hecho de que esta secuencia sea degenerada, en cuanto a la variedad de simetrías que puede adoptar (*Figura.7.a*), hace que existan varias hipótesis que plantean que, dependiendo de la presencia o no del espaciador y de la simetría interna de la secuencia, tendrá lugar un determinado tipo de respuesta por parte de p53 regulando algunos de los genes implicados en control del ciclo celular, apoptosis o reparación del ADN [Ma et al., 2007].

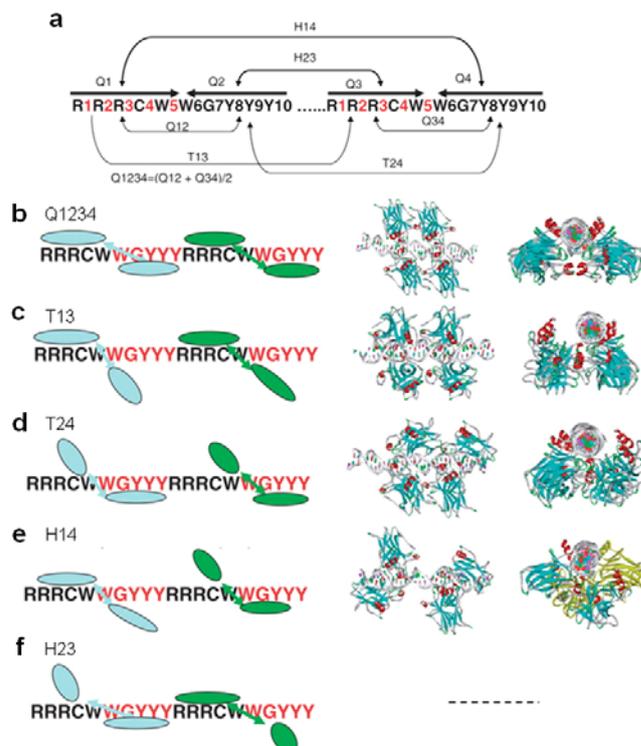


Figura.7. Simetrías de la secuencia específica de ADN. **(a)** Esquema de los posibles tipos de simetrías. **(b)** El modo Q1234 permite la interacción de cada uno de los cuatro monómeros con 5 pares de bases del ADN de la misma forma. **(c)** El modo T13 une los monómeros 1 y 3 al ADN. **(d)** El modo T24 une los monómeros 2 y 4. **(e)** El modo H14 une los monómeros 1 y 4. **(f)** El modo H23 une los monómeros 2 y 3. Estas conformaciones se han obtenido mediante simulaciones de dinámica molecular. Datos adaptados de Buyong et al., 2007.

El tipo de simetría más común en la región promotora de los genes regulados por p53 dentro del genoma humano es la de tipo H14 (simetría inversa entre los cuartos 1 y 4 de de la secuencia específica de ADN) (Figura.7.a-f) [Ma et al., 2007]. Este hecho parece estar relacionado con la alta afinidad por parte de p53 hacia este tipo de simetría. La flexibilidad de las conexiones entre el dominio central y el de tetramerización (aproximadamente 30 aminoácidos) le permite a la molécula adaptarse a cada una de las muy diversas secuencias específicas y así ejercer su función como factor de transcripción (Figura.7.b-f).

Los efectos causados por las mutaciones en cuanto a su estabilidad, han sido estudiados usando espectroscopía de fluorescencia y calorimetría, mientras que los efectos estructurales provocados por ciertas mutaciones, han sido estudiados por RMN y cristalografía de Rayos-X [Bullock et al., 1997, 2000, 2001; Wong et al., 1999] (Figura.8.a-c)

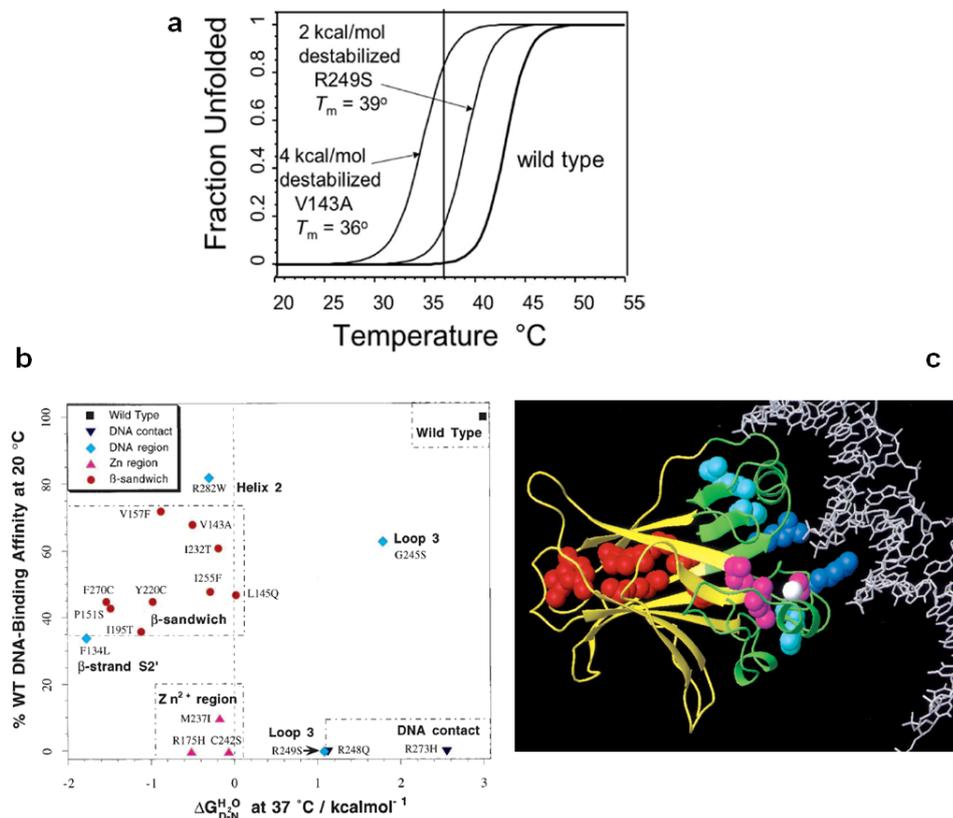


Figura.8. Estudio de los efectos causados por las mutaciones en p53. **(a)** Curvas de desnaturalización con urea del dominio central, tanto de p53 silvestre o “wild type” como de mutantes a pH = 7.2. Los datos se obtienen midiendo la fluorescencia emitida al desnaturalizar a 356nm. Se aprecia que p53 “wild type” es más estable, ya que los mutantes a partir de 39° se desnaturalizan. Datos tomados de Bullock et al. 1997. **(b)** Gráfica que representa el % de afinidad de unión de ADN en comparación con el “wild type” (eje Y) frente a la estabilidad (eje X), siendo $\Delta G=0$ el punto donde el 50% de la proteína estaría desnaturalizada. Se pueden agrupar cuatro tipos distintos de mutaciones: (∇ , mutación de aminoácidos que interaccionan con el ADN), (\blacklozenge , mutación en una región cercana al ADN), (\blacktriangle , mutación en la región del átomo de Zn^{2+}), (\bullet , mutación en aminoácidos de la lámina- β). **(c)** Representación del dominio central, donde la región de unión de ADN y la lámina- β están coloreadas en verde y amarillo respectivamente. Algunas de las mutaciones estudiadas en la gráfica (b) aparecen coloreadas. El átomo de Zn^{2+} está representado como una bola blanca. Datos tomados de Bullock et al. 2000

Tanto p53 silvestre como p53 mutante son termodinámicamente inestables, lo que se traduce en una gran complejidad en cuanto a su estudio por métodos biofísicos y estructurales [Bullock et al., 1997; Hansen et al., 1996].

El problema en cuanto a la inestabilidad ha sido resuelto por el grupo de Alan Fersth, que han diseñado un mutante “superestable”, sin afectar a los aminoácidos que intervienen en la unión de ADN.

Este mutante se generó mediante comparación de homólogos de p53 de más de 20 especies que presentaban alguna mutación puntual en el dominio central [Nikolova et al., 1998]. Las cuatro modificaciones puntuales (M133L, V203A, N239Y y N268D) estabilizaron a la proteína sin afectar en su unión por el ADN (*Figura.9*).

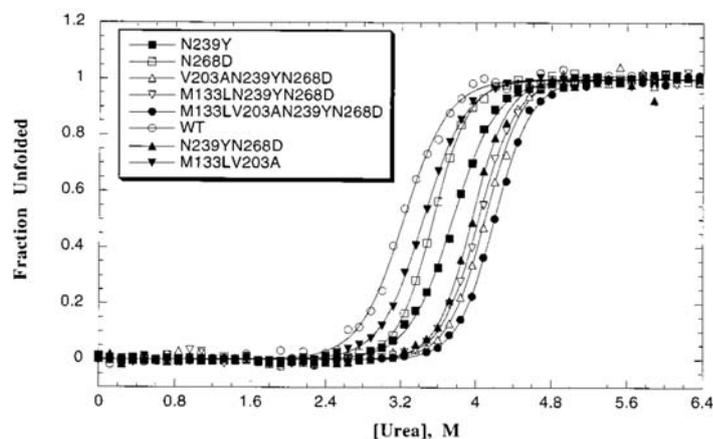


Figura.9 Curvas de desnaturalización con urea del dominio central, tanto de p53 silvestre (WT) como de mutantes a pH = 7.2. Los datos se obtienen midiendo la fluorescencia emitida al desnaturalizar a 356nm. Se aprecia que p53 WT es la menos estable (○) y la más estable es el cuádruple mutante (●) (M133L-V203A-N239Y-N268D). Datos tomados de Nikolova et al. 1998

Para comprobar los cambios estructurales, se han comparado las estructuras cristalográficas del dominio central de p53 silvestre y la del mutante “superestable”, concluyendo que apenas hay cambios significativos [Joerger et al., 2004] (*Figura.10*).

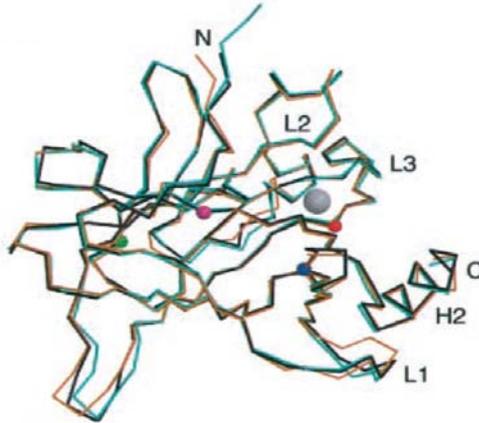


Figura.10 Comparación de las distintas estructuras cristalinas del dominio de unión de ADN. Se han superpuesto el mutante cuádruple (**M133L-V203A-N239Y-N268D**), (PDB 1UOL) en negro, el “wild-type” libre de ADN (PDB 1TSR, cadena A) en azul y el “wild-type” unido a ADN (PDB 1TSR, cadena B) en naranja. El átomo de Zn^{2+} está representado con una gran bola gris y los sitios de mutación del cuádruple mutante con bolas de colores (M133L-azul, V203A-verde, N239Y-rojo y N268D-rosa). Datos tomados de Joerger et al. 2004

c. Dominio de tetramerización (Tet)

Está localizado entre los aminoácidos 320-360 y es responsable de la formación del tetrámero [Chène et al., 2001]. Mediante cristalografía de rayos X y RMN se ha visto que está formado por un dímero de dímeros, con dos laminas- β y dos α -hélices entrecruzadas [Jeffrey et al., 1995] (*Figura.11*).

Una región flexible de unos 30 aminoácidos, se encarga de conectar el dominio de tetramerización con cada uno de los dominios centrales.

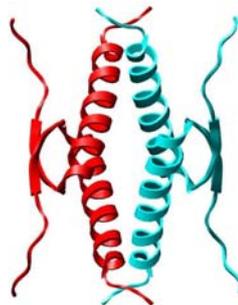


Figura.11. Estructura atómica del dominio de tetramerización (código PDB, 1C26) (Jeffrey et al., 1995). Este dominio está formado por un dímero de dímeros, coloreados en rojo y azul. Cada dos dominios centrales, que forman el dímero, interaccionan con 10 pares de bases de la secuencia específica de ADN.

d. Dominio carboxilo terminal (C-terminal) o dominio Básico.

Esta ubicado entre los aminoácidos 363-393. Contiene múltiples sitios que son modificados postraduccionalmente ante una señal de estrés (ubiquitinación, sumolación, fosforilación, acetilación y glicosilación) [Appella et al., 2001.a,b].

El C-terminal gracias a su alto contenido de residuos básicos (arginina y lisina) es capaz de unir ADN de forma inespecífica, mediante interacciones de tipo electrostático. Esta capacidad le permite regular la actividad del dominio central en su unión a secuencias específicas de ADN. La ausencia de este extremo (p53 Δ 30) o modificaciones que anulen su carga, hacen que solo sea posible la unión de ADN por el dominio central [Ahn et al., 2001; Sauer et al., 2008] (*Figura.12*).

| Protein | total number of amino acids | isoelectric point (pI) | | |
|-----------------|-----------------------------|------------------------|-------|------------|
| | | total | CT | last 30 aa |
| p53 | 393 | 6.50 | 10.00 | 10.00 |
| p53 Δ 30 | 363 | 5.79 | - | 5.19 |

Figura.12. Distribución del punto isoeléctrico de p53. Los últimos 30 aminoácidos que forman el c-terminal, tienen un Punto Isoeléctrico de 10, lo suficientemente básico como para poder unir ADN. Datos adaptados de Sauer et al.,2008

Se ha resuelto mediante RMN una región del dominio (367-386) apenas estructurado [Mujtaba et al., 2004].

I.3.4 Regulación de la unión de p53 al ADN

El C-terminal se une a distintos tipos de estructuras de ADN, tales como: ADN de cadena sencilla [Bakalkin et al., 1995], inserciones o deleciones “mismatches” [Lee et al., 1995], ADN dañado por radiación- γ [Miyashita et al., 1995], “four-way junctions” [Lee et al., 1997] y ADN superenrollado [Mazur et al., 1999].

El papel que ejerce el extremo C-terminal de p53 en la regulación de la unión específica de ADN ha sido objeto de estudio [Ahn et al., 2001].

Se cree que en condiciones normales, el dominio C-terminal hace que p53 no pueda ejercer su papel como factor de transcripción, ya que impide la unión del core con la secuencia específica de ADN, pero no está del todo claro.

Ante una señal de daño celular, los sistemas de reparación modifican postraduccionalmente el C-terminal, provocando la disminución de su punto isoeléctrico (*Figura.12*) y por tanto su capacidad para llevar a cabo interacciones de tipo electrostático con secuencias no específicas de ADN (Ejemplo: la fosforilación de la ser392 favorece la formación del tetrámero [Sakaguchi et al., 1998]). Ante esta situación el dominio central es capaz de unirse a secuencias específicas de ADN.

Han sido muchos los modelos propuestos de cómo se llevaría a cabo la regulación de p53 por parte del dominio C-terminal (*Figura.13*).

En 1995 [Hupp et al., 1994.b-1995; Jayaraman et al., 1995], se propuso que p53 permanecía en una forma latente sin interactuar con el ADN. En este modelo, el dominio C-terminal ejerce su papel regulador interactuando con el dominio central, impidiendo de esta forma que éste pueda interactuar con secuencias específicas de ADN. Por tanto, mientras no se produjese la modificación postraduccional del C-terminal (*Figura.13.a.ii*) o la interacción del C-terminal con otro ligando que le liberase de la unión al core (*Figura.13.a.i*), no podría pasar p53 a interactuar con el ADN. Esta hipótesis se ha desechado al ver por RMN que no existe interacción entre el C-terminal y el dominio central [Ayed et al., 2001] (*Figura.13.a*).

Ese mismo año surge una segunda hipótesis [Bayle et al., 1995]. Este modelo propone que la unión del C-terminal al ADN genómico (de forma no específica) previene la interacción del core con las secuencias específicas de ADN. Este modelo está basado en que ante una molécula de ADN inespecífico largo, se produce su unión con el dominio C-terminal mientras el core es incapaz de llevar a cabo la unión con el ADN de secuencia específica (*Figura13.b*). Solo en el caso de que se produjese la acetilación y/o fosforilación del C-terminal (activación de p53) (*Figura.13.b*) sería posible dicha interacción.

En el año 2002 surge otra hipótesis, opuesta a la anterior, que propone un modelo con dos sitios de unión, en el cual p53 en su estado latente o inactivo tiene múltiples interacciones con el ADN, a través de su core y del C-terminal [Yakovleva et al., 2001-2002]. La unión del C-terminal a las secuencias no específicas de ADN impide la unión del core a las secuencias específicas, pero no a las inespecíficas (*Figura.13.c*). Esta hipótesis está en contra de las antes planteadas, ya que es la única que plantea la posibilidad de unión del core al ADN de forma inespecífica, en un estado latente o inactivo de p53. Este trabajo plantea la necesidad de p53 de poder reconocer ADN de forma inespecífica. Una primera función de este tipo de unión sería la búsqueda de

secuencias específicas (mediante el deslizamiento) a lo largo del ADN genómico. La segunda función podría estar relacionada el reconociendo directo y reparación del ADN dañado por parte de p53 [Albrechtsen et al., 1999]. Una vez p53 reconoce secuencias de ADN dañado, es capaz de unir otros componentes de la maquinaria de reparación del ADN. Este modelo estaría de acuerdo con el hecho de que en general los factores de transcripción se unen al ADN por su secuencia, no por su estructura. Sin embargo las proteínas que reparan ADN generalmente se unen a ADN con una estructura específica sin importar la secuencia [Liu et al., 2006].

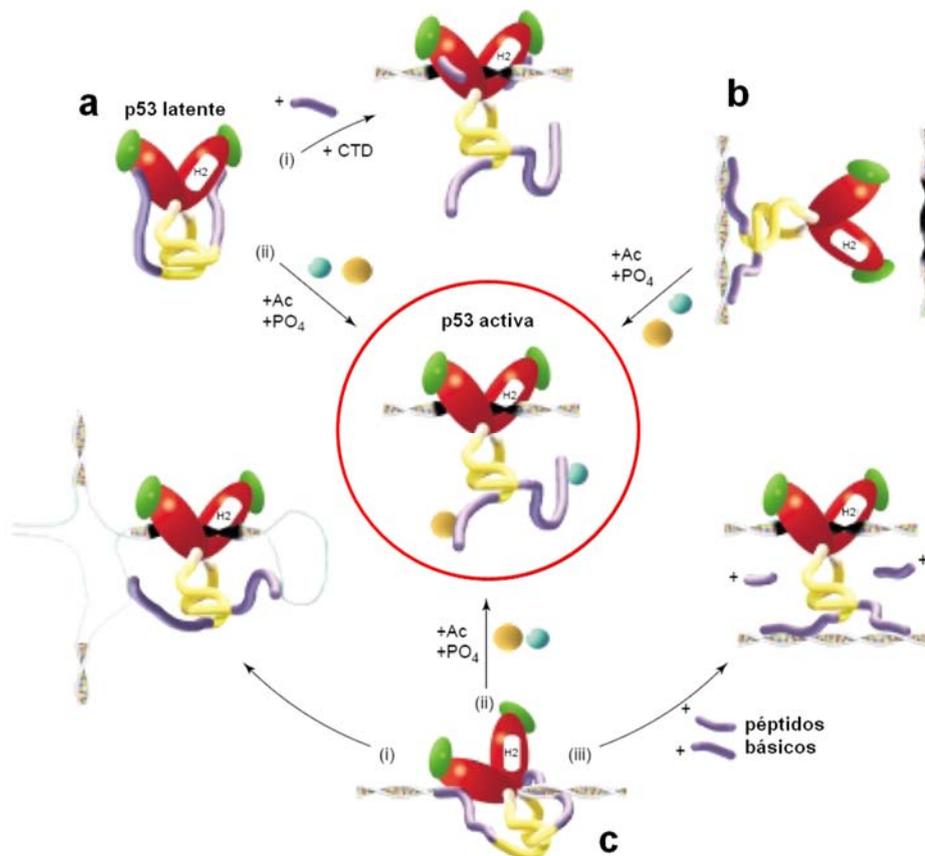


Figura.13 Modelos de activación de p53 en su unión a una secuencia específica de ADN. En el esquema están representados los dímeros de p53 en lugar de tetrámeros. El N-terminal, el C-terminal, el core y los dominios de tetramerización están coloreados en verde, violeta, rojo y amarillo, respectivamente. La secuencia específica del ADN está coloreada en negro. La α -hélice H2 de un monómero en el dímero de p53 se muestra como un cilindro blanco. En los distintos modelos se puede apreciar cuando H2 está o no en contacto con la secuencia específica de ADN. **(a)** En este modelo el C-Terminal interactúa con el core en la forma latente, impidiendo la interacción de p53 con el ADN. El desplazamiento del C-terminal de su unión con el core mediante péptidos básicos (i) o la modificación posttraduccional del C-Terminal por acetilación (Ac) o fosforilación (PO₄) (ii) induce cambios conformacionales en p53, permitiendo la interacción entre el core y la secuencia específica del ADN (p53 activa). **(b)** En este modelo se plantea que mientras el C-terminal esté interactuando con el ADN inespecífico, no es posible la unión del core con el ADN específico. La Acetilación y la fosforilación del C-terminal hace que no pueda interactuar con el ADN y por lo tanto permite que el core se una al ADN (p53 activa). **(c)** Este modelo propone dos modos de unión al ADN. La proteína p53 es capaz de unirse al ADN mediante el C-terminal y el core. En este caso, la unión del core con la secuencia específica de ADN se ve impedido por la interacción del C-terminal con el ADN. En este caso sería posible la unión del core a secuencias específicas de ADN localizadas en estructuras cruciformes de ADN (i). La Acetilación y la fosforilación del C-terminal hace que no pueda interactuar con el ADN y por lo tanto permite que el core se una al ADN (p53 activa) (ii). La presencia de un ADN adyacente interactúa con el dominio C-terminal sin modificar, mientras hace posible la unión del core con la secuencia específica de ADN (iii). Datos tomados de Yakovleva et al., 2002.

Existe otro modelo que propone que p53, gracias al C-terminal, es capaz de deslizarse a lo largo del ADN hasta localizar una secuencia específica a donde unirse por el dominio central. Modificaciones postraduccionales (PTMs), el bloqueo del c-terminal por la unión con anticuerpos (PAb421) o su delección (p53 Δ 30), hacen que la única manera de unión posible entre p53 y el ADN sea entre el dominio central y una secuencia específica [Sauer et al., 2008; Hupp et al., 1992] (*Figura.14*).

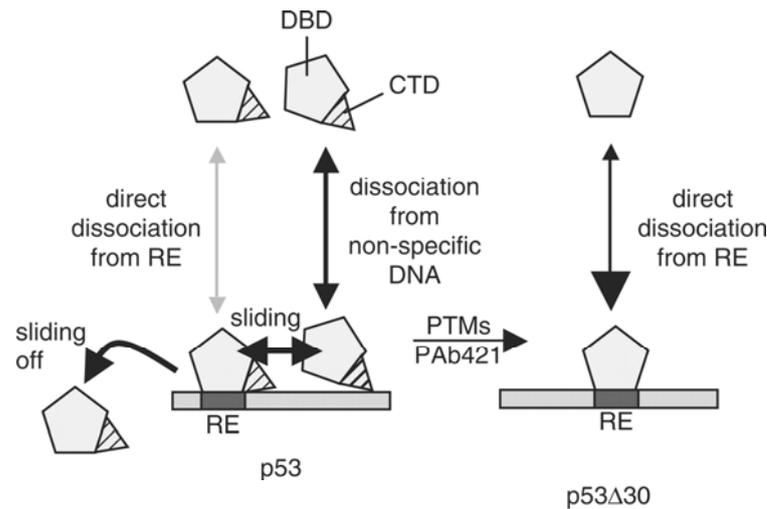


Figura.14. Modelo del mecanismo de unión p53-ADN. Mediante el dominio C-terminal (CTD) es capaz de interactuar con secuencias no específicas y deslizarse hasta localizar la secuencia específica (RE). En el caso de disponer de un mutante sin CTD (p53 Δ 30) o por estar bloqueado dicho dominio por la presencia de un anticuerpo (PAb421) o neutralizada su carga mediante modificaciones postraduccionales (PTM), la única opción de interacción p53-ADN es por el dominio central (DBD). Datos tomados de Sauer et al., 2008.

El papel atribuido al dominio C-terminal en la regulación de la unión a ADN ha ido cambiando. En un principio se pensaba que había una inhibición de la unión del core por parte del dominio C-terminal. Lo que parece que sucede es que se pueden unir al ADN los dos, C-terminal y core. Si hay mucho ADN no específico (ADN largo), la unión vía C-terminal desplaza a la unión específica, ADN-core.

En general, a nivel estructural se podría decir que el supresor tumoral p53 tiene una mezcla de dominios más o menos plegados y estables (el core y el dominio de tetramerización) y otros flexibles y desordenados (el N-terminal, C-terminal y las conexiones entre cada core con el dominio de tetramerización). La flexibilidad en algunos de sus dominios hace que p53 se pueda adaptar a cada una de las distintas secuencias específicas que reconoce de ADN y a una gran multitud de ligandos con los que interacciona para poder ejercer su papel como factor de transcripción. Esta flexibilidad plantea un serio inconveniente para, mediante las herramientas de rayos X, poder resolver su estructura. Mediante cristalografía de rayos X y/o RMN, han sido resueltos algunos de los dominios parciales de p53, pero no existen datos claros de su organización tetramérica. Una buena alternativa para el estudio de proteínas que presentan conformaciones heterogéneas podría ser la microscopía electrónica.

Los únicos datos que se habían descrito para el tetrámero completo provienen de un estudio de crio-microscopía electrónica en el que se propone una organización molecular para p53 de ratón incapaz de acoplar el dominio de tetramerización y los cuatro dominios centrales unidos al ADN [Okovorov et al., 2006] (*Figura.15*).

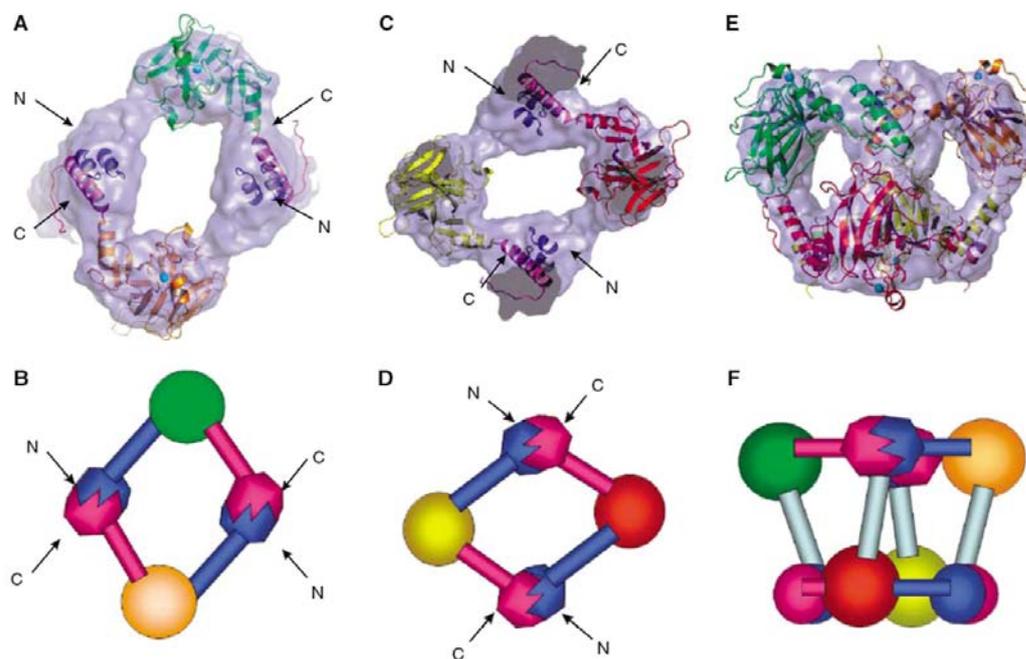


Figura.15. Organización estructural de p53 de ratón. (A-C-E) Mapas de crio-microscopía electrónica con la estructura atómica encajada de los dominios de unión de ADN, N-terminal (N) y cada una de las α -hélices del dominio de tetramerización por separado (C). (B-D-F) Esquema representativo. Cada uno de los monómeros está representado por una bola grande (amarilla, verde, roja y naranja). El resto de bolas que forman los vértices del “cubo” están formadas por la conexión entre el amino y el carboxilo terminal. La delección del N o C haría imposible la formación del tetrámero. Cada una de las hélices que constituyen el dominio de tetramerización (C) se han separado y aparecen localizadas unidas a los N-terminales (N) en algunos vértices. Datos tomados de Okovorov et al., 2006.

I.4 Microscopía electrónica

Para comprender el funcionamiento celular, es necesario estudiar la estructura de las distintas máquinas moleculares que llevan a cabo los procesos biológicos. Dichas máquinas, están constituidas principalmente por proteínas, que son capaces de funcionar por sí solas o por interacción con otras formando complejos macromoleculares. La funcionalidad de dichas máquinas está asociada con la estructura y las distintas conformaciones que pueden adoptar.

La cristalografía de rayos X (RX), la resonancia magnética nuclear (RMN), la microscopía electrónica de transmisión (MET), la microscopía de efecto túnel (STM) y la microscopía de fuerzas atómicas (AFM) son, entre otras, algunas de las técnicas que abordan el estudio estructural de proteínas.

La cristalografía de rayos X y la resonancia magnética nuclear son las herramientas que aportan la mayor información estructural (alcanzan la mayor resolución), pero ambas presentan inconvenientes.

En el caso de la cristalografía de rayos X, la limitación más importante es que se formen buenos cristales, esto implica una alta pureza y concentración en la obtención de la muestra, en ocasiones imposible de conseguir; mientras que la resonancia magnética nuclear se ve limitada por el tamaño de la macromolécula, ya que conforme aumenta también lo hace la complejidad de los espectros obtenidos.

I.4.1 El microscopio electrónico de transmisión

El microscopio electrónico se compone de una fuente de emisión, una columna de aceleración de electrones, un sistema de detección y lentes electromagnéticas. La fuente de iluminación es un filamento o cátodo, generalmente de tungsteno, wolframio o un cristal de hexaboruro de lantano (LaB_6), que emite electrones acelerados por la acción de un campo magnético.

Para evitar la dispersión de electrones al chocar con partículas de gas, se utilizan bombas de vacío que eliminan el aire dentro de la columna.

Mediante el uso de bobinas o lentes electromagnéticas, cuya función es similar a las lentes de vidrio de los microscopios ópticos, los electrones son focalizados.

El porta-muestras se localiza aproximadamente en la mitad de la columna y

permite introducir la muestra, depositada sobre una rejilla de metal, en la trayectoria del haz sin perder el vacío.

Los electrones al incidir en la muestra sufren una perturbación en la onda, lo que permite generar una imagen que se recoge en un sistema de detección (placa fotográfica, pantalla fluorescente o cámara digital CCD). La imagen generada por el microscopio electrónico se considera como una imagen de proyección del espécimen estudiado.

I.4.2 Estudio de muestras biológicas en un microscopio electrónico

La información estructural que alcanza el microscopio electrónico de transmisión oscila en un rango desde escala atómica (1\AA) a tamaños propios de estructuras sub-celulares ($10\mu\text{m}$). Esto hace que sea una potente herramienta para el estudio estructural de los diferentes componentes celulares (desde una proteína hasta un orgánulo celular). (Figura.16)

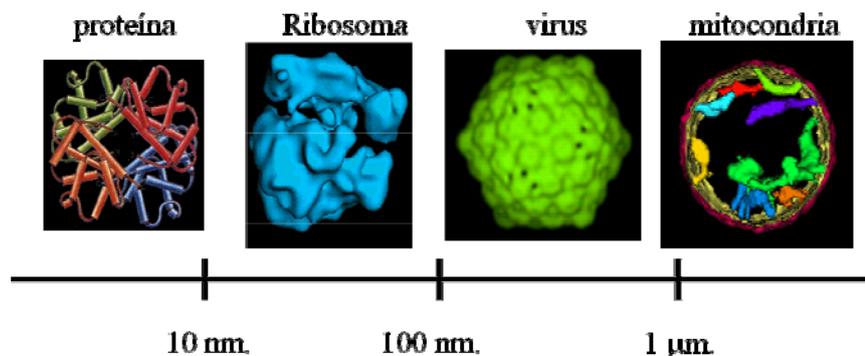


Figura.16. Rango de estructuras macromoleculares estudiadas por microscopía electrónica.

En el caso del estudio estructural de proteínas y complejos proteicos, es necesario un preproceso de purificación e identificación del espécimen de estudio, proceso constituido por una serie de análisis bioquímicos que varían en función del tipo de muestra y de sus propiedades físico-químicas.

La microscopía electrónica, frente a técnicas antes mencionadas, presenta una serie de ventajas: permite trabajar con especímenes de muy diferentes tamaños (proteínas, orgánulos, virus, células, etc.), requiere de cantidades muy pequeñas de muestra (como para poder preparar una rejilla, aproximadamente de $1-2\mu\text{M}$, dependiendo del espécimen) y además no es necesario que esté muy concentrada (en

comparación con lo necesario en cristalografía). Además, permite obtener una estructura global del complejo sobre la que se puede encajar las diferentes estructuras resueltas a resolución atómica, si las hubiera, provenientes de la cristalografía de rayos X o de la resonancia magnética nuclear.

Por el contrario, la microscopía electrónica también presenta una serie de inconvenientes: las condiciones de alto vacío que se generan en la columna hacen que la muestra se vea sometida a un proceso de deshidratación y la alta radiación a la que se somete es capaz de deformar su estructura rompiendo algunos de los enlaces.

Para el estudio de muestras biológicas por microscopía electrónica se utilizan principalmente dos técnicas. La más empleada es la tinción negativa, en la que la muestra se deposita sobre una rejilla y se embebe en una fina capa de sal de un metal pesado (acetato de uranilo, formiato de uranilo, etc.). La sal rodea a las partículas en forma de molde aumentando el contraste de la imagen final. Las imágenes que se obtienen se deben al molde generado por la sal, por lo que solo es posible obtener información de la superficie del espécimen. En función del tamaño del grano de la sal empleada para la tinción, se obtiene mejor o peor resolución en los resultados (en cualquier caso alrededor de los 20-25Å). La otra técnica es la crio-microscopía donde la muestra, depositada sobre una rejilla, se congela rápidamente para que su contenido en agua alcance un estado amorfo (hielo vítreo) evitando la formación de hielo cristalino, preservando así la muestra a temperaturas de nitrógeno líquido (aproximadamente -180 °C). En este caso, debido a que el hielo y la muestra presentan una densidad muy parecida, el contraste de las imágenes es mucho menor que en la tinción negativa.

Para disminuir la deformación ocasionada por la radiación se trabaja, tanto en tinción negativa como en crio-microscopía, en unas condiciones de dosis de electrones mínima (menos de $10e^-/\text{Å}^2$), que consiste en evitar la exposición prolongada a la radiación de la zona donde se toman las imágenes, de tal forma que se irradia y se ajustan las condiciones de enfoque en una zona de la rejilla próxima a la que interesa. [Williams et al., 1970]. Una de las ventajas de la crio-microscopía frente a la tinción negativa es que la muestra está protegida por la congelación y las bajas temperaturas a las que se trabaja.

I.4.3 Procesamiento de imagen

El hecho de reducir al mínimo la dosis de electrones empleada hace que la relación señal-ruido de las imágenes obtenidas sea muy baja. Para aumentar esta relación se recurre a técnicas de procesamiento que combinan la información contenida en multitud de imágenes individuales mediante su alineamiento y promediado, obteniendo medias en 2 y 3 dimensiones. Obviamente, esto solo tiene sentido en el caso de que las imágenes sean de la misma muestra. A pesar de partir de una muestra proteica de alta pureza es posible obtener una gran heterogeneidad debido a que la proteína cae con distintas orientaciones sobre la rejilla o pueda adoptar distintas conformaciones estructurales.

Las herramientas de clasificación de imágenes individuales (Ejemplo: *ML2D*, *CL2D*, *refine2d.py*, *startnclases*) y de proyecciones en 3 dimensiones de muestras heterogéneas (Ejemplo: *ML3D*, *multirefine*), son necesarias para poder entender y resolver la estructura de la proteína [Frank et al., 1996].

I.4.4 Reconstrucción tridimensional

Las imágenes obtenidas en microscopía electrónica de transmisión son proyecciones bidimensionales de los objetos tridimensionales observados. La información de la tercera dimensión se puede recuperar a partir de la información contenida en cada proyección del objeto, siempre y cuando el conjunto de proyecciones se haya obtenido desde un amplio rango de puntos de vista del objeto tridimensional [Radón., 1917]. Esto se puede conseguir bien mediante inclinación del porta-muestras dentro de la columna del microscopio o bien por la capacidad intrínseca que tenga el espécimen de exhibir diferentes vistas de sí mismo al ser adsorbido en diferentes orientaciones.

En el caso que la proteína al depositarse sobre la rejilla adopte multitud de orientaciones bien distribuidas en el espacio angular, es posible reconstruir su estructura tridimensional (3D), haciendo uso de una serie de algoritmos que buscan la orientación angular para cada imagen y la comparan con las distintas proyecciones de un volumen modelo que se usa como referencia. En el método conocido como “Projection Matching” se proyecta el volumen que se usa como referencia y se comparan las proyecciones con las imágenes experimentales. Cuando el coeficiente de correlación

entre la imagen y la proyección es elevado, se asigna a la imagen los ángulos correspondientes a la proyección del modelo. Una vez se conoce la distribución angular de las imágenes, se puede retro-proyectar y obtener un volumen 3D a partir de las imágenes experimentales. Haciendo este proceso de forma iterativa, se puede conseguir un mapa cada vez más preciso y alejado del modelo utilizado como referencia [Marabini et al., 1998].

El inconveniente de esta técnica es que, a pesar de hacer multitud de iteraciones, el mapa final mantiene mucha dependencia del modelo usado de referencia, comúnmente llamado “model bias”. Para evitarlo se han desarrollado el método de reconstrucción angular o líneas comunes y el llamado esquema del cónico aleatorio o “random conical tilt”, que permiten generar un primer mapa a partir de las propias imágenes, evitando así tener que usar una primera referencia inadecuada.

El método de líneas comunes está basado en el “teorema de la sección central”, que determina la existencia de una línea común en el espacio de Fourier entre dos proyecciones de un mismo objeto. Es necesario partir de una clasificación que cubra de forma significativa todas las posibles orientaciones, y a partir de la imagen media de cada una de las clases es posible generar un primer mapa 3D [Crowther et al., 1970].

El método de esquema del cónico aleatorio se utiliza generalmente cuando la muestra adopta una orientación preferente sobre la rejilla y consiste en inclinar el porta-muestras dentro del microscopio. Al girar se puede comparar, para cada imagen, la vista en el plano horizontal y a un determinado ángulo, que dependerá de lo que se gire el goniómetro. De esta forma se obtiene la información para cada imagen de dos vistas y un ángulo que las relaciona. El inconveniente del método es que el giro al que se puede someter a la muestra está limitado (en torno a los 60°), dando lugar a lo que se llama el cono perdido, con la consecuente falta de información o anisotropía en algunas direcciones [Radermacher et al., 1987-1994].

Objetivos

El objetivo general de este trabajo es avanzar en el conocimiento estructural de p53 libre y formando complejo con ADN. Para ello el estudio se dirigió hacia los siguientes puntos:

- Obtener la estructura tridimensional de p53 libre mediante el procesamiento computacional de imágenes.
- Obtener la estructura tridimensional del complejo p53-ADN mediante el procesamiento computacional de imágenes.
- Analizar tridimensionalmente el fenómeno del polimorfismo cuaternario de p53 en su interacción con el ADN, mediante el estudio por microscopía electrónica tridimensional haciendo uso de las herramientas de clasificación de imágenes.

En concreto:

- Calcular las reconstrucciones tridimensionales de las distintas formas de p53 ante dos moléculas de ADN de distinta longitud con la misma secuencia específica de ADN de la región promotora del gen *GADD45*.
- Localizar los distintos dominios resueltos en los mapas reconstruidos para las diferentes arquitecturas.

Capítulo 2

Materiales y Métodos

M.1 Muestra de p53

Para el análisis estructural de la proteína p53 humana y el complejo p53-ADN se ha usado la microscopía electrónica de transmisión y herramientas de procesamiento de imagen.

La expresión y purificación de la muestra p53 se realizó en el laboratorio del Dr. Alan Fersht en el Centre for Protein Engineering (MRC) en Cambridge.

La construcción p53 (1-393) fue clonada en el plásmido pET24A-HLTV con una cola de 6 Histidinas en su N-terminal, un dominio lipoamyl y un sitio de corte para la proteasa TEV. La construcción contenía cuatro mutaciones puntuales en el dominio central (M133L/V203A/N239Y/N268D) [Nikolova et al., 1998]. Fue expresada en células BL21 de *E.coli* a 20°C durante 16 horas. Antes de la inducción se agregó 0.1 mM de Zn²⁺.

Para la purificación se siguieron los protocolos estándares de His-tag, seguido de una digestión por la proteasa del virus del tabaco, una cromatografía de heparina y un paso final de filtración en gel. Para prevenir la agregación se utilizó glicerol al 10% (v/v). La proteína se mantiene en buffer fosfato de sodio 25mM, pH 7.2, NaCl 150mM, DTT 5mM y glicerol al 5% (v/v).

M.2 Muestras de ADN

Para estudiar los complejos formados por p53 y el ADN, se trabajó con distintos tipos de ADN que contenían la secuencia específica (marcada en negrita) de reconocimiento por el dominio central de p53 (*Tabla.2*).

Para la formación de la doble hebra, se resuspendió cada uno del oligos en agua mili-Q, quedando a una concentración final de 1mg/ml. Para la hibridación se les

sometió 2 minutos a 90°C, seguidos de 15 minutos a temperatura ambiente y por último 15 minutos a 4°C en nevera o hielo. Finalmente se comprobó la formación correcta de la doble hebra, de cada uno de ellos mediante un gel de agarosa al 3% (Nusieve GTG Agarosa Low-melting).

| Gen | Oligo | Secuencia |
|---------------|---------------|---|
| GADD45 | p53 forw 60bp | 5'-CC TAC AGA ATC GCT CTA CAG AAC ATG TCT AAG CAT GCT GGG GAC TG ATG CTG GGG ACT GG-3' |
| | p53 rev 60bp | 5'-CC AGT CCC CAG CAT CAG TCC CCA GCA TGC TTA GAC ATG TTC TGT AG AGC GAT TCT GTA GG-3' |
| GADD45 | p53 forw 44bp | 5'-CCT ACA CTA CAG AAC ATG TCT AAG CAT GCT GGG GAC TGG ACT GG-3' |
| | p53 rev 44bp | 5'-CCA GTC CAG TCC CCA GCA TGC TTA GAC ATG TTC TGT AGT GTA GG-3' |

Tabla.2 Oligos utilizados en la hibridación del ADN.

M.3 Formación de complejos p53 - ADN

Para la formación del complejo p53-ADN, se incubó p53 con ADN durante 1 hora a temperatura ambiente en una relación molar 1:4 (ADN:p53).

Para el estudio de la formación del complejo se realizaron ensayos EMSA, que permite ver la movilidad, tanto del ADN libre y/o acomplexado, como de la proteína. Estos ensayos se realizaron con una agarosa de bajo punto de fusión a un porcentaje del 3% en TAE y SYBR-safe 1x10.000 (agente que se intercala en las hebras de ADN y emite luz al ser irradiado con una luz ultravioleta). Las muestras se diluyen en buffer loading dye 6X y se cargan en el gel. La electroforesis se corre a 80V en buffer TAE1X a 4°C. Los resultados de movilidad de ADN se analizan en una lámpara UV o GelDoc.

La movilidad de la proteína se observa tiñendo el gel toda la noche con Coomassie y destiñendo con 15% ácido acético glacial- 10% metanol.

M.4 Preparación de muestras para microscopía electrónica

Con el fin de optimizar el análisis por microscopía electrónica de la estructura cuaternaria de p53 y de su complejo con ADN, se han seguido dos metodologías distintas. Una en la que se analizó la muestra en su totalidad y otra de reciente implantación (GRAFIX, [kastner et al., 2008]) que permitió aislar los tetrámeros de p53 del resto de los estados oligoméricos.

M.4.1 Análisis de la muestra total

Una vez purificadas y caracterizadas cada una de las muestras de p53 y sus complejos con ADN se realizó un estudio de concentraciones, a base de sucesivas diluciones, hasta encontrar la idónea para la microscopía (aproximadamente ente 1-2 μM). Se emplearon rejillas de cobre recubiertas por una película de plástico (collodion) sobre la cual se evaporó una fina capa de carbón. Antes de su utilización fueron sometidas a descarga iónica en vacío moderado durante 40 segundos con objeto de hidrofilar la superficie de carbón favoreciendo así la adsorción de la muestra.

A continuación se incubaron durante 3 minutos en la muestra (p53 o p53-ADN). Pasados los 3 minutos, se realizaron dos lavados consecutivos en agua Mili-Q. Finalmente, las rejillas fueron teñidas con acetato de uranilo al 2% (p/v) durante 1 minuto.

M.4.2 GRAFIX

El método Grafix [Kastner et al., 2008] permite separar en distintas fracciones los distintos estados oligoméricos que pueda adoptar una proteína en función de su densidad y propiedades hidrodinámicas. Al aplicar este método se mejora considerablemente la calidad de las imágenes individuales obtenidas para el estudio de la muestra mediante microscopía electrónica (*Figura.17*).

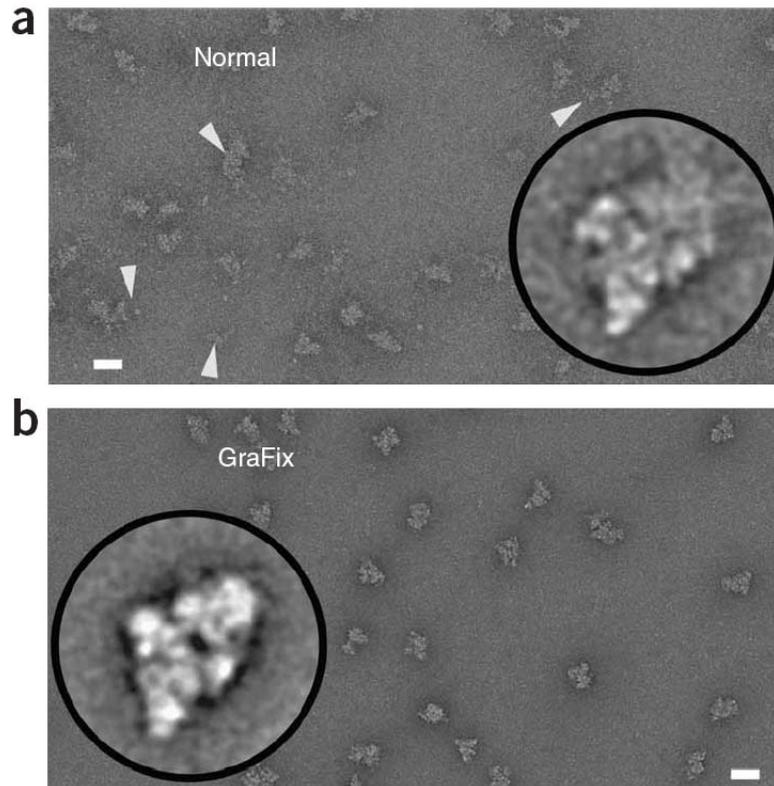


Figura.17. Comparación de imágenes de microscopía electrónica mediante, **(a)** el método normal (análisis de la muestra total) y **(b)** el método Grafix.

El complejo p53-ADN, se centrifugó en un gradiente formado por concentraciones crecientes de glicerol y de un agente químico fijador (glutaraldehído). Dicho agente fija aminas primarias disponibles (en proteínas, fundamentalmente lisinas) y estabiliza las moléculas individuales estableciendo enlaces covalentes con los grupos amino libres presentes en las proteínas, formando entrecruzamientos. El glutaraldehído al polimerizar puede establecer puentes de distinta longitud.

Para este método es necesario evitar el uso de tampón Tris (tris(hidroximetil)aminometano) de fórmula $(\text{HOCH}_2)_3\text{CNH}_2$, ya que el glutaraldehído fija grupos aminas presentes en las lisinas de la proteína pero también en el Tris. Es decir, el Tris compite con la proteína por el glutaraldehído y sería necesario añadir una gran cantidad de agente fijador, por lo que se trabaja en tampón fosfato (tampón en el que se mantiene la muestra p53).

En este sistema la concentración de glutaraldehído es relativamente baja en las fracciones que analizamos para la microscopía (aproximadamente el 0.05 %), lo que reduce la formación de artefactos y de grandes agregados.

En la centrifugación, las moléculas avanzan en el gradiente dependiendo de su tamaño, de tal forma que es posible separar cada uno de los diferentes estados oligoméricos que pueda adoptar el complejo p53-ADN.

Para facilitar la puesta a punto del método se usan dos instrumentos: un formador de gradientes (Gradient Station, Biocomp) y un fraccionador.

En el formador de gradientes se genera un gradiente lineal de glicerol y glutaraldehido a lo largo de un tubo de ensayo. Se partió de dos soluciones de distinto porcentaje de glicerol (5% y 25%) y se añadió el agente fijador (glutaraldehido al 0.15%) en la de mayor porcentaje de glicerol. Una vez obtenido el gradiente, se colocó la muestra en la parte superior del tubo y se centrifugó a 25.000 rpm 18 horas a 4°C.

El fraccionador nos permitió separar en distintas fracciones los distintos estados oligoméricos que pueda adoptar la muestra.

El glutaraldehido presenta la ventaja de mantener fijada la estructura en cada fracción, mediante enlaces covalentes con la proteína, de tal forma que nos permitió separar, del resto de los estados oligoméricos, los tetrámeros de p53. Además el hecho de estar fijado impide que se pueda volver a recuperar una situación de equilibrio.

Mediante geles de acrilamida SDS-PAGE teñidos con Coomasie se analizó el estado oligomérico de p53 en cada una de las fracciones. Con geles de agarosa “low melting” se analizó el retardo en la movilidad del ADN, que se traduce en la formación o no del complejo p53-ADN.

Al comparar ambos tipos de geles, podremos conocer que estado oligomérico presenta en cada fracción p53 al interaccionar con el ADN.

Para los ensayos de microscopía electrónica, se utilizaron las fracciones que presentaron un estado de oligomerización tetramérico y que formaron complejos con ADN. Se realizó un estudio de concentraciones, a base de sucesivas diluciones, hasta encontrar la idónea para la microscopía.

En este caso se emplearon rejillas de Quantifoil, sobre las que se evaporó una fina capa de carbón. Al igual que en el caso anterior, se sometieron a descarga iónica y se incubaron durante 3 minutos en presencia de la solución de p53. Tras dos lavados consecutivos en agua Mili-Q, se metió la rejilla en una gran gota de formiato de uranilo 1% (p/v) donde previamente se había flotado una fina capa de carbón. Al sacar la rejilla se capturó el carbón de tal forma que la muestra queda comprendida entre las dos capas de carbón. Este método se llama “sándwich” de doble carbón. El hecho de utilizar formiato de uranilo en lugar de acetato de uranilo se debe a que el acetato de uranilo y

su pH hacen que el fosfato precipite. En cambio, en el caso del formiato de uranilo, al ser el pH más alto (aproximadamente 5) hace que el fosfato no precipite.

M.5 Obtención de imágenes en el microscopio electrónico

Las muestras fueron visualizadas en un microscopio electrónico JEOL 1200 EX II de 80KV (Centro Nacional de Biotecnología-CSIC) y en un JEOL 200KV JEM-2200FS (CICBIOGUNE).

Con el JEOL 80KV se capturaron micrografías a una magnificación de 60.000X y posteriormente fueron digitalizadas en un escáner (Zeiss/Imaging) resultando un tamaño de píxel final de 3.5 Å. La calidad óptica de las micrografías fue evaluada mediante el examen del espectro de frecuencias de las imágenes, eliminándose aquellas que presentaron defectos como deriva o astigmatismo.

Con el JEOL 200KV se tomaron las imágenes con cámara CCD (Charged-Coupled-Device), a una magnificación nominal del microscopio de 50.000X resultando un tamaño de píxel final de 2.1 Å.

De igual forma que en el caso de las micrografías se analizó el espectro de frecuencias de las imágenes, eliminándose aquellas que presentaron defectos como deriva o astigmatismo.

M.6 Procesamiento de imagen

Para el procesamiento de imagen se utilizaron los paquetes informáticos EMAN [Ludtke et al., 1999] y XMIPP [Marabini et al., 1996; Sorzano et al., 2004; Scheres et al., 2008].

Las partículas individuales fueron seleccionadas manualmente a partir de las micrografías o de las imágenes capturadas por la cámara CCD, dependiendo de cada uno de los casos estudiados.

M.6.1 Clasificación 2D

Una vez seleccionadas las partículas de cada una de las muestras, se llevó a cabo una clasificación 2D de las imágenes, con la intención de comprobar el rango de vistas de las que se disponía.

Para la clasificación se utilizaron los programas *startnclasses* y *refine2d.py* de EMAN [Ludtke et al., 1999] (en sus versiones 1.7 y 1.8 respectivamente) y el programa *ML2D* de XMIPP [Scheres et al., 2005.a,b-2007].

Estas herramientas permiten analizar la heterogeneidad presente en la muestra, mediante sucesivas iteraciones en las que se alinean las imágenes a nivel translacional y rotacional. Gracias a estas herramientas es posible separar un conjunto inicial de imágenes en clases homogéneas, que representen las distintas orientaciones que pueda adoptar la muestra al depositarse sobre la rejilla [Frank et al., 1996].

M.6.2 Reconstrucción 3D

En cada uno de los casos estudiados se usó el método de líneas comunes. Se partió de las clases generadas libres de referencia (*ML2D*, *startnclasses* y *refine2d.py*) para obtener un primer volumen, aplicando distintos tipos de simetría (C1, C2 y C4) con la herramienta *startAny* de EMAN [Ludtke et al., 1999]

Durante el proceso iterativo de refinamiento, la búsqueda de orientaciones en la proyección angular del volumen inicial fue inicialmente en incrementos de 15° y terminó siendo de entre 8-10°. Las imágenes individuales con la menor similitud a la media dentro de cada clase, fueron eliminadas.

Se estimó, al término del refinamiento angular, la resolución alcanzada mediante el método de correlación en los anillos de Fourier [Saxton et al., 1982], para lo que se dividió el conjunto inicial de partículas en dos grupos. Se calculó la estructura de cada uno de ellos mediante la herramienta *eotest* de EMAN. Las dos estructuras obtenidas se compararon en el espacio de Fourier calculando la correlación entre ambas en función de la frecuencia espacial o “Fourier Shell Correlation” FSC (Van Heel et al., 1987).

Un criterio para definir la resolución alcanzada es aquel que se corresponde a un valor de FSC = 0,5 [Bottcher et al., 1997].

Los umbrales de la densidad, que definieron los límites para los mapas 3D, fueron asignados para representar aproximadamente el 100% del volumen calculado para un tetrámero y/o que contuvieran estructuras previamente resueltas de p53.

M.6.3 Clasificación 3D

El método de máxima verosimilitud (*ML3D*) incluido en el paquete XMIPP [Scheres et al., 2007.b], permite separar en clases las proyecciones de distintas estructuras (conformaciones) 3D que pueda adoptar la muestra de estudio, a partir de un mapa muy filtrado usado como referencia. Para este método no es necesario tener ningún conocimiento previo acerca de la variabilidad estructural de los datos.

Esta técnica nos permitió explorar la variabilidad estructural que presenta p53 en cada uno de los casos estudiados, a partir de un set de imágenes de una misma muestra sin saber de antemano que tipo de variabilidad conformacional presentaría la molécula.

M.6.4 Ajuste de estructuras atómicas

Las estructuras atómicas de los distintos dominios de p53 fueron ajustados manualmente dentro de los volúmenes como cuerpos rígidos utilizando el programa CHIMERA de UCSF [Pettersen et al., 2004].

Utilizamos las estructuras previamente resueltas, bien por cristalografía de rayos X o por Resonancia magnética nuclear del complejo tetramérico formado por cuatro dominios centrales (cores) unidos a una molécula de ADN, del dominio de tetramerización y del dominio C-terminal (*Tabla.3*).

| Dominio estructural | Código PDB |
|--------------------------------|------------------------------|
| Central de unión al ADN o core | 2ATA [Kitayner et al., 2006] |
| Tetramerización | 3SAK [Jeffrey PD, 1995] |
| Carboxilo terminal | 1JSP [Mujtaba et al., 2004] |

Tabla.3. Estructuras atómicas encajadas en los mapas tridimensionales obtenidos por microscopía electrónica de p53.

Capítulo 3

Resultados

R.1 p53 libre

a. Caracterización bioquímica

La muestra de p53 se encuentra a una concentración aproximada de $165\mu\text{M}$ en tampón fosfato de sodio 25mM , pH 7.2, NaCl 150mM , DTT 5mM y glicerol al 5%. Se analizó la muestra por espectrometría de masas dando como resultado la presencia mayoritaria de p53 humana. Se pasó la muestra por una columna de intercambio iónico (Hitrap QFF) acoplada a un sistema ÄKTA-PRIME (Amersham Biosciences). Esto permitió añadir un paso más al proceso de purificación y cambiar a un tampón tris. Para ello, se diluyó la muestra en tampón A (Tris-HCl 20mM , pH 7.5, NaCl 100mM , 2mM DTT) y se pasó por la columna previamente equilibrada en dicho tampón. Las proteínas unidas a la resina de la columna se eluyeron utilizando un gradiente lineal de NaCl (100mM – 1M) (Figura.18.línea verde) en tampón B (Tris-HCl 20mM , pH 7.5, NaCl 1M , 2mM DTT) midiendo en todo momento la absorbancia a 280nm (Figura.18.línea azul).

Las fracciones recogidas se analizaron en gel SDS-PAGE (Figura.18)

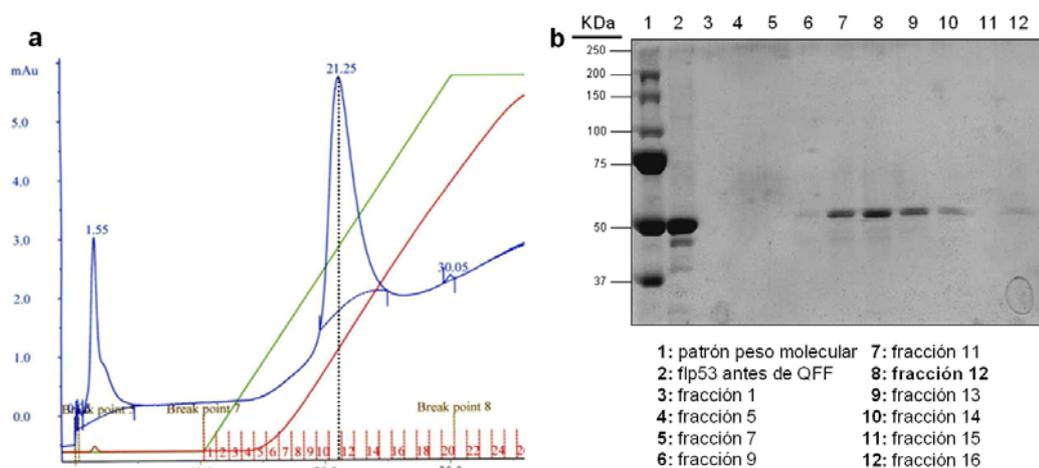


Figura.18. (a) Perfil de elución de p53 (línea azul) en una columna de intercambio iónico QFF 1ml con un gradiente lineal de NaCl (100mM - 1M) (línea verde). (b) Gel SDS-PAGE de las fracciones recogidas. Las fracciones 11-13 (calles 7-9) presentan una pureza mayor a la que presentaba la muestra antes de pasar por la columna (calle 2).

b. Reconstrucción tridimensional

Para la microscopía electrónica se utilizaron las fracciones de mayor concentración y pureza de p53 (fracciones 11, 12 y 13), que se correspondieron con el pico de elución de la columna. Al ser observadas en el microscopio electrónico, contrastadas negativamente con acetato de uranilo al 2% (análisis total de la muestra) se observaron partículas que tenían forma globular. La heterogeneidad en cuanto a forma y tamaño de las imágenes obtenidas nos sugería que p53 se depositaba sobre la superficie de la rejilla de muy diversas formas, dando lugar a imágenes de proyección muy variables (*Figura.19.a,b*). Se seleccionaron manualmente las partículas que tenían un tamaño similar hasta obtener un total de 4.798 partículas. Las partículas seleccionadas no representaron más que el 20% de las partículas totales fijadas por adsorción a la capa del carbón. A partir de estas imágenes y utilizando los algoritmos de clasificación de imágenes 2D libres de referencia *startnclasses* de EMAN [Ludke et al., 1999] y *ML2D* de XMIPP [Scheres et al., 2005], se obtuvieron cada una de las clases que se correspondían con las imágenes de proyección de p53 (*Figura.19.c,d*)

Tanto las imágenes individuales como las clases obtenidas, en su mayoría, presentaron una forma globular. En algunas de ellas (encuadradas en rojo), se pueden apreciar un tipo de orientación característica de la molécula donde se distinguen dos masas de distinto tamaño unidas por conexiones débiles (*Figura.19*).

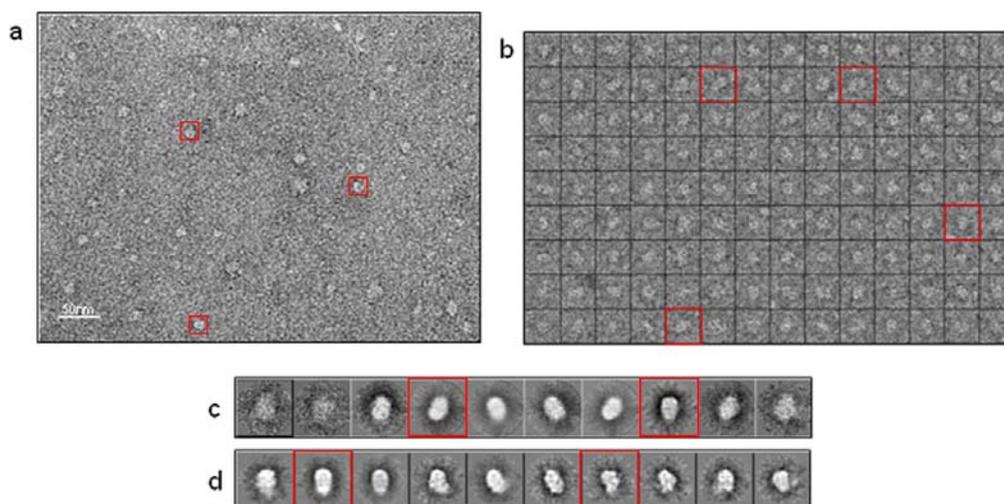


Figura.19. Microscopía electrónica de p53 libre. **(a)** Micrografía de la muestra p53 contrastada negativamente con acetato de uranilo al 2%. **(b)** Galería de imágenes de microscopía electrónica de p53. **(c)** Clasificación de imágenes mediante *ML2D* (XMIPP). **(d)** Clasificación de imágenes mediante *startnclasses* (EMAN). Las clases obtenidas mediante XMIPP **(c)** e EMAN **(d)** presentaron características en común. Tanto en las partículas individuales como en algunas clases libres de referencia, se aprecia un tipo de vista característico con dos masas de distinto tamaño unidas por conexiones débiles (encuadradas en rojo).

Para llevar a cabo una reconstrucción tridimensional es necesario partir de un primer volumen de referencia. Este volumen es importante que proceda de las propias imágenes y no de un modelo, ya que el “model bias” que acompaña a los programas de refinamiento hace que los resultados sean muy parecidos al usado como referencia.

Para generar un modelo 3D inicial de este grupo de partículas se utilizó el método de líneas comunes *startAny* de EMAN, partiendo de las imágenes medias calculadas para cada una de las clases y usando distintos tipos de simetría (C1, C2 y C4). La forma y el tamaño totales de los mapas 3D convergieron al modelo inicial C2, que se corresponde con la formación tetramérica a partir de un dímero de dímeros. Esta simetría fue seleccionada y utilizada a lo largo del refinamiento de la reconstrucción 3D.

El modelo inicial resultante se refinó mediante sucesivas iteraciones comparando cada una de las proyecciones con los datos originales hasta obtener un resultado convergente (*Figura.20*). Las proyecciones generadas a partir del mapa 3D final son compatibles con imágenes medias obtenidas de los algoritmos de clasificación libres de referencia *startnClasses* y *ML2D* (*comparación Figura.19.c,d y Figura.20*).

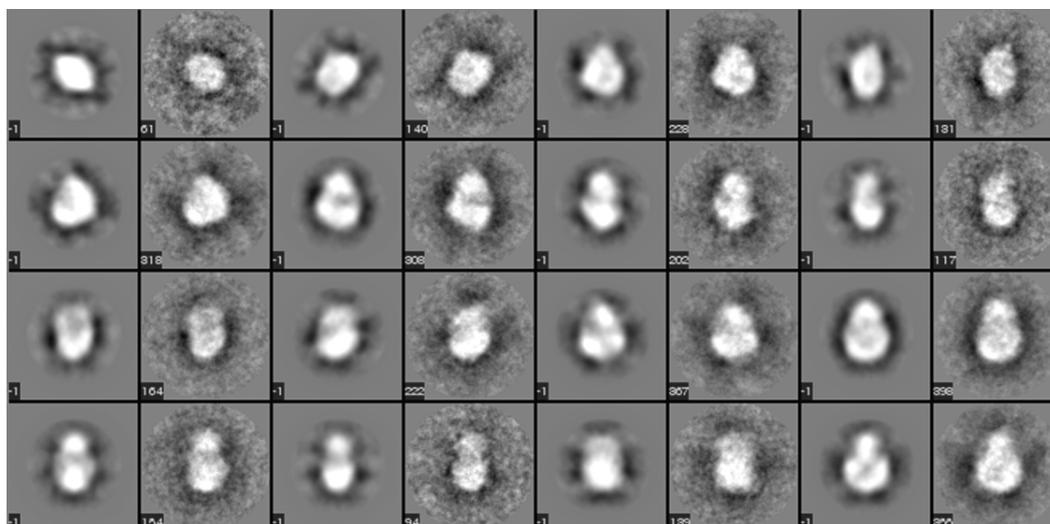


Figura.20. Galería de proyecciones y medias equivalentes de p53 libre. Obtenidas en el proceso de refinamiento angular iterativo. Están representados dos a dos, primero (con “-1”) la proyección del mapa y al lado la clase media y el número de partículas que corresponden a cada clase.

El mapa resultante (*Figura.21.a-c*) está formado por dos masas (una arriba y otra abajo) conectadas por dos uniones débiles. A pesar de las limitaciones de la técnica de tinción negativa y de la moderada resolución obtenida (estimada entre 25-30 Å), los

límites del mapa permiten encajar, como cuerpos rígidos, las coordenadas atómicas de los cuatro dominios centrales o cores, resuelta por cristalografía de rayos X (Código PDB: 2ATA, sin el ADN [Kitayner et al., 2006]) y la estructura del dominio de tetramerización, resuelta por RMN (Código PDB: 3SAK [Jeffrey et al., 1995]) (Figura.21.d-f). La estructura tetramérica que presentan tanto los cuatro dominios centrales, como el dominio de tetramerización, facilita poderlos encajar dentro del mapa de microscopía electrónica sin la necesidad de cambios importantes.

Es importante destacar que la simetría global del mapa es C2, con una región de simetría D2 en el dominio de tetramerización y simetría C2 en la región central. Esta organización fue propuesta en algunos de los datos obtenidos por cristalografía de rayos X y RMN [Cho et al., 1994; Jeffrey et al., 1995].

Las estructuras encajadas en el mapa 3D (dominios estables) representan solo el 60% de la estructura global del tetrámero p53, el resto de la proteína, formada por segmentos flexibles desaparecen al hacer el promediado en 3D. En la parte inferior del mapa (marcado con “*”, Figura.21.e) hay un espacio libre donde podrían encajarse los 4 dominios N-terminal de cada cadena, con sus correspondientes dominios de transactivación y la región rica en prolina.

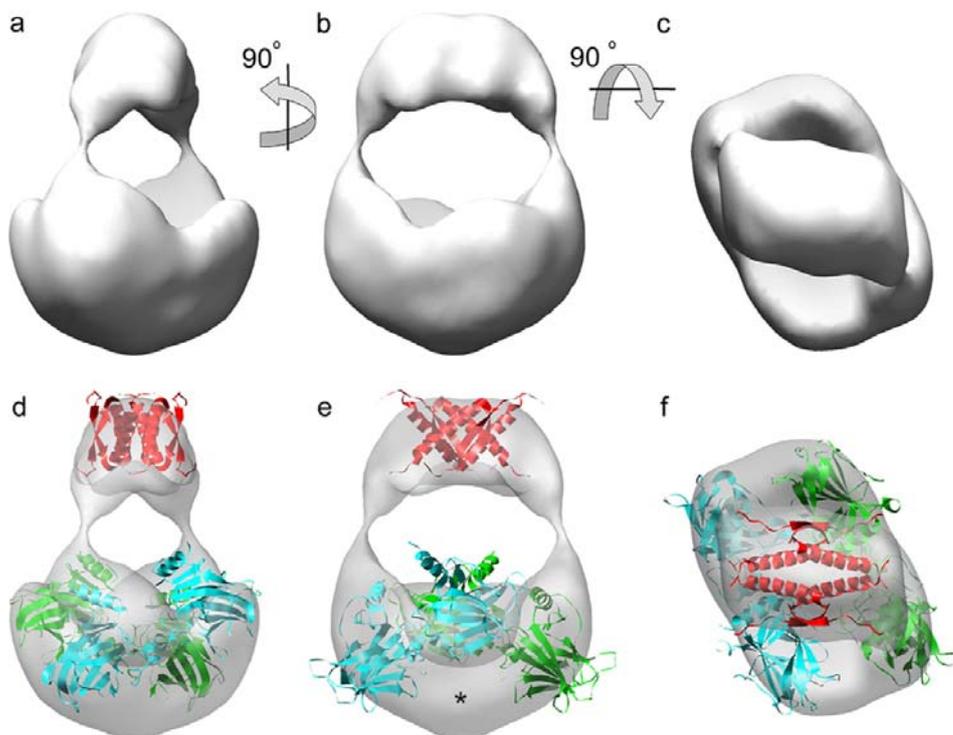


Figura.21. Estructura tridimensional del tetrámero p53 libre. (a-c) Diferentes vistas en diferentes orientaciones. (d-f) Una representación semitransparente del mapa permite ajustar las coordenadas atómicas de los dominios estructurales resueltos por cristalografía de rayos X (dominio central o de unión a ADN (verde y azul) (código PDB: 2ATA), y RMN (dominio de tetramerización (rojo) (código PDB: 1SAK). Con un “*” aparece marcada la zona donde estarían localizados los cuatro dominios N-terminales de cada monómero. Datos tomados de Melero et al., 2007.

Una vez concluido el proceso de refinamiento angular se llevó a cabo la estimación de la resolución de la estructura 3D resuelta. El volumen obtenido tiene una resolución aproximada de 27Å para un valor de FSC de 0,5 (*Figura.22*) y sus dimensiones aproximadas son 110Å en el eje longitudinal, 125Å de altura y un grosor de unos 105Å

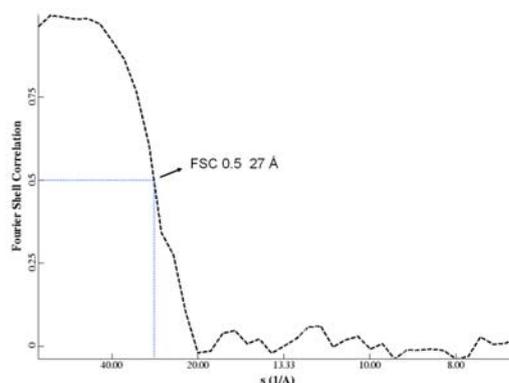


Figura.22. Resolución de mapa 3D de p53 libre. El volumen obtenido tiene una resolución aproximada de 27Å para un valor de FSC de 0,5.

R.2 El complejo p53-ADN

Para el estudio de la estructura tridimensional del complejo p53-ADN por microscopía electrónica se utilizaron los dos métodos mencionados en el capítulo anterior: el análisis de la muestra total y Grafix. Este último permite analizar una población de tamaño uniforme previamente fijada.

Se estudió el papel que ejerce el tamaño del ADN en el reconocimiento y formación del complejo p53-ADN mediante ambos métodos, incubando p53 con dos muestras de ADN de distinta longitud (60 y 44 pares de bases) pero con la misma secuencia específica de reconocimiento del gen *GADD45*.

R.2.1 Análisis de la muestra total

a. Caracterización bioquímica

Para la formación del complejo p53-ADN, se incubó la muestra de p53 (4.3µM) con ADN a distinta concentración (0.4, 0.8, 1.2 y 1.6 µM). Se usó ADN doble hebra de 60

pares de bases que contenía la secuencia del promotor de transcripción para el gen *GADD45* [Ang et al., 2006; Weinberg et al., 2004]. La formación del complejo se analizó en un gel de agarosa EMSA (*Figura.23*), cuya tinción con bromuro de etidio, permitió ver el retardo en el ADN en presencia de p53. (*Figura.23.a*).

Tras teñir y desteñir el gel con Coomassie se observó que p53 cambia su movilidad en presencia de ADN (*calles 2 (p53 libre) y calles 3-6 (p53-ADN)*, *Figura.23.b*). El hecho de que ambas señales (la del ADN y la de p53) coincidan indica la formación del complejo p53-ADN.

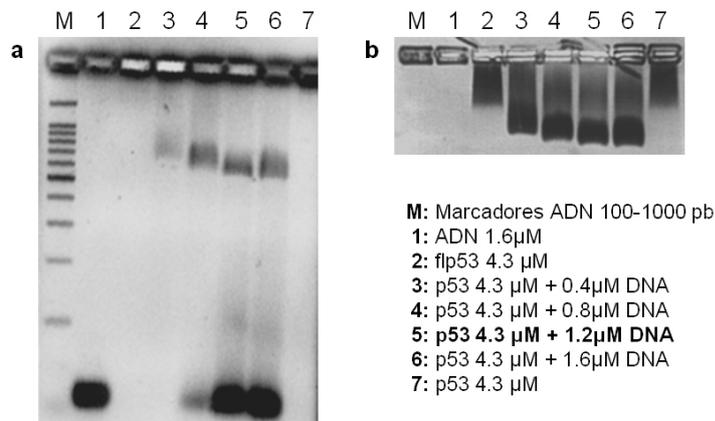


Figura.23. Caracterización bioquímica del complejo p53-ADN. Gel EMSA de interacción p53 y ADN de 60 pares de bases. **(a)** Gel teñido con bromuro de etidio. Se aprecia un cambio en la movilidad del ADN libre (calle 1) respecto a cuando está formando complejo con p53 (calles 3-6). En las calles 4-6 hay un exceso de ADN (corre en el gel como ADN libre de 60 pares de bases). **(b)** Gel teñido/desteñido con Coomassie. El cambio de movilidad de p53 (calle 2) en presencia de ADN (calles 3-6) indica la formación del complejo p53-ADN.

b. Reconstrucción tridimensional

Para la microscopía electrónica se utilizaron las condiciones donde la relación molar entre el ADN y cada monómero de p53 era aproximadamente 1:4 (ADN:p53) (*Figura.23.a.calle 5*). En estas condiciones se estaría formando el complejo p53-ADN, quedando ADN libre, en exceso.

Las imágenes de microscopía electrónica teñidas negativamente con acetato de uranilo al 2%, presentaron una distribución más uniforme, en tamaño y forma, que las partículas individuales de p53 libre. Se seleccionaron manualmente un total de 5.517 partículas (*Figura.24.a-b*).

La clasificación de imágenes de partículas individuales del complejo p53-ADN, al igual que en caso anterior, se realizó utilizando la herramienta de clasificación *startnclasses* de EMAN y *ML2D* de XMIPP (*Figura.24.c-d*).

En este caso, tanto en las imágenes individuales como en las clases libres de referencia obtenidas (encuadradas en rojo) se pueden diferenciar dos masas de distinto tamaño unidas por conexiones algo más fuertes que en el caso de p53 libre. Cabe destacar, en algunas de las vistas, la presencia de una especie de canal o falta de densidad en el centro de la molécula, posiblemente debido al paso del ADN a su través.

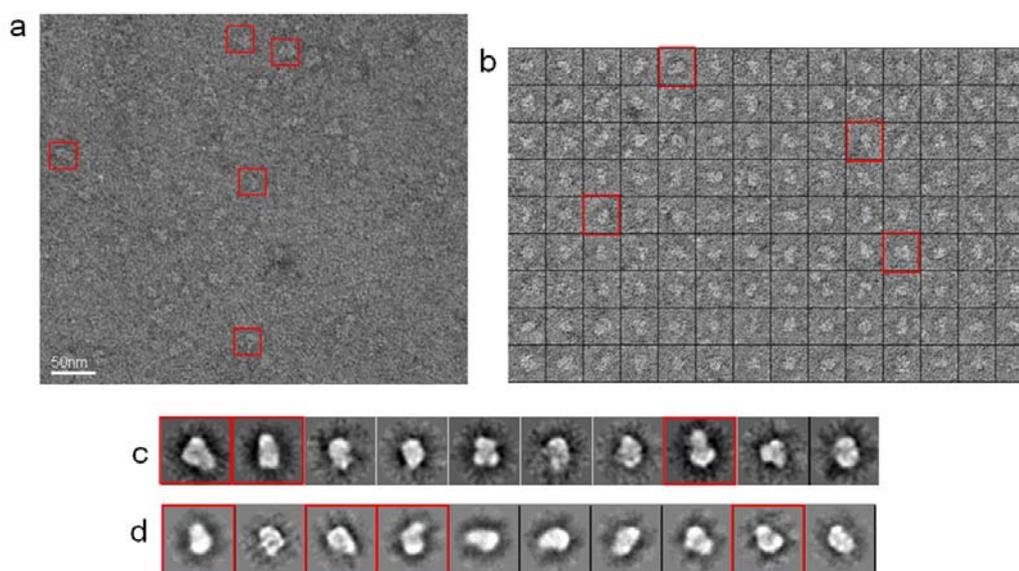


Figura.24. Microscopía electrónica del complejo p53-ADN. **(a)** Micrografía de la muestra p53-ADN teñida negativamente con acetato de uranilo al 2%. **(b)** Galería de imágenes seleccionadas de p53-ADN. **(c)** Clasificación de imágenes mediante *ML2D*. **(d)** Clasificación de imágenes mediante *startnclasses*. Las clases obtenidas por EMAN **(c)** y con XMIPP **(d)** presentaron características en común. Tanto en las partículas individuales como en algunas clases libres de referencia, se aprecia un tipo de vista característico con dos masas de distinto tamaño (encuadradas en rojo). En este caso se aprecia en algunas de ellas la existencia de un canal o falta de densidad en el centro de la molécula.

Siguiendo la misma estrategia usada para p53 libre (*startAny* de EMAN), se generó el mapa 3D inicial correspondiente del complejo p53-ADN. El modelo inicial se refinó mediante sucesivas iteraciones comparando cada una de las proyecciones con los datos originales hasta obtener un resultado convergente (*Figura.25*).

Las proyecciones generadas a partir del mapa 3D final son compatibles con imágenes medias obtenidas de los algoritmos de clasificación libres de referencia *startnclasses* y *ML2D* (*comparación Figura.24.c-d* y *Figura.25.b*).

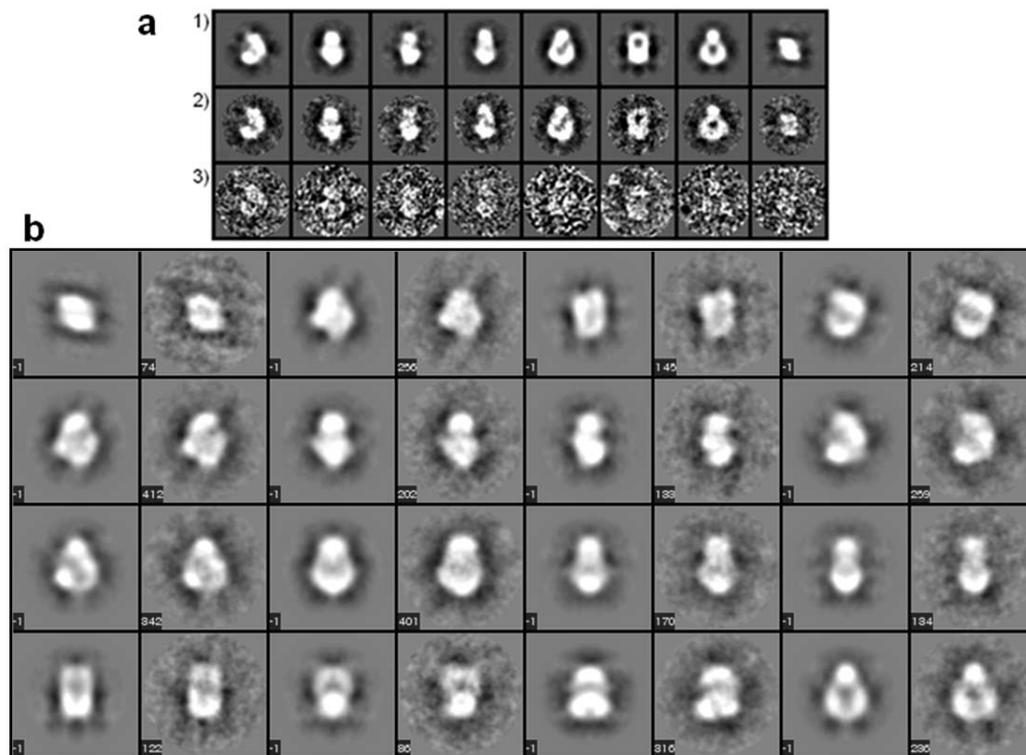


Figura.25. Refinamiento angular de p53-ADN. **(a.1)** Ejemplo de algunas proyecciones de las diferentes orientaciones de los mapas 3D. **(a.2)** Media de las partículas asignadas a cada proyección. **(a.3)** Partícula ejemplo de cada una de las clases. **(b)** Galería de proyecciones y medias equivalentes del complejo p53-ADN mediante el análisis de la muestra total, obtenidas en el proceso de refinamiento angular iterativo. Están representados dos a dos, primero (con “-1”) la proyección del mapa y al lado la clase media y el número de partículas que corresponden a cada clase.

El mapa resultante del complejo p53-ADN (*Figura.26.a-c*) es similar al obtenido por microscopía electrónica del tetrámero libre (*Figura.21.a-c*), es decir, con dos densidades (una arriba y otra abajo) unidas por dos conexiones. En este caso, las conexiones presentan una mayor densidad que en el caso de p53 libre y se atribuyen a las cuatro regiones flexibles que unen el dominio de tetramerización con el Core, lo que sugiere que el tetrámero de p53 está organizado por un dímero de dímeros.

En la vista lateral del mapa (*Figura.26.a*), en el eje longitudinal de la molécula, se aprecia claramente la formación de un canal en forma de círculo. Aunque la formación del complejo p53-ADN fue confirmada por los análisis de EMSA (*Figura.23.a-b*), la presencia del ácido nucleico no se puede ver en el mapa. Esta es una desventaja típica en tinción negativa de muestras que tienen ácidos nucleicos. La carga y forma del ADN, junto al bajo poder de resolución del agente de tinción empleado (acetato de uranilo), hacen que el ADN apenas tenga contraste y desaparezca en el mapa 3D. Al encajar en el mapa los cuatro cores unidos al ADN [Kitayner et al., 2006] se

aprecia que la doble hebra de ADN se adapta perfectamente al tamaño y sección del canal presente en el mapa (*Figura.26.e-f*). De lo que se deduce que la presencia del canal en el mapa 3D está relacionado con la posición del ADN en la molécula. El mapa de microscopía electrónica se puede interpretar como un tetrámero de p53 que envuelve a un invisible ADN (debido a la técnica de tinción negativa empleada para su estudio).

Por otra parte, a pesar de las limitaciones de la técnica de tinción negativa y de la resolución moderada alcanzada del mapa 3D (estimada entre 25-30 Å), los límites permitieron encajar las coordenadas atómicas disponibles para cada uno de los cuatro dominios centrales unidos al ADN (Código PDB: 2ATA [Kitayner et al., 2006]) y el dominio de tetramerización de p53 (Código PDB: 3SAK [Jeffrey et al., 1995]) (*Figura.26.d-f*). Al encajar los cuatro cores dentro del mapa hay dos de ellos (el par simétrico de monómeros verde y azul de arriba) (*Figura.26.e*) que sobresalen fuera del mapa 3D. Una posible explicación es que en la estructura resuelta de Kitayner, el tetrámero está formado por la unión de p53 a dos moléculas de ADN, mientras que en el mapa de microscopía electrónica hay una única molécula de ADN. Esto hace que, al haber encajado las coordenadas atómicas como cuerpos rígidos sobre el mapa, no se pueda modelar y no se ajuste perfectamente el mapa a la estructura de Rayos X. Otra posibilidad es que debido a la naturaleza agresiva de la técnica de tinción negativa exista una distorsión parcial de la estructura del complejo p53-ADN.

Al igual que en el caso de p53 libre, se han encajado solo las estructuras estables (los 4 dominios core unidos al ADN y el dominio de tetramerización) debido a que el resto de los dominios flexibles que forman el complejo desaparecen al hacer el promediado en 3D. En la parte inferior del mapa (*marcado con “*”, Figura.26.e*) hay un espacio libre donde podrían encajarse los 4 dominios N-terminal de cada cadena.

En una vista desde arriba del mapa (*Figura.26.f*) se aprecia otra característica interesante. Uno de los tres ejes de simetría del dominio de Tetramerización, el longitudinal que divide la estructura en dos dímeros, corre paralelo al eje de la hélice de ADN de doble cadena.

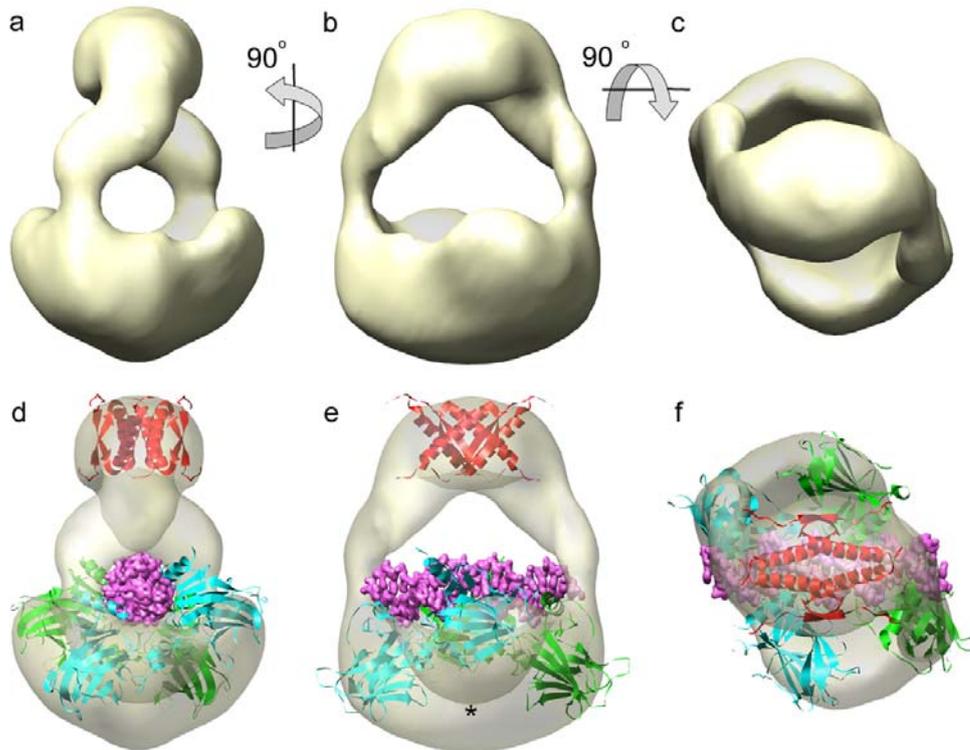


Figura.26. Estructura tridimensional del complejo p53-ADN. **(a-c)** Representación del 100% del volumen esperado del complejo p53-ADN en distintas orientaciones. **(d-f)** En un mapa semitransparente de estos volúmenes se pueden encajar algunos de los dominios resueltos por cristalografía de rayos X y RMN (dominio central (verde y azul) unido a ADN (rosa) (código PDB: 2ATA), dominio de tetramerización (rojo) (código PDB: 1SAK). El canal generado da lugar a un espacio adecuado para el paso del ADN. Con un "*" aparece marcada la zona donde estarían localizados los cuatro dominios N-terminales de cada monómero. Datos tomados de Melero et al., 2007.

Una vez concluido el proceso de refinamiento angular se llevó a cabo la estimación de la resolución de la estructura 3D resuelta. El volumen obtenido tiene una resolución aproximada de 26Å para un valor de FSC de 0,5 (*Figura.27*). Sus dimensiones aproximadas son 110Å en el eje longitudinal, 125Å de altura y un grosor de unos 105Å.

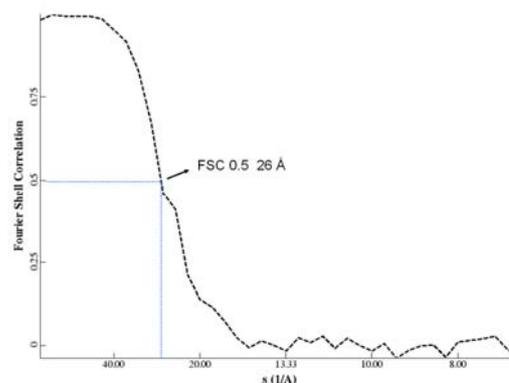


Figura.27. Resolución de mapa 3D del complejo p53-ADN. El volumen obtenido tiene una resolución aproximada de 26Å para un valor de FSC de 0,5.

R.2.2 Grafix: Purificación en gradiente de densidad/fijación

En los estudios realizados hasta el momento, la gran heterogeneidad que presenta la muestra total nos hace pensar en la mezcla de agregados, p53 libre y otras opciones.

Para mejorar estos aspectos se realizaron una serie de cambios en la preparación de las muestras siguiendo el protocolo de Grafix [Kastner et al., 2008]. Este método propone un aumento en cuanto a la calidad de las imágenes de microscopía electrónica debido, en parte, a que se trabajaba con una fracción homogénea de la muestra ayudada por la fijación. Para este método:

- Se prepararon los complejos con ADN a alta concentración de p53 (10 μ M) y en relación molar 1:4 (ADN:p53) para evitar, en la medida de lo posible, tener p53 libre.
- Se sometió a la muestra a un gradiente de densidad de glicerol (5-25%)-glutaraldehído (0-0.15%), que permitió separar los diferentes estados oligoméricos de p53 fijada en distintas fracciones. En la centrifugación, a la vez que avanza la muestra en el gradiente de densidad, se produce simultáneamente la fijación de dominios y/o conformaciones accesibles para que al separar las poblaciones en distintas fracciones no haya un re-equilibrio, que daría lugar de nuevo a la mezcla. El hecho de trabajar con una concentración de agente fijador tan baja (gradiente 0-0.15 % de glutaraldehído), hace que la fijación sea suave (la concentración en glutaraldehído de la fracción que contiene los tetrámeros es baja, aproximadamente del 0.05%).
- Se analizó en las diferentes fracciones la formación y el tamaño del complejo p53-ADN mediante geles EMSA (nos informa de la formación del complejo) y geles SDS-PAGE (nos sirve para estimar el tamaño del complejo).
- Para la microscopía electrónica se usaron las fracciones que presentaron tetrámeros puros de p53-ADN. La tinción negativa se realizó con un agente de menor tamaño de grano (formiato de uranilo). En este caso las rejillas se prepararon por el método del doble carbón o “sándwich”, en el que la muestra se queda entre dos láminas de carbón asegurando así el contraste en toda la estructura.

Para estudiar el papel que ejerce el tamaño del ADN en la formación del complejo, se utilizaron dos muestras de ADN de distinta longitud (60 y 44 pares de bases) pero con la misma secuencia específica de reconocimiento del gen *GADD45*.

R.2.2.1 Purificación de tetrámeros de p53 unidos a ADN de 60 pb en gradiente de densidad/fijación

a. Caracterización bioquímica

Para llevar a cabo el método Grafix se utilizó la muestra de p53 en el tampón fosfato, para evitar la interacción del glutaraldehído con los grupos aminos del tris.

La muestra de p53, en tampón fosfato, se incubó 1h a temperatura ambiente con ADN de 60pb que contenía (en negrita) la secuencia específica de la región promotora del gen *GADD45*: (5'-CC TAC AGA ATC GCT CTA CAG **AAC ATG TCT AAG CAT GCT** GGG GAC TG ATG CTG GGG ACT GG-3'), en una relación molar 1:4 (ADN:p53).

Una vez formado el complejo se centrifugó a 25.000 r.p.m 18 horas a 4°C en un gradiente de glicerol (5-25%) y glutaraldehído (0-0.15%). En la centrifugación, las moléculas avanzan en el gradiente dependiendo de su tamaño, de tal forma que es posible separar cada uno de los diferentes estados oligoméricos que pueda adoptar la proteína en la formación del complejo p53-ADN.

Se analizaron las diferentes fracciones en geles SDS-PAGE (para ver el estado oligomérico de p53) y EMSA (para estudiar la presencia de ADN en cada uno de los distintos estados oligoméricos que pueda adoptar la proteína). Al comparar ambos tipos de geles se puede deducir en que estado oligomérico esta interaccionando p53 con el ADN. En la calle S del EMSA (muestra antes del GRAFIX) (*Figura.28.a*) se aprecia la formación del complejo p53-ADN (banda retardada) y el ADN libre, en exceso. En las calles 4-5, p53 forma complejo con el ADN en un estado oligomérico tetramérico, mientras que en las calles 6-9 su estado oligomérico es mayor (octámeros, agregados,...) (*Figura.28.a-b*). Mediante el análisis por microscopía electrónica de las fracciones 7-9 se confirmó la presencia de estados oligoméricos superiores al tetrámero,

al obtenerse imágenes de cuerpos bilobulares, pertenecientes a dímeros de tetrámeros y/o agregados grandes.

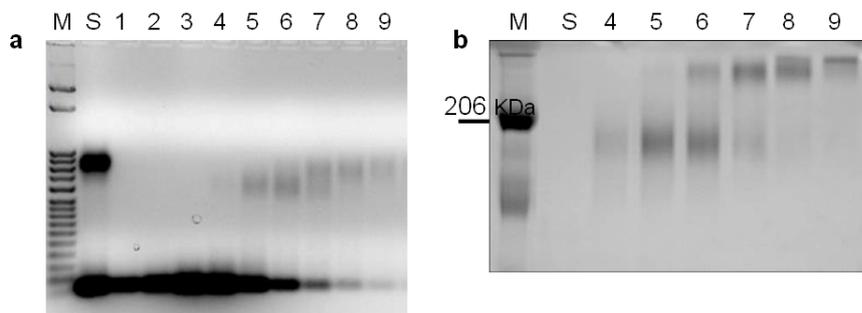


Figura.28. Caracterización bioquímica tras el gradiente de densidad/fijación de p53 y ADN de 60pb. Análisis de las diferentes fracciones del gradiente de densidad/fijación. **(a)** Gel EMSA de las fracciones de la centrifugación teñido con bromuro de etidio. **(b)** Gel SDS-PAGE de las fracciones de la centrifugación teñido con Coomassie.

b. Reconstrucción tridimensional

Para los ensayos de microscopía electrónica, se utilizaron las fracciones que presentaron un estado de oligomerización tetramérico con el ADN.

Con las fracciones correspondientes al complejo tetramérico p53-ADN (*calles 4-5, Figura.28*) se realizaron rejillas de tinción negativa con formiato de uranilo al 1% mediante el método de sándwich de doble carbón. La homogeneidad en el tamaño de los complejos mejoró la uniformidad en cuanto al tamaño de las partículas (*Figura.29.a-b*), lo que permitió seleccionar manualmente hasta un total de 14.000 partículas. Para la clasificación de imágenes de partículas individuales, de igual forma que en casos anteriores, se utilizaron las herramientas *refine2d.py* de EMAN y *ML2D* de XMIPP (*Figura.29.c-d*). Al comparar estas clases con las obtenidas mediante el análisis de la muestra total de p53 libre (*Figura.19.c-d*) y del complejo p53-ADN (*Figura.24.c-d*), se apreció en este caso una gran mejoría en cuanto a la uniformidad, calidad y definición de las imágenes (*Figura.29.a-d*). Esta mejoría permitió diferenciar, dentro de algunas de las clases y partículas individuales (encuadradas en rojo, *Figura.29.a-d*), cada uno de los cuatro cores y el dominio de tetramerización. Además, la presencia de un canal en el centro de la molécula, vuelve a sugerir la existencia de un paso para el ADN.

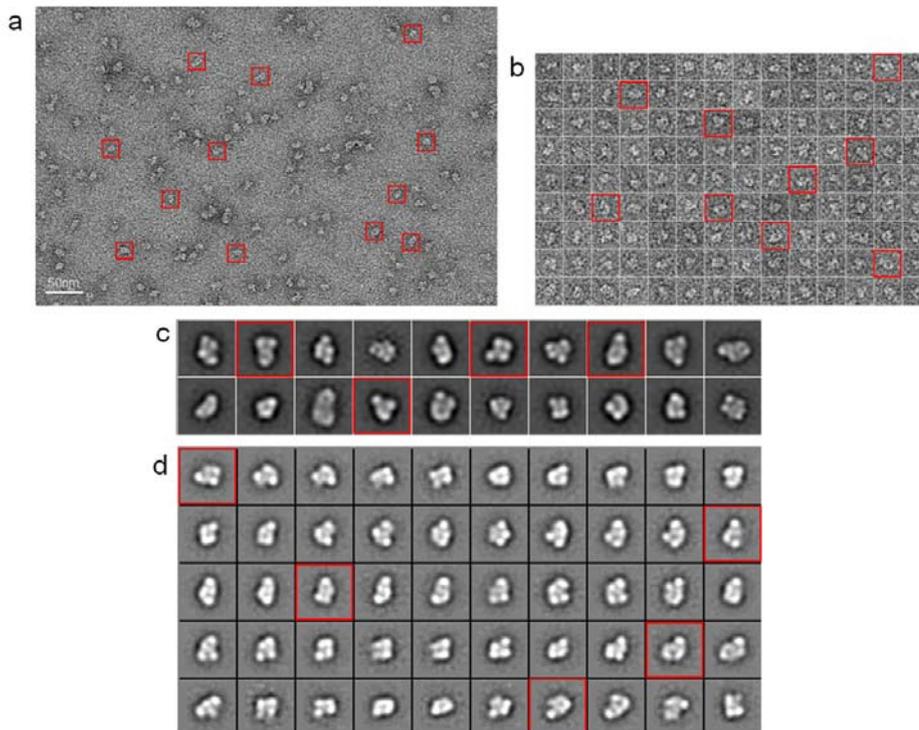


Figura.29. Microscopía electrónica del complejo tetramérico p53 con ADN de 60 pares de bases por el método Grafix. **(a)** Micrografía de p53-ADN teñida negativamente con formiato de uranilo al 1%. **(b)** Galería de imágenes seleccionadas de p53-ADN. **(c)** Clasificación de imágenes mediante *ML2D* (XMIPP). **(d)** Clasificación de imágenes mediante *refine2d.py* (EMAN 1.8). Si comparamos las clases obtenidas por EMAN y con XMIPP, veremos que ambas tienen características en común. Tanto en las partículas individuales como en algunas clases libres de referencia, se aprecia un tipo de vista característico donde se pueden distinguir las masas correspondientes a los cores y al dominio de tetramerización (encuadradas en rojo). Se aprecia en algunas de ellas la existencia de un canal o falta de densidad en el centro de la molécula.

Se generó, como en casos anteriores, un modelo inicial 3D partiendo de las imágenes medias calculadas para cada una de las clases, mediante el método de líneas comunes (*startAny* de EMAN). El modelo inicial resultante se refinó mediante sucesivas iteraciones comparando cada una de las proyecciones con los datos originales hasta obtener un resultado convergente (*Figura.30*). Las proyecciones generadas a partir del mapa 3D final son compatibles con imágenes medias obtenidas de los algoritmos de clasificación libres de referencia *ML2D* y *refine2d.py* (*comparación Figura.29.c-d y Figura.30*)

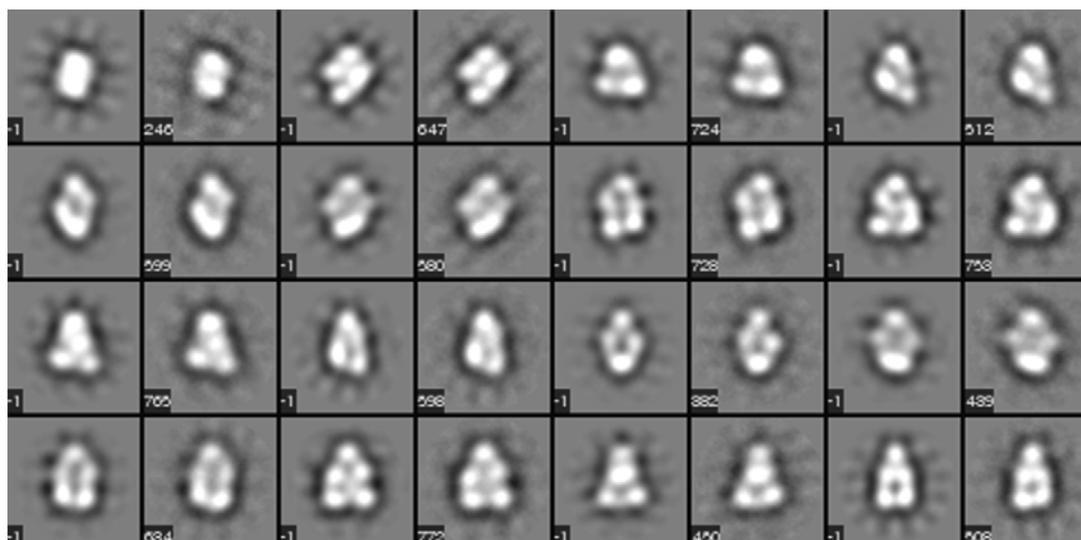


Figura.30. Galería de proyecciones y medias equivalentes del complejo p53-ADN (de 60pb) tras el método Grafix, obtenidas en el proceso de refinamiento angular iterativo. Están representados dos a dos, primero (con “-1”) la proyección del mapa y al lado la clase media y el número de partículas que corresponden a cada clase.

Con la nueva preparación (Grafix), el mapa final 3D obtenido (*Figura.31.a-d*) presenta características generales en común con el obtenido anteriormente del complejo p53-ADN (mediante el método de análisis de la muestra total) (*Figura.26.a-c*).

El mapa 3D muestra una estructura con el dominio de tetramerización arriba y los cuatros dominios centrales o cores abajo (*Figura.31.a*). Las dos densidades fuertes que unen la parte superior con la inferior se atribuyen a las cuatro regiones flexibles que unen el dominio de tetramerización con los cuatro cores, lo que sugiere que el tetrámero de p53 está organizado por un dímero de dímeros.

Se aprecia claramente la formación de un canal en forma de círculo en la vista lateral del mapa (*Figura.31.a*).

En este caso hay una importante mejoría, con respecto a los mapas anteriores, ya que se distinguen los cuatro cores abajo (*Figura.31.a-c*), mientras que antes era solo una gran masa (*Figura.26.a-c*). Sin embargo, al igual que ocurrió con el método de análisis de la muestra total del complejo p53-ADN, en este caso tampoco se aprecia ni la densidad del ácido nucleico (resultado típico de la tinción negativa de muestras con ADN) ni la señal de las regiones flexibles (debido al promediado en 3D).

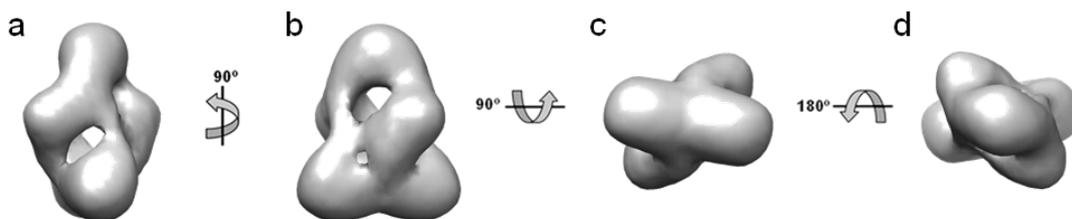


Figura.31. Mapa de microscopía electrónica de p53-ADN de 60 pares de bases tras el método Grafix. **(a-d)** Diferentes vistas del mapa 3D. La masa de arriba corresponde al dominio de tetramerización y cuatro masas de abajo corresponden a los cuatro cores que unen el ADN.

Una vez concluido el proceso de refinamiento angular se llevó a cabo la estimación de la resolución de la estructura 3D resuelta. El volumen obtenido tiene una resolución aproximada de 28Å para un valor de FSC de 0,5 (*Figura.32*). Sus dimensiones aproximadas son 100Å en el eje longitudinal, 125Å de altura y un grosor de unos 105Å

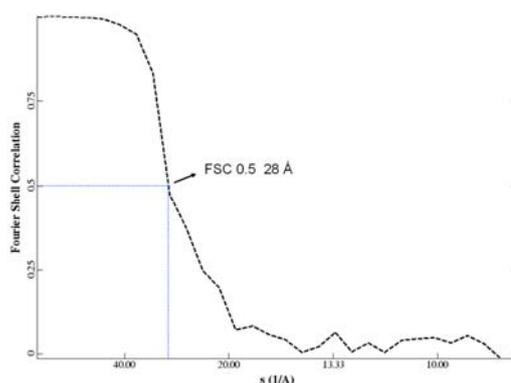


Figura.32. Resolución del mapa 3D de p53-ADN de 60 pb tras el método Grafix. El volumen obtenido tiene una resolución aproximada de 28Å para un valor de FSC de 0,5.

c. Polimorfismo cuaternario de p53 en su interacción con ADN de 60pb

La flexibilidad de las regiones que unen el dominio de tetramerización con cada uno de los cuatro dominios centrales, le permiten a la molécula p53 adaptarse en el reconocimiento e interacción con el ADN, para la formación de los tetrámeros p53-ADN. Esta flexibilidad explica la distinta posición de los cores (debido a su movilidad) dentro del tetrámero unido al ADN en los cristales resueltos por Rayos X [Cho et al., 1994; Malecka et al., 2009; Kitayner et al., 2006; Chen et al., 2010]. Para estudiar esta movilidad (o cambios conformacionales) que pueda adoptar p53 en su interacción con

el ADN, se usó la herramienta de clasificación *ML3D* [Scheres et al., 2007] dividiendo el conjunto de datos dentro de cuatro grupos diferentes.

Los Mapas de cada una de las clases (I-IV) obtenidos por *ML3D* muestran grandes diferencias en cuanto a la posición relativa de los 4 cores entre si y con respecto al dominio de tetramerización (*Figura.33.b*). Al alinear los cuatro mapas por los dos cores inferiores, las dos densidades correspondientes a los dominios superiores (marcados con asterisco “*”) cambian su posición en las distintas clases. Al comparar los 4 mapas entre si se ve que hay dos cores (abajo) que se mantienen, más o menos, en la misma posición y dos (arriba) que oscilan. Estos últimos aparecen arriba en las clases I, II y IV y abajo en la clase III (*Figura.33.b*).

Las conexiones entre los dominios centrales y entre dímeros, son diferentes para cada clase. El mapa de la clase III es el único donde están interaccionando los cuatro cores entre si (conformación cerrada), mientras que en el resto de las clases solo hay interacción dos a dos (conformación abierta).

Al comparar la clase I (*Figura.33.b.claseI*) y la clase II (*Figura.33.b.claseII*), se observa que los cores móviles (de arriba, marcados con “*”) cambian la forma de conexión con los cores fijos (de abajo). Esto sugiere que los cores pueden cambiar la forma de interaccionar entre sí para la formación del tetrámero.

En una vista lateral de los mapas (*Figura.33.c*) se aprecia, en cada una de las clases, la presencia de canales a lo largo de la molécula, que como ya se ha visto anteriormente, podría estar relacionado con la localización del ADN en la molécula.

En ninguno de los mapas se aprecia, ni la densidad del ácido nucleico (resultado típico de la tinción negativa de muestras con ADN), ni la señal de las regiones flexibles (debido al promediado en 3D).

Cabe destacar, que al hacer la clasificación dividiendo el conjunto de datos en un mayor número de grupos (hasta doce) se obtuvo un resultado similar de variabilidad conformacional a la obtenida al hacer la separación en 4 clases.

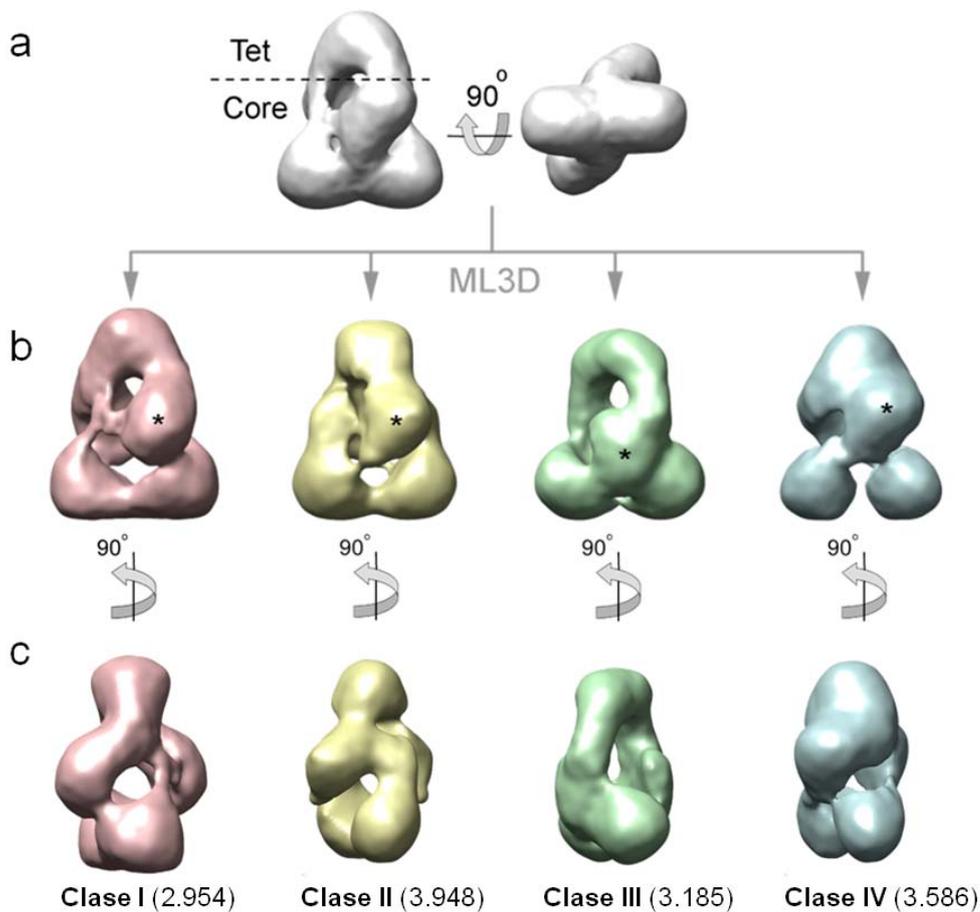


Figura.33. Clasificación mediante *ML3D* del complejo p53-ADN de 60 pb. **(a)** Representación del 100% de la masa esperada del complejo p53-ADN. **(b-c)** Diferentes vistas de los mapas de cada una de las clases (I-IV) obtenidos por *ML3D*. Están marcados con “*” las dos densidades correspondientes a los dominios superiores, cuya posición cambia en las distintas clases. Entre paréntesis aparece indicado en número de imágenes perteneciente a cada clase.

El mapa de la clase III (*Figura.33.b.claseIII*) forma un tetrámero real, con interacciones intra- e inter- diméricas similares a las del tetrámero de Kitayner [Kitayner et al., 2006], por lo que es posible encajar dentro del mapa, cada uno de los cuatro dominios centrales unidos al ADN (Código PDB: 2ATA [Kitayner et al., 2006]) y el dominio de tetramerización (Código PDB: 3SAK [Jeffrey et al., 1995]) (*Figura.34.c*) como cuerpos rígidos (*Figura.34.a*).

En el resto de las clases no se dan estas interacciones, debido a su conformación abierta. El hecho de que en los mapas de las clases I, II y IV tengan los monómeros muy alejados, en ocasiones sin interactuar entre sí, hace que sea imposible encajar ninguna de las estructuras atómicas resueltas para tetrámeros del core. Al igual que ocurrió en los casos anteriores, en ninguno de los mapas se ve el ADN. Sin embargo, la presencia de canales en el mapa está relacionada con la posible localización del ADN dentro de la molécula. De tal forma que se puede encajar una molécula de ADN dentro de cada uno

de los mapas que muestren un canal. Siguiendo este criterio se ha encajado un ADN de doble hebra y el dominio de tetramerización en los mapas de la clase I y III con la intención de identificar distintos modos de unión.

El mapa de la clase I muestra perfectamente un canal donde se puede encajar el ADN (*Figura.34.b*). Al girar 50° el mapa de la clase III, se ve el canal con el ADN encajado (*Figura.34.c*). Al comparar entre ambos mapas la dirección del ADN con respecto al dominio de tetramerización, se aprecia que en la clase I corren paralelos, mientras que en la clase III el dominio de tetramerización y el ADN están girados (aproximadamente 50°).

Cuando el ADN se localiza paralelo al dominio de tetramerización lo hemos llamado ADN I, mientras que cuando esta girado lo hemos llamado ADN II (*Figura.34.d-f*).

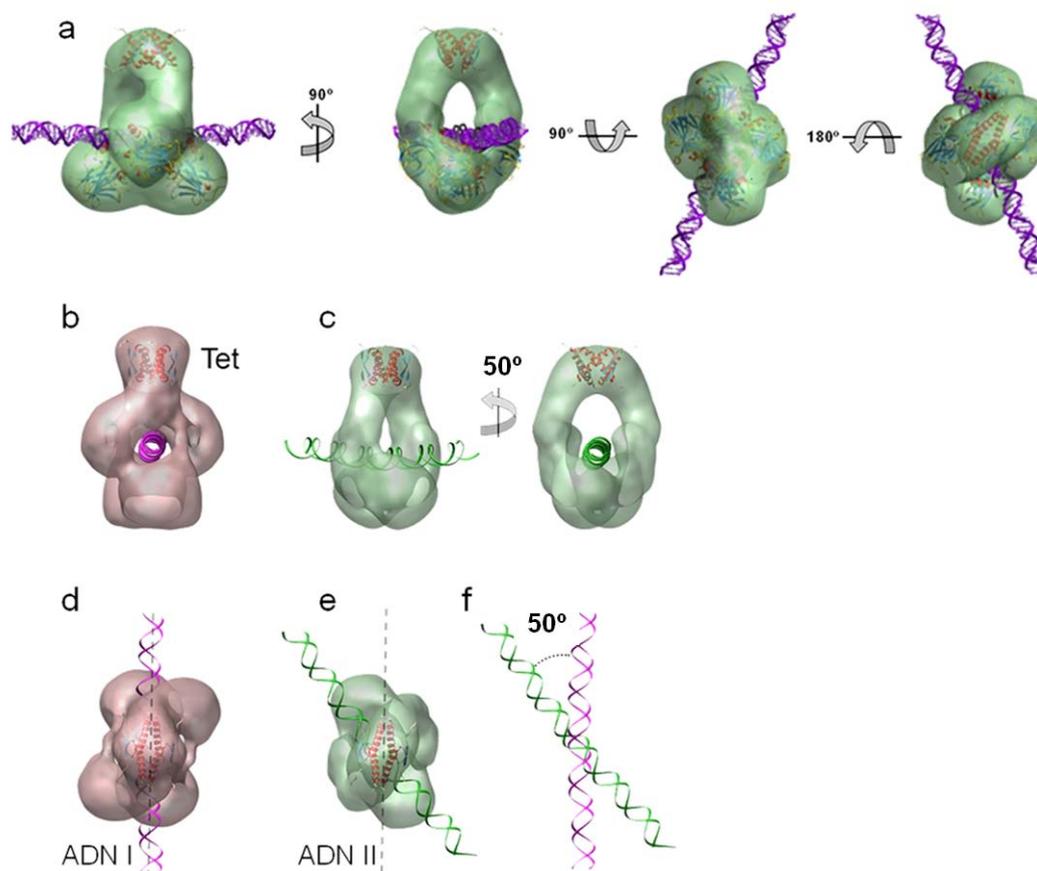


Figura.34. Ajuste de estructuras atómicas en los mapas 3D de las clases I y III del complejo p53-ADN (60pb). **(a)** Mapa de la clase III con los cuatro cores, el ADN y el dominio de tetramerización encajados. El canal generado da lugar a un espacio adecuado para el paso del ADN. **(b)** Mapa de la clase I con el ADN y el dominio de tetramerización encajados. En este caso no hay forma de encajar ninguna de las estructuras cristalinas resueltas de los cores, ya que al ser una estructura abierta, estos no llegan a interactuar entre si. **(c)** Un giro de 50° en el eje C2 de la molécula expone una mejor vista del canal. **(d)** En el mapa de la clase I el ADN corre paralelo al dominio de tetramerización (ADN I). **(e)** En el mapa de la clase III el ADN está girado 50° con respecto al dominio de tetramerización (ADN II). **(f)** Los ADN encajados en los mapas de la clase I y III presentan un giro entre sí de 50° .

Una vista lateral de la clase I muestra una característica interesante, se ven cuatro conexiones (dos fuertes y dos débiles) entre el dominio de tetramerización y los cuatro cores (*Figura.35.a*), mientras que en la clase III solo hay dos conexiones fuertes (*Figura.35.c*). Estas dos conexiones débiles de la clase I nos hacen sospechar que se tratan de dos de los dominios C-terminales que interactúan con el ADN de forma inespecífica. También podrían ser dos de las cuatro conexiones entre el dominio de tetramerización y el core. Al encajar en ambos mapas, los 4 dominios C-terminales siguiendo la dirección (marcada con “*”) de cada una de sus α -hélices correspondientes del dominio de tetramerización (dos hacia arriba y dos hacia abajo), se observa que en el caso de la clase I, los dos C-terminal que bajan hacia los core, llegan a interactuar con el ADN (*Figura.35.a-b*), mientras que en el caso de la clase III no hay densidad para ninguno de ellos (*Figura.35.c*).

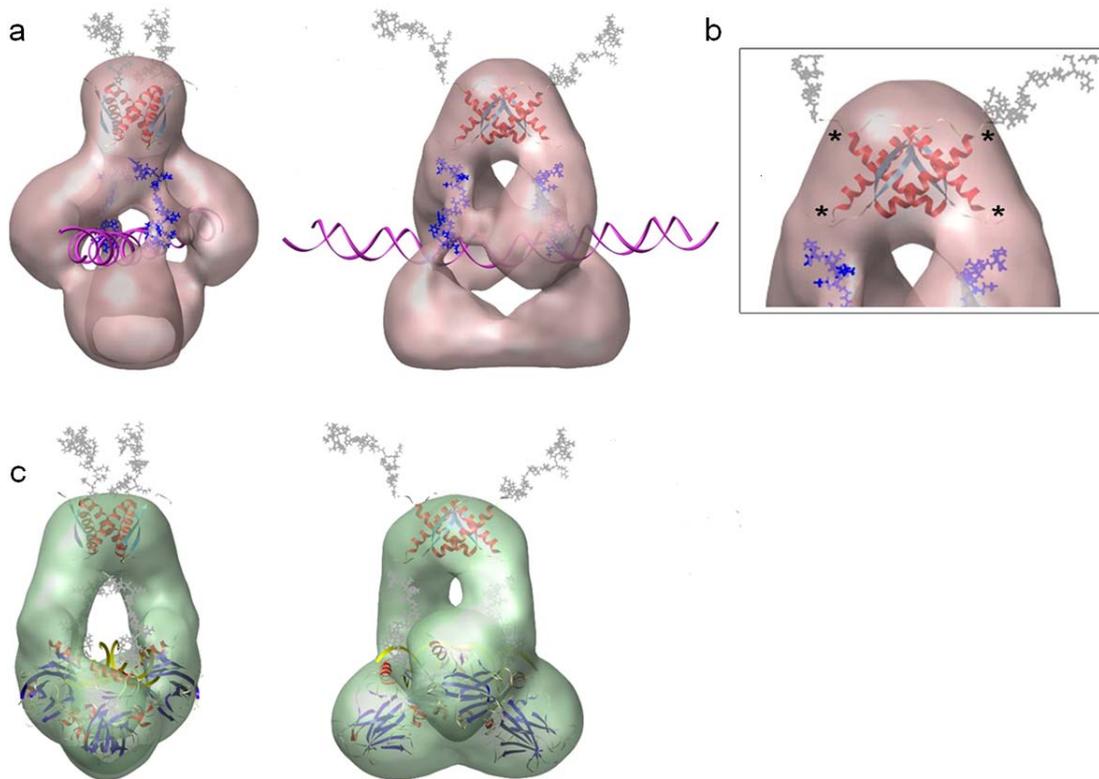


Figura.35. Ajuste del dominio C-terminal en los mapas 3D de las clases I y III del complejo p53-ADN (60pb). **(a)** Diferentes vistas del mapa de la clase I con el dominio de tetramerización, el ADN y los cuatro dominios C-terminales encajados. Los dos C-terminales que se dirigen hacia el core están coloreados en azul. **(b)** Ampliación de la zona del dominio de tetramerización en el mapa de la clase I. Con “*” están marcadas las conexiones de cada α -hélice del dominio de tetramerización con su correspondiente dominio C-terminal **(c)** Diferentes vistas del mapa de la clase III con el dominio de tetramerización, el ADN y los cuatro dominios C-terminales encajados.

R.2.2.2 Purificación de tetrámeros de p53 unidos a ADN de 44 pb en gradiente de densidad/fijación

a. Caracterización bioquímica

Para la determinación de la organización tridimensional del complejo formado entre p53 y ADN de 44 pares de bases por microscopía electrónica se empleó el mismo procedimiento que el utilizado para el cálculo de los mapas de p53 y ADN de 60pb.

La muestra de p53, en tampón fosfato se incubó 1h a temperatura ambiente con ADN de 44pb que contenía (en negrita) la secuencia específica de la región promotora del gen *GADD45*: (5'-CCT ACA CTA CAG AAC ATG TCT AAG CAT GCT GGG GAC TGG ACT GG-3'), en una relación molar 1:4 (ADN:p53).

Una vez formado el complejo se centrifugó a 25.000 r.p.m 18 horas a 4°C en un gradiente de glicerol (5-25%) y glutaraldehído (0-0.15%). Las diferentes fracciones fueron analizadas en gel SDS-PAGE y gel EMSA (*Figura.36*) Al comparar ambos tipos de geles se puede deducir en que estado oligomérico está interaccionando p53 con el ADN. En la calle S (muestra antes del GRAFIX) del EMSA (**a**) se aprecia la formación del complejo p53-ADN (banda retardada) y el ADN en exceso (libre), mientras que en el gel SDS-PAGE (**b**) aparece la banda de la proteína desnaturalizada (aprox. 50KDa). En la calle 3, p53 forma complejo con el ADN (**a**) en un estado oligomérico tetramérico (**b**), mientras que en las calles 4-7 su estado oligomérico es mayor (octámeros, agregados,...) (*Figura.36.a,b*).

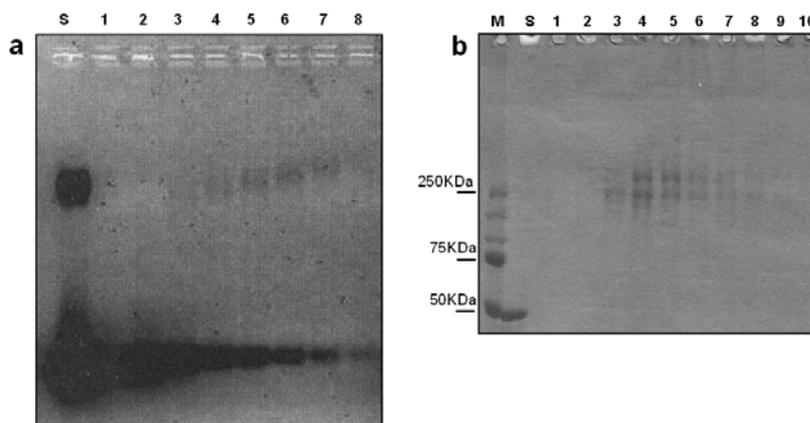


Figura.36. Caracterización bioquímica de p53 y ADN de 44pb tras el gradiente de densidad/fijación. Análisis de las diferentes fracciones del gradiente de densidad/fijación. (**a**) Gel EMSA de las fracciones teñido con bromuro de etidio. (**b**) Gel SDS-PAGE de las fracciones teñido con Coomassie.

b. Reconstrucción tridimensional

Con las fracciones que contienen los tetrámeros de p53-ADN (*Figura.36.a.calle 3*) se realizaron rejillas de tinción negativa con formiato de uranilo al 1% mediante el método de “sándwich” de doble carbón.

Fueron seleccionadas manualmente un total de 24.670 partículas y clasificadas mediante las herramientas libres de referencia *refine2d.py* de EMAN) y *ML2D* de XMIPP (*Figura.37.a-d*). Las clases obtenidas por EMAN y XMIPP presentaron características en común. Una mejoría en la calidad de las imágenes permitió diferenciar los dominios core y el dominio de tetramerización en algunas de las partículas individuales y en algunas clases (encuadradas en rojo). También en este caso se aprecia la presencia de un canal en el centro de la molécula en alguna de ellas que nos llevó a pensar en el paso del ADN a su través.

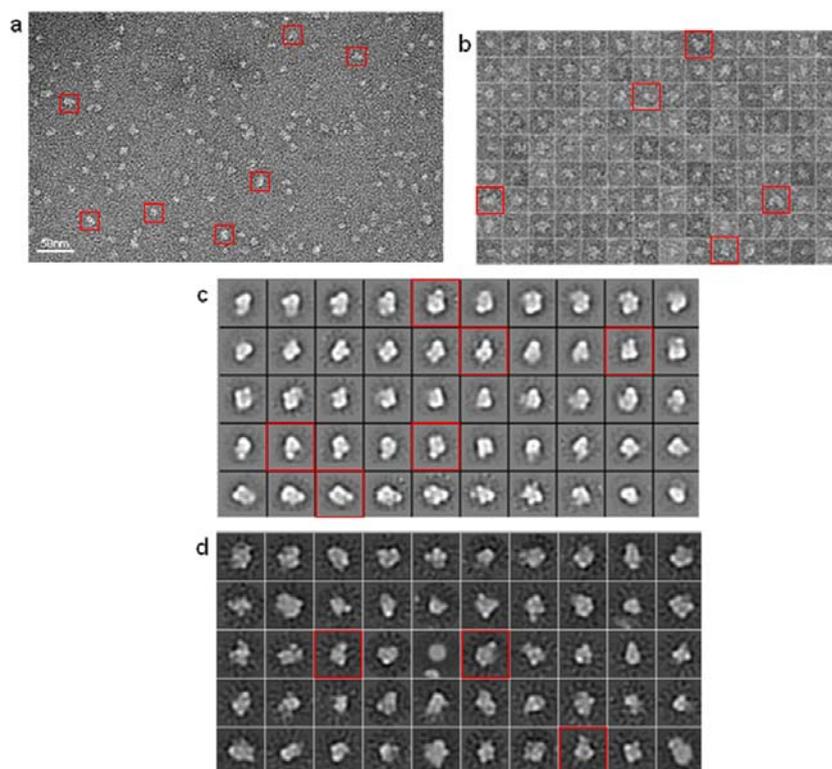


Figura.37. Microscopía electrónica del complejo tetramérico p53 con ADN de 44 pares de bases por el método Grafix. **(a)** Micrografía de p53-ADN teñida negativamente con formiato de uranilo al 1%. **(b)** Galería de imágenes seleccionadas de p53-ADN (44pb). **(c)** Clasificación de imágenes mediante *refine2d.py* (EMAN 1.8). **(d)** Clasificación de imágenes mediante *ML2D* (XMIPP). Tanto en las partículas individuales como en algunas clases libres de referencia, se aprecia un tipo de vista característico donde se pueden distinguir las masas correspondientes a los cores y al dominio de tetramerización (encuadradas en rojo). Se aprecia en algunas de ellas la existencia de un canal o falta de densidad en el centro de la molécula.

Las clases resultantes se utilizaron para generar un primer volumen mediante el método de líneas comunes con la herramienta *startAny* de EMAN. El modelo inicial resultante se refinó mediante sucesivas iteraciones comparando cada una de las proyecciones con los datos originales hasta obtener un resultado convergente (*Figura.38*).

Las proyecciones generadas a partir del mapa 3D final son compatibles con imágenes medias obtenidas de los algoritmos de clasificación libres de referencia *ML2D* y *refine2d.py* (comparación *Figura.37.c-d* y *Figura.38*)

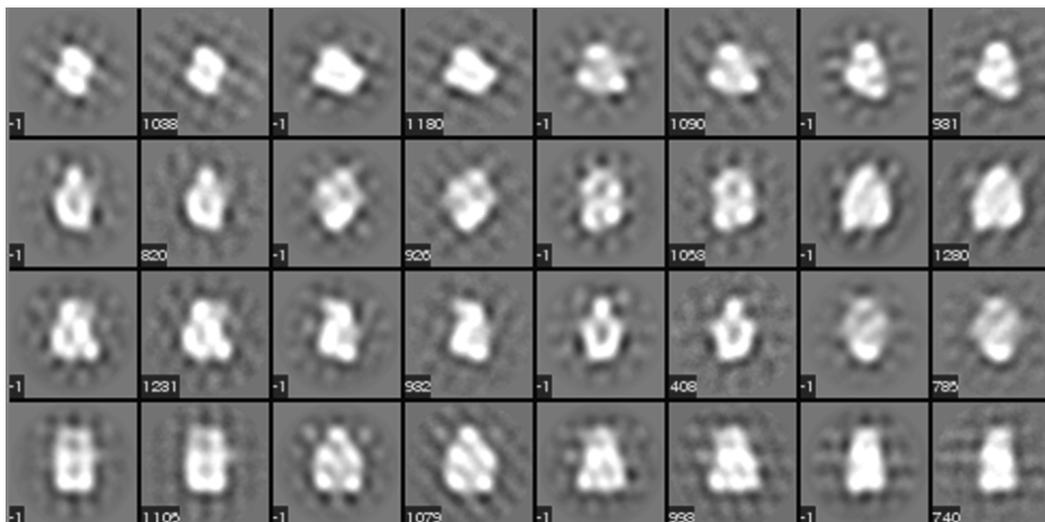


Figura.38. Galería de proyecciones y medias equivalentes del complejo p53-ADN (de 44pb) tras el método Grafix, obtenidas en el proceso de refinamiento angular iterativo. Están representados dos a dos, primero (con “-1”) la proyección del mapa y al lado la clase media y el número de partículas que corresponden a cada clase.

El mapa 3D final muestra una estructura con el dominio de tetramerización arriba y los cuatros dominios centrales o cores abajo (*Figura.39.a-d*). En este caso se aprecian dos densidades fuertes que unen la parte superior con la inferior, que se atribuyen a las cuatro regiones flexibles que unen el dominio de tetramerización con los cuatro cores (lo que sugiere que el tetrámero de p53 está organizado por un dímero de dímeros) y dos densidades débiles que salen del dominio de tetramerización hacia los core. En una vista lateral del mapa se aprecia la formación de un canal en forma de círculo (*Figura.39.a*). Al igual que ocurrió en los casos antes estudiados, en este caso tampoco se apreció la densidad del ácido nucleico, ni de las regiones flexibles.

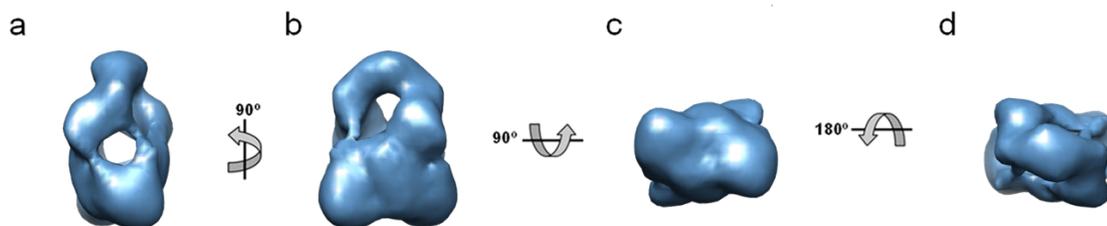


Figura.39. Mapa de microscopía electrónica de p53-ADN de 44 pares de bases tras el método Grafix. **(a-d)** Diferentes vistas del mapa 3D. La masa de arriba corresponde al dominio de tetramerización y cuatro masas de abajo corresponden a los cuatro cores que unen el ADN.

Una vez concluido el proceso de refinamiento angular se llevó a cabo la estimación de la resolución de la estructura 3D resuelta. El volumen obtenido tiene una resolución aproximada de 26 Å para un valor de FSC de 0,5 (*Figura.40*).

Sus dimensiones aproximadas son 110-115 Å en el eje longitudinal, 125 Å de altura y un grosor de unos 105 Å

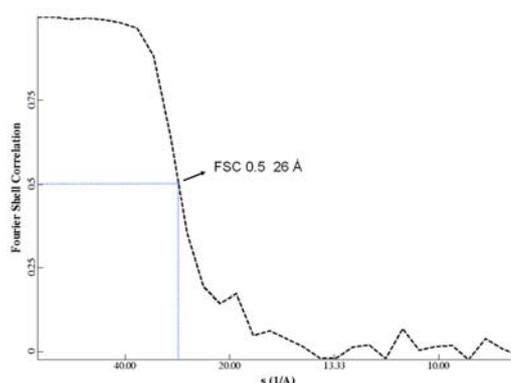


Figura.40. Resolución del mapa 3D de p53-ADN (de 44 pb) tras el método Grafix. El volumen obtenido tiene una resolución aproximada de 26Å para un valor de FSC de 0,5.

c. Polimorfismo cuaternario de p53 en su interacción con ADN de 44pb

Para analizar la variabilidad estructural de las imágenes obtenidas se usó la herramienta de clasificación *ML3D* de XMIPP [Scheres et al., 2007] dividiendo el conjunto de datos dentro de tres grupos diferentes. En este caso, ante un ADN más corto, la variación conformacional fue mucho menor que la obtenida para los complejos de p53 con ADN de 60 pares de bases.

En los mapas de cada una de las clases (I-III) obtenidas por *ML3D*, se aprecia que el 70% de las partículas totales van a la clase II, mientras que la clase I se queda prácticamente vacía (*Figura.41.b*)

Ante un ADN más corto, los tetrámeros formados en ambas clases (II y III) son mucho más compactos (cerrados) que los que presentó la muestra con un ADN más largo, ninguna de las clases obtenidas adopta una conformación abierta. Una conformación cerrada hace que aumente la probabilidad de unión específica entre los cores con el ADN, debido a que permite interacciones inter- e intradiméricas.

Al comparar los dos mapas entre sí (clase II y III) se ve que ambos presentan características en común. Los dos muestran una estructura con el dominio de tetramerización arriba y los cuatro dominios centrales abajo interaccionado entre sí (conformación cerrada) (*Figura.41.b.claseII-III*). En una vista lateral de ambos mapas (*Figura.41.b*) se aprecia, en cada una de las clases, la presencia de canales a lo largo de la molécula, lo que se relaciona, por su tamaño y forma, con la posible localización del ADN en la molécula.

En ambos casos hay cuatro densidades (dos fuertes y dos débiles) que bajan del dominio de tetramerización hacia los core. Las dos densidades fuertes se atribuyen a las cuatro regiones flexibles que unen el dominio de tetramerización con los cuatro cores, lo que sugiere que el tetrámero de p53 está organizado por un dímero de dímeros, mientras que las dos débiles, por su orientación de salida con respecto al dominio de tetramerización, podrían estar relacionadas con los dominios C-terminal.

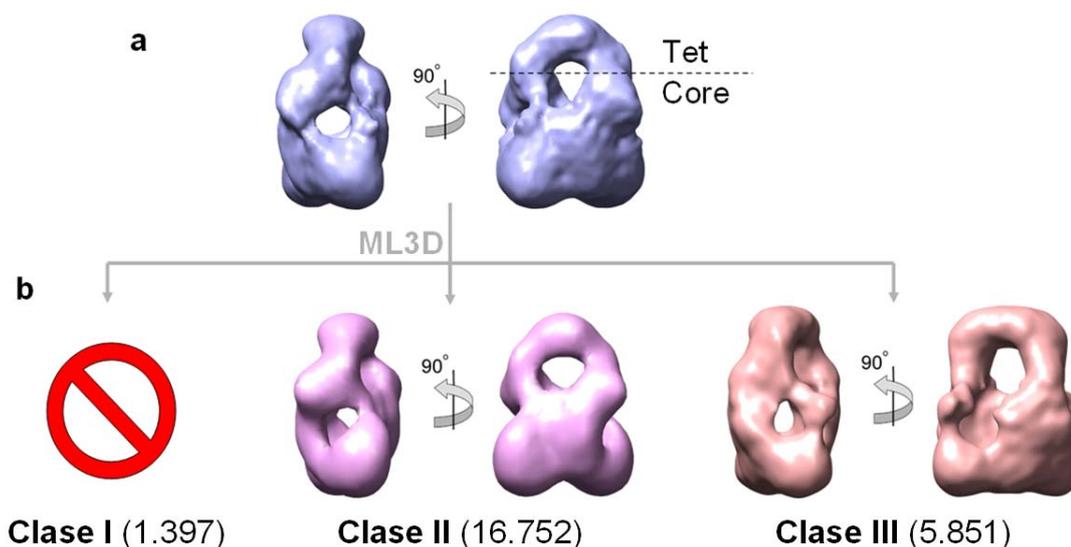


Figura.41. Clasificación mediante *ML3D* del complejo p53-ADN (de 44 pb) tras el método *Grafix* **(a)** Representación del 100% de la masa esperada del complejo p53-ADN. **(b)** Mapas de cada una de las clases (I-III) obtenidas por *ML3D*. Entre paréntesis aparece indicado en número de imágenes perteneciente a cada clase.

El mapa de la clase II (*Figura.42.a*) forma un tetrámero con todos sus cores interaccionando entre sí donde se podría encajar la estructura de los cuatro cores unidos al ADN (Código PDB: 2ATA [Kitayner et al., 2006]) y el dominio de tetramerización (Código PDB: 3SAK [Jeffrey et al., 1995]) como cuerpos rígidos (*Figura.42.a*). Sin embargo, a pesar de que los cuatro cores encajan perfectamente en el mapa, el ADN no encajaría con la posición del canal dentro del mapa. Por tanto no es posible encajar la estructura de Kitayner sin que el ADN choque con la densidad del mapa 3D (*Figura.42.a*).

Al encajar solo el ADN (por la dirección del canal) y el dominio de tetramerización (*Figura.42.b*), en este caso se ve que corren paralelos (habría una disposición del tipo ADN I) (*Figura.42.b.4*). Con el ADN encajado se aprecia que las conexiones débiles que bajan del dominio de tetramerización hacia el core llegan a interaccionar con el ADN, lo que hace pensar que se trata del C-terminal.

El mapa de la Clase III, permite encajar cada uno de los cuatro dominios centrales unidos al ADN (Código PDB: 2ATA [Kitayner et al., 2006]) y el dominio de tetramerización (Código PDB: 3SAK [Jeffrey et al., 1995]) como cuerpos rígidos (*Figura.42.c*). Este mapa tiene dos canales en su vista lateral (*Figura.c.2* y *Figura.d.2*), al intentar encajar la estructura de kitayner guiándose por la dirección del canal con el ADN, solo es correcta (de acuerdo a la estructura atómica) una de las opciones (*Figura.d*). A pesar de que los cuatro cores encajan perfectamente en el mapa, el ADN no encajaría con el resto de los cores dentro de la estructura atómica (*Figura.42.c*).

Al encajar el ADN de acuerdo a la estructura atómica correcta, se observa que el ADN y el dominio de tetramerización están girados (habría una disposición del tipo ADN II) (*Figura.d.4*) Además este mapa, al igual que el anterior, presenta una densidad extra que baja del dominio de tetramerización hasta la zona por donde iría el ADN, lo que hace suponer que se trata del dominio C-terminal.

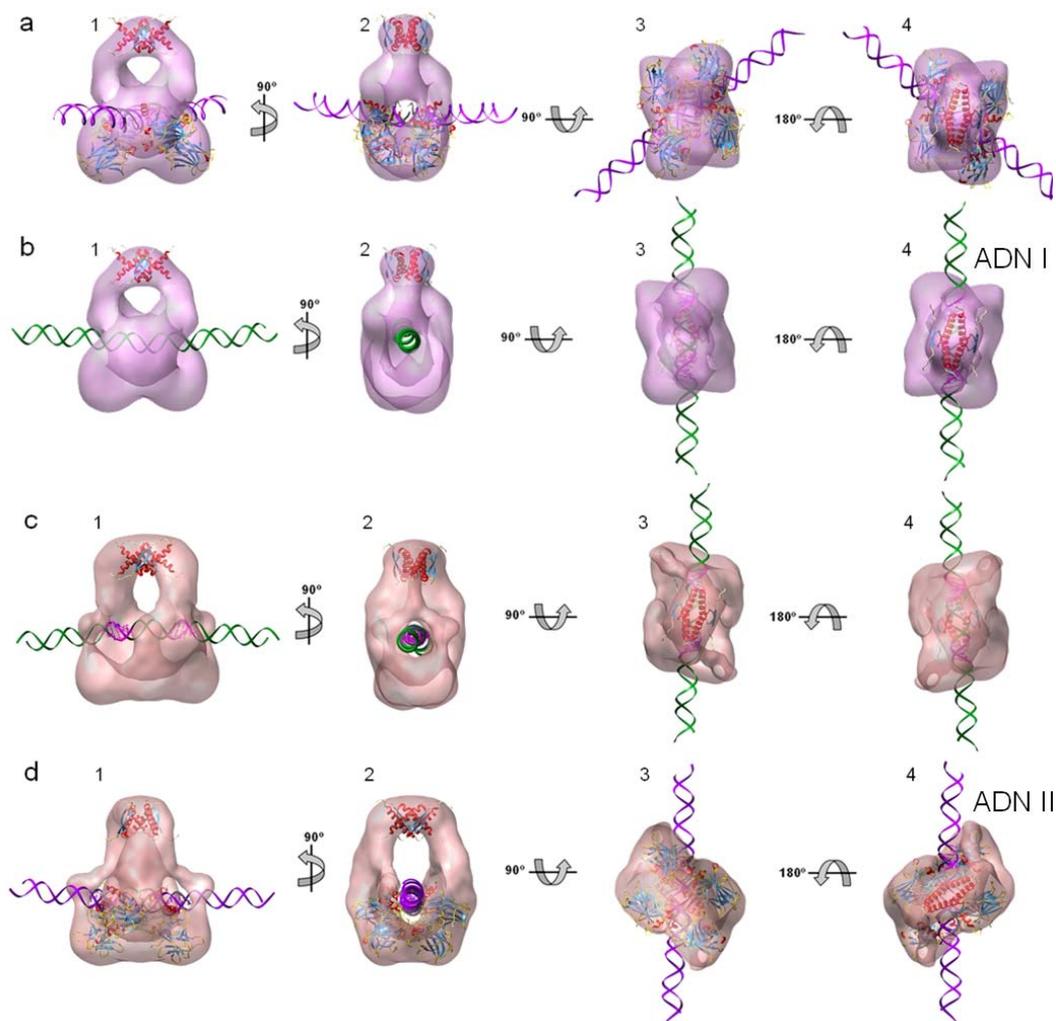


Figura.42. Comparación de los mapas de las clases II y III con ADN de 44 pares de bases. **(a.1-4)** Diferentes orientaciones del mapa de la clase II con los cuatro cores, el ADN y el dominio de tetramerización encajados. El canal generado (a.2) no coincide con el paso del ADN en la estructura atómica. **(b.1-4)** Diferentes orientaciones del mapa de la clase II con el ADN y el dominio de tetramerización encajados. En este caso no hay forma de encajar ninguna de las estructuras cristalinas resueltas de los cores sin que el ADN choque con la densidad del mapa. **(c.1-4)** Diferentes orientaciones del mapa de la clase III con el dominio de tetramerización encajados y el ADN encajado en uno de los dos canales (c.2 y c.3) que expone el mapa 3D. El canal generado no coincide con el paso del ADN en la estructura atómica, por lo que no se puede encajar ninguna de las estructuras resueltas de los cuatro cores unidos al ADN. **(d.1-4)** Diferentes orientaciones del mapa de la clase III con el ADN encajado en el otro canal del mapa (d.2 y d.3). Esta conformación permite encajar sin problemas el dominio de tetramerización y los cuatro cores unidos al ADN. En el mapa correcto de la clase II (b.4), el ADN corre paralelo al dominio de tetramerización (ADN I), mientras que en el mapa correcto de la clase III (d.4), el ADN está girado 50° con respecto al dominio de tetramerización (ADN II).

Al superponer el mapa de la clase III con un ADN de 44 pb, con el mapa de la clase III obtenido con un ADN de 60 pb, se observa que son casi idénticos, excepto en la presencia de dos masas adicionales en el mapa de 44 pb (*Figura.43*)

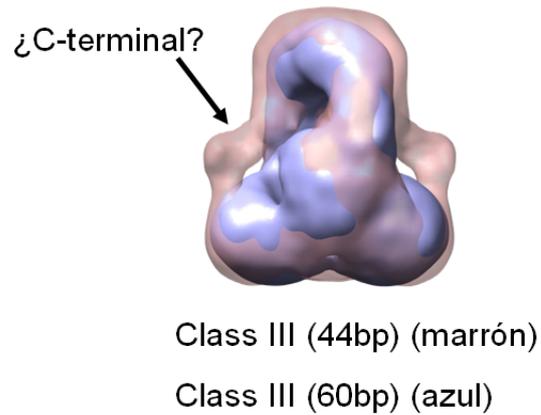


Figura.43. Comparación de los mapas obtenidos por *ML3D* de p53 con dos muestras ADN de distinta longitud (60 y 44pb). Al superponer clase III (ADN de 44 pb, en marrón) con la clase III (ADN de 60 pb, en azul), se ve que la única diferencia entre ambos radica en la densidad extra que presenta el mapa marrón respecto al azul. Esta densidad podría ser el C-terminal.

Capítulo 4

Discusión

p53 pertenece a una gran clase de proteínas eucariotas que contienen una mezcla de dominios estructurales ordenados y estables (el core y el dominio de tetramerización) y otros flexibles y desordenados (N-terminal, C-terminal y conexiones entre core y dominio de tetramerización). Este tipo de conformación hace que p53 pueda funcionar en distintos escenarios e interactuar con diferentes ligandos. Esta flexibilidad plantea un problema en trabajos de cristalografía de rayos X. Una buena opción en el estudio de poblaciones heterogéneas podría ser la microscopía electrónica.

En este trabajo se ha realizado un análisis de la estructura cuaternaria del tetrámero de p53 libre y formando complejo con ADN, mediante las herramientas de microscopía electrónica y procesamiento de imagen. Los resultados obtenidos por esta técnica se han complementado mediante otras técnicas de análisis estructural de proteínas (SAXS y RMN) en colaboración con otros laboratorios (MRC, EMBL).

Para resolver el problema de inestabilidad que presenta la proteína p53 silvestre, en este trabajo se ha utilizado una variante de p53 con cuatro mutaciones puntuales en el dominio de unión al ADN que no cambian la estructura global de la molécula [Nikolova et al., 1998].

La información aportada por la combinación de estas técnicas ha permitido analizar el problema de la heterogeneidad estructural de p53.

Los resultados revelan la existencia de un equilibrio entre dos conformaciones (abierta y cerrada) en el caso de la proteína libre, equilibrio que se desplaza a la forma cerrada cuando p53 se une a su secuencia específica.

Con la intención de comprender la estructura tridimensional que adopta p53 en cada uno de sus pasos, desde que se encuentra libre hasta que reconoce y forma complejo con una secuencia específica del ADN, se ha realizado en esta memoria el estudio estructural de las diferentes etapas.

D.1 p53 libre, antes de su interacción con el ADN

Por microscopía electrónica se estudió la estructura tridimensional que presentaba la proteína p53 libre, y los datos se compararon con los obtenidos mediante SAXS y espectroscopía por Resonancia Magnética Nuclear (Melero et al., 2007)

La estrategia seguida de encajar las estructuras de alta resolución de los dominios individuales resueltos por cristalografía de rayos X y/o RMN en los modelos de baja resolución obtenidos por EM y SAXS hace que en conjunto sea una buena forma de resolver la estructura de proteínas. Tanto SAXS como la espectroscopía de RMN tienen la ventaja de poder analizar estructura de proteínas en solución, obviando los artefactos asociados tanto a la congelación de las muestras como a la tinción negativa usadas en la microscopía electrónica. La técnica SAXS optimiza la disposición espacial de las diferentes subunidades estructurales encajando los datos de dispersión de las estructuras resueltas de alta resolución para poder formar complejos macromoleculares de un determinado tamaño y simetría, siendo posible más de una solución. Los datos obtenidos por SAXS son particularmente útiles para poder discriminar entre dos modelos aquel que no se ajuste a los perfiles de dispersión obtenidos.

Los resultados obtenidos de p53 libre por microscopía electrónica presentaron una estructura cerrada. La inmovilización de la muestra sobre el soporte de carbón o el propio agente de tinción puede conducir a la selección de una conformación particular, dando lugar, en el caso del estudio de p53, a una conformación compacta (*Figura.44.b*).

En el caso de los estudios realizados de p53 libre por espectroscopía de RMN y SAXS mostraron una forma mayoritaria abierta (*Figura.44.a*).

Las diferencias entre las estructuras de p53 libre en solución por SAXS y por EM muestran que es necesario realizar el estudio estructural de una proteína mediante distintos métodos.

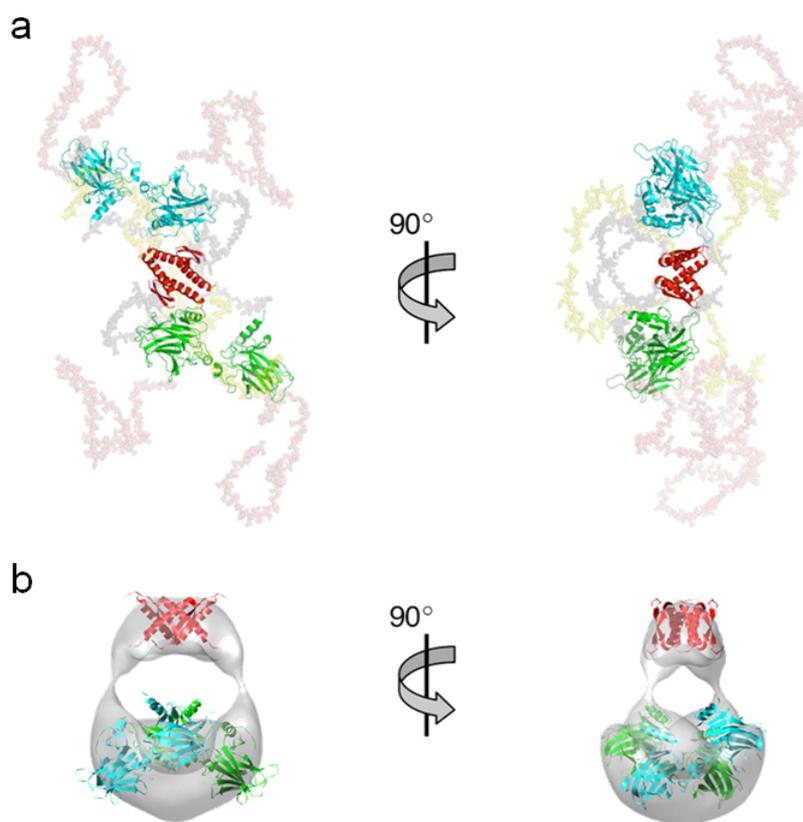


Figura.44. Comparación de mapas de p53 libre obtenidos mediante dos técnicas de análisis de estructura de macromoléculas distintos: SAXS y microscopía electrónica. **(a)** Modelo de SAXS de p53 libre. Los dominios core (verdes y azules) y el dominio de tetramerización (rojo) están unidos mediante conexiones flexibles (gris). El N-terminal (rosa) y el C-terminal (amarillo) están representados de forma semitransparente. **(b)** Una representación semitransparente del mapa de microscopía electrónica de p53 permite observar las coordinadas atómicas ajustadas de los dominios estructurales resueltos por cristalografía de rayos X y RMN.

El hecho de que por microscopía se obtenga solo la forma cerrada puede deberse a que la forma abierta se colapsaría en la rejilla. Esta es la razón por la que en las micrografías de p53 libre la muestra presentase tanta heterogeneidad (*Figura.19.a-b*) y que solo se seleccionasen el 20% de las partículas totales presentes en la rejilla (partículas “cerradas”). La gran heterogeneidad mostrada por las partículas “abiertas” en cuanto a su forma, debido a la gran posibilidad de caer sobre la rejilla, impediría calcular una media 3D, por tratarse de una estructura muy variable.

Los resultados tanto de SAXS como de microscopía electrónica de p53 libre [Melero et al., 2007], se compararon con los obtenidos por crió microscopía electrónica de p53 de ratón [Okovorov et al., 2006].

El mapa obtenido de crio-microscopía de p53 de ratón tiene una estructura compacta donde se ha encajado el dominio de tetramerización con las 4 α -hélices por separado (Tet, flecha roja) en contacto con los dominios N terminal (N-t, flecha negra) (*Figura.45.c*). Esta separación está en contra de la estructura resuelta del dominio de tetramerización mediante rayos X y RMN [Jeffrey et al., 1995], además hace que se pierda la simetría D2 que caracteriza a este dominio.

Los resultados obtenidos por SAXS y microscopía electrónica muestran una situación de equilibrio entre una conformación abierta y cerrada, debido a la flexibilidad de las regiones que unen el dominio de tetramerización con cada uno de los cores. Este equilibrio está desplazado hacia una conformación abierta, cuando p53 esta libre (*Figura.45.a-b*). Las estructuras obtenidas por microscopía electrónica y SAXS de p53 humano, no tiene nada que ver con la obtenida para p53 de ratón.

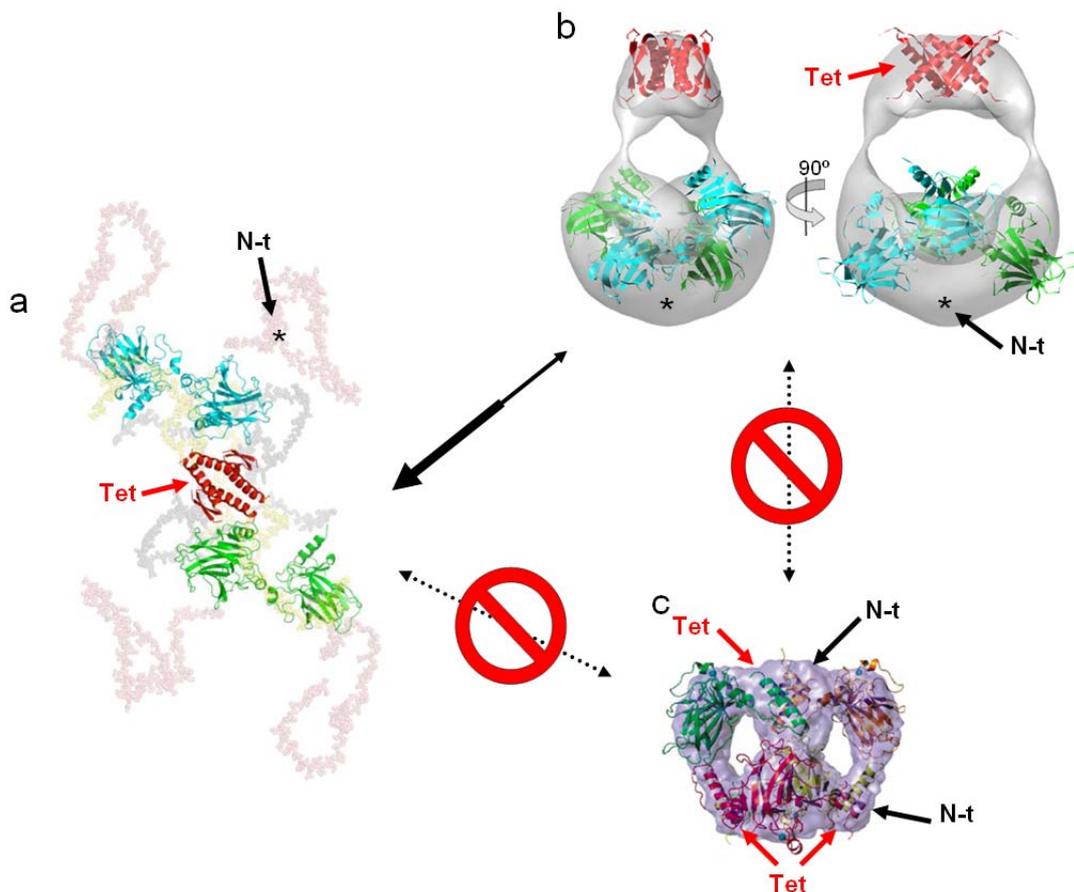


Figura.45. Comparación de los resultados obtenidos por microscopía electrónica y SAXS para p53 humano respecto a los resultados obtenidos por crio-microscopía electrónica para p53 de ratón. (a) Estructura de p53 libre de humano, mediante SAXS. (b) Estructura de p53 libre de humano, mediante microscopía electrónica. (c) Estructura de p53 libre de ratón, mediante crio-microscopía electrónica. Este resultado está en desacuerdo con los obtenidos por SAXS y microscopía electrónica para p53 humana.

En los resultados obtenidos por SAXS y microscopía electrónica de p53, el dominio de tetramerización, en todo momento mantiene su simetría D2.

En la forma abierta (SAXS) el dominio de tetramerización tiene simetría D2. Al cerrarse, las conexiones flexibles hacen que los cores se unan en forma de tetrámero, por lo que la simetría D2 general se rompe (solo queda D2 en el dominio de tetramerización). La estructura cerrada tiene un eje de simetría C2 (*Figura.46*), mientras que en el mapa de crio-microscopía electrónica de p53 de ratón se ha aplicado simetría D2 [Okovorov et al., 2006].

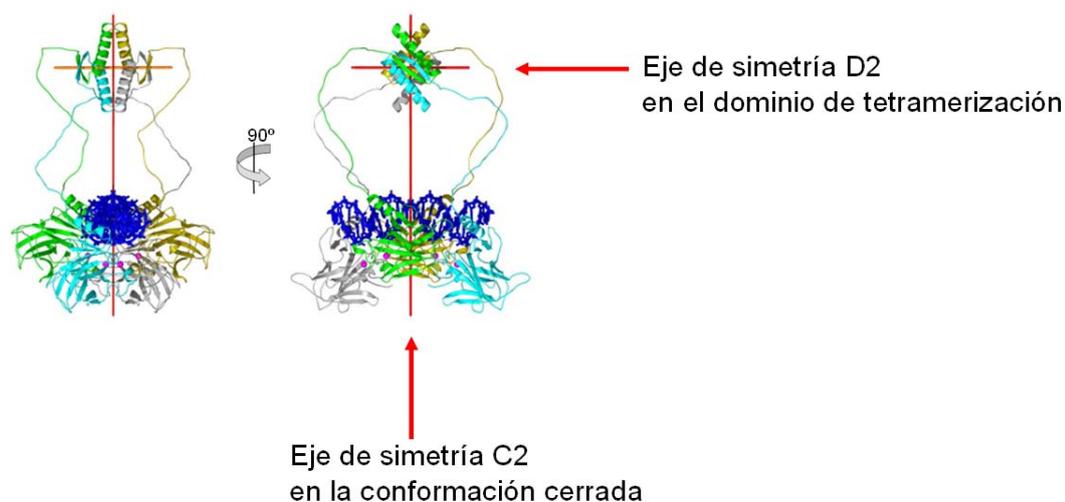


Figura.46. Ejes de simetría. El dominio de tetramerización por sí solo tiene dos ejes de simetría D2 (ejes en rojo). EN la forma abierta los mantiene, sin embargo en su forma cerrada (bien sea libre en formando complejo con el ADN) se mantiene solo la simetría D2 en el propio dominio de tetramerización, mientras que en la molécula completa se genera un eje de simetría C2. Datos tomados de Kitayner et al., 2006.

D.2 p53 formando complejo con el ADN

El estudio estructural que se ha realizado en este trabajo para el complejo p53-ADN proporciona la primera información tridimensional sobre la organización cuaternaria de esta proteína completa unida al ADN.

Para este estudio, en primer lugar se identificó la formación del complejo entre p53 y una sonda de ADN de 60 pares de bases que contenía la secuencia específica de reconocimiento por el dominio central de la proteína. Una vez confirmada la formación del complejo (*Figura.23.a-b*) se analizó la muestra por microscopía electrónica. En las micrografías, las imágenes de las partículas individuales del complejo p53-ADN presentaron una distribución heterogénea, pero más compacta y uniforme que en el

caso de p53 libre, dando como resultado en este caso una estructura 3D cerrada (*Figura.26.a*).

Estos resultados fueron comparados con los obtenidos por espectroscopía de RMN y SAXS. Tanto la microscopía electrónica como SAXS generan un modelo muy similar para el complejo p53-ADN, en el que la proteína p53 adopta una conformación cerrada capturando en su interior el ADN (*Figura.47*). Debido a la metodología de tinción negativa empleada para la preparación de la muestra en la microscopía electrónica, es normal que en el mapa 3D no se vea el ADN. El bajo poder de resolución del agente de tinción empleado (acetato de uranilo) y la carga y forma del ADN hacen que apenas tenga contraste.

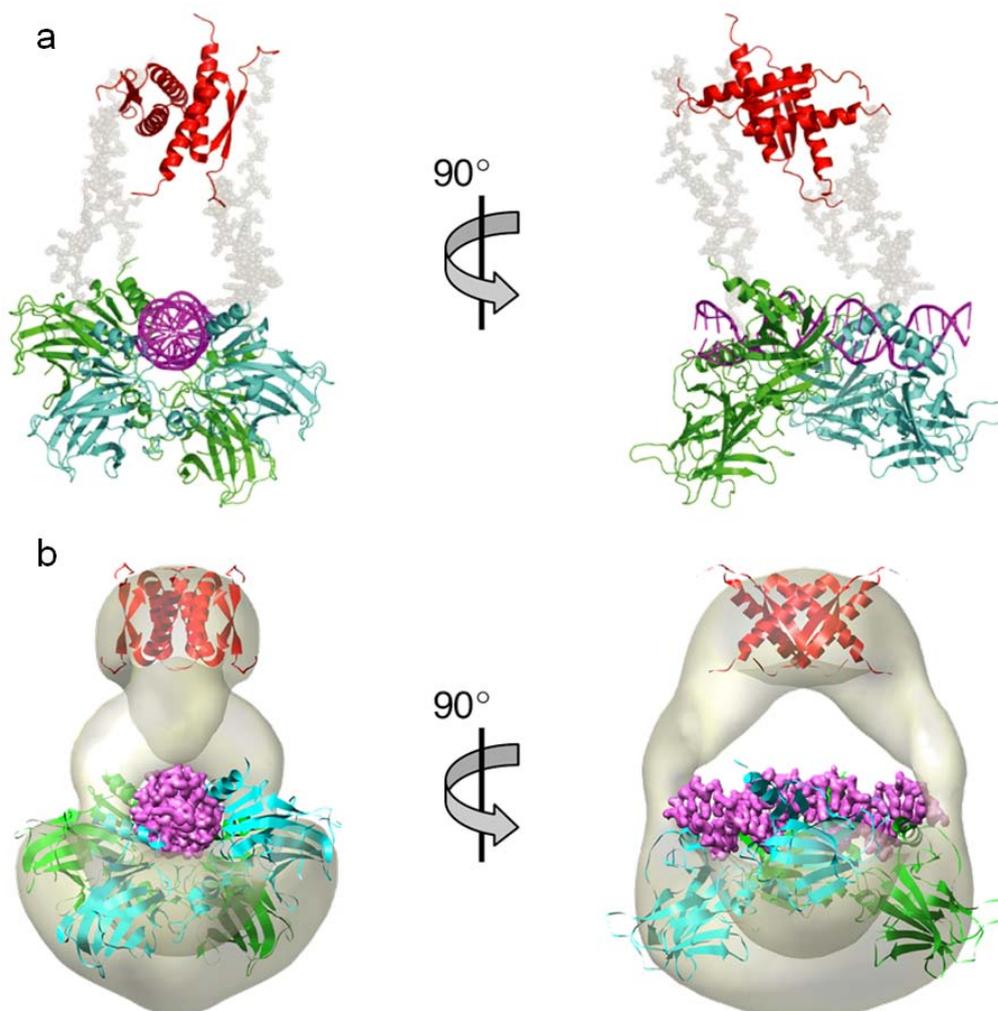


Figura.47. Comparación de mapas del complejo p53-ADN obtenidos mediante SAXS y microscopía electrónica. **(a)** Modelo de SAXS del complejo p53-ADN. Los dominios core (verdes y azules) y el dominio de tetramerización (rojo) están unidos mediante conexiones flexibles (gris) “atrapando” en su interior una molécula de ADN. **(b)** Modelo 3D de microscopía electrónica del complejo p53-ADN. Ante la presencia de ADN p53 se cierra formando un dímero de dímeros, puesto que solo hay dos conexiones entre el dominio de tetramerización (rojo) y los cuatro dominios centrales (dímero verde y dímero azul).

En vista de los resultados obtenidos hasta el momento se puede concluir que el supresor tumoral p53 en su forma mayoritariamente abierta (libre), al interactuar con el ADN se cierra formando una estructura compacta y estable con el ADN en su interior. El hecho de que solo se vean dos conexiones entre el dominio de tetramerización y los cuatro dominios centrales que reconocen el ADN, nos hace pensar que el tetrámero está formado por un dímero de dímeros.

D.3 Mecanismo de formación del complejo tetramérico p53 - ADN

Con el fin de entender el mecanismo y los cambios conformacionales que experimenta p53 en el reconocimiento e interacción con el ADN para la formación del complejo, se usó la herramienta de clasificación *ML3D* a partir de las imágenes obtenidas del complejo tetramérico p53 y un ADN de 60 pares de bases tras su purificación en un gradiente de densidad/fijación [Scheres et al., 2007].

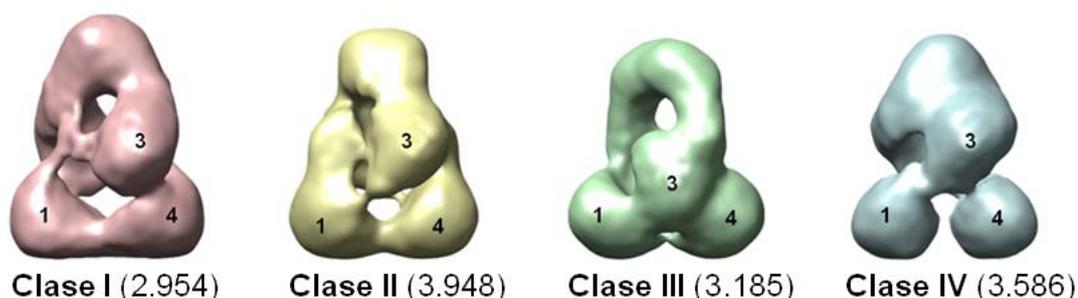


Figura.48. Mapas resultantes de la clasificación por *ML3D* del complejo p53-ADN (de 60 pb) tras el método *Grafix*. Con los números (1, 2, 3 y 4) están marcados cada uno de los cuatro cores (en los distintos mapas el core 2 no se ve por estar en simetría C2 con el core 3). Entre paréntesis está indicado el número de partículas totales que tras la clasificación, entran en cada una de las clases (I-IV).

En cada una de las cuatro clases resultantes del *ML3D* (*Figura.48*) se puede ver que hay dos cores abajo (1 y 4) que se mantienen, más o menos, en la misma posición y dos que oscilan (2 y 3). Estos últimos aparecen arriba en las clases I, II y IV y abajo en la clase III.

Esta variabilidad conformacional apoya los resultados y modelos planteados en otros trabajos [Ma et al., 2007] donde se propone que en la formación del complejo tetramérico (p53-ADN) hay dos cores, de dos dímeros distintos, que son los primeros en interactuar con el ADN. Una vez unidos al ADN se mantienen estables mientras se produce posteriormente la entrada de los otros dos cores para formar el tetrámero.

Que se produzca primero la entrada o interacción con el ADN de uno u otro core puede depender de la simetría de la secuencia específica de ADN (*Figura.49.a*). Se ha sugerido que esta simetría está estrechamente relacionada con los mecanismos moleculares y respuestas biológicas que puede ejercer p53 en la regulación de la expresión de ciertos genes (*Tabla.4*). El tipo de simetría más común en la región promotora de los genes regulados por p53 dentro del genoma humano es la de tipo H14 (*Figura.49.a-b* y *Tabla.4*) [Ma et al., 2007]. Este tipo de simetría induce una interacción fuerte entre p53 y el ADN (con una K_D de 7.7 nM) (*Tabla.4*).

La simetría H14 hace que sean los cores **1** y **4** (de distintos dímeros) los primeros en entrar a interactuar con el ADN, a partir de este momento el ensamblaje del tetrámero es cooperativo, es decir los cores “móviles” (**2** y **3**) pueden interactuar con uno u otro de los cores fijos (**1** y **4**) para la formación final del tetrámero (*Figura.49.b*).

El tetrámero cerrado se forma cuando los cuatro cores interactúan entre sí y a su vez con el ADN (*Figura.49.b*). Este hecho hace que en los mapas obtenidos anteriormente no se pudiese encajar perfectamente los dos cores (**2** y **3**) de la estructura atómica de Kitayner [Kitayner et al., 2006], ya que estos se mantienen móviles para poderse adaptar al ADN en la formación del tetrámero. Esta movilidad hace que al encajarlos en el mapa p53-ADN quedasen algunas de las regiones de los cores fuera del mapa 3D (*Figura.26.d-f*).

El haber trabajado con un ADN con simetría H14 (secuencia específica del gen *GADD45*) explica los resultados obtenidos en la clasificación.

En las clases I, II y IV se ha producido la interacción con el ADN por los cores **1** y **4** (*como propone el modelo de la Figura.49.b*), sin embargo los cores “móviles” **2** y **3** aún no están interactuando con el ADN. Al comparar cada una de estas clases entre sí se puede apreciar el ensamblaje cooperativo en la formación del tetrámero. En la clase I los dímeros están formados por los cores (1-2) y (3-4) mientras que en las clases II y IV los dímeros los forman los cores (1-3) y (2-4). En estas clases aún no está formado el tetrámero con el ADN, debido a que los cores están muy alejados entre sí, lo que hace imposible el poder encajar dentro de estos mapas las estructuras atómicas resueltas por rayos X.

En la clase III los cuatro cores adoptan una conformación cerrada interactuando entre sí y con el ADN. Las interacciones entre los core (dentro de cada dímero y entre dímeros) son estrechas y hay mucha superficie de contacto. En este mapa

es posible encajar la estructura atómica de los cuatro cores unidos al ADN resuelta por rayos X [kitayner et al., 2006]. En esta estructura el ADN esta formado por dos fragmentos, no es un continuo. Si el ADN fuese un continuo, al encajarlo en el mapa de la clase III, este debería de estar severamente distorsionado [Ma et al., 2007] (Figura.49.b). El hecho de no ver el ADN hace que no tengamos datos directos y no podamos modelar el ADN.

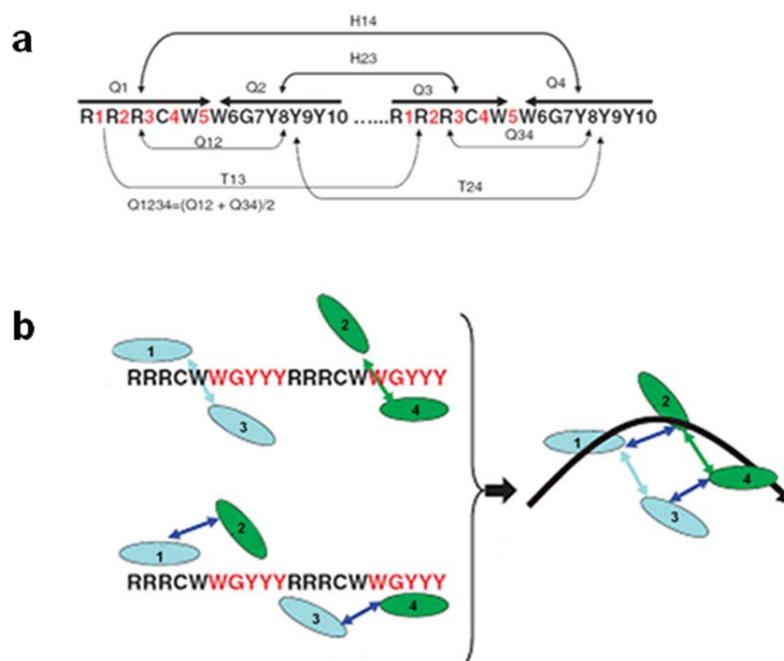


Figura.49. Simetría H14. (a) Tipos de simetrías (Q1234, T13, T24, H14, H23) que puede presentar la secuencia específica (RE) de ADN reconocida por p53. (b) Esquema del mecanismo cooperativo en la formación del tetrámero cooperativo ante un ADN con simetría H14. Los cores 1 y 4 se mantienen unidos al ADN mientras el 2 y el 3 se mueven hasta formar una estructura cerrada, momento en el que los cuatro cores interaccionan entre sí y con un ADN doblado.

| Group | P53REs | Q1234 | T13 | T24 | H14 | H23 | K_D (nM) ^a |
|---------------------|--------------------|-------|------|------|------|------|-------------------------|
| Very strong binding | 14-3-3 σ s1 | 13 | 13.4 | 20.8 | 44.7 | 8.1 | 9.9 \pm 2.1 |
| | 14-3-3 σ s2 | 16.1 | 16.7 | 23.6 | 30.5 | 13.1 | 10 \pm 2.1 |
| | GADD45 | 18.1 | 21.7 | 20.1 | 24.5 | 15.4 | 7.7 \pm 1.2 |
| | Maspin | 18.9 | 18.4 | 22.2 | 22.4 | 18.1 | 9.2 \pm 1.7 |
| | MDM2 | 13.7 | 21.8 | 16.5 | 33.6 | 14.4 | 12.3 \pm 2.9 |
| | NOXA | 16.3 | 18.8 | 19.9 | 31.3 | 13.8 | 8.6 \pm 1.6 |
| | P21-5' site | 17.7 | 21.2 | 24.5 | 20 | 16.6 | 4.9 \pm 0.6 |
| | P21-3' site | 16.3 | 24.3 | 15.8 | 30.9 | 12.8 | 12 \pm 7.0 |
| | P53AIP1 | 16.5 | 17.8 | 27.1 | 21.5 | 17 | 11.0 \pm 1.8 |
| | P53R2 | 15.9 | 16 | 21.6 | 33.1 | 13.5 | 5.7 \pm 0.8 |
| | PCNA | 17.1 | 26.2 | 20.2 | 21.5 | 15.1 | 6.6 \pm 1.4 |
| | PIDD | 19 | 17.2 | 23.9 | 19.8 | 20.1 | 14.7 \pm 5.0 |
| | PUMA-BS2 | 14.4 | 14.7 | 24.5 | 33.3 | 13.1 | 7.1 \pm 1.8 |
| | Cyclin G | 14.5 | 15.4 | 20.8 | 37.3 | 12 | 7.8 \pm 1.6 |
| | Group average | 16.3 | 18.8 | 21.6 | 28.9 | 14.5 | 9.1 |
| Strong binding | KAI 1 | 18.9 | 19 | 22 | 24.7 | 15.4 | 33 \pm 10.9 |
| | Group average | 18.9 | 19 | 22 | 24.7 | 15.4 | 33 \pm 10.9 |
| Weak binding | Bax | 18.9 | 18.1 | 22.1 | 19 | 21.8 | 73 \pm 33 |
| | IGF-BP3 box A | 18.6 | 18.8 | 21.8 | 23 | 17.7 | 81 \pm 25 |
| | P53DINP1 | 17 | 26.2 | 18.1 | 25.6 | 13 | 98 \pm 23 |
| | Group average | 18.2 | 21 | 20.7 | 22.5 | 17.5 | 84 |
| Very weak binding | P2XM | 18.7 | 21.3 | 19 | 24.3 | 16.7 | 221 \pm 41 |
| | PUMA-BS1 | 18.2 | 15.9 | 26.4 | 12.6 | 26.9 | 260 \pm 50 |
| | Group average | 18.5 | 18.6 | 22.7 | 18.5 | 21.8 | 241 |

Tabla.4. Tabla de probabilidad del tipo de simetría para algunos de los genes que regula p53 y valores de estabilidad (K_D en nM) de la formación del complejo (p53-ADN)

D.4 Papel regulador del dominio C-terminal en la formación del tetrámero p53 - ADN

La proteína p53 tiene dos sitios de unión de ADN, el dominio central (conservado evolutivamente) y el dominio básico C-terminal. En experimentos *in vitro* se ha visto que parte de la regulación llevada por parte del dominio C-terminal es el resultado de la competición entre la unión de forma inespecífica (por parte del C-terminal) y específica (por parte del core) y el ADN [Weinberg et al., 2004]. Según el modelo propuesto en otros trabajos, [Sauer et al., 2008; Liu et al., 2006], el dominio básico C-terminal hace que p53 pueda interactuar con secuencias no específicas de ADN, permitiendo así su deslizamiento “sliding” a lo largo del ADN para facilitar la búsqueda de la secuencia específica (RE) reconocida por el dominio central (*Figura.50.a*). De esta manera el dominio C-terminal sería capaz de regular la unión y actividad del dominio central.

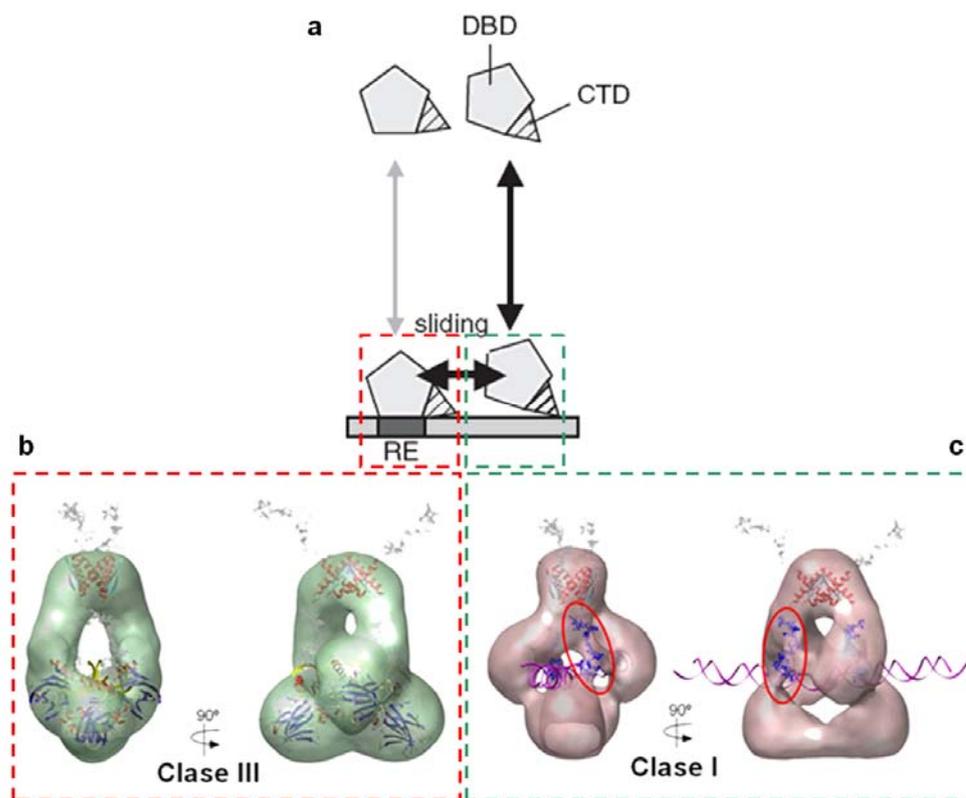


Figura.50. Papel del dominio C-terminal en los mapas 3D de las clases I y III del complejo p53-ADN (de 60pb) (a) Modelo del mecanismo de unión de p53 a secuencias no específicas y específicas (RE) de ADN. Mediante el dominio C-terminal (CTD), p53 es capaz de interactuar con secuencias no específicas y deslizarse “sliding” a lo largo del ADN hasta unirse por los cores (DBD) a la secuencia específica (RE). Datos tomados de Sauer et al., 2008. (b) Mapa de la clase III con los cuatro cores, el ADN, el dominio de tetramerización y los cuatro dominios C-terminales encajados. En este caso no hay densidad en el mapa para ningún C-terminal. (c) Mapa de la clase I con el ADN, el dominio de tetramerización y los cuatro dominios C-terminales encajados. En este caso no hay forma de encajar ninguna de las estructuras cristalinas resueltas de los cores, ya que al ser una estructura abierta, estos no llegan a interactuar entre si. En este clase si hay densidad en el mapa para dos de los dominios C-terminal (círculo rojo) que parecen estar interactuando con el ADN.

En el estudio estructural realizado, las clases abiertas (I, II y IV) son compatibles con estructuras que podrían deslizarse sobre el ADN. Debido a que no existe una interacción fuerte entre los cores y el ADN, es posible que lo estén rodeando permitiendo así su deslizamiento.

En el mapa de la clase I se ven cuatro conexiones entre el dominio de tetramerización y los cuatro cores (mientras que en la clase III solo hay dos) (*Figura.50.b-c*). El cambio conformacional hace que en el mapa de la clase I se vean dos regiones más que antes no veíamos. Esto nos hacen sospechar que se trata de los dominios C-terminales que interaccionan con el ADN de forma inespecífica (*Figura.50.c*). En este caso el ADN estaría sujeto por ambos dominios C-terminal facilitando así su deslizamiento como un “modelo de monorraíl”. Este modelo le proporciona a p53 la capacidad estructural para su difusión lineal a lo largo del ADN [McKinney et al., 2004], hasta localizar la secuencia específica (RE) de ADN con la que interacciona el dominio central.

La activación de p53 ante la acción de agentes que provocan la modificación postraduccional de algunos de los aminoácidos cargados (sobre todo lisinas) del dominio C-terminal, hace que disminuya su afinidad por el ADN. Esta alteración obstacularizaría la difusión lineal de p53 a lo largo del ADN favoreciéndose la unión del core a secuencias específicas de ADN (clase III). En nuestra interpretación de la estructura que presenta la clase III la interacción es de forma específica entre la secuencia específica del ADN y los cuatro cores de p53. En este caso al encajar la estructura de los cuatro dominios C-terminales estos no se interponen en la interacción del ADN por el dominio central (*Figura.50.b*).

El cambio conformacional que experimenta p53 inducido por el ADN, se ha estudiado por SM-FRET (Transmisión de energía de resonancia por fluorescencia de una molécula individual). Los estudios por SM-FRET son útiles para estudiar si una molécula es capaz de adoptar múltiples conformaciones. Para estos experimentos se marcó con Alexa Fluor 546 y 647 las cisteínas 292 (del core) y 372 (del C-terminal) de p53. Los datos indican un acercamiento del C-terminal al core cuando hay ADN. De los resultados obtenidos por la combinación de la microscopía electrónica y SM-FRET interpretamos que el C-terminal, en presencia de ADN, interacciona con el ADN a la vez que se aproxima a la región de los core. Lo que podemos concluir es que existe una variedad de conformaciones en equilibrio, unas en las que los cores no forman un tetrámero cerrado (clases I, II y IV), y otras en las que sí (clase III). Las clases más

abiertas son de mayor tamaño y candidatas a deslizamiento, la cerrada no sabemos si está unida al ADN de forma específica (a nivel atómico) o no, pero es compatible con un reconocimiento específico del core [Kitayner et al., 2006].

D.5 Estudio de la influencia del tamaño del ADN en la formación del complejo con p53

Con la intención de reducir la probabilidad de encontrar a p53 en una conformación estructural de deslizamiento sobre el ADN (estructura abierta) y favorecer, en la medida de lo posible, una conformación de interacción entre la secuencia específica y los cores (estructura cerrada), se realizó el mismo tipo de experimento que en el caso anterior pero con una muestra de ADN de 44 pares de bases, con la misma secuencia específica de unión del gen *GADD45*. El hecho de utilizar un ADN de 44 pares de bases impide espacialmente la interacción de p53 solo con ADN de secuencia no específica, ya que por tamaño necesitaría interactuar con parte de la específica. En este caso, los 12 pares de bases de secuencia no específica localizados a ambos lados de la específica, creemos que no son suficientes para el deslizamiento y por tanto cabría esperar una menor variabilidad conformacional. (Figura.51.d-f). El haber utilizado un ADN de 44pb y no uno menor se debe a que se pretendía que el ADN sobresaliese de la molécula y así poderlo ver por microscopía electrónica.

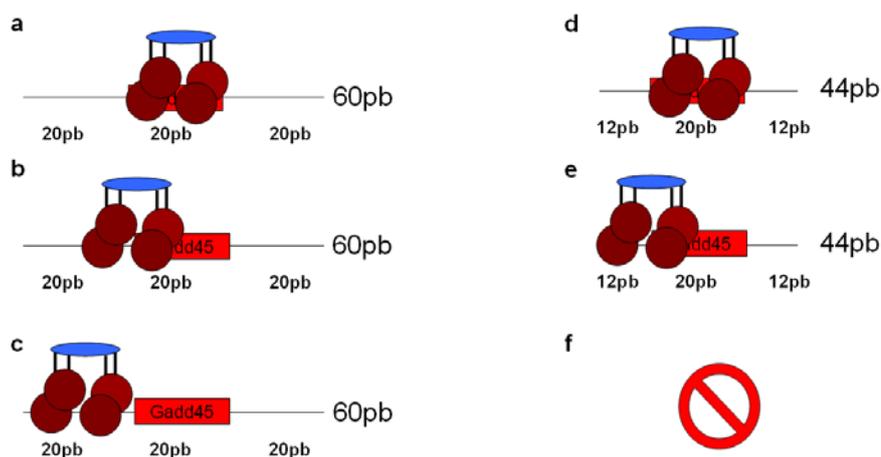


Figura.51. Esquema comparativo de las distintas conformaciones posibles que puede adoptar p53 en su interacción con dos muestras de ADN de distinta longitud (60 y 44 pb) con la misma secuencia específica del gen *GADD45*. **(a-d)** p53 unido directamente a la secuencia específica del gen *GADD45* en un ADN de 60pb (a) y 44pb (d). **(b-e)** p53 unido por dos cores a la secuencia específica y por los otros dos a la no específica en un ADN de 60 pb (b) y 44pb (e). En estas condiciones no se produce deslizamiento sobre el ADN. **(c)** p53 unido solo a la secuencia no específica. En estas condiciones se produce el deslizamiento a lo largo del ADN. Esta situación es solo posible en el caso de un ADN con más de 20 pb de secuencia no específica. **(f)** En el caso de un ADN de 44 pb no es posible espacialmente que p53 pueda interactuar con los 12 pb de secuencia no específica.

De igual modo, los ensayos de microscopía se realizaron tras el gradiente de densidad/fijación, utilizándose solo las fracciones que presentaron complejo con el ADN de 44 pares de bases en un estado de oligomerización tetramérico. El mapa resultante muestra una estructura compacta donde los cuatro cores (1, 2, 3 y 4) se localizan próximos e interaccionando entre sí (*Figura.52.a-d*).

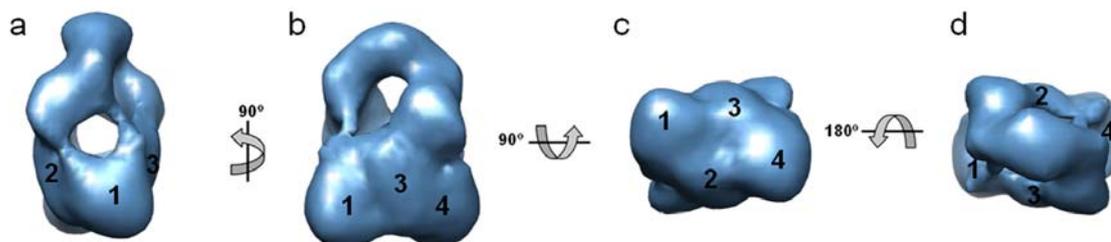


Figura.52. Mapa de microscopía electrónica de p53-ADN de 44 pares de bases tras el método GRAFIX. (a-d) Diferentes vistas del mapa 3D. La masa de arriba corresponde al dominio de tetramerización y cuatro masas marcadas (1, 2, 3 y 4) de abajo corresponden a los cuatro cores que unen el ADN.

Se estudió los cambios conformacionales que experimenta p53 en su interacción con el ADN de 44 pares de bases mediante la herramienta de clasificación *ML3D*. Los resultados obtenidos muestran que al dividir el conjunto de imágenes en tres clases, una de ellas se queda casi vacía (clase I), mientras la mayor parte de las imágenes van a parar a la clase II (aproximadamente el 70%). En ambos casos, tanto la clase II como la III, muestran una estructura cerrada, con los cuatro cores (1, 2, 3 y 4) en la misma posición (*Figura.53*).

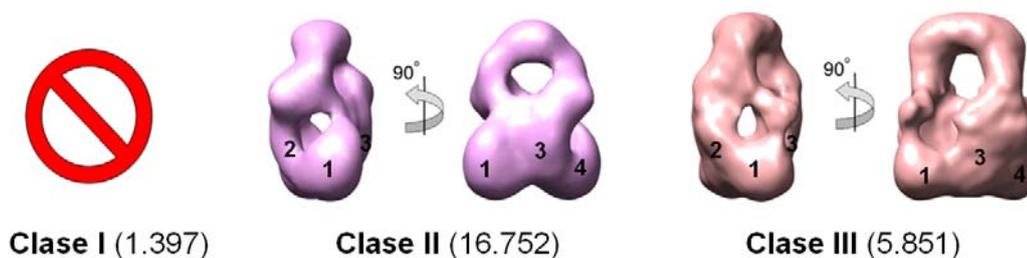


Figura.53. Mapas resultantes de la clasificación por *ML3D* del complejo p53-ADN (de 44 pb) tras el método Grafix. Con los números (1, 2, 3 y 4) están marcados cada uno de los cuatro cores. Entre paréntesis está indicado en número de partículas totales que tras la clasificación por *ML3D*, entran en cada una de las clases (I-III).

Los resultados obtenidos al trabajar con un ADN de 44 pares de bases, indican que, al no disponer de segmentos de secuencia no específica lo suficientemente largos (solo 12 pares de bases), hace que se impida el deslizamiento y se favorece la

interacción directa con la secuencia específica. Los mapas obtenidos en todos los casos son de estructuras cerradas, han desaparecido las formas abiertas. Las estructuras obtenidas son muy parecidas a la clase III, obtenida de la interacción de p53 con el ADN de 60pb (*Figura.33.b*). Solo en una de ellas (clase III (ADN 44pb)) se puede encajar la estructura atómica del tetrámero formado por los cuatro cores y el ADN [kitayner et al., 2006]. Esto sugiere que a pesar de tener los cores compactos es posible que la unión no sea específica (es lo que ocurre en la clase II (ADN 44pb)). Por tanto hay que distinguir entre estructura cerrada (compacta) y abierta e interacción específica e inespecífica.

En el caso de la clase II (ADN 44pb) habría una estructura compacta con unión no específica (al no ser compatible con la estructura de Kitayner) (*Figura.42.b*).

En el caso de la clase III (ADN 44pb) habría también una estructura compacta, pero en este caso la unión es específica (encaja perfectamente la estructura de Kitayner) (*Figura.42.d*).

Los resultados obtenidos nos hacen suponer que la dirección del ADN con respecto al dominio de tetramerización también juega un papel importante en cuanto al tipo de unión (específica o inespecífica). De tal forma que en los mapas con el dominio de tetramerización paralelo al ADN la unión es inespecífica (clase I (ADN 60pb) y clase II (ADN 44pb)). Mientras que cuando entre ambos forman un ángulo aproximado de 50° la unión es específica (clase III (ADN 60pb) y clase II (ADN 44pb)) (*Tabla.5*).

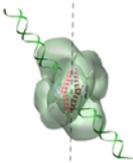
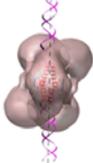
| | | |
|---|---|---|
|  <p>Clase III ADN 60pb</p> |  <p>Clase III ADN 44pb</p> | <p>Unión específica core-ADN</p> |
|  <p>Clase I ADN 60pb</p> |  <p>Clase II ADN 44pb</p> | |

Tabla.5. Unión específica e inespecífica entre p53 y ADN. El dominio de tetramerización de p53 adopta distintas orientaciones respecto al ADN. La unión es específica cuando existe un giro entre el ADN con respecto al dominio de tetramerización (aproximadamente 50°). La unión es inespecífica cuando el ADN corre paralelo al dominio de tetramerización.

Al comparar los resultados obtenidos al trabajar con ADN de 60 y 44 pares de bases, podemos sugerir que con el ADN corto (44pb) el deslizamiento de p53 sobre el ADN (sliding) se ha eliminado prácticamente, ya que no hay ninguna estructura abierta y compatible con esa función.

En ambos casos estudiados, la población que encaja con la estructura atómica de Kitayner (unión específica de ADN), es de aproximadamente el 25% del total. Es decir, que el equilibrio entre unión específica y no específica parece que no cambia (25% unión específica frente al 75% de unión inespecífica). Dentro del 75% restante cabe la posibilidad de que se produzca, o no, deslizamiento sobre el ADN.

D.6 Papel del dominio C-terminal cuando no hay deslizamiento sobre el ADN

En los mapas obtenidos se aprecia una densidad entre el dominio de tetramerización y los cuatro cores que se puede atribuir al C-terminal. Nuestra interpretación es que pueden interaccionar simultáneamente con el ADN los cuatro cores y los C-terminal (*Figura.54.b*). Los experimentos de SM-FRET se realizaron para descartar que esas densidades sean parte de las conexiones entre el dominio de tetramerización y los cores. De los resultados obtenidos por SM-FRET interpretamos que el C-terminal, en presencia de ADN, es capaz de interaccionar con el ADN a la vez que se aproxima a la región de los core.

Experimentos *in vitro* muestran que, ante la necesidad de una respuesta rápida por parte de p53, se produce su interacción directa con la secuencia específica (RE) del ADN (*Figura.54.a*). En esta situación, p53 estaría regulado negativamente por la acción del C-terminal. Ante la acción de señales externas se produce la modificación postraduccional del C-terminal, p53 se activa e induce la transcripción.

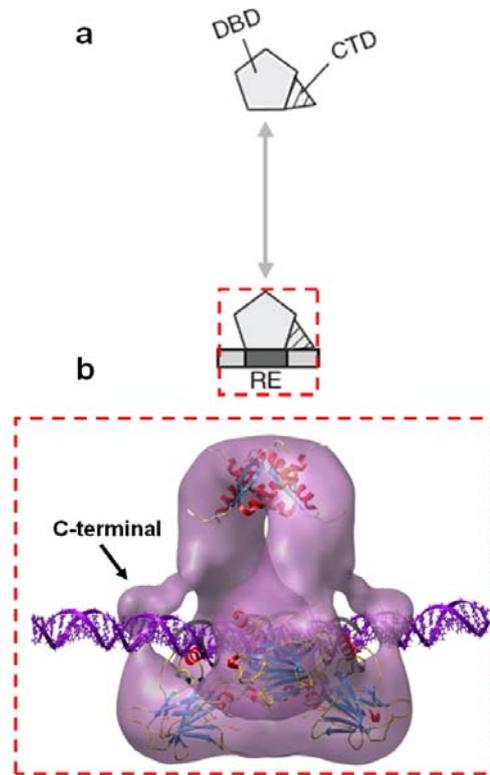


Figura.54. Unión directa de p53 a una secuencia específica de ADN. Un ADN corto hace que el tetramero se vea obligado a unirse directamente a la secuencia específica **(a)** Modelo del mecanismo de unión directa de p53 a un RE. Datos tomados de Sauer et al., 2008. **(b)** Mapa de microscopía electrónica de la clase III de p53 unido de forma específica a ADN de 44 pares de bases. El mapa permite encajar el dominio de tetramerización y los cuatro cores unidos al ADN. En este caso hay densidad en el mapa para dos de los dominios C-terminal que parecen estar interaccionando con el ADN.

En definitiva, en función del estado de la cromatina en el núcleo, p53 puede tener disponibles segmentos largos de ADN y deslizarse en busca de las secuencias específicas (RE) (situación que nos encontramos al trabajar con ADN de 60pb). Pero también puede tener localmente segmentos cortos que contengan RE con los que p53 interactuará de una manera más directa esperando a que su modificación le permita producir una respuesta rápida de inducción de la transcripción (situación que nos encontramos al trabajar con ADN de 44pb). En ambos casos el dominio C-terminal regula que p53 pueda estructuralmente formar un tetramero funcional, capaz de inducir la transcripción de genes (*Figura.55*).

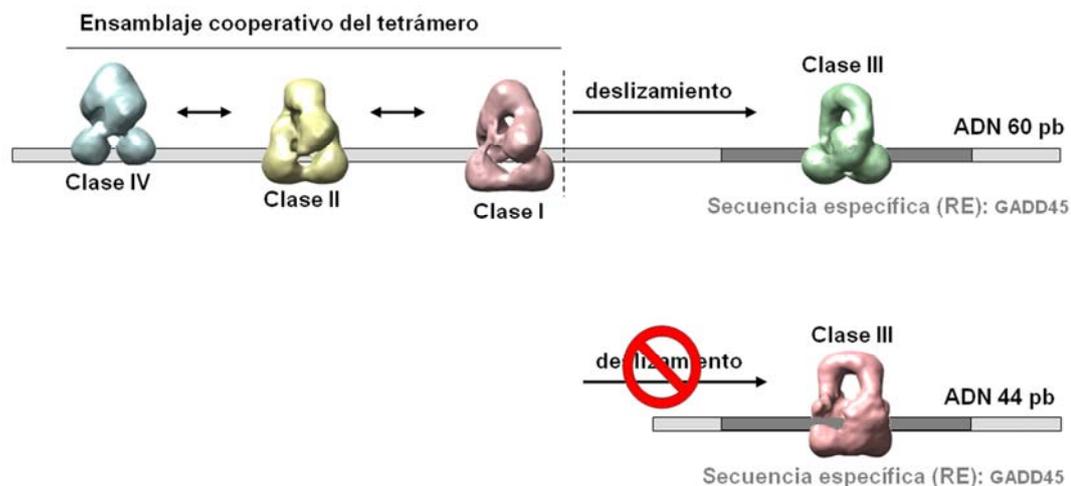


Figura.55. Esquema comparativo de las distintas conformaciones posibles que adopta p53 en su interacción con ADN de 60 y 44 pb. En el caso de un ADN de 60 pb, p53 puede unirse y deslizarse hasta localizar la secuencia específica. En el caso de un ADN de 44 pb, al no haber espacio suficiente para el deslizamiento p53 ha de interactuar directamente con la secuencia específica.

D.7 Implicaciones Biológicas

La flexibilidad y variabilidad que presenta la proteína p53, lleva implícito multitud de implicaciones biológicas.

Para poder ejercer su papel en el control del ciclo celular, como factor de transcripción, es necesaria dicha flexibilidad que le permite a la molécula adaptarse a la gran variedad de secuencias consenso dentro del genoma.

Además, los modelos planteados en este trabajo proponen dos modos de unión de p53 con el ADN, por el dominio central y por el dominio C-terminal, lo que permite a la proteína ejercer multitud de diferentes interacciones con el ADN genómico.

El hecho de que el dominio C-terminal pueda interactuar con el ADN hace que p53 pueda realizar un deslizamiento en la búsqueda de secuencias específicas localizadas en la región promotora de algunos de los genes que regula. Cabe la posibilidad que p53 pueda llevar a cabo una activación (deslizamiento hasta unión del core con la secuencia específica de ADN y transcripción de genes) o desactivación (una vez unido el core al ADN, el C-terminal reinicia el deslizamiento soltando a p53 de la secuencia específica y parando o desactivando la transcripción del gen).

Este mecanismo de deslizamiento sobre el ADN para la activación y desactivación de genes se ha visto también en otras proteínas que interaccionan con ADN (ejemplo: BamHI, que mantiene una disposición estructural abierta mientras está realizando la búsqueda de la secuencia específica y una vez localizada, adopta un cambio conformacional pasando a tener una estructura cerrada o compacta alrededor del ADN) [Viadiu et al., 2000].

El hecho de que el deslizamiento este realizado por el dominio C-terminal hace que las modificaciones que se produzcan sobre este dominio y/o la multitud de factores o ligandos que se unan al C-Terminal puedan actuar como reguladores de la función de p53. De igual forma, las modificaciones que se ocasionan en el dominio C-terminal pueden influir en el reclutamiento de otros factores (ejemplo: acetilasas de histonas, donde p53, como factor de transcripción, para poderse unir a la secuencia específica de ADN necesita que se produzcan una serie de modificaciones en las proteínas histonas, alrededor de las cuales se enrolla el ADN. Las enzimas acetilasas de histonas alteran la estructura de la cromatina (aflojando la interacción ADN-histona) permitiendo que los factores de transcripción (p53, entre otros) se unan al ADN [Gu et al., 1997; Struhl et al., 1998].

Cabe destacar la importancia del dominio N-terminal en la respuesta y funcionalidad de p53 como supresor tumoral y factor de transcripción. La flexibilidad que presenta la molécula, hace que los core se muevan, lo que implica a su vez el movimiento de los N-terminal. La presencia de un dominio N-terminal extendido y flexible en el modelo es consonante con su función biológica. Estos juegan un papel crucial en su unión a otros cofactores necesarios para la regulación de la transcripción de genes. El hecho de estar expuestos en todo momento hace que puedan interaccionar con dichos cofactores (esto no ocurre en los datos de Okovorov, donde los N-terminal están interaccionando con las α -hélices del dominio de tetramerización, [Okovorov et al., 2006]. El dominio N-terminal de p53 es objeto de múltiples modificaciones postraduccionales que potencialmente regulan la actividad transcripcional de p53 [Banin et al., 1998; Shieh et al., 1997], uniéndose a una variedad de proteínas que promueven la modulación de la actividad de p53, como MDM2 [Kussie et al., 1996], p300 [Grossman et al., 2001], varios activadores transcripcionales del complejo TFIID [Lu et al., 1995; Naar et al., 2001; Thut et al., 1995] y a RPA [Bochkareva et al., 2005].

En la unión con otras proteínas, intervienen pequeñas regiones con estructura helicoidal dentro de un intrínsecamente desordenado dominio N-terminal [Kussie et al., 1996; Bochkareva et al., 2005], fenómeno que también se observa en otros factores de transcripción [Bowers et al., 1999; Graham et al., 2000; Spolar et al., 1994].

La conformación abierta de p53 libre presentada en este trabajo, con los dos pares de dominios centrales separados, permite a proteínas reguladoras del tipo Bcl-xL [Chipuk et al., 2005], Bcl2 [Tomita et al., 2006] o 53BP2 [Gorina et al., 1996; Tidow et al., 2006] unirse a los cores. Esta conformación abierta también permite el acceso al ADN.

Por lo tanto, p53 es una proteína multifuncional y su habilidad de regular múltiples procesos esta presuntamente basada en su plasticidad conformacional, que permite a la proteína asumir distintas conformaciones (abiertas y cerradas), interaccionar con diversas secuencias de ADN (específicas y no específicas) y con multitud de factores y proteínas que regulan su actividad.

Conclusiones

1. La información aportada por la combinación de la Microscopía Electrónica de Transmisión, Small-angle X-ray scattering y la Resonancia Magnética Nuclear han permitido resolver el problema de la heterogeneidad estructural de p53.
2. Los resultados revelan la existencia de un equilibrio entre dos conformaciones (abierta y cerrada) en el caso de la proteína libre, equilibrio que se desplaza a la forma cerrada cuando p53 se une a la secuencia específica de ADN.
3. En los modelos tridimensionales obtenidos mediante microscopia electrónica se han podido encajar las estructuras tanto del dominio central o de unión al ADN, como del dominio de tetramerización resueltas por cristalografía de rayos X y/o RMN.
4. El dominio extendido N-Terminal esta completamente expuesto en la estructura cuaternaria. Esta conformación le permite ser modificado fácilmente a nivel postraduccional y a su vez unirse a varias proteínas que regulan su actividad, como Mdm2, p300, y RPA.
5. Los resultados obtenidos en este trabajo constituyen la primera información estructural disponible acerca del complejo tetramérico completo entre p53 y el ADN, y sugieren la unión simultánea de dominios de reconocimiento específico y no específico.
6. El comportamiento dinámico de los tetrámeros sobre el ADN parece estar relacionado con la habilidad de p53 de deslizarse sobre el ADN para localizar los sitios específicos de reconocimiento, permitiendo a p53 ejercer su papel como factor de transcripción.

Bibliografía

Ahn, J, Prives, C (2001). The C-terminus of p53: the more you learn the less you know. *Nat. Struct. Biol.*, 8, 9:730-2

Albrechtsen, N, Dornreiter, I, Grosse, F, Kim, E, Wiesmüller, L, Deppert, W (1999). Maintenance of genomic integrity by p53: complementary roles for activated and non-activated p53. *Oncogene*, 18, 53:7706-17.

Ang, HC, Joerger, AC, Mayer, S, Fersht, AR (2006). Effects of common cancer mutations on stability and DNA binding of full-length p53 compared with isolated core domains. *J. Biol. Chem.*, 281, 31:21934-41.

Appella, E, Anderson, CW (2001.a). Post-translational modifications and activation of p53 by genotoxic stresses. *Eur. J. Biochem.*, 268, 10:2764-72.

Appella, E (2001.b). Modulation of p53 function in cellular regulation. *Eur. J. Biochem.*, 268, 10:2763.

Ayed, A, Mulder, FA, Yi, GS, Lu, Y, Kay, LE, Arrowsmith, CH (2001). Latent and active p53 are identical in conformation. *Nat. Struct. Biol.*, 8, 9:756-60.

Bakalkin, G, Selivanova, G, Yakovleva, T, Kiseleva, E, Kashuba, E, Magnusson, KP, Szekely, L, Klein, G, Terenius, L, Wiman, KG (1995). p53 binds single-stranded DNA ends through the C-terminal domain and internal DNA segments via the middle domain. *Nucleic Acids Res.*, 23, 3:362-9.

Baker S.J., Fearon E.R., Nigro J.M., Hamilton S.R., Preisinger A.C., Jessup J.M., vanTuinen P., Ledbetter D.H., Barker D.F., Nakamura Y (1989). Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science.*, 244: 217-221.

Banin, S, Moyal, L, Shieh, S, Taya, Y, Anderson, CW, Chessa, L, Smorodinsky, NI, Prives, C, Reiss, Y, Shiloh, Y, Ziv, Y (1998). Enhanced phosphorylation of p53 by ATM in response to DNA damage. *Science*, 281, 5383:1674-7.

Baptiste, N, Friedlander, P, Chen, X, Prives, C (2002). The proline-rich domain of p53 is required for cooperation with anti-neoplastic agents to promote apoptosis of tumor cells. *Oncogene*, 21, 1:9-21.

Barak, Y, Juven, T, Haffner, R, Oren, M (1993). Mdm2 expression is induced by wild type p53 activity. *EMBO J.*, 12, 2:461-8.

Bargonetti, J, Manfredi, JJ, Chen, X, Marshak, DR, Prives, C (1993). A proteolytic fragment from the central region of p53 has marked sequence-specific DNA-binding activity when generated from wild-type but not from oncogenic mutant p53 protein. *Genes Dev.*, 7, 12B:2565-74.

Bayle, JH, Elenbaas, B, Levine, AJ (1995). The carboxyl-terminal domain of the p53 protein regulates sequence-specific DNA binding through its nonspecific nucleic acid-binding activity. *Proc. Natl. Acad. Sci. U.S.A.*, 92, 12:5729-33.

Bishop, JM (1991). Molecular themes in oncogenesis. *Cell*, 64, 2:235-48.

Bochkareva, E, Kaustov, L, Ayed, A, Yi, GS, Lu, Y, Pineda-Lucena, A, Liao, JC, Okorokov, AL, Milner, J, Arrowsmith, CH, Bochkarev, A (2005). Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A. *Proc. Natl. Acad. Sci. U.S.A.*, 102, 43:15412-7.

Bottcher, B, Wynne, SA, Crowther, R. A. (1997). Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature* 386, 88-91.

Bowers, PM, Schaufler, LE, Klevit, RE (1999). A folding transition and novel zinc finger accessory domain in the transcription factor ADR1. *Nat. Struct. Biol.*, 6, 5:478-85

Bullock, AN, Henckel, J, DeDecker, BS, Johnson, CM, Nikolova, PV, Proctor, MR, Lane, DP, Fersht, AR (1997). Thermodynamic stability of wild-type and mutant p53 core domain. *Proc. Natl. Acad. Sci. U.S.A.*, 94, 26:14338-42.

Bullock, AN, Henckel, J, Fersht, AR (2000). Quantitative analysis of residual folding and DNA binding in mutant p53 core domain: definition of mutant states for rescue in cancer therapy. *Oncogene*, 19, 10:1245-56.

Bullock, AN, Fersht, AR (2001). Rescuing the function of mutant p53. *Nat. Rev. Cancer*, 1, 1:68-76.

Candau, R, Scolnick, DM, Darpino, P, Ying, CY, Halazonetis, TD, Berger, SL (1997). Two tandem and independent sub-activation domains in the amino terminus of p53 require the adaptor complex for activity. *Oncogene*, 15, 7:807-16

Chang, J, Kim, DH, Lee, SW, Choi, KY, Sung, YC (1995). Transactivation ability of p53 transcriptional activation domain is directly related to the binding affinity to TATA-binding protein. *J. Biol. Chem.*, 270, 42:25014-9.

Chen, Y, Dey, R, Chen, L (2010). Crystal structure of the p53 core domain bound to a full consensus site as a self-assembled tetramer. *Structure.*, 18, 246–256

Chène, P (2001). The role of tetramerization in p53 function. *Oncogene*, 20, 21:2611-7.

Chipuk, JE, Bouchier-Hayes, L, Kuwana, T, Newmeyer, DD, Green, DR (2005). PUMA couples the nuclear and cytoplasmic proapoptotic function of p53. *Science*, 309, 5741:1732-5.

- Cho, Y., Gorina, S., Jeffrey, P. D. & Pavletich, N. P (1994). Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science* 265, 346-55
- Crowther, R.A., DeRosier, D.J., Klug, A. (1970) The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. *Proc. R. Soc. Lond.*, 317, 319-340.
- el-Deiry, WS, Kern, SE, Pietenpol, JA, Kinzler, KW, Vogelstein, B (1992). Definition of a consensus binding site for p53. *Nat. Genet.*, 1, 1:45-9.
- el-Deiry, WS, Tokino, T, Velculescu, VE, Levy, DB, Parsons, R, Trent, JM, Lin, D, Mercer, WE, Kinzler, KW, Vogelstein, B (1993). WAF1, a potential mediator of p53 tumor suppression. *Cell*, 75, 4:817-25.
- Finlay C.A., Hinds P.W., Levine A.J (1989). The p53 proto-oncogene can act as a suppressor of transformation. *Cell*, 57: 1083-1093.
- Frank, J. (1996) Three-dimensional electron microscopy of macromolecular assemblies. Academic Press. Ltd, London, UK.
- Friedlander, P, Legros, Y, Soussi, T, Prives, C (1996). Regulation of mutant p53 temperature-sensitive DNA binding. *J. Biol. Chem.*, 271, 41:25468-78.
- Gorina, S, Pavletich, NP (1996). Structure of the p53 tumor suppressor bound to the ankyrin and SH3 domains of 53BP2. *Science*, 274, 5289:1001-5.
- Graham, TA, Weaver, C, Mao, F, Kimelman, D, Xu, W (2000). Crystal structure of a beta-catenin/Tcf complex. *Cell*, 103, 6:885-96.
- Grossman, SR (2001). p300/CBP/p53 interaction and regulation of the p53 response. *Eur. J. Biochem.*, 268, 10:2773-8.
- Gu, W, Roeder, RG (1997). Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell*, 90, 4:595-606.
- Hainaut, P. & Wiman, K. G. e. 25 years of p53 research, Hainaut P, Wiman K.G. (eds), Springer. (2005).
- Hansen, S, Hupp, TR, Lane, DP (1996). Allosteric regulation of the thermostability and DNA binding activity of human p53 by specific interacting proteins. CRC Cell Transformation Group. *J. Biol. Chem.*, 271, 7:3917-24.
- Ho, W. C., Fitzgerald, M. X. & Marmorstein, R (2006). Structure of the p53 core domain dimer bound to DNA. *J Biol Chem* 281, 20494-502.
- Hollstein M., Sidransky D., Vogelstein B., Harris C.C (1991). p53 mutations in human cancers. *Science*, 253: 49-53.
- Hupp, TR., Meek, D.W., Midgley, C.A. and Lane, D.P (1992) Regulation of the specific DNA binding function of p53. *Cell*, 71, 875-886.

Hupp, TR, Lane, DP (1994.a). Regulation of the cryptic sequence-specific DNA-binding function of p53 by protein kinases. *Cold Spring Harb. Symp. Quant. Biol.*, 59:195-206.

Hupp, TR, Lane DP (1994.b). Allosteric activation of latent p53 tetramers. *Current Biol.*, 4: 865–875.

Hupp, TR, Sparks, A, Lane, DP (1995). Small peptides activate the latent sequence-specific DNA binding function of p53. *Cell*, 83, 2:237-45.

Jayaraman, J, Prives, C (1995). Activation of p53 sequence-specific DNA binding by short single strands of DNA requires the p53 C-terminus. *Cell*, 81, 7:1021-9.

Jayaraman, L, Prives, C (1999). Covalent and noncovalent modifiers of the p53 protein. *Cell. Mol. Life Sci.*, 55, 1:76-87.

Jeffrey, PD, Gorina, S, Pavletich, NP (1995). Crystal structure of the tetramerization domain of the p53 tumor suppressor at 1.7 angstroms. *Science*, 267, 5203:1498-502

Joerger, A. C., Allen, M. D. & Fersht, A. R. (2004). Crystal structure of a superstable mutant of human p53 core domain. Insights into the mechanism of rescuing oncogenic mutations. *J Biol Chem* 279, 1291-6.

Kastan M.B., Onyekwere O., Sidransky D., Vogelstein B., Craig R.W (1991). Participation of p53 protein in the cellular response to DNA damage. *Cancer Research*, 51: 6304-6311.

Kastner, B, Fischer, N, Golas, MM, Sander, B, Dube, P, Boehringer, D, Hartmuth, K, Deckert, J, Hauer, F, Wolf, E, Uchtenhagen, H, Urlaub, H, Herzog, F, Peters, JM, Poerschke, D, Lührmann, R, Stark, H (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods*, 5, 1:53-5.

Kern, SE, Kinzler, KW, Bruskin, A, Jarosz, D, Friedman, P, Prives, C, Vogelstein, B (1991). Identification of p53 as a sequence-specific DNA-binding protein. *Science*, 252, 5013:1708-11.

Kitayner, M, Rozenberg, H, Kessler, N, Rabinovich, D, Shaulov, L, Haran, TE, Shakked, Z (2006). Structural basis of DNA recognition by p53 tetramers. *Mol. Cell*, 22, 6:741-53.

Kluck, RM, Bossy-Wetzel, E, Green, DR, Newmeyer, DD (1997). The release of cytochrome c from mitochondria: a primary site for Bcl-2 regulation of apoptosis. *Science*, 275, 5303:1132-6.

Kussie, PH, Gorina, S, Marechal, V, Elenbaas, B, Moreau, J, Levine, AJ, Pavletich, NP (1996). Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science*, 274, 5289:948-53.

Lane D.P., Crawford L.V (1979) T antigen is bound to a host protein in SV40-transformed cells. *Nature*, 278: 261-263.

- Lane D.P (1992) p53, guardian of the genome. *Nature*; 358:15-16.
- Latchman, DS (1997). Transcription factors: an overview. *Int. J. Biochem. Cell Biol.*, 29, 12:1305-12.
- Lee, S, Elenbaas, B, Levine, A, Griffith, J (1995). p53 and its 14 kDa C-terminal domain recognize primary DNA damage in the form of insertion/deletion mismatches. *Cell*, 81, 7:1013-20.
- Lee, S, Cavallo, L, Griffith, J (1997). Human p53 binds Holliday junctions strongly and facilitates their cleavage. *J. Biol. Chem.*, 272, 11:7532-9.
- Levine A.J., Momand J., Finlay C.A (1991). The p53 tumour suppressor gene. *Nature*; 351:453-456.
- Levine A.J (1997). p53, the cellular gatekeeper for growth and division. *Cell*; 88:323-331.
- Libermann, TA, Zerbini, LF (2006). Targeting transcription factors for cancer gene therapy. *Curr Gene Ther*, 6, 1:17-33.
- Linzer, DI, Levine, AJ (1979). Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. *Cell*, 17, 1:43-52.
- Liu, WF, Yu, SS, Chen, GJ, Li, YZ (2006). DNA damage checkpoint, damage repair, and genome stability. *Yi Chuan Xue Bao*, 33, 5:381-90.
- Lu, H, Levine, AJ (1995). Human TAFII31 protein is a transcriptional coactivator of the p53 protein. *Proc. Natl. Acad. Sci. U.S.A.*, 92, 11:5154-8.
- Ludtke, SJ, Baldwin, PR, Chiu, W (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.*, 128, 1:82-97.
- Ma, B, Levine, AJ (2007). Probing potential binding modes of the p53 tetramer to DNA based on the symmetries encoded in p53 response elements. *Nucleic Acids Res.*, 35, 22:7733-47.
- Malecka, KA, Ho, WC, Marmorstein, R (2009). Crystal structure of a p53 core tetramer bound to DNA. *Oncogene*, 28, 3:325-33.
- Marabini, R, Herman, GT, Carazo, JM (1998). 3D reconstruction in electron microscopy using ART with smooth spherically symmetric volume elements (blobs). *Ultramicroscopy*, 72, 1-2:53-65.
- Marabini, R., Masegosa, I.M., San Martín, M.C., Marco, S., Fernandez, J.J., de la Fraga, L.G., Vaquerizo, C. and Carazo, J.M. (1996) Xmipp: An Image Processing Package for Electron Microscopy. *J Struct Biol*, 116, 237-240.

Mazur, SJ, Sakaguchi, K, Appella, E, Wang, XW, Harris, CC, Bohr, VA (1999). Preferential binding of tumor suppressor p53 to positively or negatively supercoiled DNA involves the C-terminal domain. *J. Mol. Biol.*, 292, 2:241-9.

Melero, R, Tidow, H, Mylonas, E, Freund, SM, Grossmann, JG, Carazo, JM, Svergun, DI, Valle, M, Fersht, AR (2007). Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex. *Proc. Natl. Acad. Sci. U.S.A.*, 104, 30:12324-9.

McKinney, K, Mattia, M, Gottifredi, V, Prives, C (2004). p53 linear diffusion along DNA requires its C terminus. *Mol. Cell*, 16, 3:413-24.

Miyashita, T, Reed, JC (1995). Tumor suppressor p53 is a direct transcriptional activator of the human bax gene. *Cell*, 80, 2:293-9.

Mujtaba, S, He, Y, Zeng, L, Yan, S, Plotnikova, O, Sachchidanand, R, Sanchez, NJ, Zeleznik-Le, Z, Ronai, MM, Zhou, (2004). Structural mechanism of the bromodomain of the coactivator CBP in p53 transcriptional activation. *Mol. Cell*, 13, 2:251-63.

Näär, AM, Lemon, BD, Tjian, R (2001). Transcriptional coactivator complexes. *Annu. Rev. Biochem.*, 70:475-501.

Nikolova, PV, Henckel, J, Lane, DP, Fersht, AR (1998). Semirational design of active tumor suppressor p53 DNA binding domain with enhanced stability. *Proc. Natl. Acad. Sci. U.S.A.*, 95, 25:14675-80.

Nurse, P, Masui, Y, Hartwell, L (1998). Understanding the cell cycle. *Nat. Med.*, 4, 10:1103-6.

Oda, E, Ohki, R, Murasawa, H, Nemoto, J, Shibue, T, Yamashita, T, Tokino, T, Taniguchi, T, Tanaka, N (2000). Noxa, a BH3-only member of the Bcl-2 family and candidate mediator of p53-induced apoptosis. *Science*, 288, 5468:1053-8.

Okamoto K., Hongyun L., Jensen M.R., Zhang T., Taya Y., Thorgeirsson S.S., Prives C (2002). Cyclin G recruits PP2A to dephosphorylate Mdm2. *Mol Cell*, 9: 761-771.

Okorokov, AL, Sherman, MB, Plisson, C, Grinkevich, V, Sigmundsson, K, Selivanova, G, Milner, J, Orlova, EV (2006). The structure of p53 tumour suppressor protein reveals the basis for its functional plasticity. *EMBO J.*, 25, 21:5191-200.

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, 13:1605-12.

- Radermacher, M., Wagenknecht, T., Verschoor, A. and Frank, J. (1987) Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of *Escherichia coli*. *J Microsc*, 146 (Pt 2), 113-136.
- Radermacher, M. (1994) Three-dimensional reconstruction from random projections: orientational alignment via Radon transforms. *Ultramicroscopy*, 53, 121-136.
- Radon, J. (1917) Ueber die Bestimmung von Funktionen durch ihre Integralwerte laengs gewisser Mannigfaltigkeiten. *Berichte ueber die Verhandlungen der Koeniglich Saechsischen Gesellschaft der Wissenschaften zu Leipzig. Math. Phys., Klasse 69*, 262-277.
- Robles, AI, Bemmels, NA, Foraker, AB, Harris, CC (2001). APAF-1 is a transcriptional target of p53 in DNA damage-induced apoptosis. *Cancer Res.*, 61, 18:6660-4.
- Rosal, R, Pincus, MR, Brandt-Rauf, PW, Fine, RL, Michl, J, Wang, H (2004). NMR solution structure of a peptide from the mdm-2 binding domain of the p53 protein that is selectively cytotoxic to cancer cells. *Biochemistry*, 43, 7:1854-61.
- Rotter, V, Aloni-Grinstein, R, Schwartz, D, Elkind, NB, Simons, A, Wolkowicz, R, Lavigne, M, Beserman, P, Kapon, A, Goldfinger, N (1994). Does wild-type p53 play a role in normal cell differentiation?. *Semin. Cancer Biol.*, 5, 3:229-36.
- Sakaguchi, K, Sakamoto, H, Lewis, MS, Anderson, CW, Erickson, JW, Appella, E, Xie, D (1997). Phosphorylation of serine 392 stabilizes the tetramer formation of tumor suppressor protein p53. *Biochemistry*, 36, 33:10117-24.
- Sakamuro, D, Sabbatini, P, White, E, Prendergast, GC (1997). The polyproline region of p53 is required to activate apoptosis but not growth arrest. *Oncogene*, 15, 8:887-98.
- Sarnow P., Ho Y.S., Williams J., Levine A.J. (1982) Adenovirus E1b-58kd tumor antigen and SV40 large tumor antigen are physically associated with the same 54kd cellular protein in transformed cells. *Cell*, 28: 387-394.
- Sauer, M, Bretz, AC, Beinoraviciute-Kellner, R, Beitzinger, M, Burek, C, Rosenwald, A, Harms, GS, Stiewe, T (2008). C-terminal diversity within the p53 family accounts for differences in DNA binding and transcriptional activity. *Nucleic Acids Res.*, 36, 6:1900-12.
- Saxton, W.O. and Baumeister, W. (1982) The correlation averaging of a regularly arranged bacterial cell envelope protein. *J Microsc*, 127 (Pt 2), 127-138
- Scheres, SH, Valle, M, Nuñez, R, Sorzano, CO, Marabini, R, Herman, GT, Carazo, JM (2005.a). Maximum-likelihood multi-reference refinement for electron microscopy images. *J. Mol. Biol.*, 348, 1:139-49.

Scheres, SH, Valle, M, Carazo, JM (2005.b). Fast maximum-likelihood refinement of electron microscopy images. *Bioinformatics*, 21 Suppl 2:ii243-4.

Scheres, SH, Núñez-Ramírez, R, Gómez-Llorente, Y, San Martín, C, Eggermont, PP, Carazo, JM (2007.a). Modeling experimental image formation for likelihood-based classification of electron microscopy data. *Structure*, 15, 10:1167-77.

Scheres SH, Gao H, Valle M, Herman GT, Eggermont PP, Frank J, Carazo JM (2007.b) Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nat Methods* 4:27–29

Scheres, SH, Núñez-Ramírez, R, Sorzano, CO, Carazo, JM, Marabini, R (2008). Image processing for electron microscopy single-particle analysis using XMIPP. *Nat Protoc*, 3, 6:977-90.

Shieh, SY, Ikeda, M, Taya, Y, Prives, C (1997). DNA damage-induced phosphorylation of p53 alleviates inhibition by MDM2. *Cell*, 91, 3:325-34.

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.

Spolar, RS, Record, MT (1994). Coupling of local folding to site-specific binding of proteins to DNA. *Science*, 263, 5148:777-84.

Struhl, K (1998). Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev.*, 12, 5:599-606.

Takekawa M., Adachi M., Nakahata A., Nakayama I., Itoh F., Tsukauda H., Taya Y., Imai K (2000). p53-inducible wip1 phosphatase mediates a negative feedback regulation of p38 MAPKp53 signaling in response to UV radiation. *EMBO J*, 19: 6517-6526.

Tanaka, H, Arakawa, H, Yamaguchi, T, Shiraishi, K, Fukuda, S, Matsui, K, Takei, Y, Nakamura, Y (2000). A ribonucleotide reductase gene involved in a p53-dependent cell-cycle checkpoint for DNA damage. *Nature*, 404, 6773:42-9.

Thut, CJ, Chen, JL, Klemm, R, Tjian, R (1995). p53 transcriptional activation mediated by coactivators TAFII40 and TAFII60. *Science*, 267, 5194:100-4.

Tidow, H, Veprintsev, DB, Freund, SM, Fersht, AR (2006). Effects of oncogenic mutations and DNA response elements on the binding of p53 to p53-binding protein 2 (53BP2). *J. Biol. Chem.*, 281, 43:32526-33.

Tomita, Y, Marchenko, N, Erster, S, Nemajerova, A, Dehner, A, Klein, C, Pan, H, Kessler, H, Pancoska, P, Moll, UM (2006). WT p53, but not tumor-derived mutants, bind to Bcl2 via the DNA binding domain and induce mitochondrial permeabilization. *J. Biol. Chem.*, 281, 13:8600-6.

- Viadiu, H, Aggarwal, AK (2000). Structure of BamHI bound to nonspecific DNA: a model for DNA sliding. *Mol. Cell*, 5, 5:889-95.
- Wei, CL, Wu, Q, Vega, VB, Chiu, KP, Ng, P, Zhang, T, Shahab, A, Yong, HC, Fu, Y, Weng, Z, Liu, J, Zhao, XD, Chew, JL, Lee, YL, Kuznetsov, VA, Sung, WK, Miller, LD, Lim, B, Liu, ET, Yu, Q, Ng, HH, Ruan, Y (2006). A global map of p53 transcription-factor binding sites in the human genome. *Cell*, 124, 1:207-19.
- Weinberg, RA (1991). Tumor suppressor genes. *Science*, 254, 5035:1138-46.
- Weinberg, RL, Veprintsev, DB, Fersht, AR (2004). Cooperative binding of tetrameric p53 to DNA. *J. Mol. Biol.*, 341, 5:1145-59.
- Werness B.A., Levine A.J., Howley P.M (1990). Association of human papillomavirus types 16 and 18 E6 proteins with p53. *Science*, 248: 76-79.
- Williams, R.C. and Fisher, H.W (1970). Electron microscopy of tobacco mosaic virus under conditions of minimal beam exposure. *J Mol Biol* 52(1): 121-123.
- Wong, KB, DeDecker, BS, Freund, SM, Proctor, MR, Bycroft, M, Fersht, AR (1999). Hot-spot mutants of p53 core domain evince characteristic local structural changes. *Proc. Natl. Acad. Sci. U.S.A.*, 96, 15:8438-42.
- Xiong Y., Hannon G.J., Zhang H., Casso D., Kobayashi R., Beach D (1993). p21 is a universal inhibitor of cyclin kinases. *Nature*, 366: 701-704.
- Yakovleva, T, Pramanik, A, Kawasaki, T, Tan-No, K, Gileva, I, Lindegren, H, Langel, U, Ekstrom, TJ, Rigler, R, Terenius, L, Bakalkin, G (2001). p53 Latency. C-terminal domain prevents binding of p53 core to target but not to nonspecific DNA sequences. *J. Biol. Chem.*, 276, 19:15650-8.
- Yakovleva, T, Pramanik, A, Terenius, L, Ekström, TJ, Bakalkin, G (2002). p53 latency--out of the blind alley. *Trends Biochem. Sci.*, 27, 12:612-8.
- Yang, A, Kaghad, M, Caput, D, McKeon, F (2002). On the shoulders of giants: p63, p73 and the rise of p53. *Trends Genet.*, 18, 2:90-5.

Glosario

Listado de abreviaturas

2D: Bidimensional

3D: Tridimensional

ADN: Ácido desoxirribonucleico

AFM: Microscopio de fuerzas atómicas (Atomic Force Microscopy)

ARN: Ácido ribonucleico

°C: Grados Celsius

NaCl: Cloruro sódico

CCD: Charged Coupled Device

C-terminal: Dominio(s) C-terminal

N-terminal: Dominio(s) N-terminal

DTT: Ditiotreitól

EMSA: Electroforetic Movility Shift Assay

E.coli: Escherichia coli

g: Gramo

kV: Kilovoltios

MET: Microscopio electrónico de transmisión

ml: mililitro

mM: Milimolar

μM: Micromolar

nm: Nanómetro

pb: pares de bases

RE: Elementos de respuesta (Response Element)

r.p.m: Revoluciones por minuto

SDS: Dodecilsulfato sódico

Tris: Tris (hidroximetil) aminometano

WT: Wild Type, silvestre

WWW: World Wide Web

Direcciones Web:

(Proveedores de productos, software y bases de datos)

EMAN: <http://blake.bcm.tmc.edu/eman/>

Fei Company: <http://www.fei.com/>

Gatan, Inc: <http://www.gatan.com/>

General Electric Healthcare; GE: <http://gelifesciences.com>

Invitrogen Corporation: <http://invitrogen.com>

Jeol Ltd.: <http://jeol.com>

Kodak; Eastman Kodak Company: <http://kodak.com>

Millipore Corporation: <http://millipore.com>

PDB, RCSB Protein Data Bank: <http://www.rcsb.org/pdb/home/home.do>

Quantifoil Micro Tools GmbH: <http://quantifoil.com>

Roche Diagnostics Corporation: <http://roche-applied-science.com>

Xmipp, X-Window-based Microscopy Image Processing Package:

<http://xmipp.cnb.csic.es/>

Zeiss; Carl Zeiss, Inc.: <http://www.zeiss.com/explore>

Biocomp Instruments: <http://www.biocompinstruments.com/>

Apéndice

Publicaciones y manuscritos enviados del autor de la presente memoria de tesis doctoral:

Roberto Melero, Sridharan Rajagopalan, Melisa Lázaro, David Gil, S.H.W. Scheres, Andreas C. Joerger, Gorika Lasso, José María Carazo, Alan Fersht and Mikel Valle. “Dynamic p53 tetramers on DNA”. En preparación

C.O.S. Sorzano, J.R. Bilbao-Castro, Y. Shkolnisky, M. Alcorlo, **R. Melero**, G. Caffarena-Fernández, M. Li, G. Xu R. Marabini, J.M. Carazo. “A clustering approach to multireference alignment of single-particle projections in electron microscopy”. Aceptado en *Journal of Structural Biology*.

Scheres, SH, **Melero, R**, Valle, M, Carazo, JM (2009). “Averaging of electron subtomograms and random conical tilt reconstructions through likelihood optimization”. *Structure*, 17, 12:1563-72.

Melero, R, Tidow, H, Mylonas, E, Freund, SM, Grossmann, JG, Carazo, JM, Svergun, DI, Valle, M, Fersht, AR (2007). “Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex”. *Proc. Natl. Acad. Sci. U.S.A.*, 104, 30:12324-9.

Quaternary structures of tumor suppressor p53 and a specific p53–DNA complex

Henning Tidow*, Roberto Melero†, Efstratios Mylonas‡, Stefan M. V. Freund*, J. Guenter Grossmann§, José María Carazo†, Dmitri I. Svergun¶||, Mikel Valle†||**, and Alan R. Fersht*||

*Medical Research Council Centre for Protein Engineering, Hills Road, Cambridge CB2 0QH, United Kingdom; †Centro Nacional de Biotecnología, Darwin 3, Cantoblanco 28049 Madrid, Spain; **CIC-bioGUNE, Parque Tecnológico de Bizkaia, 48160 Derio, Spain; ‡European Molecular Biology Laboratory, Hamburg Outstation, Notkestrasse 85, 22603 Hamburg, Germany; †Institute of Crystallography, Russian Academy of Sciences, Leninsky pr. 59, 117333 Moscow, Russia; and §Molecular Biophysics Group, Council for the Central Laboratory of the Research Councils (CCLRC) Daresbury Laboratory, Warrington, Cheshire WA4 4AD, United Kingdom

Contributed by Alan R. Fersht, May 30, 2007 (sent for review May 8, 2007)

The homotetrameric tumor suppressor p53 consists of folded core and tetramerization domains, linked and flanked by intrinsically disordered segments that impede structure analysis by x-ray crystallography and NMR. Here, we solved the quaternary structure of human p53 in solution by a combination of small-angle x-ray scattering, which defined its shape, and NMR, which identified the core domain interfaces and showed that the folded domains had the same structure in the intact protein as in fragments. We combined the solution data with electron microscopy on immobilized samples that provided medium resolution 3D maps. *Ab initio* and rigid body modeling of scattering data revealed an elongated cross-shaped structure with a pair of loosely coupled core domain dimers at the ends, which are accessible for binding to DNA and partner proteins. The core domains in that open conformation closed around a specific DNA response element to form a compact complex whose structure was independently determined by electron microscopy. The structure of the DNA complex is consistent with that of the complex of four separate core domains and response element fragments solved by x-ray crystallography and contacts identified by NMR. Electron microscopy on the conformationally mobile, unbound p53 selected a minor compact conformation, which resembled the closed conformation, from the ensemble of predominantly open conformations. A multipronged structural approach could be generally useful for the structural characterization of the rapidly growing number of multidomain proteins with intrinsically disordered regions.

DNA binding | intrinsically unfolded | modular | natively disordered | protein

The tumor suppressor p53 is a tetrameric, multidomain transcription factor that plays a central role in the cell cycle and maintaining genomic integrity (1, 2). It binds to specific DNA response elements, is integrated in various signaling networks by a multitude of protein–protein interactions, and is controlled by extensive posttranslational modifications (1, 3, 4). p53 protein is a homotetramer of 4×393 residues. Each chain consists of two folded domains [the core, or DNA-binding, domain (94–294) and the tetramerization domain (323–360)] that are linked by an intrinsically disordered sequence. The transactivation domain (1–67) (5), proline-rich region (67–94), nuclear localization signal (NLS)-containing region (303–323) (6), and C-terminal negative regulatory domain (360–393) are also intrinsically disordered (7–9) (see refs. 10 and 11 for reviews). The DNA-binding core domain (residues 94–294) binds to sequence-specific response elements associated with p53 target gene promoters (12–14). The structures of the core domain complexes with DNA have been solved by crystallography (15–17), and in solution in the absence of DNA by NMR (18). The structure of the tetramerization domain has been solved by both NMR and x-ray crystallography (19–21).

Structural studies on full-length p53 have been impeded both by its intrinsic instability and the presence of disordered regions

(8, 18, 22, 23). There is an increasing number of proteins being discovered to have globular domains linked by intrinsically disordered regions (24), and so the determination of such structures will be a recurring problem. Recently, the first structure of full-length murine p53, obtained by cryoelectron microscopy, was reported (25). It gave a radically new concept of the molecular organization of p53. We have solved the instability problem of human p53 by designing a biologically active mutant with a superstable core domain (26, 27) that is suitable for NMR studies. NMR studies have given a picture of the p53 tetramer as a dimer of loosely tethered core dimers of appropriate symmetry to be poised to bind target DNA with well resolved transverse relaxation optimized spectroscopy (TROSY) NMR spectra (23). But, on addition of DNA, the protein rigidifies with poorly resolved spectra (23).

Here, we used small-angle x-ray scattering (SAXS), electron microscopy (EM), and NMR spectroscopy in a multitechnique approach to determine the structures of tetrameric full-length human p53 and various truncation mutants, both unligated and as a complex with a DNA response element.

Results and Discussion

Quaternary Structure of p53 from SAXS. SAXS experiments were performed on solutions of three p53 constructs: p53 core plus tetramerization domain (94–360) (CTetD), p53 core plus tetramerization plus C-terminal domain (94–393) (CTetCD), and full-length p53 (1–393) (flp53). All had four stabilizing mutations in the core domain (M133L/V203A/N239Y/N268D) (27) that do not perturb the structure (26). The processed experimental patterns (scattering intensity I versus momentum transfer $s = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle and $\lambda = 1.5 \text{ \AA}$ is the wavelength) are displayed in Fig. 1*a*, and the overall parameters are summarized in Table 1. The solutions are monodisperse [supporting information (SI) Fig. 6], and the molecular weights (M_r) of the solutes determined from the extrapolated zero angle intensity $I(0)$ consistently point to a tetrameric assembly for each construct. The radius of gyration (R_g) increases from CTetD (52.2 Å) to CTetCD (54.0 Å) to flp53 (68.2 Å) (Table 1). The

Author contributions: H.T. and R.M. contributed equally to this work; H.T., R.M., E.M., S.M.V.F., J.M.C., M.V., and A.R.F. designed research; H.T., R.M., E.M., S.M.V.F., J.G.G., and M.V. performed research; H.T., R.M., E.M., S.M.V.F., J.M.C., D.I.S., M.V., and A.R.F. analyzed data; and H.T., D.I.S., M.V., and A.R.F. wrote the paper.

The authors declare no conflict of interest.

Abbreviations: CTetD, p53 core and tetramerization domains; CTetCD, p53 core, tetramerization and C-terminal domain; flp53, full-length p53; SAXS, small-angle X-ray scattering; TROSY, transverse relaxation optimized spectroscopy.

See Commentary on page 12231.

¶To whom correspondence may be addressed. E-mail: arf25@cam.ac.uk, mvalle@cicbiogune.es, or svergun@embl-hamburg.de.

This article contains supporting information online at www.pnas.org/cgi/content/full/0705069104/DC1.

© 2007 by The National Academy of Sciences of the USA

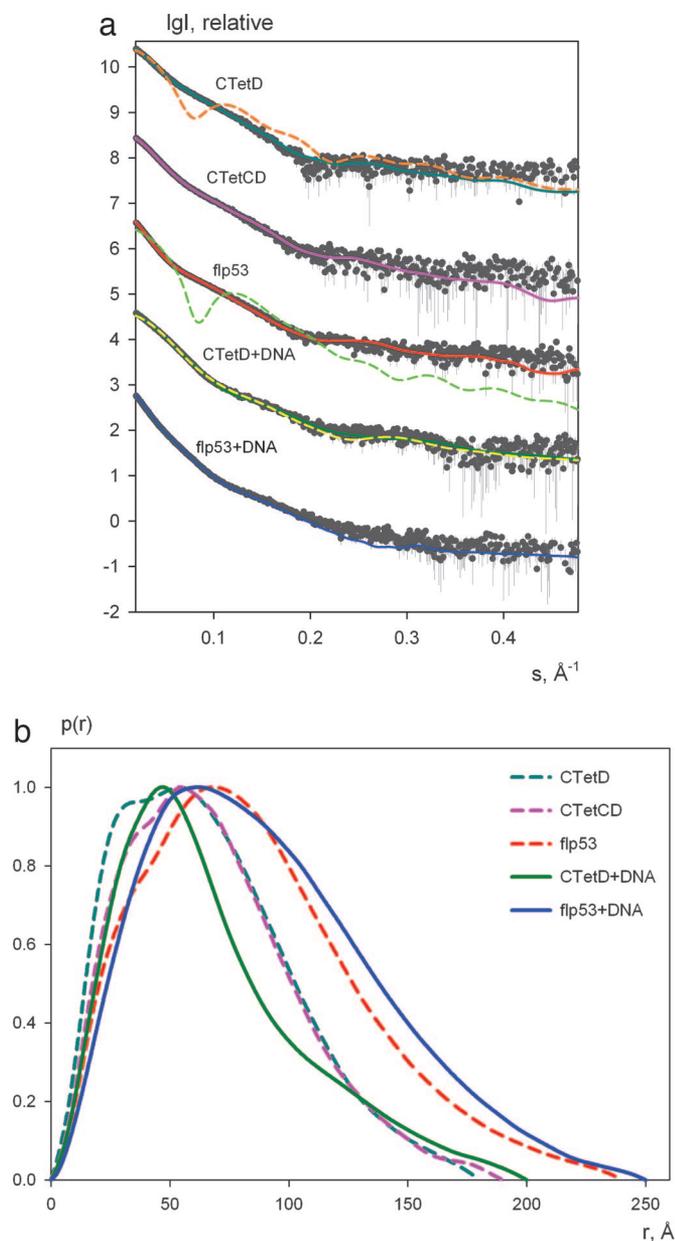


Fig. 1. SAXS analysis of p53 constructs. (a) Experimental intensities (dots with error bars) and fits computed from the structural models by rigid body modeling: CTetD (cyan), CTetCD (magenta), flp53 (red), CTetD+DNA (dark green). For the flp53–DNA complex the fit from the *ab initio* model (SI Fig. 9) is displayed in blue. The scattering profiles are displaced along the ordinate for better visualization. The fit to the CTetD data from the model proposed by Okorokov *et al.* (25) and the fit to flp53 data from the EM-map by Okorokov *et al.* (25) are shown in dashed orange and dashed light green, respectively. The fit to our EM-based model is shown in dashed yellow. (b) The distance distribution plots computed from the experimental data and normalized to the maximum value of unity.

distance distribution functions of all of the constructs display peaks or shoulders at shorter distances (20–50 Å), indicating the presence of separated subdomains (Fig. 1b). CTetCD reveals only small expansion of the structure compared with CTetD, whereas flp53 demonstrates a significant increase in the overall dimensions (Table 1). This increase indicates that the N-terminal domain is not folded back onto the core protein but instead is largely extended and potentially flexible (although, as evidenced by the so-called Kratky plot in SI Fig. 7, not completely un-

Table 1. Overall parameters of p53 constructs determined by SAXS

| Construct | R_g , Å | D_{max} , Å | Mass, kDa* |
|-----------------|----------------|---------------|--------------------|
| p53 CTetD | 52.2 ± 0.4 | 180 ± 10 | 105 ± 10 (120) |
| p53 CTetCD | 54.0 ± 0.3 | 190 ± 15 | 118 ± 12 (134) |
| Full-length p53 | 68.2 ± 0.3 | 240 ± 30 | 199 ± 20 (175) |
| p53 CTetD + DNA | 52.8 ± 0.5 | 200 ± 20 | 140 ± 15 (135) |
| flp53 + DNA | 72.3 ± 0.8 | 250 ± 30 | 290 ± 40 (190) |

R_g , radius of gyration; D_{max} , maximum size of particle.

*Mass values were evaluated by normalization against the BSA solution accounting for the contrast reduction due to 5% glycerol and (for the DNA-containing constructs) also for the higher DNA contrast compared with that of the protein. In parentheses, the expected masses of tetrameric constructs are given.

folded). This finding agrees with NMR results showing that the N-terminal region within the full-length protein is flexible (S.M.V.F. and A.R.F., unpublished results).

Low-resolution DAMMIN (28) revealed elongated cross-shaped structures (a typical model is displayed in SI Fig. 8). There are only small changes in chemical shifts of the core and tetramerization domains in flp53 compared with separate domains in solution (23). Accordingly, the known high-resolution structures of p53 core (PDB ID code: 1TUP, 2AC0) and tetramerization domains (PDB ID code: 1C26) were used to build models of p53 constructs by using the programs SASREF and BUNCH (29). These programs perform rigid-body refinement against single or multiple data sets but also allow one to model the missing linkers as dummy residue chains. The configuration of the tetramerization domain was fixed as in the crystal structure, and the linkers and core domains were added to fit the experimental data assuming the P222 symmetry of the constructs. Independently determined models of all constructs consistently display extended cross-like structures, their overall shape being compatible with the *ab initio* reconstructions (SI Fig. 8). Importantly, the tetramerization domain in the middle of these models is surrounded by two pairs of associated core domains, which is consistent with previous NMR results on the mobility of core domains (23).

To constrain the modeling further, we fitted the three scattering patterns from CTetD, CTetCD, and flp53 simultaneously and added information about the interacting surface of the core dimers. The residues likely to be involved in domain interfaces are found from comparing chemical shifts in the tetrameric constructs with those in isolated core domains (23). The interface found previously (23) has now been reassigned to that of the interaction of core domain in flp53 with its N-terminal extension (S.M.V.F., C. Baloglu, and A.R.F., unpublished results). The interface between the core domains in the crystal structure of the complex of p53 monomers with DNA has self-complementarity of residues 177, 178, 180, 181, 243, and 244 (17). These residues are also implicated in the NMR spectra of dimeric constructs of two core domains with half-site DNA (30, 31). Signals for residues 181, 243, and 244 are in clear isolated positions in the 800 and 900 MHz ^1H - ^{15}N TROSY spectrum of core domain (Fig. 2). Although there are insignificant changes in chemical shifts of nearly all of the remaining 147 assigned residues of the core domain in CTetD relative to core, the signals from 181, 243, and 244 are missing. We attribute the absence of those signals to line broadening from chemical exchange as each pair of core domains associates and dissociates, lowering the signals below the threshold for detection. Further, the normalized changes in ^1H - ^{15}N chemical shifts for 178 and 179 are among the two greatest in the spectrum at 0.034 and 0.059 ppm, respectively. These data are compatible with a transient inter-

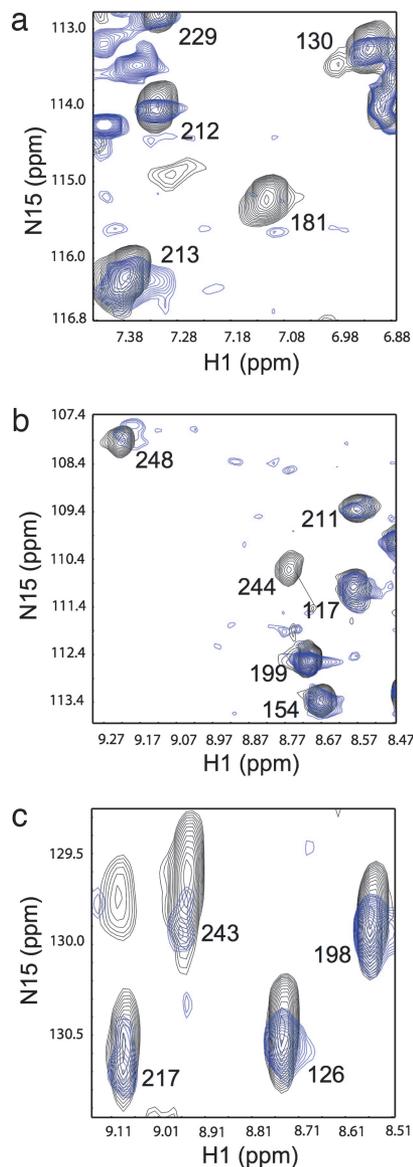


Fig. 2. Overlay of TROSY spectra acquired for p53 core domain (black) and p53 CTetD (blue) of 100- μ M samples. Chemical shift deviations for the p53 CTetD spectra were estimated from the overlay and partly confirmed with TROSY-based triple resonance spectra (data not shown). Most of the core resonances showed insignificant chemical shift deviations between the two spectra. The expansions *a*, *b*, and *c* show signals that are strong and well isolated in the spectrum of isolated core domains but are not seen in the spectrum of CTetD most likely because of substantial line broadening associated with residues in a protein–protein interface. In *a* 181 and in *b* 244 are lost, and in *c* 243 is severely diminished. Spectra were plotted close to the noise level; additional signals in *a* are from the linker region between core and tetramerization domain.

action between core domains involving the same interface observed in the crystal structure of the core domain–DNA complex (17).

We included the interface detected by NMR constraints, using the crystal structure coordinates (17). A typical constrained model displayed in Fig. 3 fits simultaneously the scattering patterns from CTetD, CTetCD, and flp53 with discrepancy $\chi = 0.82, 0.83,$ and $1.24,$ respectively (Fig. 3). The overall shape of this model is compatible with those of the unconstrained model and with the *ab initio* model (SI Fig. 8), but the constrained model displays a somewhat different orientation of the core domains.

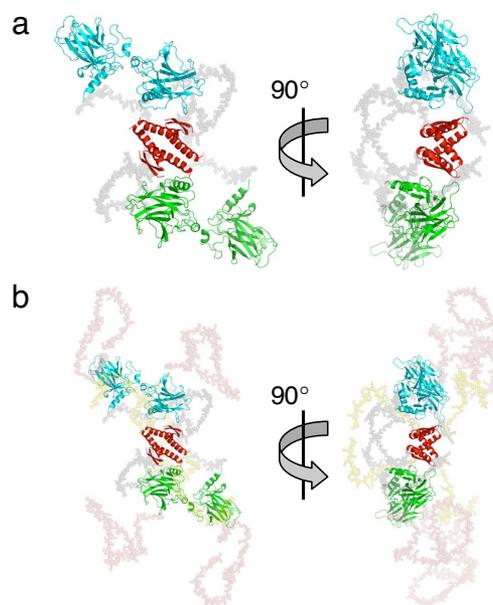


Fig. 3. SAXS models of free p53 in solution from rigid body analysis and addition of missing fragments. (a) CTetD portion of flp53 shown in *b*. Core domains and tetramerization domain are displayed in cartoon representation, connecting linkers (gray), N termini (salmon), and C termini (yellow) in semitransparent spacefill mode. Both models are presented in two orthogonal views. The model of flp53 and its CTetD and CTetCD portions fit simultaneously the experimental SAXS data from appropriate constructs in Fig. 1*a*.

The C termini are located in the central part surrounding the tetramerization domain, whereas the N-terminal domains are rather extended and point toward the periphery of the molecule.

It cannot be excluded that p53 has a considerable degree of conformational flexibility in solution and that the SAXS-generated models represent an average over such an ensemble. The conformation of potentially flexible linkers and termini, in particular in Fig. 3*b*, should be taken only as an indication of the volume occupied by these chains. On the other hand, the configuration of the tetramerization and core domains in CTetD appears rather rigid and reproducibly reconstructed in the independent modeling of different constructs. Summarizing, the model in Fig. 3 is compatible with a large body of SAXS and NMR evidence. In contrast, the calculated scattering from the recently published EM structure of p53 (25) does not agree with experimental SAXS data, yielding a poor fit with $\chi = 9.2$ (Fig. 1*a*).

Structure of p53–DNA Complex from SAXS. Tetrameric flp53 and CTetD bind 24-residue *gadd45* or *p21* response element DNA with K_d in the low nanomolar range (32, 33). The scattering profiles from the DNA complexes are displayed in Fig. 1*a* and the overall parameters are summarized in Table 1. The CTetD–DNA complex is clearly compatible with tetrameric CTetD and a single DNA molecule, whereas the M_r of the flp53–DNA complex somewhat exceeds the expected value, which may point to some aggregation or high flexibility. The distance distributions of DNA complexes display no “fine-structure” at small distances observed for free p53 constructs, suggesting that the internal structure of the complexes is more compact than that of free protein (Fig. 1*b*). This compaction is further confirmed by comparison of the Kratky plots of the free and DNA-bound p53 constructs (SI Fig. 7).

We generated the structural model for the CTetD–DNA complex by rigid body modeling using the core domains, tetramerization domain, and DNA. This procedure yielded two

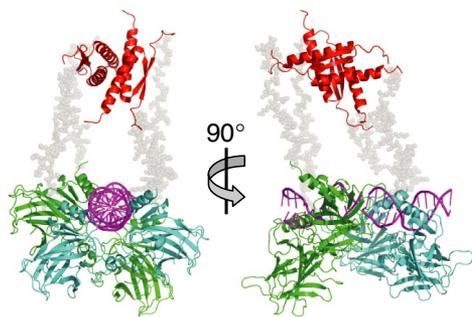


Fig. 4. Rigid body model of a p53–DNA complex from SAXS data. Core domains, tetramerization domain and DNA are shown in cartoon representation, connecting linkers in semitransparent spacefill mode. The model is displayed in two orthogonal views.

types of alternative models both showing four core domains binding to DNA and a tetramerization domain in its tetrameric form. These two structures fit the experimental data equally well and differ only by the positioning of the tetramerization domain with respect to the core domains–DNA complex (data not shown). Because one SAXS model was very similar to our EM model (see *Structure of p53–DNA Complex from EM*), we used this latter model as a starting point to add missing linkers and termini that could not be visualized in the EM map. The final model of the CTetD–DNA complex fitting the experimental data with $\chi = 1.48$ (Fig. 1*a*) reveals two superdomains: four core domains bound to DNA as described by Kitayner *et al.* (17), and the tetramerization domain (Fig. 4). These superdomains are separated by 40–120 Å and connected by linkers. The full-length p53–DNA complex appears much more extended than the CTetD–DNA complex (Table 1), suggesting that the N terminus is extremely extended. This extended structure agrees with the TROSY-NMR spectra on these complexes, which show that the N-terminal resonances remain unchanged (ref. 23). Attempts to add the missing termini yielded poor fits to the scattering pattern from the flp53–DNA complex, suggesting that the N termini may be even more flexible than those in the free-full length p53 (as indicated above, aggregation could also not be excluded). A tentative model of the flp53–DNA complex constructed *ab initio* by DAMMIN accommodates the rigid-body model of the CTetD–DNA complex and leaves ample space for the missing termini (SI Fig. 9).

Structure of p53 from EM. The EM images for negative-stained flp53 were heterogeneous in shape and size, and so the global set was unsuitable for 3D averaging techniques. A fraction, <20%, of the total particles adsorbed to the carbon support layer, however, was compact, with uniform structures of a size that was compatible with 2D projections of flp53 tetramers. We calculated the 3D map for this group following a reference-free scheme and imposing an overall C2 symmetry (see *SI Materials and Methods* and SI Fig. 10). At the defined density threshold, the map reveals a structure with two main masses linked by two faint connecting regions (SI Fig. 10*a* and *b*). Despite the limitations of the negative staining approach and the moderate resolution of the 3D map (estimated in the range of 25–30 Å), the boundaries allowed us to fit available atomic coordinates for p53 core (15, 17) and tetramerization (19, 20) domains. The 3D map (SI Fig. 10*d–f*) resembles and accommodates both structures with no need of major rearrangements, and the C termini of the core domains point toward the tetramerization domains for their connection. This structure is not the major one in solution that was found from the SAXS studies, and the computed scattering pattern does not match the experimental profile from CTetD ($\chi = 4.37$). It may well be that this is a genuine low

occupancy conformation that has to open to allow access of DNA to the binding site.

Structure of p53–DNA Complex from EM. EM images of stained flp53 in a complex with a 60-bp dsDNA probe containing a sequence-specific binding site DNA had a more uniform size distribution than the unligated protein. The volume, rendered from the calculated 3D map (Fig. 5*a–c*), does not show the dsDNA. This invisibility is a typical drawback of negative stained protein–nucleic acid samples, where nucleic acids (thin and charged strings) are rarely visible and sometimes positively stained. The EM map, however, contains structural features that confirm the presence of the DNA probe, and we independently confirmed the presence of the flp53–DNA complex by EMSA (data not shown).

As with the EM structure of free flp53, there were two regions of density, at the top and bottom, ≈ 40 Å apart, connected by two linkers (Fig. 5*a* and *b*). Importantly, the connecting densities in the map for p53–DNA complex were more obvious than for the free flp53 (see SI Fig. 10*a* and *b* and Fig. 5*a* and *b*) and were retained at high cut-off density thresholds. Those densities clearly delineate a see-through channel with the shape of a circle in one of the side views (Fig. 5*a*). The crystal structure of four core domains bound to dsDNA (17) fits very well around the channel of missing density of DNA (Fig. 5*d*). The EM map is understood as of a stained flp53 tetramer wrapped around unstained dsDNA.

The fitted atomic structures for the core (17) and tetramerization domains (19) are consistent with the EM envelope for the p53–DNA complex (Fig. 5*d–f*). Importantly, the SAXS pattern computed from the complete EM-based model, including the core and tetramerization domains and the DNA, is fully consistent with the experimentally measured scattering data from CTetD+DNA construct (Fig. 1*a*, dashed yellow curve; the fit neatly is graphically indistinguishable from the dark green curve). A local disagreement is seen for two of the core domains that protrude outside the 3D map exposing their C-terminal region (apparent for the right cyan monomer in Fig. 5*d*). We have used the atomic coordinates as rigid bodies, and no alternative arrangement is presented. Some differences between our p53–DNA complex and the crystallographic structure of the core domain tetramer bound to DNA (17) could explain this discrepancy. The x-ray structure was determined for the isolated core domain tetramer and in complex with two DNA molecules, whereas our p53–DNA sample covers the entire protein bound to a unique DNA probe. However, we cannot rule out either a partial distortion of the p53–DNA structure because of the aggressive nature of the negative staining technique.

Kitayner *et al.* (17) modeled the structure of flp53 bound to DNA based on their crystal structure of four core domains bound to two segments of DNA and the positions of where the linker sections to the tetramerization domain would exit the truncated structure. The relative positions of core and tetramerization domains are in excellent agreement with the structures determined here by both SAXS and EM.

General Implications for Structure Determination. p53 is typical of the class of proteins containing a mixture of ordered and disordered domains. The strategy used here of fitting the high resolution structures of individual domains, solved by x-ray crystallography or NMR, into low resolution SAXS and EM data, guided by TROSY NMR constraints on interfaces, may well be a useful general approach to solving their structures. SAXS, like conventional NMR spectroscopy, has the advantage of analyzing structures in solution, obviating any artifacts associated with freezing of samples or immobilizing them on a grid, as used in EM. However, it does not generate unique models. SAXS optimizes the spatial arrangement of subunits by direct

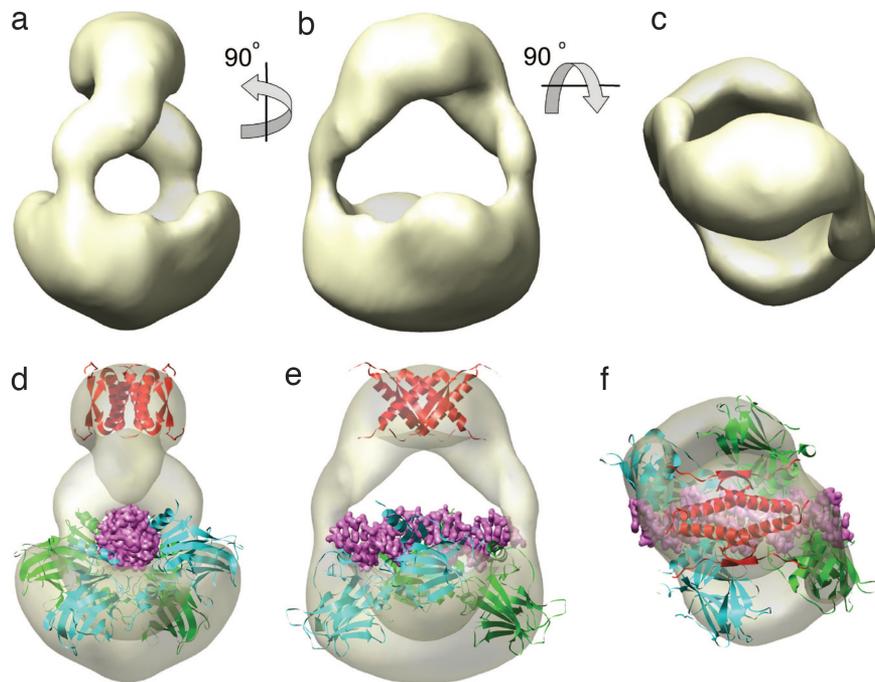


Fig. 5. 3D-EM map for flp53-DNA complex. The 3D map for the p53-DNA complex is rendered with a volume that covers 100% of the expected volume for a p53 tetramer in solid (a–c) and semitransparent (d–f) views. The fitted coordinates place the representation for the dsDNA in the see-through channel of one of the side views (see a and d).

fitting of experimental scattering data on docking high-resolution structures into shapes of macromolecular complexes, but there may be more than one solution (29). EM of negative stained samples uses an envelope obtained after 3D image reconstruction for docking of high-resolution models of subunits, and the overall shape should not be ambiguous if the high-resolution structures of domains are assigned correctly in the envelope. However, immobilization on the carbon support and differential staining may lead to selection of a particular conformation. SAXS data are particularly useful for discriminating between independently determined models by comparing calculated and observed scattering profiles because SAXS can eliminate incorrect models (34), and differences between SAXS solution models and EM-generated models have been reported (35). The differences between the structures of unligated p53 determined in solution and by EM show that a multimethod approach is necessary.

It is striking, here, that both EM and SAXS generate a very similar model for a p53-DNA complex, which is a relatively rigid structure. For the free p53, however, the solution studies using SAXS and NMR indicate a structure of loosely tethered pairs of core domains, with the possibilities of facile conformational rearrangement. Our EM studies, in contrast, show a structure that is preformed for binding the DNA response element. However, the calculated scattering profile is inconsistent with that found from the overall population in solution. It is likely that the experimental conditions for EM have selected a minor closed conformation that has to open for DNA to have access to the binding site. The model proposed for murine p53 (25) may also be a minor closed conformation. Key features of the murine model are not consistent with our data: for example, the murine model has the tetramerization domain dissociated and the N and C termini in contact. We find from SAXS that unligated p53 and its truncation variants all have the same quaternary arrangements of core and tetramerization domains, showing that the structure does not depend on the interaction of the N and C termini. Further, the dissociation constants of flp53 and CTetD

lacking the C and N termini are quite similar (23). For the conformationally flexible free p53, the structures from SAXS, with supplementary information from NMR, must have precedence over the structures purely from EM.

Biological Relevance of the Quaternary Structure. The presence of an extended N-terminal domain in our model is consonant with its biological function. The N-terminal domain of p53 is subject to extensive posttranslational modifications that potentially regulate transactivational activity (36, 37). It also binds to a variety of proteins that further modulate p53 activity, like MDM2 (38), p300 (39), several transcription activators of the TFIID complex (40–42), and RPA (43). Upon binding to other proteins, small regions within the intrinsically unfolded p53 N terminus form helical structures (38, 43), a phenomenon also observed for other transcription factors (44–46). Further, our model of loosely coupled core domains assembled in an elongated cross-shaped structure allows binding of regulatory protein like Bcl-xL (47), Bcl2 (48), or 53BP2 (49, 50) to p53 core domain.

The open arrangement in flp53 of two separate pairs of core domains (Fig. 3) allows access of DNA to the binding site: first, one pair of core domains binds to two of the four binding site on a response element; then, the flexible linkers between the core and tetramerization domains allow the second pair of core domains to bind to the remaining two sites, thus burying the DNA within the protein.

Methods

Human p53 CTetD, CTetCD, and full-length p53 were expressed in *Escherichia coli* and purified as described (23). SAXS data were collected at the X33 beamline at European Molecular Biology Laboratory/Deutsches Elektronen Synchrotron (EMBL/DESY), Hamburg and at station 2.1 of the Daresbury Synchrotron Radiation Source (SRS) following standard procedures. Electron microscopy using negative staining involved manual selection of particles and reference-free classification (see SI Fig. 11). TROSY NMR spectra were acquired on 800

MHz and 900 MHz spectrometers by using perdeuterated, isotopically ^2H , ^{15}N -labeled samples. Details of sample preparation, SAXS measurements and modeling, electron microscopy, image processing and domain fitting, and NMR spectroscopy are given in *SI Materials and Methods*.

We thank Dr. A. Joerger for helpful discussions, C. Blair (MRC Centre for Protein Engineering, Cambridge, U.K.) for TEV protease and protein preparation, Dr. E. Orlova for communication of coordinates of murine p53 and kind comments, Dr. J. Milner for kind comments, and Dr. R. Nuñez for help using EMAN. M.V. is supported by Spanish

Ministry of Education and Science Grant BFU2006-09648, by Etorrek Research Programs 2005/2006 (The Department of Industry, Tourism and Trade of the Government of the Autonomous Community of the Basque Country), and by the Innovation Technology Department of the Bizkaia County. H.T. is supported by a fellowship from the Boehringer Ingelheim Fonds. We acknowledge support from European Community–Research Infrastructure Action under the FP6 (RII3/CT/2004/5060008) for access to EMBL/DESY, Hamburg, and access to the European Large Scale Facility for NMR Spectroscopy in Utrecht, The Netherlands. D.I.S. and J.G.G. acknowledge support from the European Union FP6 Design Study Small-Angle X-Ray Scattering Initiative for Europe (SAXIER), RIDS 011934.

- Hainaut P, Wiman KG, eds (2005) *25 Years of p53 Research* (Springer, New York).
- Vogelstein B, Lane D, Levine AJ (2000) *Nature* 408:307–310.
- Hupp TR, Lane DP (1994) *Cold Spring Harbor Symp Quant Biol* 59:195–206.
- Hupp TR, Sparks A, Lane DP (1995) *Cell* 83:237–245.
- Dawson R, Muller L, Dehner A, Klein C, Kessler H, Buchner J (2003) *J Mol Biol* 332:1131–1141.
- Shaalsky G, Goldfinger N, Ben-Ze'ev A, Rotter V (1990) *Mol Cell Biol* 10:6565–6577.
- Ayed A, Mulder FA, Yi GS, Lu Y, Kay LE, Arrowsmith CH (2001) *Nat Struct Biol* 8:756–760.
- Bell S, Klein C, Muller L, Hansen S, Buchner J (2002) *J Mol Biol* 322:917–927.
- Lee H, Mok KH, Muhandiram R, Park KH, Suk JE, Kim DH, Chang J, Sung YC, Choi KY, Han KH (2000) *J Biol Chem* 275:29426–29432.
- Joerger AC, Fersht AR (2007) *Oncogene* 26:2226–2242.
- Roemer L, Klein C, Dehner A, Kessler H, Buchner J (2006) *Angew Chem Int Ed Engl* 45:6440–6460.
- el-Deiry WS, Kern SE, Pietenpol JA, Kinzler KW, Vogelstein B (1992) *Nat Genet* 1:45–49.
- Kern SE, Kinzler KW, Bruskin A, Jarosz D, Friedman P, Prives C, Vogelstein B (1991) *Science* 252:1708–1711.
- Wei CL, Wu Q, Vega VB, Chiu KP, Ng P, Zhang T, Shahab A, Yong HC, Fu Y, Weng Z, et al. (2006) *Cell* 124:207–219.
- Cho Y, Gorina S, Jeffrey PD, Pavletich NP (1994) *Science* 265:346–355.
- Ho WC, Fitzgerald MX, Marmorstein R (2006) *J Biol Chem* 281:20494–20502.
- Kitayner M, Rozenberg H, Kessler N, Rabinovich D, Shaulov L, Haran TE, Shakked Z (2006) *Mol Cell* 22:741–753.
- Canadillas JM, Tidow H, Freund SM, Rutherford TJ, Ang HC, Fersht AR (2006) *Proc Natl Acad Sci USA* 103:2109–2114.
- Clore GM, Omichinski JG, Sakaguchi K, Zambrano N, Sakamoto H, Appella E, Gronenborn AM (1995) *Science* 267:1515–1516.
- Jeffrey PD, Gorina S, Pavletich NP (1995) *Science* 267:1498–1502.
- Lee W, Harvey TS, Yin Y, Yau P, Litchfield D, Arrowsmith CH (1994) *Nat Struct Biol* 1:877–890.
- Bullock AN, Henckel J, DeDecker BS, Johnson CM, Nikolova PV, Proctor MR, Lane DP, Fersht AR (1997) *Proc Natl Acad Sci USA* 94:14338–14342.
- Veprintsev DB, Freund SM, Andreeva A, Rutledge SE, Tidow H, Canadillas JM, Blair CM, Fersht AR (2006) *Proc Natl Acad Sci USA* 103:2115–2119.
- Dyson HJ, Wright PE (2002) *Curr Opin Struct Biol* 12:54–60.
- Okorokov AL, Sherman MB, Plisson C, Grinkevich V, Sigmundsson K, Selivanova G, Milner J, Orlova EV (2006) *EMBO J* 25:5191–5200.
- Joerger AC, Allen MD, Fersht AR (2004) *J Biol Chem* 279:1291–1296.
- Nikolova PV, Henckel J, Lane DP, Fersht AR (1998) *Proc Natl Acad Sci USA* 95:14675–14680.
- Svergun DI (1999) *Biophys J* 76:2879–2886.
- Petoukhov MV, Svergun DI (2005) *Biophys J* 89:1237–1250.
- Klein C, Planker E, Diercks T, Kessler H, Kunkele KP, Lang K, Hansen S, Schwaiger M (2001) *J Biol Chem* 276:49020–49027.
- Rippin TM, Freund SM, Veprintsev DB, Fersht AR (2002) *J Mol Biol* 319:351–358.
- Ang HC, Joerger AC, Mayer S, Fersht AR (2006) *J Biol Chem* 281:21934–21941.
- Weinberg RL, Veprintsev DB, Fersht AR (2004) *J Mol Biol* 341:1145–1159.
- Vestergaard B, Sanyal S, Roessle M, Mora L, Buckingham RH, Kastrup JS, Gajhede M, Svergun DI, Ehrenberg M (2005) *Mol Cell* 20:929–938.
- Pioletti M, Findeisen F, Hura GL, Minor DL, Jr (2006) *Nat Struct Mol Biol* 13:987–995.
- Banin S, Moyall L, Shieh S, Taya Y, Anderson CW, Chessa L, Smorodinsky NI, Prives C, Reiss Y, Shiloh Y, Ziv Y (1998) *Science* 281:1674–1677.
- Shieh SY, Ikeda M, Taya Y, Prives C (1997) *Cell* 91:325–334.
- Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, Levine AJ, Pavletich NP (1996) *Science* 274:948–953.
- Grossman SR (2001) *Eur J Biochem* 268:2773–2778.
- Lu H, Levine AJ (1995) *Proc Natl Acad Sci USA* 92:5154–5158.
- Naar AM, Lemon BD, Tjian R (2001) *Annu Rev Biochem* 70:475–501.
- Thut CJ, Chen JL, Klemm R, Tjian R (1995) *Science* 267:100–104.
- Bochkareva E, Kaustov L, Ayed A, Yi GS, Lu Y, Pineda-Lucena A, Liao JC, Okorokov AL, Milner J, Arrowsmith CH, Bochkarev A (2005) *Proc Natl Acad Sci USA* 102:15412–15417.
- Bowers PM, Schaufler LE, Klevit RE (1999) *Nat Struct Biol* 6:478–485.
- Graham TA, Weaver C, Mao F, Kimelman D, Xu W (2000) *Cell* 103:885–896.
- Spolar RS, Record MT, Jr (1994) *Science* 263:777–784.
- Chipuk JE, Bouchier-Hayes L, Kuwana T, Newmeyer DD, Green DR (2005) *Science* 309:1732–1735.
- Tomita Y, Marchenko N, Erster S, Nemajerova A, Dehner A, Klein C, Pan H, Kessler H, Pancoska P, Moll UM (2006) *J Biol Chem* 281:8600–8606.
- Gorina S, Pavletich NP (1996) *Science* 274:1001–1005.
- Tidow H, Veprintsev DB, Freund SM, Fersht AR (2006) *J Biol Chem* 281:32526–32533.

Averaging of Electron Subtomograms and Random Conical Tilt Reconstructions through Likelihood Optimization

Sjors H.W. Scheres,^{1,*} Roberto Melero,¹ Mikel Valle,² and Jose-Maria Carazo¹

¹Biocomputing Unit, Centro Nacional de Biotecnología - CSIC, Darwin 3, Cantoblanco, 28049, Madrid, Spain

²CICBiogune, Parque Tecnológico de Bizkaia, 48160, Derio-Bizkaia, Spain

*Correspondence: scheres@cnb.csic.es

DOI 10.1016/j.str.2009.10.009

SUMMARY

The reference-free averaging of three-dimensional electron microscopy (3D-EM) reconstructions with empty regions in Fourier space represents a pressing problem in electron tomography and single-particle analysis. We present a maximum likelihood algorithm for the simultaneous alignment and classification of subtomograms or random conical tilt (RCT) reconstructions, where the Fourier components in the missing data regions are treated as hidden variables. The behavior of this algorithm was explored using tests on simulated data, while application to experimental data was shown to yield unsupervised class averages for subtomograms of groEL/groES complexes and RCT reconstructions of p53. The latter application served to obtain a reliable *de novo* structure for p53 that may resolve uncertainties about its quaternary structure.

INTRODUCTION

Modern electron microscopes allow the visualization of individual copies of biological macromolecular complexes in a hydrated environment (Stahlberg and Walz, 2008). Because the electron dose needs to be strictly limited to prevent radiation damage, the resulting images are extremely noisy and averaging over multiple copies of identical complexes is often a prerequisite to obtain reliable structural information. However, because of the high noise levels, averaging approaches in three-dimensional electron microscopy (3D-EM) often rely on comparison with a reference structure, and thereby may become sensitive to bias toward this reference. This problem may be particularly aggravated when the data are not measured in its totality, which is the case for subtomograms or RCT reconstructions. Because in both types of experiment the tilt angle in the microscope is physically limited, one cannot collect the data to completeness and the resulting reconstructions contain empty regions in Fourier space. In this contribution, we present an unsupervised averaging approach that can simultaneously align and classify such 3D reconstructions with missing data regions in Fourier space.

In electron tomography, a 3D density map of a unique object (e.g., an entire cell) is obtained by tilting the sample over a controlled series of angles and combining the corresponding

projection images in a single reconstruction (Hoenger and McIntosh, 2009; Leis et al., 2009). In the case of single-axis tilting, the missing region in Fourier space is wedge-shaped, whereas dual-axis tilting reduces this region to a pyramid. The incompleteness of the data in Fourier space results in serious artifacts in the 3D density maps. However, if multiple copies of the same macromolecular complex can be identified and extracted from this reconstruction, one can average over these so-called subtomograms, provided that their relative orientations can be determined and they can be separated into structurally homogeneous classes. Apart from reducing the high levels of noise, averaging also reduces the real-space artifacts due to the incompleteness of the data by “filling in” the missing data regions (Forster and Hegerl, 2007). The latter is true because those data points that are missing in a particle in a given orientation may be observed for another particle in a different orientation, and vice versa. Moreover, by placing the averaged structures back into the cellular reconstruction, one could obtain the “molecular atlas” of the cell, i.e., the spatial distribution of the different functional states of multiple macromolecular complexes throughout the cell. However, because of the high noise levels of individual subtomograms, their incomplete character, and problems related to model bias, the unsupervised alignment and classification of subtomograms currently still represents an important hurdle for what has been called the *visual proteomics* application of cellular tomography (Nickell et al., 2006).

In single-particle analysis, isolated and purified complexes are imaged in a thin layer of vitreous ice, and multiple projections of assumedly identical copies of the complexes are combined in a single 3D reconstruction (Frank, 2006). Because the molecules adopt unknown orientations on the experimental support, their relative projection directions are typically determined a posteriori by comparison of the experimental images with computer-simulated projections of a 3D reference structure. However, the generation of a suitable initial reference map is a highly complicated task, especially for small (i.e., < 400 kDa) or flexible molecules with no or low symmetry (Llorca, 2005). A proven way to generate 3D reconstructions *de novo* is the random conical tilt (RCT) method, where one records image pairs at two different tilt angles in the microscope (Radermacher, 1988; Radermacher et al., 1986). In the general case that the molecules adopt multiple orientations on the support, RCT processing will produce multiple reconstructions in different orientations and, if the molecules are structurally variable, also in distinct conformations. Moreover, the geometrical design of the RCT experiment results in reconstructions with missing regions in Fourier space that are

cone shaped. Thereby, averaging over RCT reconstructions again involves alignment and classification of incomplete reconstructions, and is conceptually very similar to the problem of subtomogram averaging.

Early work on aligning RCT reconstructions includes Penczek et al. (1994). More recently, the potentials of averaging RCT reconstructions have been illustrated by pioneering work on highly flexible spliceosome complexes by Stark and co-workers (Sander et al., 2006), and also Radermacher employed the RCT technique to study highly flexible complexes (Radermacher, 2009). Nevertheless, to date no specific algorithms for the averaging of RCT reconstructions that take the missing cones into account have been described. More contributions are available on subtomogram averaging, and several efforts to obtain molecular atlases of cells have already been reported (e.g., Brandt et al., 2009; Malmstrom et al., 2009). Initial approaches to subtomogram averaging that did not take the missing wedges into account were observed to suffer from the alignment or classification of artifacts (Walz et al., 1997). More recently, the treatment of the missing wedges has played an increasingly important role. The Baumeister group was the first to propose a scoring function that accounts for missing wedges, the constrained cross-correlation coefficient, which limits the comparison of two maps to those regions in Fourier space where both maps contain data. Although initially only used for subtomogram alignment (Beck et al., 2004; Forster et al., 2005; Frangakis et al., 2002), more recently pairwise calculation of this similarity measure for aligned subtomograms was also employed for classification purposes (Forster et al., 2008). Similar approaches to take the missing wedge into account were also used by other groups. Subramanian and coworkers used a multiplicative mask for the missing wedge in Fourier space and proposed an iterative procedure of efficient subtomogram alignment based on spherical harmonics analysis, and classification based on hierarchical clustering (Bartesaghi et al., 2008; Liu et al., 2008). Winkler et al. classify and align subtomograms based on constrained cross-correlation using either “alignment by classification” (Dube et al., 1993), or iterative schemes involving multivariate statistical analysis, hierarchical clustering, and realignment of the classes (Winkler et al., 2009). Schmid and Booth proposed a simpler scoring function that also restricts the comparison to the mutually observed regions in Fourier space, but is much faster to calculate (Schmid and Booth, 2008).

Common characteristics of existing approaches for subtomogram averaging are that alignment and classification are often performed in separate steps, and that the resulting averages are calculated in Fourier space by dividing the sum of all images by the number of times that each Fourier component was observed. (The latter has been called *weighted averaging*, and we will use the same term throughout this paper.) Both characteristics may behave suboptimally in specific situations. Depending on the type of structural heterogeneity and the overall shape of the particles, the problems of alignment and classification may be strongly intertwined, while weighted averaging in Fourier space may put too strong emphasis on noisy Fourier components that were observed only a few times in cases where the particles adopt a preferred orientation and the missing regions are not filled up.

Our proposal is to tackle alignment and classification simultaneously through a multireference refinement, and to obtain the

parameters of a statistical data model by maximum likelihood estimation. Similar approaches have proven highly effective for the alignment and classification of structurally heterogeneous two-dimensional (2D) projection data sets in single-particle analysis (Scheres et al., 2005 2007a.). A major benefit of the maximum-likelihood approach is the explicit, statistical description of the abundant experimental noise in the data, which allows one to “marginalize,” i.e., calculate probability-weighted assignments, over the so-called hidden variables. These variables are necessary to solve the problem, but were not observed in the experiment. Here, obvious hidden variables are the class and orientational assignments of every particle. In addition, to explicitly reflect the unobserved character of the missing data in our statistical model, we also treat all Fourier components in the missing data regions as hidden variables. We will derive the corresponding algorithm, where the missing data regions are complemented with the current estimates for the model parameters, and we will show that this algorithm may be employed in a completely unsupervised manner for the alignment and classification of subtomograms or RCT reconstructions.

Results

The Data Model

We model our data set to consist of N 3D reconstructions (particles) X_i , with 3D Fourier transforms X_i , which are described according to the following model in Fourier space:

$$X_i = R_{\Phi_i} A_{\kappa_i} + G_i \text{ for } i = 1, 2, \dots, N \quad (1)$$

where:

- $X_i \in \mathfrak{N}^J$ is a J -dimensional vector of complex values $(X_i)_j$ or X_{ij} in short, which can be divided into a vector X_{ij}^o of observed components X_{ij}^o and a vector X_{ij}^m of missing components X_{ij}^m . Each data vector X_i may have a unique pattern of missing components, but the pattern is known for every X_i . We formalize the notion of a missing data generation mechanism by defining a J -dimensional missing data indicator vector w_i , such that $w_{ij} = 1$ for observed elements, and $w_{ij} = 0$ for missing elements.

- A_{κ_i} is one of K unknown 3D structures in Fourier space $A_1, A_2, \dots, A_K \in \mathfrak{N}^J$. These are the objects we wish to estimate from the data. κ_i is an unknown, random integer with possible values 1, 2, ..., K , indicating which of the unknown structures corresponds to particle X_i .

- R_{Φ_i} is a transformation matrix, which maps the unknown structure A_{κ_i} on particle X_i . The actual transformation Φ_i for particle X_i is unknown. We parameterize this transformation by ϕ_i , a 6D vector (corresponding to three Euler angles q_{rot} , q_{tilt} , q_{psis} , and three real-space translation coordinates, q_x , q_y , q_z).

- $G_i \in \mathfrak{N}^J$ is a J -dimensional vector of unknown, independent Gaussian noise with mean zero and unknown standard deviation σ .

In this model, the data space comprises the entire J -dimensional vector X_i , including the part that remained unobserved, which contrasts with alternative approaches where the data space is restricted to only the observed part of X_i , (e.g., Forster and Hegerl, 2007). We note here that also in the latter case one may devise a maximum-likelihood algorithm, and investigations into alternative ways to deal with preferred orientations within such a data model are currently underway (unpublished results). Also, the assumption of white Gaussian noise in this model may

Structure

Maximum-Likelihood Subvolume Averaging in 3D-EM

only be an approximate description of experimental noise in 3D-EM. Still, similar assumptions in maximum-likelihood (ML) classification approaches for single-particle data have been shown to be highly effective (Scheres et al., 2005, 2007a). Moreover, the mathematical framework presented here could easily be adapted to describe colored noise by introducing a resolution-dependent estimate for the standard deviation of the noise. Again, for single-particle analysis such an approach has previously been shown to be effective (Scheres et al., 2007b), but we chose not to implement a similar approach here because of its increased computational complexity.

The Optimization Problem

The task at hand is to estimate A_1, A_2, \dots, A_K from the partially observed data vectors X_1, X_2, \dots, X_N . We view this estimation problem as an incomplete data problem, where the missing data associated with X_i are their unknown transformation Φ_i and class assignment κ_i , as well as the vector of missing Fourier components X_{ij}^m . Note that if the data would have been complete, i.e., if Φ_i , κ_i and X_{ij}^m would have been observed, this problem would be extremely easy to solve by straightforward averaging. Because this is not the case, we deal with the incompleteness of the data by expectation maximization (Dempster et al., 1977). This method provides a general solution for finding maximum likelihood estimates of parameters in probabilistic models that depend on unobserved, or so-called “hidden variables,” which are added to the observed data forming a complete-data vector. The algorithm forms a likelihood function on the complete-data vector, and then overcomes the fact that part of the data is missing by iteratively working with the conditional expectation of the complete-data log-likelihood given the observed data, which is effected using the current fit for the missing parameters.

In this case, the complete data set corresponds to

$$(X_i^o, X_{ij}^m, \Phi_i, \kappa_i) \text{ for } i = 1, 2, \dots, N, \quad (3)$$

and we aim to find those parameters that maximize the joint probability of observing the entire set of observed data vectors $X_1^o, X_2^o, \dots, X_N^o$:

$$L(\Theta) = \sum_{i=1}^N \log \sum_{k=1}^K \int_{\phi} \int_{M_i} P(X_i^o | k, \phi, X_{ij}^m, \Theta) P(k, \phi, X_{ij}^m | \Theta) d\phi dM_i \quad (4)$$

where $\int_{M_i} dM_i$ is a shorthand notation for the integrals from $-\infty$ to ∞ for every missing Fourier component X_{ij}^m in X_i : $\int_{M_i} dM_i = \int_{X_{i1}^m = -\infty}^{\infty} \int_{X_{i2}^m = -\infty}^{\infty} \dots dX_{i1}^m dX_{i2}^m \dots$

According to the data model in (1), the probability of observing X_i^o for a given combination of the hidden and model parameters is expressed as:

$$P(X_i^o | k, \phi, X_{ij}^m, \Theta) = \left(\frac{1}{2\pi\sigma^2} \right)^J \exp \left\{ \frac{1}{-2\sigma^2} \sum_{j=1}^J (w_{ij} \| (R_{\phi} A_k)_j - X_{ij}^o \|^2 + (1 - w_{ij}) \| (R_{\phi} A_k)_j - X_{ij}^m \|^2) \right\} \quad (5)$$

We assume that the prior probabilities of all k , ϕ and X_{ij}^m are independent, and that all rotations and all values of the missing Fourier components are equally likely. In addition, we assume that the translational components of ϕ are distributed according to a 3D Gaussian, centered at the origin and with standard deviation ξ , and the proportion of particles belonging to the class is described by values α_k (with $\alpha_k \geq 0$, and $\sum_{k=1}^K \alpha_k = 1$). Thereby, $P(k, \phi, X_{ij}^m | \Theta)$ is expressed as:

$$P(k, \phi, X_{ij}^m | \Theta) = \frac{C\alpha_k}{\xi^3} \exp \left\{ \frac{q_x^2 + q_y^2 + q_z^2}{-2\xi^2} \right\}, \quad (6)$$

where C is a constant to ensure that $\sum_{k=1}^K \int_{\phi} \int_{M_i} P(k, \phi, X_{ij}^m | \Theta) d\phi dM_i = 1$.

From (5) and (6) we see that model parameter set Θ comprises estimates for A_1, A_2, \dots, A_K , $\alpha_1, \alpha_2, \dots, \alpha_K$, σ , and ξ , while as stated above, the hidden variables consist of k , ϕ for every particle and all X_{ij}^m .

The Optimization Algorithm

In the Supplemental Data (available online), we derive the following expectation-maximization algorithm for the optimization of the log-likelihood function in (4). The estimates for the underlying 3D structures may be updated by:

$$A_{kj}^{\text{new}} = \frac{\sum_{i=1}^N \int_{\phi} P(k, \phi | X_i^o, \Theta^{\text{old}}) \left\{ (R_{\phi}^{-1} w_{ij}) (R_{\phi}^{-1} X_i^o)_j + (1 - (R_{\phi}^{-1} w_{ij})) A_{kj}^{\text{old}} \right\} d\phi}{\sum_{i=1}^N \int_{\phi} P(k, \phi | X_i^o, \Theta^{\text{old}}) d\phi}, \quad (7)$$

where R_{ϕ}^{-1} denotes the inverse transformation of R_{ϕ} , and $P(k, \phi | X_i^o, \Theta)$ is the conditional probability of k and ϕ in terms of the current model estimates and calculated only over the observed Fourier components X_i^o .

Similarly, the estimate for the standard deviation in the noise may be updated using:

$$(\sigma^{\text{new}})^2 = \frac{1}{NJ} \sum_{i=1}^N \sum_{j=1}^J \sum_{k=1}^K \int_{\phi} P(k, \phi | X_i^o, \Theta^{\text{old}}) \times \left\{ w_{ij} \| (R_{\phi} A_k)_j - X_{ij}^o \|^2 + (1 - w_{ij}) (\sigma^{\text{old}})^2 \right\} d\phi \quad (8)$$

The updates of α_k and ξ are analogous to their updates in the 2D case (Scheres et al., 2005), and are given by:

$$\alpha_k^{\text{new}} = \sum_{i=1}^N \int_{\phi} P(k, \phi | X_i^o, \Theta^{\text{old}}) d\phi \quad (9)$$

$$(\xi^{\text{new}})^2 = \frac{1}{3N} \sum_{i=1}^N \sum_{k=1}^K \int_{\phi} P(k, \phi | X_i^o, \Theta^{\text{old}}) \left\{ q_x^2 + q_y^2 + q_z^2 \right\} d\phi \quad (10)$$

Comparison with the ML Algorithm for Single Particles

It is interesting to compare the estimation problem and the optimization algorithm at hand with those for the simultaneous alignment and classification of 2D single-particle projections (Scheres et al., 2005). Although the latter concerned a 2D case

in real space, Θ comprised the same set of model parameters. The main differences between the two cases lie in the hidden variables. In the 2D case, ϕ comprised translational coordinates in only two directions and only a single in-plane rotation, whereas for the case at hand 3D rotations and three translations are to be evaluated. Moreover, the data vectors in the 2D case were completely observed, so that X_i^m did not form part of the hidden variables.

Consequently, also the two optimization algorithms are similar. As for the 2D case, the algorithm derived above calculates model parameter updates as probability-weighted averages over all possible assignments of the hidden variables ϕ and k . However, rather than calculating these probabilities over the entire data vectors, they are calculated only over the observed elements. In this way, the underlying similarity measure of the experimental particles and the average is similar to the constrained cross-correlation coefficient as proposed previously (Forster et al., 2005; Frangakis et al., 2002). Due to the incompleteness of the data vectors, additional terms arise in (7) and (8), compared to their counterparts in the 2D case. To be more concise, for those regions in Fourier space that were unobserved, the data vectors are complemented with the corresponding values of the model at hand (A_{kj}^{old} or σ^{old}). That is, one “fills in” the missing wedges of the incomplete data vectors with the current model estimates in order to generate a pseudo-complete data set. In this way, the iterative algorithm functions in a way that is very similar to the complete-data case and that does not involve any division by the number of times a Fourier component has been observed, as would be the case in approaches based on weighted averaging.

A Regularized Optimization Target

We also implemented an algorithm to optimize a regularized log-likelihood function. Rather than maximizing (4), this algorithm optimizes the following target, which contains an additional empirical term that restraints the differences between the distinct class averages by minimizing the square L2-norms of all respective difference maps:

$$\Lambda(\Theta) = L(\Theta) - \frac{\vartheta}{2\sigma^2} \sum_{k=1}^K \sum_{l=1}^K v_{kl} \sum_{j=1}^J \|A_{kj} - A_{lj}\|^2. \quad (11)$$

Here, θ and v_{kl} are free parameters that determine to what extent the similarity among the different class averages is imposed. We note that this regularization is very similar to the neighbor restraints in the kernel-density self-organizing map (Pascual-Montano et al., 2002). Actually, although our current implementation only employs $v_{kl} = 1$, we note that for a special choice of v_{kl} a neural network-like algorithm may be obtained that is very similar to that of the self-organizing map, but that complements the missing wedges with the model estimates from the previous iteration. Because the regularization term in a way contradicts the goal of clustering (to obtain as different as possible class averages), we typically only employ this term during the first few iterations and gradually decrease the value of θ to zero.

The corresponding algorithm is very similar to the one outlined above (see [Supplemental Data](#) available online). The update formulae for the model estimates A_{kj} and σ contain addition terms compared with the updates given in (7) and (8), while the

updates for α_k and ξ are not affected by the regularization term and remain as given in (9) and (10).

RESULTS

Model Calculations

To test the potentials of the presented ML approach, we first used simulated data to compare their performance with a standard multireference refinement based on constrained cross-correlation coefficients and weighted averaging (CC). The simulated data were generated in a perhaps simplistic manner to represent the idealized case, where the data are in accordance with the underlying statistical model of our approach. Atomic models of GroEL₁₄ and GroEL₁₄GroES₇ complexes (Protein Data Bank [PDB] IDs 1GR5 and 2C7C, respectively) were converted to 2.5 nm resolution electron density maps using the `convert_pdb2vol` program from the Xmipp package (Sorzano et al., 2004). All simulated data sets contained 200 subtomograms, for which GroEL₁₄ or GroEL₁₄GroES₇ were randomly selected with equal probability. These maps were subsequently rotated and translated, various amounts of independent Gaussian noise were added, and a missing wedge was applied in the Fourier domain corresponding to a tilt series from -60° to 60° . Random translations were sampled from a uniform distribution between -3 and $+3$ voxels in all three directions. Random rotations were sampled from an approximately even distribution of three Euler angles with random perturbations. To simulate different degrees of preferred orientations, Euler triplets where the second Euler angle was between 60° and 120° or between 40° and 140° were omitted from the angular distributions. This corresponded to a moderate and a more severe preferred orientation of the particles with an increasing lack of information along the seven-fold axis of the GroEL complexes. The direction of these preferred orientations corresponds to the worst-case scenario, i.e. it coincides with the direction in which the differences between the two structures are most different.

Unsupervised alignment and classification runs with these model data sets comprised refinements with $K = 2$ references and an angular sampling of 15. All refinements were started from the weighted averages of two random subsets of the data in random orientations, and all runs were stopped after 40 iterations. Preliminary experiments without the additional regularization term on the log-likelihood function (11) showed suboptimal classification behavior for both the ML and the CC algorithms, which appeared to get stuck in local minima at the initial iterations of the optimization process. Therefore, in all subsequent ML and CC runs we employed a regularization scheme where the regularization parameter θ was initially set to $5N/K^2$, and then linearly decreased to 0 in the first 5 iterations. Then, multiple runs of the ML and the CC approaches were compared for data sets with the moderate preferred orientation and signal-to-noise ratios (SNRs) ranging from 0.002 to 0.1 (Figure 1). Except for the lowest tested SNR, where both approaches failed, the ML approach gave purer classes than the CC approach over the whole SNR range. Notably, at SNRs higher than 0.007 the ML approach gave almost completely pure classes for all runs performed, whereas even for SNRs as high as 0.1 the CC approach still failed in some of the runs. In the latter cases, typically all images were assigned to a single

Structure

Maximum-Likelihood Subvolume Averaging in 3D-EM

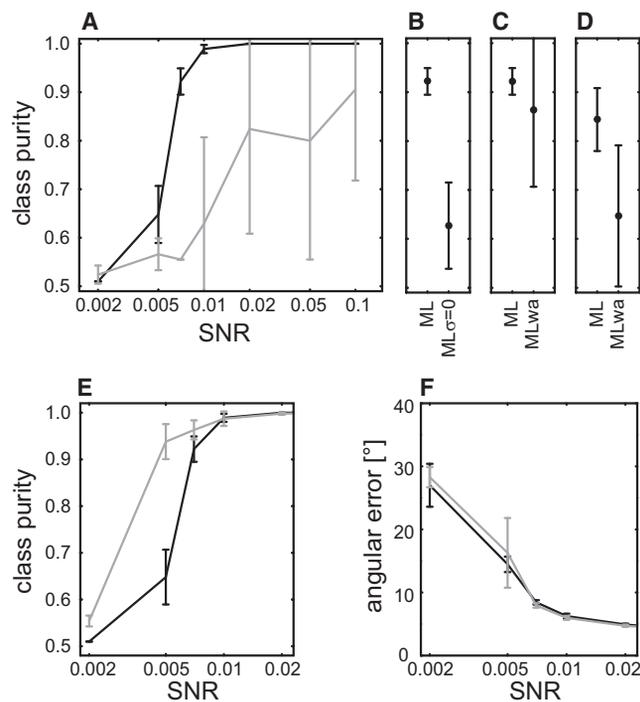


Figure 1. Model Calculations

(A) Class purities (see below) for the ML (black) and CC (gray) approach to unsupervised multireference refinement of simulated data sets with varying amounts of noise. Averages and standard deviations (error bars) are calculated over five runs that were started from different random class and orientation assignments.

(B) As in A, for simulated data with SNR = 0.07, comparing the ML approach with a modified ML approach (ML σ = 0) where the standard deviation of the noise was fixed at a value 100 times lower than the actual noise level.

(C) As in B, comparing the ML approach with a modified ML approach (MLwa) where the marginalization over the missing data regions was replaced by a weighted-averaging approach.

(D) As in C, but for data with a more severe preferred orientation of the particles.

(E) Class purities for the ML multireference refinements (as in A, black) and for the ML refinements with perfectly aligned particles (gray).

(F) Errors in the angular assignments of the groEL₁₄ complexes in the ML multireference refinements (black) and the ML refinements with structurally homogeneous subsets (gray). To calculate class purities and angular errors, class and angular assignments were based on the maximum of the probability distribution of each particle. Then, as given on p. 549 of (Tan et al., 2006), class purity was defined as

$$\frac{1}{N} \sum_{k=1}^2 \max_t m_{kt},$$

with m_{kt} being the number of particles from true structure t assigned to class k . Angular errors were calculated as the average angular distance between the three vectors of the coordinate systems defined by the perfect Euler angles and the angles obtained in the ML refinements. All 14 equivalent orientations due to the D7 symmetry of the groEL₁₄ complex were taken into account. Note that, as a consequence of this high symmetry the angular errors at low SNR (almost 30°) correspond to near-random angular assignments.

class and one class was left empty. Such behavior was not observed for the ML approach, but we note that although these solutions result in low class purities, they can be easily discerned from correct solutions.

To gain a better understanding of the ML algorithm, we then performed two additional experiments for the data set with a SNR of 0.007, which appeared to be on the limit of the amount of noise the ML approach can handle for this phantom. The first experiment was designed to test the effect of marginalizing over all orientations and classes, i.e., the calculation of probability-weighted class and orientation assignments. This was done by performing ML runs with σ fixed at a value 100 times smaller than the actual noise in the data. This effectively converts the probability distributions into approximate delta functions, yielding an algorithm that is closely related to the CC approach (cf. Sigworth, 1998), although it still complements the missing data with the current model estimates in each M-step. These runs gave significantly worse classification results than the true ML approach (Figure 1B), and detailed analyses of the convergence paths of both runs (not shown) indicated that the runs with near-delta function probability distributions have an increased tendency to get stuck into local minima at early stages of the optimization process.

The second experiment was designed to test the effect of marginalizing over the missing data regions. This was done by adapting the ML algorithm to use weighted averaging rather than marginalization over the missing data regions. That is, rather than complementing the missing wedges with the model estimates from the previous iteration, the adapted ML algorithm divides every Fourier component in the average by the number of times it was observed, while it still calculates probability-weighted averages over all orientation and class assignments. For data with the moderate preferred orientation, complementing the missing wedges yielded somewhat better class purities than weighted averaging, but the differences were smaller compared with the run with near-delta functions (Figure 1C). Much more pronounced differences were observed for data with the more severe preferred orientation (Figure 1D). For the latter data, the missing wedges upon alignment of all particles do not fill up Fourier space in the average, while for the moderate preferred orientation at least several experimental observations were available for all Fourier components. Detailed analyses of these runs revealed that, in particular for the data with the severe preferred orientation, the numerical instabilities caused by division by relatively small numbers negatively affected the optimization paths of the weighted-average algorithm. In several cases, this algorithm was even observed to diverge from reasonable solutions at intermediate stages to much worse solutions at the end of the runs.

Finally, we performed two experiments to evaluate the alignment and classification properties of our algorithm separately. First, we provided the perfect alignment parameters of the simulated data sets and applied the ML algorithm to classify these data while keeping the angular assignments fixed. As expected for the less complicated task of classifying perfectly aligned particles, purer classes could be obtained at relatively low SNRs compared with the multireference refinement schemes described above (Figure 1E). Second, we applied the ML algorithm (with $K = 1$) to align the structurally homogeneous subsets consisting of only the GroEL₁₄ complexes of the data sets described above. Remarkably, despite the less complicated task of aligning structurally homogeneous data, over the entire SNR range the resulting errors in the angular assignments

were very similar to those observed in the multireference refinements (Figure 1F). Even for SNRs where the multireference refinements gave relatively impure classes (e.g., SNR = 0.005), the errors in the angular assignments were comparable to those obtained for structurally homogenous data. Our interpretation of these results is that, in this particular case, the alignment may suffer relatively little from remaining heterogeneity in the data because of the structural resemblance of the GroEL₁₄ and GroEL₁₄GroES₇ complexes.

GroEL-GroES: Subtomogram Averaging

Unsupervised alignment and classification of experimental subtomograms was tested using a previously published test data set on purified groEL and groEL/groES complexes (Forster et al., 2008) (see [Experimental Procedures](#) for details). Several particles were discarded upon visual inspection due to remaining density for gold particles. The remaining 780 subtomograms were aligned using a single reference, which was obtained in an unsupervised manner by calculating the (weighted) average of all subtomograms in random orientations. After 25 iterations, the ML approach converged to a barrel-like structure with a clear seven-fold symmetry component (not shown). Subsequently, a second run with $K = 3$ references was started from the averages of three random subsets of the data, maintaining the angles obtained in the first run. In this run an angular sampling of 10 was employed, while the angular search range was limited to angular distances within 50 from the angles of the first run. As for the simulated data, the empirical regularization term θ was initially set to $5N/K^2$, and then linearly decreased to 0 in 5 iterations.

Central slices through the resulting maps after 25 iterations are shown in Figure 2A. Of the particles assigned to class 1, 90% came from tilt series from preparations of groEL; likewise, 86% and 100% of the particles assigned to classes 2 and 3, respectively, came from tilt series from preparations of GroELGroES complexes. The average map of class 1 showed an open, barrel-like structure with clear seven-fold symmetry and was interpreted as GroEL. Also the map of class 2 showed a barrel-like structure with seven-fold symmetry, but with partially closed, additional density on one of its ends, and this map was interpreted as GroELGroES. The map of class 3 showed a smaller structure with less clear indications for seven-fold symmetry. This class could not be directly related to GroEL or GroELGroES complexes, and possibly coincides with the classes that were tentatively interpreted as groES aggregates by (Forster et al., 2008). Because the maps of classes 1 and 2 both showed clear indications of seven-fold symmetry, C7 symmetry was enforced, and subsequent fitting of the atomic structures of these complexes (PDB IDs 1GR5 and 2C7C) in UCSF Chimera (Pettersen et al., 2004) resulted in excellent fits with cross-correlation coefficients of 0.88 and 0.81, respectively (Figures 2B and 2C).

P53-DNA: RCT Reconstruction Averaging

The usefulness of the ML approach for the unsupervised averaging of RCT reconstructions was illustrated by its application in an ongoing project of our labs on the structural characterization of human p53. Two strikingly different quaternary structures have previously been proposed for murine (Okorokov et al., 2006) and human (Tidow et al., 2007) p53 complexes. In both cases, these structures were obtained by refinement of initial

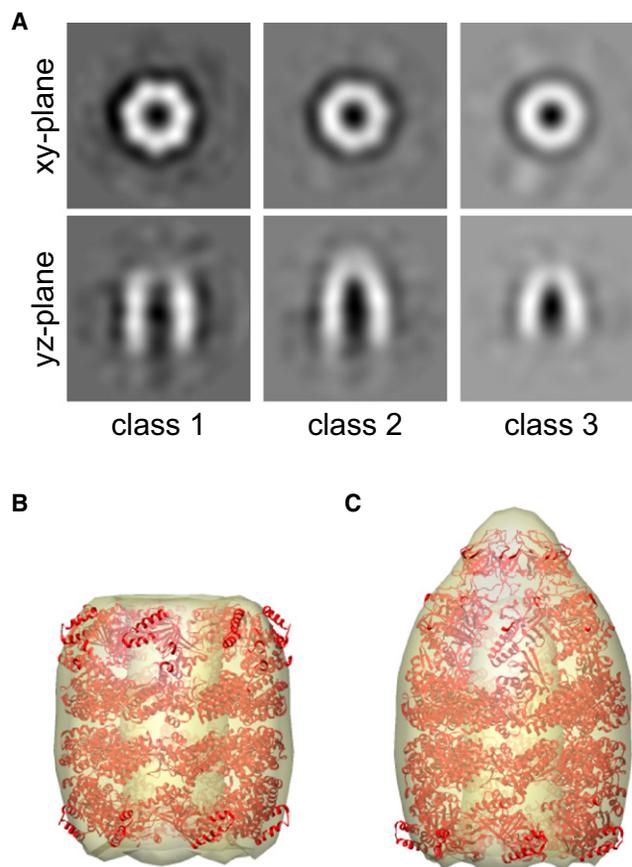


Figure 2. GroEL/GroES Subtomogram Averaging

(A) Central slices through XY and YZ of the three class averages obtained for the data set of subtomograms of groEL and groEL/groES complexes. (B) Symmetrized average subtomogram for class 2 with fitted atomic model of groEL₁₄. (C) Symmetrized average subtomogram for class 3 with fitted atomic model of groEL₁₄groES₇.

models that were based on angular reconstitution (common-lines reconstruction) (Van Heel, 1987). In this work, we set out to generate an alternative de novo model for p53 using the RCT technique, and subsequently refined this model using the EMAN package (Ludtke et al., 1999).

A data set of 2521 tilt pair particles of a stabilized p53 mutant lacking its 33 C-terminal residues (see [Experimental Procedures](#)) in complex with a dsDNA probe was processed using standard procedures in the Xmipp package (Scheres et al., 2008) to produce 40 RCT reconstructions in a range of different orientations. We then employed the ML algorithm outlined above to align these reconstructions in a completely unsupervised manner. Given the limited number of available 3D maps, the averaging of these reconstructions was performed using only a single reference (i.e., $K = 1$). Starting from a 4 nm low-pass filtered, weighted average of all RCT reconstructions in random orientations, 13 iterations of the ML algorithm were performed with an angular sampling of 15. The resulting average showed indications for two-fold symmetry (cf. Figures 3A and 3D), which was subsequently enforced on the averaged map. The symmetrized

Structure

Maximum-Likelihood Subvolume Averaging in 3D-EM

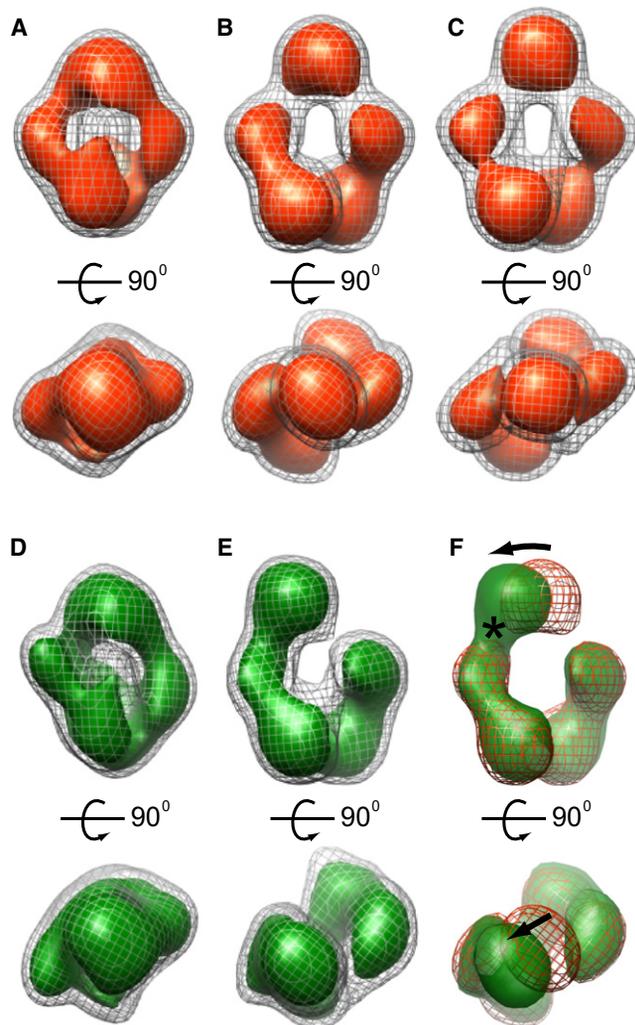


Figure 3. p53-DNA RCT Reconstruction Averaging

(A) RCT average with imposed C2 symmetry.
 (B) Refinement of the RCT average against the untilted data set, imposing C2 symmetry.
 (C) Refinement of a common-lines generated model against the untilted data set, imposing C2 symmetry.
 (D) RCT average without imposing any symmetry.
 (E) Refinement of the RCT average against the untilted data set, without imposing any symmetry.
 (F) Comparison of the refined RCT models with (orange mesh) or without (green) imposing C2 symmetry shows a departure of the top domain from the symmetry axis (arrows) and a gain in linker density with one of the core dimers (asterisk) in the asymmetric model.

All C2-symmetrized maps are shown in orange; all C1 maps are shown in green. Maps are thresholded at 50% of the expected mass in green and orange, and at 100% of the expected mass in (mesh) gray. The side views (top rows) are related to the top views (bottom rows) by a 90-degree rotation, and the putative two-fold symmetry axis is perpendicular to the plane of the top views.

map (Figure 3A) displayed a large mass at the top of the complex and four masses at the bottom, with weaker linkers connecting the top and bottom parts. This organization has remarkable similarities with the previously proposed quaternary structure for human p53, where the top part was interpreted as the tetramer-

ization domain and the bottom part as dimers of DNA-binding core domain dimers (Tidow et al., 2007).

Subsequent refinement of the average RCT reconstruction with C2 symmetry in EMAN against a second data set of 7600 (untilted) particles further increased the notion of a dimer of dimers (Figure 3B). Two additional observations validate the overall organization of this structure. First, a very similar C2-structure (Figure 3C) was obtained in a completely independent manner by refinement of a model based on common-lines reconstruction (startAny option in EMAN). Second, refinement of the asymmetric RCT model without imposing any symmetry yielded a structure that was virtually identical to the C2-structure at the bottom part, while only the mass at the top part moved slightly away from the two-fold axis (Figures 3E and 3F).

DISCUSSION

We have presented a new approach for averaging 3D maps with missing regions in Fourier space. The novelty of our approach lies in three general aspects. First, whereas many existing approaches separate alignment and classification, we tackle both problems simultaneously through multireference refinements. Second, we use maximum-likelihood estimation to calculate our model parameters as weighted averages over all possible orientation and class assignments, rather than assigning a single optimal orientation and class based on a noisy (constrained) cross-correlation function. And third, we explicitly treat the notion of incomplete experimental data in our statistical model by treating the Fourier components in the missing data regions as hidden variables. The latter results in an algorithm where one fills in the missing Fourier components with the corresponding components of the current model estimates, rather than dividing each Fourier component by the times it has been observed (as in previously proposed approaches based on weighted averaging). This gives rise to a “conservative” algorithm, which changes the model much more slowly compared with weighted averaging for those regions in Fourier space that are sparsely observed among all particles. The user may tune this degree of conservatism, because one could perform multiple iterations of marginalization over the missing data regions for every iteration of marginalization over the orientations and classes. In the limit of an infinite number of iterations of marginalization over the missing data regions, this algorithm then becomes equivalent to its weighted averaging counterpart. Although not used in the calculations presented in this paper, this option was implemented in the program and comes at virtually no additional computational cost.

The concept of complementing the missing data with the available model estimates is not restricted to the ML approach presented here. Also in approaches based on constrained cross-correlation, weighted averaging may be replaced by complementing the missing Fourier components with the current model estimates. Such changes would probably require only minor adjustments to the code of existing programs and may provide more robust algorithms for cases where the particles adopt preferred orientations. In this light, we also note the work of Bostina et al., who fill up the missing data regions of their aligned subtomograms with the corresponding values of the weighted average of all data, prior to application of standard

3D classification tools that lacked the notion of missing data (Bostina et al., 2007).

Initial tests of the presented algorithm were performed with simulated data that were in agreement with the assumptions of the underlying data model. For data with white, Gaussian noise in a wide range of SNRs, the proposed ML algorithm was shown to outperform a multireference refinement approach based on weighted averaging and constrained cross-correlation, which is in good agreement with similar observations for single-particle refinements (Scheres et al., 2005; Sigworth, 1998).

For both the ML and the CC approaches, the classification results were greatly improved using an empirical regularization term. We note that, in general, regularization may be employed to prevent overfitting or to solve ill-posed problems. Often, regularization terms tend to smooth the energy landscape of a target function, thereby lowering the number of local minima. Also in this case, our understanding of this effect is that the regularization prevents the multireference refinement runs from getting stuck in local minima at early stages of the refinement. During the initial iterations, the quality of the references is still relatively low due to suboptimal alignment parameters, and separation of the images into distinct groups at these stages may converge to mediocre solutions. The regularization term that acts during these initial iterations enforces similarity among the references, so that these local minima may be overcome. Starting from initially random assignments of rotations and classes, the ML algorithm was then shown to be capable of yielding simultaneous classification and alignment in a completely unsupervised manner.

As explained above, the main differences of the ML algorithm with conventional approaches lie in the probability-weighted averaging over all possible orientational and class assignments on one hand and the filling of the missing regions with the current model estimates on the other hand. In experiments designed to separate these two types of marginalization, in particular the probability-weighted averaging over the orientation and class assignments appeared to have a large effect on the refinement results. Complementing the missing data regions seems to be of secondary importance, although this effect became more noticeable for data with severe preferred orientations where the missing regions were not filled up in the averages.

The experimental data set of GroEL/groES subtomograms that we used was previously published by Foerster et al., who used constrained cross-correlation to align all particles against a reference based on a low-pass filtered crystal structure, and then classified these data using hierarchical clustering based on the same similarity measure (Foerster et al., 2008). We obtained similar classification results, but in a completely unsupervised manner. Moreover, the maps for the groEL₁₄ and groES₁₄groES₇ complexes obtained in our work appear to agree somewhat better with the fitted crystal structures than those presented by Foerster et al., although a quantitative comparison is not available at this point. Apart from the theoretical considerations outlined above, such an improvement could also be explained by the simultaneous alignment and classification in our refinement scheme, whereas Foerster et al. separated alignment against a single reference from subsequent classification and did not re-align the resulting classes.

The reference-free alignment of p53-DNA RCT reconstructions showed that, despite a low number of noisy RCT reconstructions, the ML approach was capable of generating an averaged map with remarkable similarities to a previously observed superdomain organization of human p53 (Tidow et al., 2007). Apart from illustrating the potentials of the ML approach in generating de novo models for subsequent single-particle refinement, these results also have relevance for p53 structural biology. Previously, two strikingly different quaternary structures were proposed based on common-lines EM-reconstructions. Tidow et al. (2007) used negative-stain EM to study a stabilized mutant of human full-length p53. Enforcing C2 symmetry, they obtained models with similar domain organizations as those shown in Figure 3, either for free p53 tetramers and for complexes with dsDNA. (Okorokov et al., 2006) used cryo-EM on full-length murine p53, and enforced D2 symmetry to obtain a model with the shape of a hollow, skewed cube. As p53 may adopt an intrinsically flexible quaternary organization to fulfill its multiple functionalities in the cell, it is possible that both EM reconstructions represent relevant oligomerization states (Okorokov and Orlova, 2009). On the other hand, some caution may be at place, because the imposition of an incorrect symmetry in the generation and refinement of common-lines models may lead to important artifacts, and the inherent flexibility combined with its small size (200 kDa) further complicate (cryo-) EM analysis of p53 samples. Our RCT averaging experiments yielded the first de novo p53 model that is not based on common lines. Moreover, whereas the previously proposed common-lines models could not be obtained without imposing symmetry, our model could be generated and refined in the absence of any symmetry.

Summarizing, this contribution provides a general framework for maximum-likelihood refinement of 3D maps with missing data, and the usefulness of this approach has been illustrated using both simulated and experimental data of subtomograms and RCT reconstructions. As has happened previously with similar approaches the field of single-particle analysis, it is likely that the underlying statistical data model may need to be adapted to describe a wide variety of experimental characteristics. Apart from the theoretical considerations about the data model outlined above, neighboring particles from the crowded environment of the cell could interfere with the correct normalization of subtomograms, as was previously observed for single-particle analyses (Scheres et al., 2009). In addition, densities in cellular subtomograms for neighboring complexes, gold particles, or membranes that are not described by the data model may lead to artifacts in the alignment and classification process, whereas real-space masks (to remove such artifacts) may lead to additional dependencies in Fourier space and to underestimation of the noise variance. To facilitate further testing and possible adaptation in the field, the presented algorithm has been made freely available through its implementation in the open-source Xmipp package (Sorzano et al., 2004). Extrapolating our experience with maximum likelihood refinement of structurally heterogeneous single-particle projection data (Scheres et al., 2005, 2007a), we expect that this implementation may provide a useful tool for unsupervised alignment and classification of electron subtomograms or RCT reconstructions.

Structure

Maximum-Likelihood Subvolume Averaging in 3D-EM

EXPERIMENTAL PROCEDURES

GroEL/groES Subtomogram Data

The tomographic data on groEL/groES mixtures used in this work was published previously (Forster et al., 2008). These authors collected single-axis tilt series on purified protein samples of GroEL and GroELGroES complexes mixed with a 10 nm BSA-colloidal gold suspension. Carbon-coated grids were vitrified by plunge freezing, and low-dose images were collected using a Tecnai G2 Polara microscope (FEI, Eindhoven, The Netherlands) equipped with a 2k × 2k CCD camera. The tilt series ranged from -65 to 65 with a 2–2.5 increment, defocus values between 4 and 7 μm were used, and the pixel size at specimen level was 0.6 nm. Manual particle selection yielded several thousand subtomograms, which were downsampled to 32 × 32 × 32 voxels with a final voxel size of 1.2 nm. From this data set, 793 subtomograms were selected based on a cross-correlation criterion with the average of all subtomograms aligned against a low-pass-filtered crystal structure of the GroELGroES complex (Xu et al., 1997).

P53 Electron Microscopy

A deletion mutant of human p53, lacking its 33 C-terminal residues (NTC, p53₁₋₃₆₀) and with four stabilizing mutations in the core domain (M133L/V203A/N239Y/N268D) (Nikolova et al., 1998), was expressed in *E. coli*, and purified as described elsewhere (Veprintsev et al., 2006). Complexes with a double-stranded DNA probe containing the TGFA response element were prepared at a high concentration (10 μM) to avoid free p53. These samples were simultaneously separated on molecular weight and cross-linked with glutaraldehyde using the Grafix method (Kastner et al., 2008). For all fractions, the interaction of the DNA probe with p53 was confirmed by electrophoretic mobility shift assay, and the size of the complex was confirmed using SDS-PAGE. Only fractions containing p53 tetramers in complex with DNA were selected for EM. Negative staining was performed with uranyl formiate in carbon sandwich technique over Quantifoil grids. Micrographs were taken in tilted pairs, at tilt angles of 0° and 45°, under low-dose conditions on a JEOL JEM-2200FS electron microscope. Images were recorded on a 4k × 4k CCD camera at a magnification of 64,305 ×, corresponding to a 2.1 Å pixel size. A total of 2521 pairs of particles were manually selected, downsampled to a pixel size of 8.4 Å, and processed using standardized RCT protocols in Xmipp (Scheres et al., 2008) to generate 40 reconstructions of 30 × 30 × 30 voxels. Similarly, a second data set consisting of 7600 untilted particles was obtained using the same protein sample, but incubated with a dsDNA probe containing the GADD45 response element.

Implementation

The ML algorithm outlined above was implemented in the *xmipp_ml_tomo* program of the Xmipp package (Sorzano et al., 2004). The same program also implements a standard multirefinement scheme based on weighted averaging and maximum constrained cross-correlations (CC). In addition, although we did not derive the corresponding formulas here, the program implements a modified maximum-likelihood algorithm that employs weighted averaging to calculate the reference structures. The latter two implementations depend on an additional free parameter in the form of a threshold value to prevent divisions by zero, which was set to 0.1 in all calculations. Although the program reads 3D maps in single-file SPIDER format only, the Xmipp package provides utilities to convert to and from other common formats like MRC (Crowther et al., 1996) or EM (Hegerl, 1996). Missing data regions may be defined as cones, wedges, or pyramids, and the program may also be used to align 3D maps without missing regions. The 6D integrations over all orientational and translational assignments are performed exhaustively, although the orientational integrations may optionally be limited to a given angular distance from a set of initial orientations for all particles. The integrations over all orientations are replaced by discrete Riemann sums over approximately even distributions of the three Euler angles. It is noteworthy that these distributions are oriented differently according to a small random perturbation at every iteration to prevent the optimization from getting stuck in local minima. The program has been parallelized with near-linear efficiency for the numbers of processors employed in this study, using a hybrid scheme of distributed and shared memory parallelization by the combined use of the message passing interface (MPI) and threads. The computational cost of the

implementation scales linearly with the number of particles, the number of classes and the number of sampled rotations. In addition, it scales approximately linearly with the number of voxels in the 3D maps. The runs with the model data took approximately 7–8 hr wall clock time each, using 20 3GHz Intel Xeon processors in parallel. The groEL runs took in total 17 hr on 128 2.3GHz IBM Power PC processors, and the alignment of the p53 maps took 20 min on 20 3GHz Intel Xeon processors.

SUPPLEMENTAL DATA

Supplemental Data include a Supplemental Approach and can be found with this article online at [http://www.cell.com/structure/supplemental/S0969-2126\(09\)00411-0](http://www.cell.com/structure/supplemental/S0969-2126(09)00411-0).

ACKNOWLEDGMENTS

We thank Michael Stoelken, and Stephan Nickell and Friedrich Foerster for useful comments, the latter for providing the groEL/groES data, Alan Fersht for providing p53 samples, and David Gil and Melissa Lazaro for help with p53 data collection. We are grateful to the Barcelona Supercomputing Center (BSC-CNS) for providing computer resources. Funding was provided by the Spanish Ministry of Science (CSD2006-00023, BIO2007-67150-C03-1/3) and the Comunidad de Madrid (S-GEN-0166-2006), the European Union (FP6-502828), the U.S. National Heart, Lung and Blood Institute (R01 HL070472). The content of this work is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung and Blood Institute.

Received: August 7, 2009

Revised: October 2, 2009

Accepted: October 22, 2009

Published: December 8, 2009

REFERENCES

- Bartesaghi, A., Sprechmann, P., Liu, J., Randall, G., Sapiro, G., and Subramaniam, S. (2008). Classification and 3D averaging with missing wedge correction in biological electron tomography. *J. Struct. Biol.* 162, 436–450.
- Beck, M., Forster, F., Ecke, M., Plitzko, J.M., Melchior, F., Gerisch, G., Baumeister, W., and Medalia, O. (2004). Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. *Science* 306, 1387–1390.
- Bostina, M., Bubeck, D., Schwartz, C., Nicastro, D., Filman, D.J., and Hogle, J.M. (2007). Single particle cryoelectron tomography characterization of the structure and structural variability of poliovirus-receptor-membrane complex at 30 Å resolution. *J. Struct. Biol.* 160, 200–210.
- Brandt, F., Etchells, S.A., Ortiz, J.O., Elcock, A.H., Hartl, F.U., and Baumeister, W. (2009). The native 3D organization of bacterial polysomes. *Cell* 136, 261–271.
- Crowther, R.A., Henderson, R., and Smith, J.M. (1996). MRC image processing programs. *J. Struct. Biol.* 116, 9–16.
- Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum-likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Series B Stat. Methodol.* 39, 1–38.
- Dube, P., Tavares, P., Lurz, R., and van Heel, M. (1993). The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry. *EMBO J.* 12, 1303–1309.
- Forster, F., and Hegerl, R. (2007). Structure determination in situ by averaging of tomograms. *Methods Cell Biol.* 79, 741–767.
- Forster, F., Medalia, O., Zauberman, N., Baumeister, W., and Fass, D. (2005). Retrovirus envelope protein complex structure in situ studied by cryo-electron tomography. *Proc. Natl. Acad. Sci. USA* 102, 4729–4734.
- Forster, F., Pruggnaller, S., Seybert, A., and Frangakis, A.S. (2008). Classification of cryo-electron sub-tomograms using constrained correlation. *J. Struct. Biol.* 161, 276–286.
- Frangakis, A.S., Bohm, J., Forster, F., Nickell, S., Nicastro, D., Typke, D., Hegerl, R., and Baumeister, W. (2002). Identification of macromolecular

- complexes in cryoelectron tomograms of phantom cells. *Proc. Natl. Acad. Sci. USA* **99**, 14153–14158.
- Frank, J. (2006). *Three-dimensional Electron Microscopy of Macromolecular Assemblies* (New York: Oxford University Press).
- Hegerl, R. (1996). The EM Program Package: A Platform for Image Processing in Biological Electron Microscopy. *J. Struct. Biol.* **116**, 30–34.
- Hoenger, A., and McIntosh, J.R. (2009). Probing the macromolecular organization of cells by electron tomography. *Curr. Opin. Cell Biol.* **21**, 89–96.
- Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., et al. (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5**, 53–55.
- Leis, A., Rockel, B., Andrees, L., and Baumeister, W. (2009). Visualizing cells at the nanoscale. *Trends Biochem. Sci.* **34**, 60–70.
- Liu, J., Bartesaghi, A., Borgnia, M.J., Sapiro, G., and Subramaniam, S. (2008). Molecular architecture of native HIV-1 gp120 trimers. *Nature* **455**, 109–113.
- Llorca, O. (2005). Introduction to 3D reconstruction of macromolecules using single particle electron microscopy. *Acta Pharmacol. Sin.* **26**, 1153–1164.
- 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.
- Malmstrom, J., Beck, M., Schmidt, A., Lange, V., Deutsch, E.W., and Aebersold, R. (2009). Proteome-wide cellular protein concentrations of the human pathogen *Leptospira interrogans*. *Nature* **460**, 762–765.
- Nickell, S., Kofler, C., Leis, A.P., and Baumeister, W. (2006). A visual approach to proteomics. *Nat. Rev. Mol. Cell Biol.* **7**, 225–230.
- Nikolova, P.V., Henckel, J., Lane, D.P., and Fersht, A.R. (1998). Semirational design of active tumor suppressor p53 DNA binding domain with enhanced stability. *Proc. Natl. Acad. Sci. USA* **95**, 14675–14680.
- Okorokov, A.L., and Orlova, E.V. (2009). Structural biology of the p53 tumour suppressor. *Curr. Opin. Struct. Biol.* **19**, 197–202.
- Okorokov, A.L., Sherman, M.B., Plisson, C., Grinkevich, V., Sigmundsson, K., Selivanova, G., Milner, J., and Orlova, E.V. (2006). The structure of p53 tumour suppressor protein reveals the basis for its functional plasticity. *EMBO J.* **25**, 5191–5200.
- Pascual-Montano, A., Taylor, K.A., Winkler, H., Pascual-Marqui, R.D., and Carazo, J.M. (2002). Quantitative self-organizing maps for clustering electron tomograms. *J. Struct. Biol.* **138**, 114–122.
- Penczek, P.A., Grassucci, R.A., and 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.
- 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.
- Radermacher, M. (2009). Chapter 1 Visualizing functional flexibility by three-dimensional electron microscopy reconstructing complex I of the mitochondrial respiratory chain. *Methods Enzymol.* **456**, 3–27.
- Radermacher, M., Wagenknecht, T., Verschoor, A., and Frank, J. (1986). A new 3-D reconstruction scheme applied to the 50S ribosomal subunit of *E. coli*. *J. Microsc.* **141**, RP1–RP2.
- Sander, B., Golas, M.M., Makarov, E.M., Brahm, H., Kastner, B., Luhrmann, R., and Stark, H. (2006). Organization of core spliceosomal components U5 snRNA loop I and U4/U6 Di-snRNP within U4/U6.U5 Tri-snRNP as revealed by electron cryomicroscopy. *Mol. Cell* **24**, 267–278.
- Scheres, S.H., Valle, M., Nunez, R., Sorzano, C.O., Marabini, R., Herman, G.T., and Carazo, J.M. (2005). Maximum-likelihood multi-reference refinement for electron microscopy images. *J. Mol. Biol.* **348**, 139–149.
- Scheres, S.H., Gao, H., Valle, M., Herman, G.T., Eggermont, P.P., Frank, J., and Carazo, J.M. (2007a). Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nat. Methods* **4**, 27–29.
- Scheres, S.H., Nunez-Ramirez, R., Gomez-Llorente, Y., San Martin, C., Eggermont, P.P., and Carazo, J.M. (2007b). Modeling experimental image formation for likelihood-based classification of electron microscopy data. *Structure* **15**, 1167–1177.
- Scheres, S.H., Nunez-Ramirez, R., Sorzano, C.O., Carazo, J.M., and Marabini, R. (2008). Image processing for electron microscopy single-particle analysis using XMIPP. *Nat. Protoc.* **3**, 977–990.
- Scheres, S.H., Valle, M., Grob, P., Nogales, E., and Carazo, J.M. (2009). Maximum likelihood refinement of electron microscopy data with normalization errors. *J. Struct. Biol.* **166**, 234–240.
- Schmid, M.F., and Booth, C.R. (2008). Methods for aligning and for averaging 3D volumes with missing data. *J. Struct. Biol.* **161**, 243–248.
- Sigworth, F.J. (1998). A maximum-likelihood approach to single-particle image refinement. *J. Struct. Biol.* **122**, 328–339.
- 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.
- Stahlberg, H., and Walz, T. (2008). Molecular electron microscopy: state of the art and current challenges. *ACS Chem. Biol.* **3**, 268–281.
- Tan, P.N., Steinbach, M., and Kumar, V. (2006). *Introduction to Data Mining* (Addison-Wesley).
- Tidow, H., Melero, R., Mylonas, E., Freund, S.M., Grossmann, J.G., Carazo, J.M., Svergun, D.I., Valle, M., and Fersht, A.R. (2007). Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex. *Proc. Natl. Acad. Sci. USA* **104**, 12324–12329.
- Van Heel, M. (1987). Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy* **21**, 111–123.
- Veprintsev, D.B., Freund, S.M., Andreeva, A., Rutledge, S.E., Tidow, H., Canadillas, J.M., Blair, C.M., and Fersht, A.R. (2006). Core domain interactions in full-length p53 in solution. *Proc. Natl. Acad. Sci. USA* **103**, 2115–2119.
- Walz, J., Typke, D., Nitsch, M., Koster, A.J., Hegerl, R., and Baumeister, W. (1997). Electron tomography of single ice-embedded macromolecules: three-dimensional alignment and classification. *J. Struct. Biol.* **120**, 387–395.
- Winkler, H., Zhu, P., Liu, J., Ye, F., Roux, K.H., and Taylor, K.A. (2009). Tomographic subvolume alignment and subvolume classification applied to myosin V and SIV envelope spikes. *J. Struct. Biol.* **165**, 64–77.
- Xu, Z., Horwich, A.L., and Sigler, P.B. (1997). The crystal structure of the asymmetric GroEL-GroES-(ADP)₇ chaperonin complex. *Nature* **388**, 741–750.