

Universidad Complutense de Madrid  
Facultad de Ciencias Químicas  
Departamento de Bioquímica y Biología Molecular I



**Nuevos métodos para el ajuste flexible de  
estructuras macromoleculares a distintas  
resoluciones empleando modos normales de  
vibración en coordenadas internas**

Tesis Doctoral  
*José Ramón López Blanco*  
**2012**





# Universidad Complutense de Madrid

## Facultad de Ciencias Químicas

### Departamento de Bioquímica y Biología Molecular I



Memoria presentada para optar al grado de  
Doctor por

*José Ramón López Blanco*

Esta tesis doctoral ha sido llevada a cabo en los Departamentos de Biología Físico-Química del Centro de Investigaciones Biológicas y de Química Física Biológica del Instituto de Química-Física Rocasolano (ambas instituciones del CSIC) bajo la dirección del Dr. Pablo Chacón Montes.

Firma del Doctorando

Firma manuscrita de José Ramón López Blanco en tinta azul.

José Ramón López Blanco

Conformidad del Director

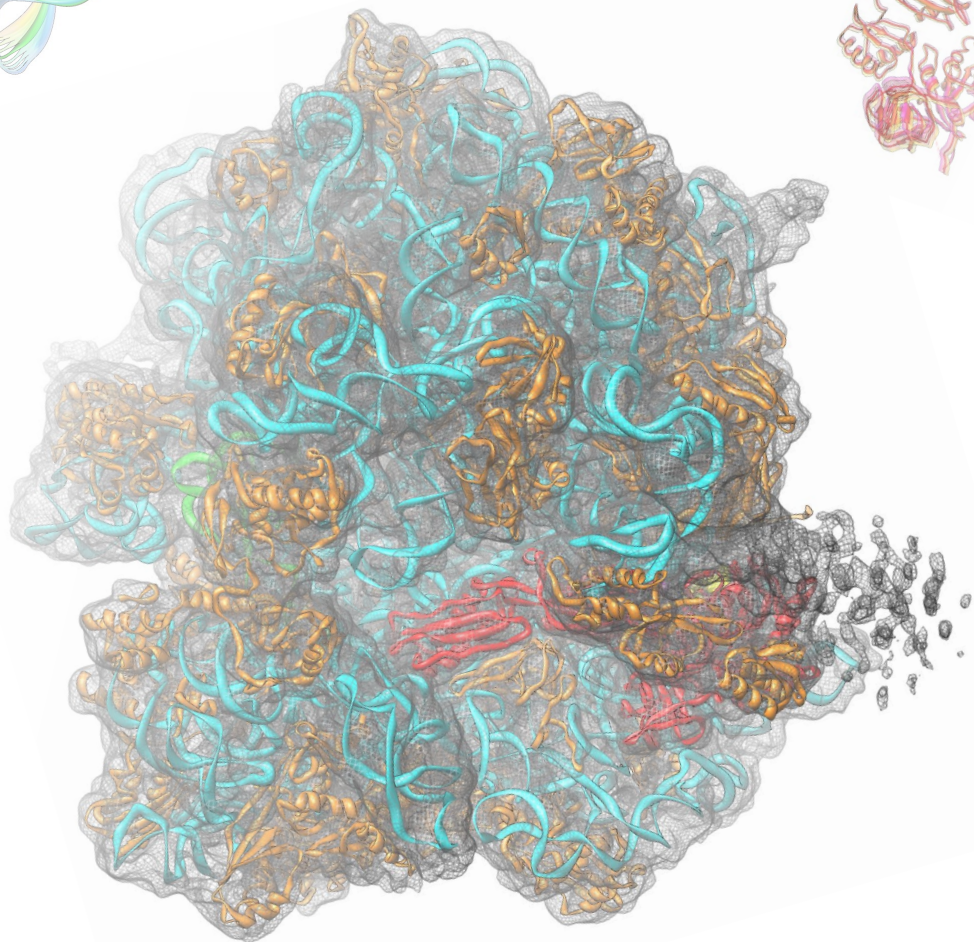
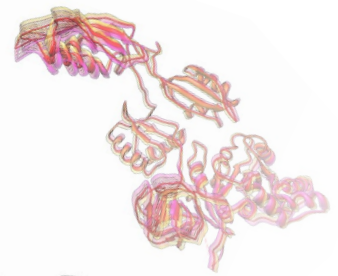
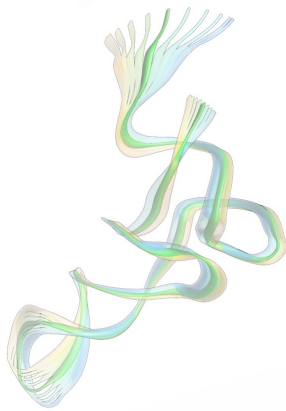
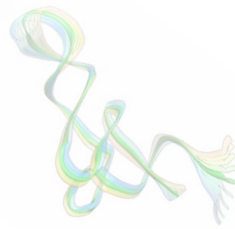
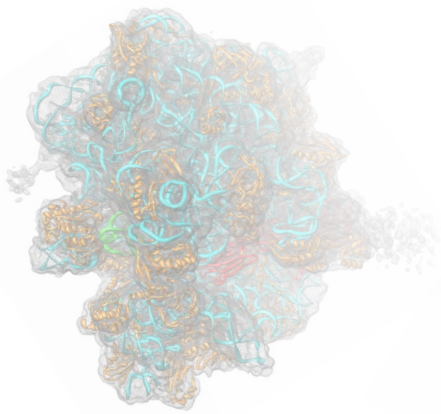
Firma manuscrita de Pablo Chacón Montes en tinta azul.

Dr. Pablo Chacón Montes





*A toda mi familia...*





---

# Índice general

<b>1. Motivación</b>	1
<b>2. Objetivos</b>	11
<b>3. Introducción</b>	13
3.1. Análisis vibracional	13
3.1.1. El modelo físico	13
3.1.2. Modelos reduccionistas	15
3.1.3. Análisis vibracional en coordenadas internas	19
3.2. Validación del análisis vibracional	19
3.2.1. Correlación con las fluctuaciones atómicas	19
3.2.2. Correspondencia con dinámica esencial.	21
3.2.3. Correlación con conjuntos de conformaciones obtenidos por RMN	22
3.3. Aplicaciones del análisis vibracional	24
3.3.1. Predicción de flexibilidad	24
3.3.2. Refinamiento estructural.	25
3.3.3. Transiciones conformacionales	26
3.3.4. Predicción de interacciones proteína-proteína	26
3.3.5. Predicción de interacciones proteína-ligando	27
3.3.6. Respuesta mecánica	28
3.3.7. Ajuste flexible de mapas de microscopía electrónica	29
3.4. Microscopía electrónica y análisis de partículas individuales	29
3.4.1. Preparación de la muestra	30
3.4.2. Toma de micrografías	31
3.4.3. Procesamiento de imagen y reconstrucción tridimensional	32
3.5. Ajuste macromolecular	34
3.5.1. Ajuste rígido	34
3.5.2. Ajuste flexible	35
3.5.3. Métodos de ajuste flexible basados en el análisis vibracional.	35

3.5.4. Métodos de ajuste flexible basados en dinámica molecular . . . . .	36
3.5.5. Métodos de ajuste flexible basados en otras aproximaciones . . . . .	40
<b>4. Materiales y métodos . . . . .</b>	<b>45</b>
4.1. Análisis vibracional . . . . .	45
4.1.1. Análisis vibracional en coordenadas generalizadas . . . . .	45
4.1.2. Análisis vibracional en coordenadas cartesianas . . . . .	46
4.1.3. Análisis vibracional en coordenadas internas . . . . .	47
4.1.4. Implementación del método rápido de Go – Cálculo del Hessiano . . . . .	49
4.1.5. Implementación del método rápido de Go – Cálculo del tensor métrico . . . . .	54
4.1.6. Mover en coordenadas internas . . . . .	55
4.2. Representaciones reduccionistas . . . . .	57
4.2.1. Modelos de coordenadas internas . . . . .	57
4.2.2. Modelos atómicos . . . . .	59
4.2.3. Modelos de red elástica . . . . .	60
4.3. Aplicaciones del análisis vibracional en coordenadas internas . . . . .	61
4.3.1. Análisis vibracional – iMode . . . . .	61
4.3.2. Visualización - iModview . . . . .	63
4.3.3. Animación de modos - iMove . . . . .	63
4.3.4. Simulación de Monte Carlo - iMC . . . . .	64
4.3.5. Simulación de transiciones conformacionales - iMorph . . . . .	65
4.3.6. Ajuste flexible de estructuras atómicas en mapas de densidad de microscopía electrónica - iModfit. . . . .	67
4.4. Reconstrucción tridimensional de RepB y su complejo con ADN . . . . .	69
4.4.1. Preparación de las muestras . . . . .	69
4.4.2. Adquisición de datos. . . . .	69
4.4.3. Procesamiento de imagen y reconstrucción tridimensional . . . . .	69
4.5. Herramientas auxiliares. . . . .	71
4.5.1. Métricas de comparación de modos normales. . . . .	71
4.5.2. Actualización del programa de análisis vibracional – DefProt . . . . .	72
4.5.3. Dinámica esencial y análisis de componentes principales – iPca . . . . .	72
4.6. Datos de validación . . . . .	73

4.6.1. Conjuntos de pruebas de estructuras atómicas . . . . .	73
4.6.2. Casos de ajuste flexible con mapas simulados . . . . .	76
4.6.3. Casos de ajuste flexible con ruido simulado. . . . .	76
4.6.4. Casos de ajuste flexible basados en modelado por homología . . . . .	77
4.7. Ajuste flexible con otras herramientas . . . . .	77
4.8. Detalles técnicos . . . . .	79
<b>5. Resultados y discusión . . . . .</b>	<b>81</b>
5.1. Diseño e implementación . . . . .	81
5.1.1. Implementación del análisis vibracional en coordenadas internas . . . . .	81
5.1.2. Eficiencia del análisis vibracional en coordenadas internas y cartesianas . . . . .	85
5.2. Validación del análisis vibracional en coordenadas internas . . . . .	88
5.3. Validación de las representaciones reduccionistas . . . . .	92
5.3.1. Validación de los modelos atómicos. . . . .	92
5.3.2. Validación de la reducción del número de coordenadas internas . . . . .	93
5.4. Validación del análisis vibracional en ARN. . . . .	94
5.5. Validación del análisis vibracional en sistemas muy grandes . . . . .	95
5.6. Simulación de trayectorias de transiciones conformacionales. . . . .	96
5.6.1. Simulación de trayectorias de transiciones conformacionales en proteínas. . . . .	96
5.6.2. Simulación de trayectorias de transiciones conformacionales con distintos modelos atómicos de proteínas . . . . .	99
5.6.3. Simulación de trayectorias de transiciones conformacionales en ARN. . . . .	100
5.6.4. Simulación de trayectorias de transiciones conformacionales reduciendo el número de coordenadas en proteínas y ARN . . . . .	100
5.7. Validación de las simulaciones de Monte Carlo . . . . .	101
5.7.1. Muestreo restringido. . . . .	103
5.8. Validación de la herramienta de ajuste flexible – iModfit . . . . .	104
5.8.1. Validación de iModfit en casos simulados de proteínas . . . . .	104
5.8.2. Validación de los modelos atómicos reduccionistas de proteínas . . . . .	106
5.8.3. Validación de la reducción del número de coordenadas internas en proteínas . . . . .	111

---

5.8.4. Dependencia de los ajustes con el número de modos . . . . .	118
5.8.5. Dependencia de los ajustes con la orientación inicial . . . . .	120
5.8.6. Validación de iModfit en casos simulados de ARN . . . . .	121
5.8.7. Comparación con otros métodos basados en análisis vibracional. . . . .	126
5.8.8. Comparación con métodos basados en dinámica molecular . . . . .	130
5.8.9. Comparación con métodos basados en dinámica molecular utilizando estructuras modeladas por homología . . . . .	136
5.8.10. Comparación con métodos basados en otras aproximaciones . . . . .	140
5.8.11. Efecto del ruido en los modelos obtenidos con iModfit, NMFF y YUPSCX .143	
5.8.12. Comparación entre los tiempos de cálculo de iModfit, NMFF y YUPSCX .	148
5.8.13. Validación de iModfit en casos experimentales de ajuste flexible . . . . .	149
5.8.14. Comparación de iModfit con NMFF y YUPSCX en casos experimentales .	171
5.8.15. Validación de iModfit con mapas experimentales del ribosoma . . . . .	174
5.9. Reconstrucciones tridimensionales del sistema RepB. . . . .	177
5.9.1. Resolviendo la ambigüedad cristalográfica . . . . .	178
5.9.2. Reconstrucción en presencia de ADN. . . . .	182
5.9.3. Ajuste flexible de las reconstrucciones de RepB con iModfit . . . . .	182
5.10. Disseminación de la metodología desarrollada . . . . .	185
<b>6. Conclusiones . . . . .</b>	<b>189</b>
<b>7. Apéndice . . . . .</b>	<b>193</b>
<b>8. Bibliografía . . . . .</b>	<b>195</b>

---

# Índice de tablas

<b>Tabla 1.</b> Solapamiento entre los espacios vectoriales descritos por NMA y por los componentes principales de conjuntos de estructuras de RMN.....	23
<b>Tabla 2.</b> Servidores web basados en NMA. ....	24
<b>Tabla 3.</b> Conjuntos de pruebas de transiciones conformacionales.....	74
<b>Tabla 4.</b> Casos de ajuste flexible basados en estructuras de modelado por homología. ....	78
<b>Tabla 5.</b> Comparación entre los tiempos de cálculo del NMA en CI y CC empleando estructuras de proteínas. ....	86
<b>Tabla 6.</b> Solapamientos de los vectores de transición de proteínas con los modos normales obtenidos en CI y CC ....	90
<b>Tabla 7.</b> Solapamientos de los vectores de transición de proteínas y ARN con los modos normales de representaciones reduccionistas.....	92
<b>Tabla 8.</b> Comparación entre los tiempos de cálculo del NMA en CI y CC con ARN. ....	96
<b>Tabla 9.</b> Resultados promedio de las simulaciones de transiciones conformacionales. ....	97
<b>Tabla 10.</b> Exploración conformacional basada en los modos normales en CI y muestreo de MC.....	102
<b>Tabla 11.</b> Resultados promedio de los ajustes flexibles de proteínas realizados con el modelo atómico HA y el 5% de los modos en función de la resolución. ....	105
<b>Tabla 12.</b> Comparación entre los ajustes flexibles de proteínas realizados con los distintos modelos atómicos y el 5% de los modos en función de la resolución.....	108
<b>Tabla 13.</b> Comparación entre los ajustes flexibles de proteínas obtenidos fijando aleatoriamente distintos porcentajes de ángulos diedros con el modelo atómico HA en función de la resolución.....	110
<b>Tabla 14.</b> Comparación entre los ajustes flexibles de proteínas obtenidos al fijar aleatoriamente distintos porcentajes de ángulos diedros con el modelo atómico C5 en función de la resolución.....	116
<b>Tabla 15.</b> Comparación entre los ajustes flexibles de proteínas obtenidos al fijar aleatoriamente distintos porcentajes de ángulos diedros con el modelo atómico C $\alpha$ en función de la resolución.....	117

---

<b>Tabla 16.</b>	Comparación entre los ajustes flexibles de proteínas obtenidos empleando distintos porcentajes de modos con el modelo atómico HA en función de la resolución..	119
<b>Tabla 17.</b>	Comparación del efecto de la orientación inicial en los ajustes flexibles de proteínas con el 5% de los modos y el modelo atómico HA.....	120
<b>Tabla 18.</b>	Resultados promedio de los ajustes flexibles de ARN realizados con el 5% de los modos en función de la resolución..	121
<b>Tabla 20.</b>	Comparación entre los ajustes flexibles de ARN obtenidos al fijar aleatoriamente distintos porcentajes de ángulos diedros en función de la resolución. ....	124
<b>Tabla 19.</b>	Resultados promedio de los ajustes flexibles de ARN realizados con el 5% de los modos y 1Å de umbral de re-diagonalización en función de la resolución.....	124
<b>Tabla 21.</b>	Comparación entre los ajustes flexibles de ARN obtenidos con distintos porcentajes de modos en función de la resolución. ....	125
<b>Tabla 22.</b>	Comparación de <i>iModfit</i> con otros métodos de ajuste flexible basados en NMA y CC. ....	126
<b>Tabla 23.</b>	Comparación entre los ajustes flexibles de proteínas llevados a cabo con <i>NMFF</i> a distintas resoluciones.....	128
<b>Tabla 24.</b>	Comparación entre los ajustes flexibles de proteínas llevados a cabo con <i>YUPSCX</i> e <i>iModfit</i> a distintas resoluciones. ....	131
<b>Tabla 25.</b>	Comparación entre <i>iModfit</i> y <i>MDFF</i> utilizando ajustes con mapas simulados. .	133
<b>Tabla 26.</b>	Comparación entre <i>iModfit</i> y el método Tama 1 utilizando ajustes con mapas simulados.....	134
<b>Tabla 27.</b>	Comparación entre <i>iModfit</i> y el método Tama 2 utilizando ajustes con mapas simulados.....	135
<b>Tabla 28.</b>	Comparación de <i>iModfit</i> con otros métodos basados en MD utilizando el conjunto de pruebas de modelado por homología.....	137
<b>Tabla 29.</b>	Comparación entre <i>iModfit</i> y el método de Zheng utilizando ajustes con mapas simulados y distintos niveles de ruido.....	142
<b>Tabla 30.</b>	Evaluación de la calidad de las estructuras obtenidas con <i>iModfit</i> utilizando mapas simulados y distintos niveles de ruido.....	144



---

<b>Tabla 31.</b> Desviaciones finales de los modelos obtenidos con <i>iModfit</i> , <i>NMFF</i> y <i>YUPSCX</i> utilizando mapas simulados y distintos niveles de ruido. . . . .	145
<b>Tabla 32.</b> Evaluación de la calidad de las estructuras obtenidas con <i>YUPSCX</i> utilizando mapas simulados y distintos niveles de ruido. . . . .	146
<b>Tabla 33.</b> Evaluación de la calidad de las estructuras obtenidas con <i>NMFF</i> utilizando mapas simulados y distintos niveles de ruido. . . . .	147
<b>Tabla 34.</b> Comparación de los tiempos de ajuste flexible con <i>iModfit</i> , <i>NMFF</i> y <i>YUPSCX</i> . .	148
<b>Tabla 35.</b> Casos de ajuste flexible con mapas experimentales. . . . .	150
<b>Tabla 36.</b> Resultados del ajuste flexible con <i>iModfit</i> utilizando mapas experimentales. . .	151
<b>Tabla 37.</b> Comparativa entre los métodos <i>YUPSCX</i> , <i>NMFF</i> e <i>iModfit</i> utilizando mapas experimentales. . . . .	172
<b>Tabla 38.</b> Comparativa entre la calidad estructural de los modelos finales obtenidos con <i>YUPSCX</i> , <i>NMFF</i> , e <i>iModfit</i> utilizando mapas experimentales. . . . .	173
<b>Tabla A1.</b> Solapamientos individuales de los vectores de transición con los modos normales de proteínas obtenidos en CI. . . . .	193
<b>Tabla A2.</b> Solapamientos individuales de los vectores de transición con los modos normales de ARN obtenidos en CI. . . . .	193
<b>Tabla A3.</b> Resultados individuales de las simulaciones de transiciones conformacionales de proteínas. . . . .	194
<b>Tabla A4.</b> Resultados individuales de las simulaciones de transiciones conformacionales de ARN. . . . .	194

---

---

# Índice de figuras

<b>Figura 1.</b> Algunas macromoléculas biológicas de gran tamaño y su localización celular. . . . .	1
<b>Figura 2.</b> Comparación de las principales técnicas experimentales para obtener información estructural de complejos macromoleculares de gran tamaño . . . . .	2
<b>Figura 3.</b> Ajuste macromolecular rígido y flexible . . . . .	4
<b>Figura 4.</b> Algunas estructuras macromoleculares a media/baja resolución que experimentan cambios conformacionales de gran amplitud . . . . .	5
<b>Figura 5.</b> Modos normales de vibración . . . . .	7
<b>Figura 6.</b> Representación esquemática del modelo físico de osciladores armónicos acoplados considerado en el NMA . . . . .	13
<b>Figura 7.</b> Modelo energético del análisis vibracional . . . . .	14
<b>Figura 8.</b> Algunas redes elásticas . . . . .	16
<b>Figura 9.</b> Comparación de los <i>B-factors</i> de MD y NMA . . . . .	20
<b>Figura 10.</b> Comparación del solapamiento entre los espacios vectoriales de ED y NMA . . . . .	22
<b>Figura 11.</b> Imágenes en proyección de partículas individuales de la proteína RepB . . . . .	30
<b>Figura 12.</b> Representación esquemática del proceso de reconstrucción 3D por microscopía electrónica de transmisión y análisis de partículas individuales . . . . .	33
<b>Figura 13.</b> Modelo de dos cuerpos de Go . . . . .	48
<b>Figura 14.</b> Modelo de movimiento para calcular el Hessiano por el método rápido de Go . . . . .	50
<b>Figura 15.</b> Modelo de unidades rígidas articuladas para obtener el Hessiano por el método rápido de Go. . . . .	52
<b>Figura 16.</b> Cálculo de la matriz <i>U</i> para obtener el Hessiano por el método rápido de Go . . . . .	53
<b>Figura 17.</b> Representación esquemática de la divergencia en la aproximación lineal de una trayectoria curva. . . . .	56
<b>Figura 18.</b> Representación esquemática de tres aproximaciones distintas para actualizar las posiciones atómicas . . . . .	57
<b>Figura 19.</b> Representación esquemática de los modelos atómicos y de CI para proteínas . . . . .	58

<b>Figura 20.</b> Representación esquemática del modelo atómico y de CI para ácidos nucleicos . . . . .	59
<b>Figura 21.</b> Modelos atómicos de proteínas . . . . .	60
<b>Figura 22.</b> Representación de modos normales con <i>iModview</i> . . . . .	62
<b>Figura 23.</b> Diagrama de flujo de <i>iMC</i> . . . . .	65
<b>Figura 24.</b> Diagrama de flujo de <i>iMorph</i> . . . . .	66
<b>Figura 25.</b> Diagrama de flujo de <i>iModfit</i> . . . . .	68
<b>Figura 26.</b> Modelo inicial de RepB obtenido por RCT . . . . .	70
<b>Figura 27.</b> Perfiles de FSC de las reconstrucciones 3D de RepB . . . . .	71
<b>Figura 28.</b> Coste computacional del cálculo de las matrices del sistema por los métodos rápido y directo . . . . .	82
<b>Figura 29.</b> Coste computacional de la fase de diagonalización . . . . .	84
<b>Figura 30.</b> Porcentajes de tiempo requeridos para efectuar las distintas tareas del NMA. . . . .	84
<b>Figura 31.</b> Consumo de memoria del análisis vibracional. . . . .	85
<b>Figura 32.</b> Comparación de los tiempos de cálculo del NMA en CI y CC . . . . .	87
<b>Figura 33.</b> Comparación del consumo de memoria del NMA en CI y CC . . . . .	88
<b>Figura 34.</b> Comparación visual del vector de transición con el primer modo normal en CI. . . . .	89
<b>Figura 35.</b> Comparación visual de los primeros modos normales de una proteína en CI y CC . . . . .	91
<b>Figura 36.</b> Simulación de una trayectoria de transición conformacional . . . . .	98
<b>Figura 37.</b> Muestreo conformacional de MC . . . . .	103
<b>Figura 38.</b> Desviaciones individuales de los ajustes flexibles con el modelo HA y mapas simulados de proteínas . . . . .	106
<b>Figura 39.</b> Tiempo de cálculo necesario para efectuar los ajustes flexibles de proteínas en función del modelo atómico y la resolución . . . . .	109
<b>Figura 40.</b> Desviaciones individuales de los ajustes flexibles de proteínas obtenidos eliminando aleatoriamente el 50% de las variables con el modelo HA en función de la resolución. . . . .	113

<b>Figura 41.</b> Desviaciones individuales de los ajustes flexibles de proteínas obtenidos eliminando aleatoriamente el 90% de las variables con el modelo HA en función de la resolución. . . . .	115
<b>Figura 42.</b> Tiempo de cálculo necesario para efectuar ajustes flexibles de proteínas con el modelo HA en función de la resolución y del porcentaje de ángulos fijado . . .	115
<b>Figura 43.</b> Tiempo de cálculo necesario para efectuar los ajustes de proteínas con el modelo C5 en función de la resolución y del porcentaje de ángulos fijado . . . . .	118
<b>Figura 44.</b> Tiempo de cálculo necesario para efectuar los ajustes de proteínas con el modelo $C\alpha$ en función de la resolución y del porcentaje de ángulos fijado. . . . .	118
<b>Figura 45.</b> Desviaciones individuales de los ajustes flexibles con mapas simulados de ARN. . . . .	122
<b>Figura 46.</b> Desviaciones individuales obtenidas por las herramientas de ajuste flexible basadas en NMA en los casos de ajuste con mapas simulados de proteínas. . . . .	128
<b>Figura 47.</b> Desviaciones individuales finales obtenidas por <i>YUPSCX</i> en los casos de ajuste flexible con mapas simulados de proteínas . . . . .	133
<b>Figura 48.</b> Desviaciones individuales obtenidas con las herramientas de ajuste flexible basadas en MD y el conjunto de pruebas de modelado por homología . . . . .	138
<b>Figura 49.</b> Mapas simulados a 10Å de resolución de la adenilato quinasa con distintos niveles de ruido . . . . .	141
<b>Figura 50.</b> Sistema GroEL/GroES. . . . .	152
<b>Figura 51.</b> Ajuste flexible de la estructura atómica de GroEL en un mapa experimental de microscopía electrónica a 8Å . . . . .	154
<b>Figura 52.</b> Ajuste flexible de la estructura atómica de GroEL en un mapa experimental de microscopía electrónica a 10Å . . . . .	155
<b>Figura 53.</b> Ajuste flexible de la estructura atómica del EFG en un mapa experimental de microscopía electrónica a 11Å . . . . .	156
<b>Figura 54.</b> Comparación de los ajustes flexibles de la estructura atómica del EFG en un mapa experimental. . . . .	157
<b>Figura 55.</b> Ajuste flexible de la estructura atómica de prefoldina en un mapa experimental de microscopía electrónica a 20Å . . . . .	158

<b>Figura 56.</b> Ajuste flexible de la estructura atómica de la proteína clamp en un mapa experimental de microscopía electrónica a 17Å . . . . .	159
<b>Figura 57.</b> Ajuste flexible de la estructura atómica de la ATPasa de calcio en un mapa experimental de microscopía electrónica de 8Å . . . . .	161
<b>Figura 58.</b> Ajuste flexible de la estructura atómica de la bomba de cobre en un mapa experimental de microscopía electrónica a 17Å . . . . .	162
<b>Figura 59.</b> Mapas experimentales del termosoma . . . . .	164
<b>Figura 60.</b> Ajuste flexible de la estructura atómica del termosoma en un mapa experimental de microscopía electrónica a 10Å . . . . .	165
<b>Figura 61.</b> Ajuste flexible de la estructura atómica de la ARN polimerasa II humana en un mapa experimental abierto de microscopía electrónica a 22Å . . . . .	166
<b>Figura 62.</b> Ajuste flexible de la estructura atómica de la ARN polimerasa II humana en un mapa experimental cerrado de microscopía electrónica a 22Å . . . . .	167
<b>Figura 63.</b> Cambios conformacionales de la ARN polimerasa II humana. . . . .	168
<b>Figura 64.</b> Estructura de los microtúbulos decorados con Ndc80/Nuf2 ilustrada sobre un mapa de microscopía electrónica a 8Å . . . . .	169
<b>Figura 65.</b> Ajuste rígido de las estructuras de tubulina y Ndc80/Nuf2 en un mapa experimental del microtúbulo decorado con Ndc80/Nuf2 a 8Å . . . . .	170
<b>Figura 66.</b> Ajuste flexible de las estructuras de tubulina y Ndc80/Nuf2 un el mapa experimental del microtúbulo decorado con Ndc80/Nuf2 a 8Å . . . . .	171
<b>Figura 67.</b> Ajuste rígido inicial del ribosoma . . . . .	175
<b>Figura 68.</b> Ajuste flexible del ribosoma 70S. . . . .	176
<b>Figura 69.</b> Función de RepB. . . . .	179
<b>Figura 70.</b> Pruebas de relajación de simetría durante la reconstrucción 3D de RepB. . . . .	180
<b>Figura 71.</b> Reconstrucción 3D por microscopía electrónica y SP de RepB libre . . . . .	181
<b>Figura 72.</b> Ajuste rígido de la estructura atómica de RepB en el mapa de microscopía . . . . .	181
<b>Figura 73.</b> Reconstrucción 3D de microscopía electrónica y SP de RepB en presencia de ADN. . . . .	182

---

<b>Figura 74.</b> Comparación entre los mapas de la proteína RepB libre y en presencia de ADN. . . . .	183
<b>Figura 75.</b> Ajuste flexible del hexámero de la proteína RepB libre utilizando <i>iModfit</i> . . .	184
<b>Figura 76.</b> Ajuste flexible del complejo de la proteína RepB en presencia de ADN utilizando <i>iModfit</i> . . . . .	185
<b>Figura 77.</b> Disseminación de <i>iMod</i> . . . . .	186
<b>Figura 78.</b> Disseminación de <i>iModfit</i> . . . . .	187

---



---

# Acrónimos

**2D:** Bidimensional.

**3D:** Tridimensional.

**ADN:** Ácido desoxirribonucleico.

**ADP:** Adenosín difosfato.

**ANM:** Modelo de red anisotrópica (del inglés, *Anisotropic Network Model*).

**ARN:** Ácido ribonucleico.

**ARNm:** ARN mensajero.

**ARNP:** ARN polimerasa.

**ARNr:** ARN ribosómico.

**ARNt:** ARN de transferencia.

**ATP:** Adenosín trifosfato.

**BPTI:** Inhibidor de la tripsina pancreática bovina (del inglés, *Bovine Pancreatic Trypsin Inhibitor*).

**CC:** Coordenadas Cartesianas.

**CCD:** Dispositivo acoplado de carga (del inglés, *Charge Coupled Device*).

**CCMV:** Virus de la judía clorótica moteada (del inglés, *Cowpea Chlorotic Mottle Virus*).

**CG:** Reduccionista (del inglés, *Coarse-Grained*).

**CI:** Coordenadas Internas.

**CIB:** Centro de Investigaciones Biológicas.

**CRE:** Cristalografía de Electrones.

**CRX:** Cristalografía de Rayos X.

**CSIC:** Consejo Superior de Investigaciones Científicas.

**Ct:** Extremo C-terminal.

**CTF:** Función de transferencia de contraste (del inglés, *Contrast Transfer Function*).

**DTT:** Ditiotreitól.

**ED:** Dinámica esencial (del inglés, *Essential Dynamics*).

**ed-ENM:** Modelo de red elástica de dinámica esencial (del inglés, *essential dynamics Elastic Network Model*).

**EDTA:** Ácido etilendiaminotetraacético (del inglés, *Ethylenediaminetetraacetic acid*).

**EMDB:** Banco de datos de microscopía electrónica (del inglés, *Electron Microscopy Data Bank*).

**FFT:** Transformada rápida de Fourier (del inglés, *Fast Fourier Transform*).

---

**FSC:** Correlación de capas de Fourier (del inglés, *Fourier Shell Correlation*).

**GDP:** Guanosín difosfato.

**GDPNP:** Guanosín diiminodifosfato.

**GNM:** Modelo de red gaussiana (del inglés, *Gaussian Network Model*).

**GTP:** Guanosín trifosfato.

**HA:** Átomos pesados (del inglés, *Heavy Atoms*).

**IBDV:** Virus de la enfermedad de la bursitis infecciosa (del inglés, *Infectious Bursal Disease Virus*).

**IPI:** Interpolación Parabólica Inversa.

**IRB:** Instituto de Investigación Biomédica (del catalán, *Institut de Recerca Biomédica*).

**MC:** Monte Carlo.

**MD:** Dinámica molecular (del inglés, *Molecular Dynamics*).

**MET:** Microscopía Electrónica de Transmisión.

**NMA:** Análisis de los modos normales de vibración (del inglés, *Normal Mode Analysis*).

**NO:** Solapamiento nativo (del inglés, *Native Overlap*).

**Nt:** Extremo N-terminal.

**PC:** Ordenador personal (del inglés, *Personal Computer*).

**PCA:** Análisis de componentes principales (del inglés, *Principal Component Analysis*).

**PCC:** Canal conductor de proteínas (del inglés, *Protein Conducting Channel*).

**PDB:** Banco de datos de proteínas (del inglés, *Protein Data Bank*).

**RAM:** Memoria de acceso aleatorio (del inglés, *Random Access Memory*).

**RC:** Círculo rodador (del inglés, *Rolling Circle*).

**RCT:** Giro cónico aleatorio (del inglés, *Random Conical Tilt*).

**RMN:** Resonancia Magnética Nuclear.

**RMSD:** Desviación cuadrática media (del inglés, *Root Mean Square Deviation*).

**RTB:** Rotación traslación de bloques (del inglés, *Rotation Translation of Blocks*).

**SCV:** Escalado de la variable colectiva (SCV, del inglés *Scaled Collective Variable*).

**SNR:** Relación señal ruido (del inglés, *Signal to Noise Ratio*).

**SP:** Análisis de partículas individuales (del inglés, *Single Particle analysis*).

**TE:** Tomografía Electrónica.

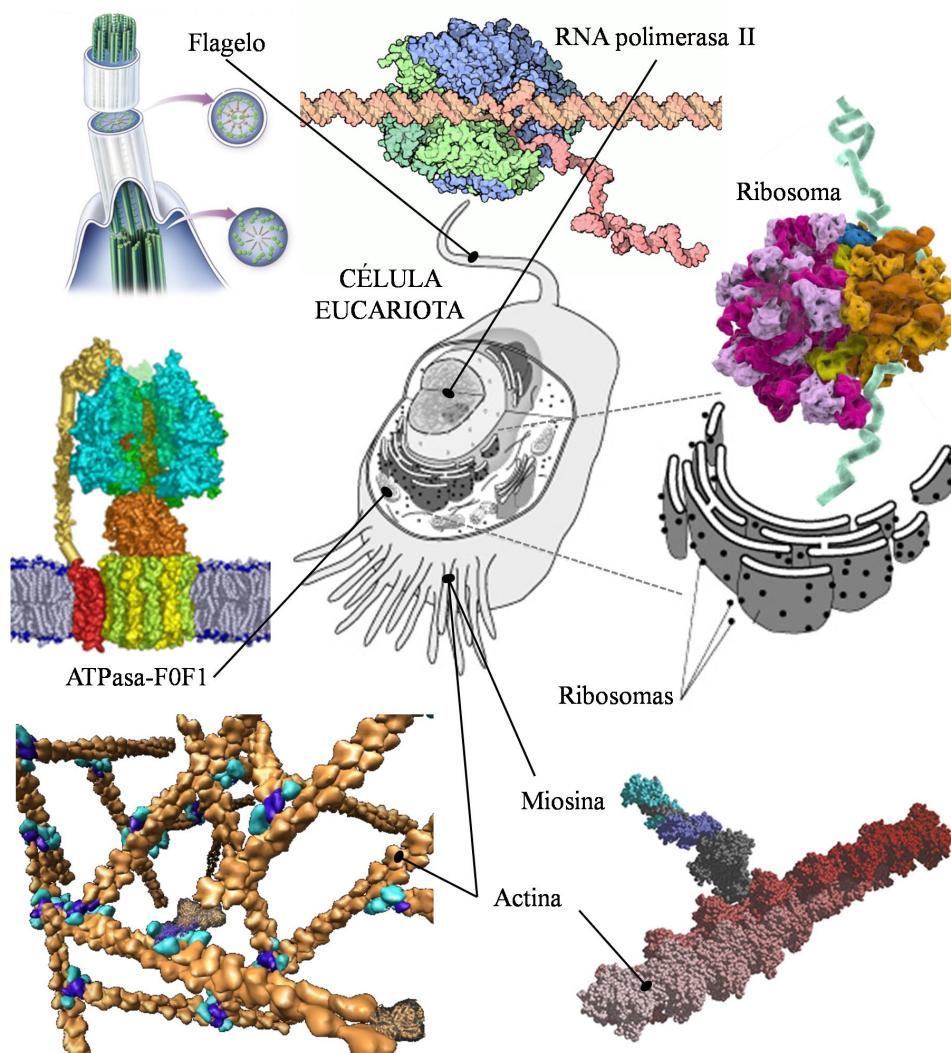
**Ti<sub>post</sub>:** Intermediario posterior a la translocación (del inglés, *post-Translocational Intermediate*).

**Ti<sub>pre</sub>:** Intermediario previo a la translocación (del inglés, *pre-Translocational Intermediate*).

**Tris:** Tris(hidroximetil)aminometano.

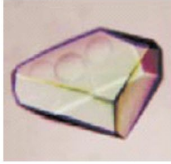

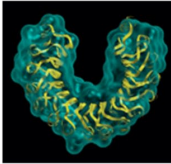
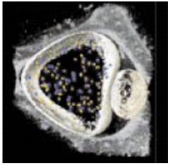
## 1. Motivación

Uno de los paradigmas más importantes en biología consiste en entender la relación entre la estructura y la función de las proteínas y los ácidos nucleicos. Estas moléculas, interaccionan unas con otras para formar grandes complejos macromoleculares que son los encargados de llevar a cabo los procesos fundamentales que tienen lugar en los organismos vivos. Como ejemplos notables de estos complejos (véase la Figura 1) destacan: la ARN polimerasa que transcribe el ADN en ARN en el núcleo de las células para posteriormente ser traducido en proteínas en el citoplasma por los ribosomas; la ATPasa-F<sub>0</sub>F<sub>1</sub> que sintetiza en las mitocondrias la principal molécula energética que utilizan todas las células (ATP) a partir de un gradiente electroquímico de protones; los microtúbulos y moléculas de dineína de cilios y flagelos o las fibras de actina y miosina de los pseudópodos que se entrecruzan y deslizan activamente entre sí para permitir los distintos mecanismos de locomoción con los que se desplazan los seres vivos.



**Figura 1.** Algunas macromoléculas biológicas de gran tamaño y su localización celular.

La mayoría de las macromoléculas actúan como nano-máquinas que cambian de forma de manera bien definida y regulada para llevar a cabo su función biológica. Estos cambios se encuentran codificados en la disposición espacial que adquieren los átomos. Para desentrañar este código es necesario disponer de información experimental directa sobre la estructura, flexibilidad y dinámica de las macromoléculas en condiciones fisiológicas. Existe una gran variedad de técnicas experimentales que capturan información estructural detallada de los complejos macromoleculares de gran tamaño en diversas condiciones, como cristalografía de rayos X (CRX) (Ilari y Savino, 2008) y de electrones (CRE) (Renault, *et al.*, 2006), microscopía electrónica y análisis de partículas individuales (SP, del inglés *Single Particle*) (Cheng y Walz, 2009), tomografía electrónica (TE) (Jonic, *et al.*, 2008; Lucic, *et al.*, 2008), o dispersión de rayos X de bajo ángulo (SAXS, del inglés *Small Angle X-ray Scattering*) (Neylon, 2008), entre otras. Aunque cada una de ellas permite conocer la estructura macromolecular a distintos niveles de resolución (véase la Figura 2), la información que proporcionan sobre su flexibilidad y dinámica suele ser insuficiente para observar cómo se producen los cambios conformacionales.

				
	CR	SP	SAXS	TE
Tamaño	Sin límites	$\geq 70$ kD	Sin límites	$> 400$ kD
Cantidad de muestra	Elevada	Mínima	Elevada	Mínima
Resolución	Atómica	10-30 Å	20-40 Å	40-60 Å
Condiciones	Restringidas a la cristalización	Condiciones más cercanas a las fisiológicas		
Ventajas	Resolución atómica	Condiciones flexibles	Condiciones flexibles + resolución temporal	Condiciones flexibles + contexto celular
Inconvenientes	Las grandes macromoléculas son difíciles de cristalizar	Resolución limitada	Resolución limitada (Modelado)	Resolución muy limitada

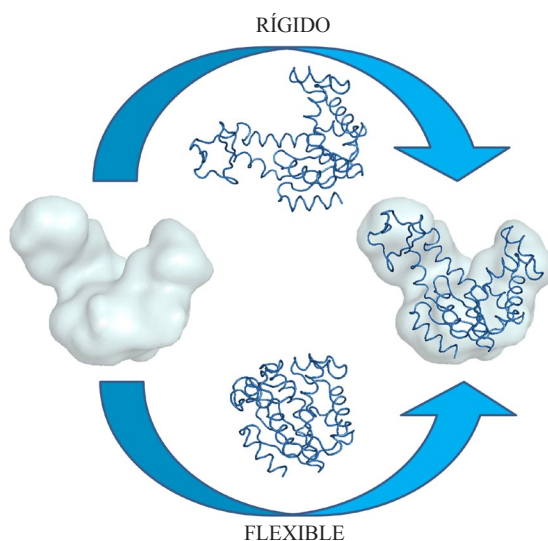
**Figura 2.** Comparación de las principales técnicas experimentales para obtener información estructural de complejos macromoleculares de gran tamaño. Estas son: cristalografía (CR), microscopía electrónica de transmisión y análisis de partículas individuales (SP), dispersión de rayos X de bajo ángulo (SAXS) y tomografía electrónica (TE).

Para estudiar la estructura de las macromoléculas más grandes se emplean principalmente las técnicas cristalográficas y de microscopía electrónica de transmisión (MET). Las primeras permiten resolver la estructura atómica a partir de sus formas cristalinas. Para ello utilizan la información estructural de alta resolución contenida en los patrones de difracción que se producen cuando la radiación electromagnética de rayos X (CRX) o un haz de electrones (CRE) interacciona con los cristales. La CRX es la técnica de referencia para obtener a resolución atómica estructuras de proteínas y ácidos nucleicos. Para hacerse una idea de su importancia relativa, de las ~76000 estructuras atómicas depositadas actualmente en el PDB (*Protein Data Bank*) (Berman, *et al.*, 2000), la amplia mayoría proceden de CRX (~67000), mientras que tan solo unas pocas (~30) son de CRE. La principal limitación de estas técnicas consiste en la dificultad para obtener cristales que difracten adecuadamente. Además, una vez conseguidos estos cristales, todavía existe la posibilidad de que las restricciones impuestas por el empaquetamiento cristalino alteren las estructuras alejándolas de la conformación fisiológicamente activa. Otra limitación es la elevada cantidad de muestra necesaria. Conseguir una cantidad suficiente de muestra es especialmente difícil cuando las moléculas son grandes y complejas. En cambio, con técnicas de MET es posible estudiar la estructura que adquieren los grandes complejos macromoleculares en condiciones cercanas a las fisiológicas utilizando cantidades muy pequeñas de muestra. Las técnicas de MET se basan en la atenuación que sufre un haz coherente de electrones cuando atraviesa la muestra en el interior del microscopio. El haz de electrones atenuado es recogido en un detector para formar imágenes en proyección con información estructural de la muestra. Para reconstruir un modelo tridimensional (3D) es necesario combinar adecuadamente suficientes imágenes en orientaciones diferentes. En TE se gira la muestra para obtener distintas proyecciones de la misma región con orientaciones conocidas. Desafortunadamente, solo es posible tomar un pequeño número de ellas debido a que el haz de electrones de alta energía (~100kV) daña rápidamente la muestra. En consecuencia, esta técnica está limitada a bajas resoluciones (40-60Å). Para minimizar el daño, en otras técnicas de MET (SP, reconstrucción helicoidal, y análisis de fibras, filamentos y cristales tubulares) las partículas se adsorben en un soporte de forma que cada una de ellas adquiere una orientación aleatoria. Al tomar cada micrografía de una región distinta de la muestra se obtienen proyecciones diferentes que tan solo han sido expuestas una vez al haz de electrones. De este modo el daño por radiación es mucho menor, lo que permite alcanzar resoluciones más altas (10-30Å) e incluso resolución atómica (Zhou, 2011).

Como se acaba de mencionar, las limitaciones de las técnicas cristalográficas reducen enormemente la variedad de condiciones experimentales diferentes (salinidad, pH, presencia

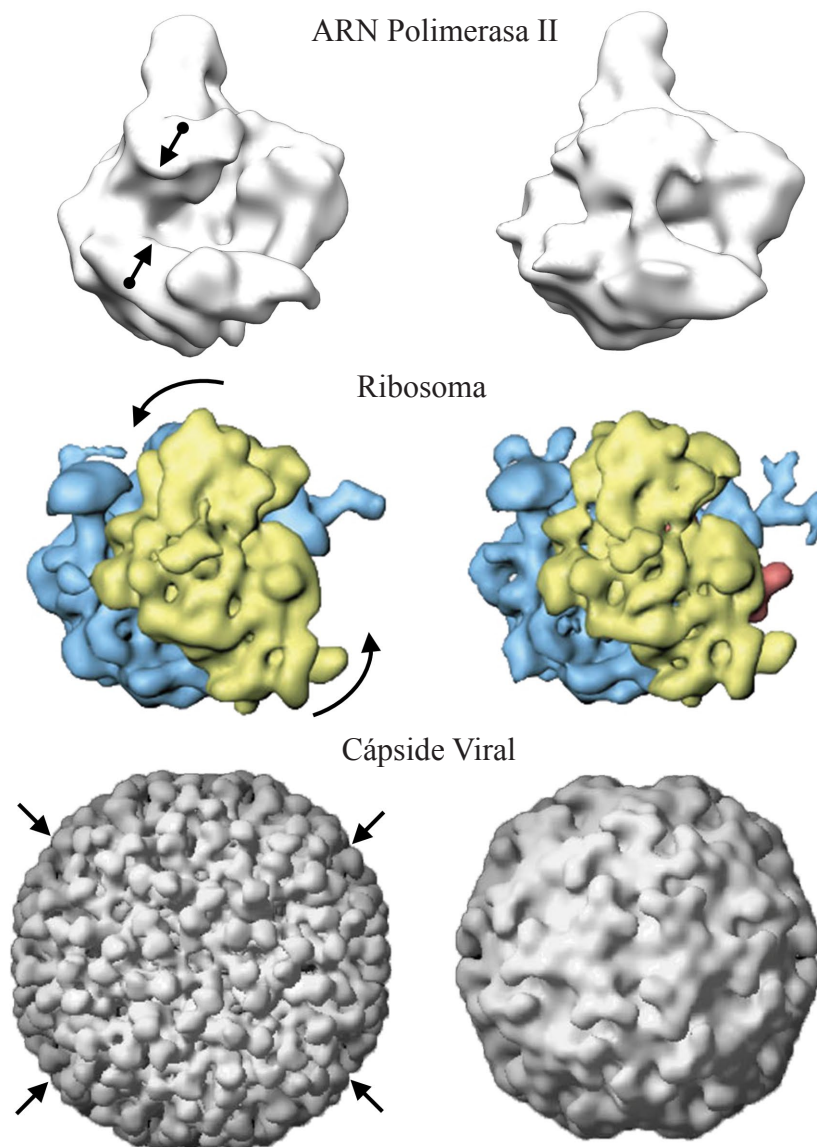


de ligandos, etc) en las que es posible conseguir estructuras a resolución atómica de grandes complejos macromoleculares; sin embargo, es relativamente fácil obtener estas estructuras a media/baja resolución utilizando técnicas de MET (generalmente SP). Por estos motivos, para desentrañar los mecanismos que rigen el funcionamiento de las macromoléculas resulta esencial integrar la información estructural a distintas resoluciones que proporcionan estas técnicas (Russel, *et al.*, 2009). Con este fin se realiza lo que se conoce como ajuste macromolecular, que consiste en localizar una estructura atómica cristalográfica en el interior de un mapa de MET a media/baja resolución para obtener así un modelo atómico en condiciones más cercanas a las fisiológicas. Aunque los ajustes pueden realizarse manualmente, es necesario utilizar métodos computacionales para evitar sesgos y obtener mejores resultados. Cuando ambas estructuras poseen una conformación similar se pueden alinear directamente utilizando algún método de ajuste rígido (Wriggers, 2010; Wriggers y Chacon, 2001) (véase la Figura 3). Sin embargo, es muy frecuente que la conformación de la estructura atómica sea diferente a la del mapa. En este caso hay que considerar la flexibilidad del modelo atómico para realizar el ajuste. Aunque ya existen herramientas eficaces para llevar a cabo el ajuste flexible, su desarrollo es relativamente reciente y todavía es posible mejorar, especialmente en lo que respecta a su eficiencia computacional y a la calidad de las estructuras obtenidas. En general, la mayor parte de los métodos disponibles en la actualidad son computacionalmente costosos; otros, en cambio, aunque consumen menos recursos, deterioran la calidad de las estructuras o se basan en aproximaciones reduccionistas que dificultan la obtención de modelos más detallados. El deterioro de las estructuras se debe principalmente al fenómeno de sobreajuste (del inglés, *overfitting*) que tiende a producirse por distintos motivos: *i*) cuando la cantidad de información presente en los mapas a media/baja resolución es mucho menor que la necesaria para determinar inequívocamente la localización de todos sus átomos, *ii*) cuando la densidad del mapa no se corresponde, tanto por exceso como por defecto, con la estructura atómica, y *iii*) por la presencia de ruido en los mapas. Por todo



**Figura 3.** Ajuste macromolecular rígido y flexible. Cuando el mapa de densidad (izquierda) y la estructura atómica (arriba) presentan aproximadamente la misma conformación que el mapa objetivo (derecha), el ajuste debe ser rígido. Si la estructura atómica disponible posee una conformación diferente (abajo) a la del mapa objetivo el ajuste ha de ser flexible.

ello, el ajuste flexible de las estructuras macromoleculares representa un reto metodológico y computacional de gran importancia en biología, especialmente cuando se trata con complejos de gran tamaño. Además, estas estructuras suelen experimentar cambios conformacionales de gran amplitud durante su funcionamiento, por lo que resulta imprescindible tener en cuenta su flexibilidad de forma eficiente. En la Figura 4 se muestran algunos cambios conformacionales de este tipo: el movimiento de pinza de la ARN polimerasa, el giro (*ratcheting*) del ribosoma y la maduración de la cápside del virus *Nudaurelia capensis omega* (NωV). En la presente tesis se pretende abordar este problema desarrollando métodos de ajuste flexible eficientes que mantengan la integridad de las estructuras, incluso durante los cambios conformacionales de gran amplitud que sufren estas enormes estructuras de más de 500 kDa de masa molecular.



**Figura 4.** Algunas estructuras macromoleculares a media/baja resolución que experimentan cambios conformacionales de gran amplitud. De arriba a abajo: la ARN polimerasa II, el ribosoma bacteriano y la cápside del virus *Nudaurelia capensis omega* (NωV).

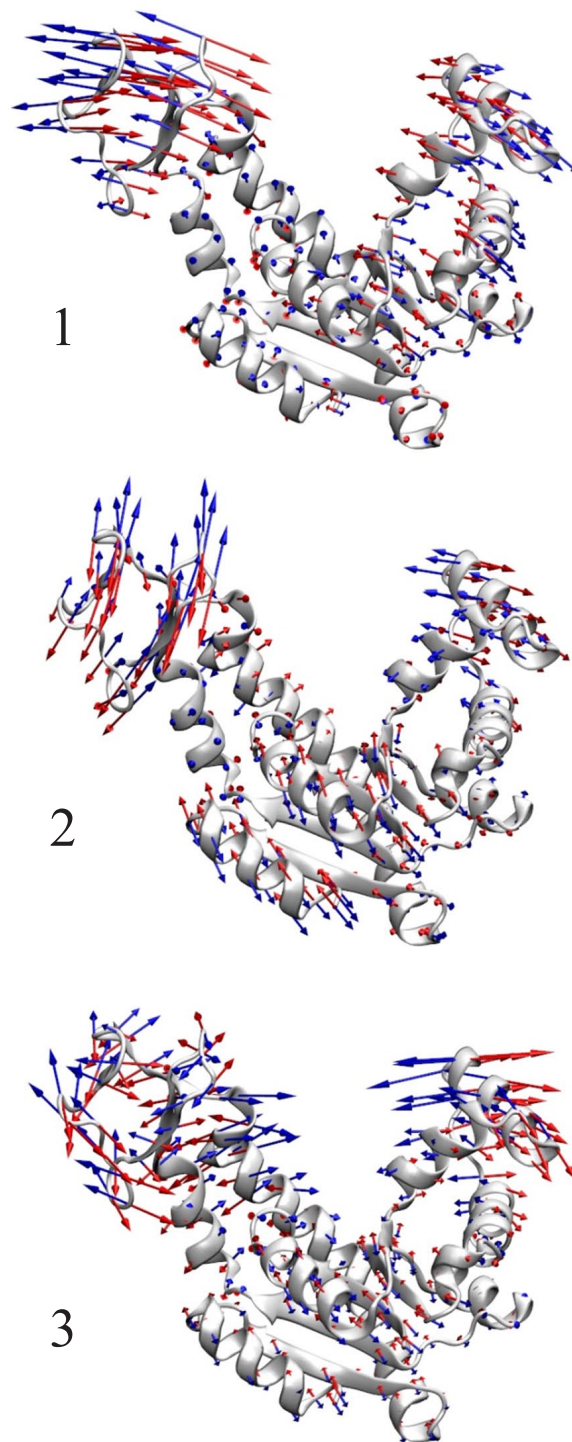
Hasta la fecha, se han desarrollado varias aproximaciones computacionales para predecir y caracterizar la flexibilidad de las proteínas a partir de su estructura. La técnica más utilizada para simular el comportamiento dinámico de las macromoléculas es, con diferencia, la dinámica molecular (MD, del inglés *Molecular Dynamics*) (Adcock y McCammon, 2006). Esta técnica consiste en la integración numérica de las ecuaciones clásicas de movimiento del sistema para obtener una trayectoria. Del análisis de estas trayectorias se puede obtener información estructural muy valiosa sobre los cambios conformacionales que ocurren durante el funcionamiento de las biomoléculas. Además, el empleo de potenciales interatómicos detallados y la consideración explícita del disolvente hacen que la información obtenida sea muy realista. Sin embargo, el tiempo de integración necesario es tan pequeño (unos pocos femto-segundos) que la simulación de trayectorias con la duración suficiente para observar gran parte de los cambios estructurales asociados a los eventos funcionalmente interesantes (micro- o mili-segundos) suele resultar una tarea muy costosa computacionalmente. A pesar de este inconveniente, desde que se llevó a cabo la primera simulación con la proteína BPTI (del inglés, *Bovine Pancreatic Trypsin Inhibitor*), de tan solo 58 aminoácidos (McCammon, *et al.*, 1977), las técnicas de MD se llevan utilizando satisfactoriamente desde hace más de 30 años en el estudio de la dinámica de proteínas y ácidos nucleicos (Adcock y McCammon, 2006; Karplus y McCammon, 2002). No obstante, la simulación con MD de máquinas moleculares de miles de amino ácidos durante períodos lo suficientemente largos como para observar cambios estructurales funcionales sigue representando un gran reto. Por ejemplo, actualmente son necesarias varios días de cálculo en un super-ordenador diseñado específicamente para alcanzar tiempos de simulación del orden de ~1 mili-segundo con la proteína BPTI (Shaw, *et al.*, 2009; Shaw, *et al.*, 2010). Por ello, a pesar del gran realismo y detalle alcanzados en las simulaciones, modelar con MD los grandes cambios conformacionales que tienen lugar en las macromoléculas en el contexto del ajuste flexible no parece ser la opción más eficiente, siempre y cuando no se empleen simplificaciones adicionales.

Una de las alternativas más interesantes a MD para estudiar la flexibilidad de las macromoléculas es el análisis vibracional o NMA (del inglés, *Normal Mode Analysis*). El NMA puede modelar con relativa facilidad los movimientos colectivos y de gran amplitud de los grandes complejos macromoleculares. Esta técnica es una aproximación de mecánica clásica que ha sido extensivamente aplicada en las ingenierías civil, mecánica y aeroespacial, así como también en otras disciplinas (Goldstein, *et al.*, 2002). Las primeras aplicaciones del NMA en biología computacional se utilizaron en los años 70 para caracterizar la flexibilidad de  $\alpha$ -hélices de unos pocos aminoácidos (Fanconi, *et al.*, 1971; Itoh y Shimanouchi, 1970; Levy y Karplus,



1979). A comienzos de los años 80 se empleó en el análisis de las vibraciones atómicas del péptido glucagón (Tasumi, *et al.*, 1982) y de la proteína BPTI (Noguti y Go, 1982). Con el tiempo se ha convertido en una técnica muy utilizada para modelar la flexibilidad de todo tipo de macromoléculas biológicas (Bahar, *et al.*, 2010a; Bahar y Rader, 2005; Cui y Bahar, 2006; Dykeman y Sankey, 2010; Eyal, *et al.*, 2011; Ma, 2004; Ma, 2005; Skjaerven, *et al.*, 2009; Tama, 2003; Yang y Chng, 2008). La principal aproximación de esta metodología consiste en que las energías potencial y cinética varían cuadráticamente alrededor de la conformación de mínima de energía del sistema. A partir de esta suposición es posible descomponer el movimiento macromolecular en una serie de modos de deformación. Estos modos forman una base ortonormal de vectores que describe todos los desplazamientos o deformaciones posibles alrededor de la conformación de equilibrio, *i.e.* cualquier movimiento puede ser expresado como una combinación lineal de dichos modos. En la Figura 5 pueden verse los tres primeros modos de la proteína adenilato quinasa. Cada uno de ellos lleva asociado una energía (o frecuencia) de forma que es posible determinar aquellos movimientos más favorables energéticamente. Nótese que a mayor frecuencia, mayor energía, y *viceversa*.

Existen muchas evidencias que soportan la aproximación armónica del movimiento molecular. La mayoría de los movimientos de proteínas documentados en el PDB corresponden a cambios conformacionales colectivos que no alteran significativamente la estructura terciaria (Gerstein y Krebs, 1998).



**Figura 5.** Modos normales de vibración. Se han representado con flechas los tres modos de menor energía de la estructura atómica de la proteína adenilato quinasa (4ake).

Asimismo, varios estudios han comprobado que tan solo unos pocos modos de baja frecuencia, los más colectivos, son capaces de justificar la mayoría de los cambios conformacionales que experimentan las proteínas (Hinsen, *et al.*, 1999; Marques y Sanejouand, 1995). Midiendo el solapamiento entre el vector de diferencia y los modos normales de 20 proteínas en conformaciones abiertas o cerradas, Tama y Sanejouand demostraron que unos pocos modos de baja frecuencia son capaces de describir más del 80% de sus diferencias conformacionales (Tama y Sanejouand, 2001). En Krebs *et al.* (Krebs, *et al.*, 2002) se comprobó que más de la mitad de un conjunto de 3800 movimientos de proteínas puede aproximarse perturbando las estructuras iniciales a lo largo de los dos modos de menor frecuencia. Además, se requiere muy poca información sobre la estructura objetivo para conducir una estructura hacia la otra mediante una combinación lineal de unos pocos modos de estos modos (Zheng y Brooks, 2005). En Rueda *et al.* (Rueda, *et al.*, 2007a) se ha determinado la buena correspondencia entre los movimientos descritos por NMA y MD utilizando simulaciones atomísticas de un conjunto representativo de 32 estructuras de proteínas. A pesar del gran número de estudios de NMA llevados a cabo, no es fácil identificar qué modos son funcionalmente relevantes sin datos experimentales adicionales. Sin embargo, de forma general, casi siempre serán uno o unos pocos de los de menor frecuencia porque representan las trayectorias de transición conformacional con menor coste energético. En principio, es posible estudiar la función de las biomoléculas filtrando los movimientos menos importantes, los de alta frecuencia, y concentrarse en aquellos dominantes, *i.e.* los de baja frecuencia y menor energía (Ma, 2005). No obstante, aunque los movimientos cooperativos de las transiciones conformacionales de las biomoléculas siguen a menudo las trayectorias de alguno (o algunos) de estos modos, sus movimientos cooperativos no siempre son descritos adecuadamente por ellos. Por ejemplo, Petrone y Pande (Petrone y Pande, 2006) utilizaron pares de estructuras en conformaciones abiertas y cerradas de cuatro proteínas (miosina, calmodulina, NtrC y hemoglobina) para evaluar si únicamente los modos de menor frecuencia eran capaces de justificar los cambios conformacionales. Estos autores observaron que al utilizar una combinación lineal de los 20 primeros modos para deformar las estructuras abiertas solo es posible reducir el reducir el RMSD con las estructuras cerradas hasta el ~50% de su valor inicial. Sin embargo, a pesar de sus limitaciones, el NMA es un método computacional efectivo para el estudio de los movimientos colectivos de gran amplitud que experimentan las macromoléculas (Bahar, *et al.*, 2010b).

En esta tesis, nos proponemos estudiar la flexibilidad de los grandes complejos macromoleculares a través de la aproximación del NMA. En primer lugar, pretendemos desarrollar una metodología para llevar a cabo eficientemente el NMA en coordenadas internas

(CI) de las principales macromoléculas biológicas. La utilización de CI no solo permitirá un incremento adicional de velocidad respecto de los métodos basados en coordenadas cartesianas (CC), sino que también mejorará la calidad estructural de los modelos generados. En segundo lugar, se aplicará la metodología desarrollada para modelar la flexibilidad de las macromoléculas en distintos problemas de bioinformática estructural. En esta línea, se han desarrollado e implementado una serie de herramientas bioinformáticas para el ajuste flexible de estructuras atómicas macromoleculares en el interior de mapas de densidad de media/baja resolución procedentes de SP, y para la simulación de trayectorias de Monte Carlo (MC) y de transiciones conformacionales.

---

## 2. Objetivos

Los objetivos principales de esta tesis son el desarrollo e implementación de una nueva metodología bioinformática para el análisis de los modos normales de vibración (NMA) en coordenadas internas (CI) y su aplicación en el ajuste flexible de estructuras atómicas macromoleculares en mapas de densidad a media/baja resolución. Con este fin se proponen los siguientes objetivos:

### 1. Desarrollo metodológico del NMA en CI

A. Desarrollo de una metodología versátil para el análisis vibracional en CI de estructuras atómicas de macromoléculas con las siguientes características:

- i) *Aplicable a las principales biomoléculas*: proteínas y ácidos nucleicos.
- ii) *Multi-escala*. Se utilizarán distintas aproximaciones reduccionistas.
- iii) *Configurable*. Las características del modelo atómico, la red elástica y los grados de libertad podrán ser modificados fácilmente por el usuario.
- iv) *Eficiente*. Para extender la aplicabilidad de los métodos a sistemas de gran tamaño se diseñarán e implementarán algoritmos eficientes.

B. Validación sistemática y optimización de la metodología NMA desarrollada. Para ello será necesario:

- i) Diseñar e implementar herramientas específicas para el cálculo, análisis, comparación y visualización de modos normales.
- ii) Desarrollar herramientas de simulación basadas en NMA para:
  - Animación de modos normales y exploración conformacional.
  - Simulación de transiciones conformacionales.
  - Simulación de Monte Carlo.
- iii) Obtener distintos conjuntos de pruebas de estructuras atómicas experimentales de proteínas y ácidos nucleicos.

### **2. Desarrollo de la metodología para el ajuste flexible de estructuras atómicas en reconstrucciones tridimensionales de microscopía electrónica empleando NMA en CI.**

- A. Diseño e implementación de la herramienta bioinformática.
- B. Obtención de bancos de pruebas de ajuste flexible con mapas simulados y experimentales.
- C. Optimización y validación de dicha herramienta con los distintos bancos de pruebas.
- D. Comparación con otros métodos de ajuste flexible.

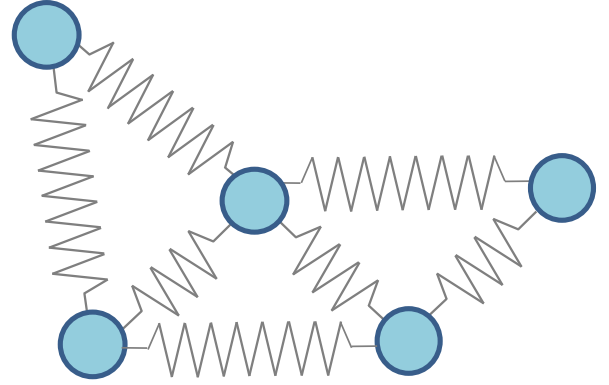
### **3. Aplicación de la herramienta de ajuste flexible.**

- A. Familiarización con las técnicas de microscopía electrónica y el análisis de partículas individuales. Reconstrucción tridimensional de la proteína RepB libre y de su complejo con ADN en tinción negativa.
- B. Aplicación de la metodología de ajuste desarrollada sobre mapas experimentales procedentes de la base de datos de microscopía electrónica, incluyendo los mapas del sistema RepB.

### 3. Introducción

#### 3.1. Análisis vibracional

La teoría del análisis de los modos normales aplicada a las vibraciones moleculares proporciona una solución analítica completa de las ecuaciones del movimiento del sistema (Wilson, *et al.*, 1955). Las soluciones de estas ecuaciones son un conjunto de vectores denominados modos normales que representan, en principio, todas las formas posibles de deformar una estructura alrededor de su conformación de equilibrio.



**Figura 6.** Representación esquemática del modelo físico de osciladores armónicos acoplados considerado en el NMA.

##### 3.1.1. El modelo físico

La principal suposición del NMA es que la molécula está representada por un conjunto de osciladores armónicos acoplados (Figura 6) cuyas energías potencial ( $V$ ) y cinética ( $K$ ) varían cuadráticamente con sus  $N$  coordenadas ( $q$ ) y sus velocidades ( $\dot{q}$ ), respectivamente. La aproximación de ambas formas de energía se obtiene por expansión de Taylor alrededor de la conformación de equilibrio:

$$V = V_0 + \sum_{\alpha=1}^N \left( \frac{\partial V}{\partial q_{\alpha}} \right)_0 q_{\alpha} + \frac{1}{2} \sum_{\alpha=1}^N \sum_{\beta=1}^N \left( \frac{\partial^2 V}{\partial q_{\alpha} \partial q_{\beta}} \right)_0 q_{\alpha} q_{\beta} + \dots \quad (3.1.1)$$

$$K = K_0 + \sum_{\alpha=1}^N \left( \frac{\partial K}{\partial \dot{q}_{\alpha}} \right)_0 \dot{q}_{\alpha} + \frac{1}{2} \sum_{\alpha=1}^N \sum_{\beta=1}^N \left( \frac{\partial^2 K}{\partial \dot{q}_{\alpha} \partial \dot{q}_{\beta}} \right)_0 \dot{q}_{\alpha} \dot{q}_{\beta} + \dots \quad (3.1.2)$$

Cuando el sistema se encuentra en reposo puede considerarse que las energías potencial ( $V_0$ ) y cinética ( $K_0$ ) son cero, al igual que la fuerza  $(\partial V / \partial q_{\alpha})_0$  y el momento conjugado  $(\partial K / \partial \dot{q}_{\alpha})_0$ . Así pues, es posible eliminar los términos de primer orden en las ecuaciones 3.1.1 y 3.1.2 para considerar que, en tales condiciones, las energías potencial y cinética dependen únicamente de las segundas derivadas parciales:

$$\mathbf{H} = H_{\alpha\beta} = \frac{\partial^2 V}{\partial q_{\alpha} \partial q_{\beta}} \quad (\text{Hessiano}) \quad \mathbf{K} = K_{\alpha\beta} = \frac{\partial^2 K}{\partial \dot{q}_{\alpha} \partial \dot{q}_{\beta}} \quad (\text{Tensor métrico}) \quad (3.1.3)$$

Las coordenadas generalizadas ( $q$ ) pueden ser CC, rotaciones/traslaciones del sólido rígido, ángulos torsionales o cualquier tipo de variable elegida para describir los  $N$  grados de libertad del sistema. Como se muestra en la Figura 7, en un sistema armónico las energías potencial y cinética oscilan entre sus valores máximos y mínimos alternativamente, *i.e.* cuando la energía potencial es máxima, la cinética es mínima, y *viceversa*.

Las ecuaciones del movimiento de Lagrange para cualquier conjunto de coordenadas se obtienen resolviendo:

$$\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_\alpha} \right) = \left( \frac{\partial L}{\partial q_\alpha} \right), \text{ donde } L = K - V \quad (3.1.4)$$

Como el potencial no depende de las velocidades ni la energía cinética de las posiciones se cumple que:

$$\frac{\partial L}{\partial \dot{q}_\alpha} = \sum_{\beta}^N K_{\alpha\beta} \dot{q}_\beta \quad (3.1.5)$$

$$\frac{\partial L}{\partial q_\alpha} = - \sum_{\beta}^N H_{\alpha\beta} q_\beta \quad (3.1.6)$$

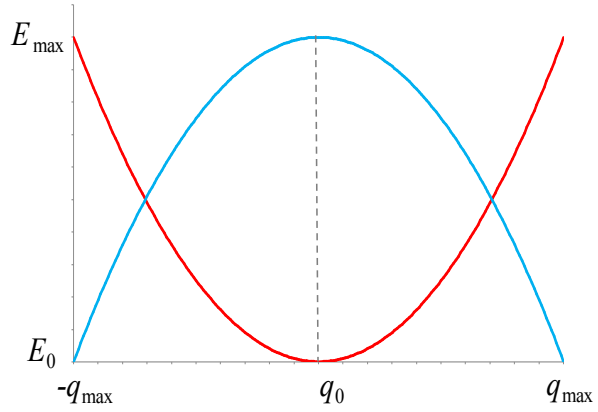
Al introducir estas expresiones en la ecuación 3.1.4 se obtienen las  $N$  ecuaciones del movimiento para cada una de las  $N$  coordenadas del sistema:

$$\sum_{\beta}^N K_{\alpha\beta} \ddot{q}_\beta = - \sum_{\beta}^N H_{\alpha\beta} q_\beta \quad (3.1.7)$$

Cuyas soluciones armónicas son de la forma:

$$q_\beta = \sum_k^N A_{\beta k} \cos(\omega_k t + \delta_k) \quad (3.1.8)$$

$$\ddot{q}_\beta = d^2 q_\beta / dt^2 = - \sum_k^N A_{\beta k} \omega_k^2 \cos(\omega_k t + \delta_k) \quad (3.1.9)$$



**Figura 7.** Modelo energético del análisis vibracional. Se ha representado esquemáticamente cómo varían las energías ( $E$ ) cinética (azul) y potencial (rojo) alrededor de la conformación de equilibrio ( $q_0$ ).



Donde  $\omega_k$  es la frecuencia angular asociada a cada una de las  $N$  soluciones. Después de introducir las fórmulas anteriores en la ecuación 3.1.7 se obtiene la expresión:

$$\sum_k^N \left( \sum_\beta^N K_{\alpha\beta} A_{\beta k} \right) \omega_k^2 \cos(\omega_k t + \delta_k) = \sum_k^N \left( \sum_\beta^N H_{\alpha\beta} A_{\beta k} \right) \cos(\omega_k t + \delta_k) \quad (3.1.10)$$

La única manera de respetar la igualdad para todo instante  $t$  es:

$$\omega_k^2 \sum_\beta^N K_{\alpha\beta} A_{\beta k} = \sum_\beta^N H_{\alpha\beta} A_{\beta k} \quad (3.1.11)$$

Este conjunto de ecuaciones constituye un problema generalizado de valores propios que puede expresarse de manera compacta utilizando notación matricial:

$$\mathbf{K}\mathbf{A}\mathbf{\Lambda} = \mathbf{H}\mathbf{A} \quad (3.1.12)$$

Finalmente, la diagonalización simultánea de las matrices de energía cinética ( $\mathbf{K}$ ) y potencial ( $\mathbf{H}$ ), permite obtener los modos normales de vibración de la macromolécula, *i.e.* los valores ( $\mathbf{\Lambda}$ ) y vectores ( $\mathbf{A}$ ) propios del sistema. Nótese que los elementos  $\Lambda_{kk}$  de la matriz diagonal  $\mathbf{\Lambda}$  se relacionan con la frecuencia de oscilación ( $\nu_k$ ) de cada modo  $k$  según la expresión:

$$\Lambda_{kk} = \omega_k^2 = (2\pi\nu_k)^2 \quad (3.1.13)$$

En la Figura 5 (véase Motivación) se ha utilizado una representación de flechas para ilustrar tanto la dirección como la magnitud relativa de las fluctuaciones atómicas codificadas en los tres modos de menor frecuencia de una proteína pequeña.

### 3.1.2. Modelos reduccionistas

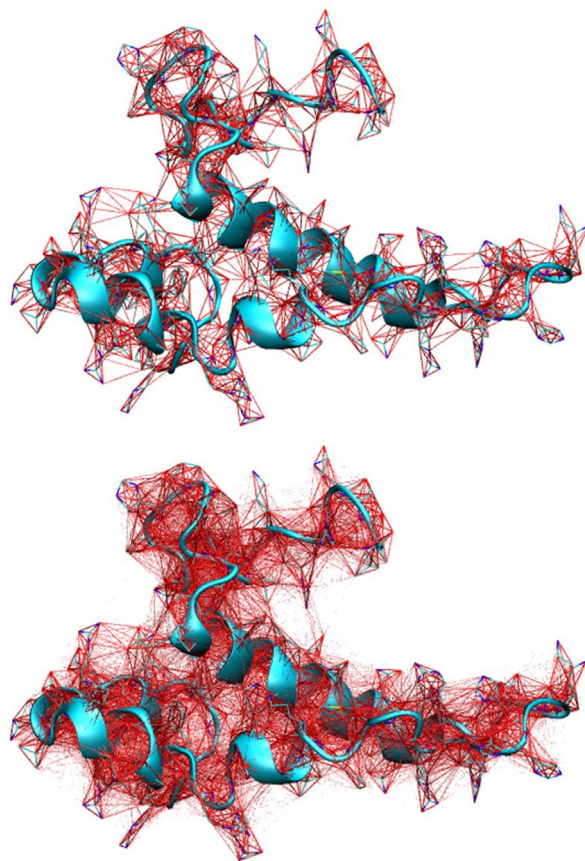
El NMA ha sido ampliamente utilizado en química para estudiar las vibraciones moleculares. La teoría de los modos normales de vibración aplicada a moléculas pequeñas ha permitido estudiar a partir de sus espectros vibracionales las fuerzas que mantienen unidos sus átomos, su estructura molecular y sus propiedades termodinámicas (Wilson, *et al.*, 1955). Las primeras aplicaciones del NMA en biología computacional datan de los años 70 y se utilizaron para caracterizar la flexibilidad de  $\alpha$ -hélices de unos pocos aminoácidos (Fanconi, *et al.*, 1971; Itoh y Shimanouchi, 1970; Levy y Karplus, 1979). A comienzos de los años 80 se estudiaron las fluctuaciones en las proximidades del equilibrio del péptido glucagón (Tasumi, *et al.*, 1982) y de una proteína globular (BPTI) (Go, *et al.*, 1983; Levitt, *et al.*, 1983; Noguti y Go, 1982). Todos estos trabajos se llevaron a cabo considerando las longitudes y ángulos de enlace fijos,

es decir, empleando únicamente los ángulos torsionales como CI del sistema. En 1983 Brooks y Karplus (Brooks y Karplus, 1983) también investigaron la BPTI, pero utilizando CC e incluyendo todos los grados de libertad con el campo de fuerzas CHARMM. Para que los términos de primer orden en las ecuaciones 3.1.1 y 3.1.2 sean cero todos estos métodos de NMA estándar requieren que la conformación inicial corresponda a un mínimo de potencial del campo de fuerzas empleado en el cálculo del Hessiano. Por este motivo, la estructura de partida se minimiza previamente utilizando dicho campo.

La utilización del NMA para describir la dinámica de las proteínas ganó popularidad a mediados de los años 90 gracias a que la aparición de modelos reduccionistas (CG, del inglés *coarse-grained*) redujo enormemente el coste computacional asociado a la obtención de los modos. Tirion (Tirion, 1996) introdujo el concepto de red elástica (Figura 8) empleando un potencial universal del tipo ley de Hooke y eliminando la costosa necesidad de minimizar inicialmente la estructura atómica. Así, la energía potencial se define de la siguiente manera en CC:

$$V = \sum_{i < j} \frac{1}{2} f_{ij} r_{ij}^2 \quad (3.1.14)$$

Donde  $f_{ij}$  es la constante de fuerza del muelle que une los átomos  $i$  y  $j$ , y  $r_{ij}$  es el cambio en el desplazamiento relativo de los átomos correspondientes desde su posición de equilibrio. Utilizando esta sencilla aproximación, los modos normales de baja frecuencia coinciden muy bien con los del NMA estándar que emplean modelos atómicos y potenciales mucho más detallados, todo ello a un coste computacional  $\sim 1000$  veces menor (Tirion, 1996). Esta cruda aproximación sugiere que la forma de la molécula, y no el nivel de detalle del campo de fuerzas, es el factor dominante que gobierna los modos de menor frecuencia (Lu y Ma, 2005).



**Figura 8.** Algunas redes elásticas. Dos modelos de red elástica: de Tirion (arriba) y función sigmoidea (abajo). El grosor de cada segmento (rojo) indica la rigidez relativa de los muelles que mantienen unidos los átomos de la proteína (azul).

De forma equivalente, pero basándose en la teoría termodinámica de redes aleatorias de polímeros (Flory, *et al.*, 1976), Bahar tuvo la idea de crear un modelo de red gaussiano, el GNM (del inglés, *Gaussian Network Model*), asumiendo que los átomos de la red elástica fluctúan siguiendo una distribución gaussiana (Bahar, *et al.*, 1997). De este modo, la energía potencial del sistema adquiere la forma:

$$V_{\text{GNM}} = \sum_{i < j} \frac{1}{2} f_{ij} (\mathbf{r}_i - \mathbf{r}_j)^2 \quad (3.1.15)$$

Donde los vectores  $\mathbf{r}_i$  y  $\mathbf{r}_j$  representan, respectivamente, los desplazamientos de los átomos  $i$  y  $j$  desde su posición de equilibrio. Según esta ecuación, el potencial del GNM no solo penaliza los cambios en las distancias interatómicas, sino que también lo hace con los cambios en la orientación relativa de los átomos. Esta modificación mejora la predicción de la magnitud de las fluctuaciones atómicas de las estructuras cristalográficas (*B-factors*) en comparación con otros modelos de red elástica convencionales. Al comparar la magnitud de las fluctuaciones atómicas predichas por este modelo con los *B-factors* experimentales se obtienen correlaciones de 0.59 para un conjunto de 64 proteínas no homólogas (Yang, *et al.*, 2007). Más aún, la correlación es incluso mayor (0.74) cuando se comparan las predicciones con los perfiles de RMSD obtenidos a partir de conjuntos de estructuras de Resonancia Magnética Nuclear (RMN) de las mismas proteínas. Sin embargo, a pesar de la elevada correspondencia con datos experimentales, los modos procedentes del GNM solo contienen información sobre la magnitud de las fluctuaciones.

Hinsen (Hinsen, 1998) substituyó la constante elástica del potencial de Tirion por una función exponencial que decae con la distancia empleando un modelo atómico a nivel de residuo. Además ajustó sus parámetros para conseguir unos modos de baja frecuencia con el mayor parecido posible a los obtenidos utilizando el NMA estándar con el campo de fuerzas clásico AMBER (Hinsen, *et al.*, 2000). Al igual que Hinsen, Atilgan (Atilgan, *et al.*, 2001) empleó este modelo con una constante similar a la de Tirion, pero asumiendo una única masa para cada residuo. Estos modelos, en los que la molécula se representa como una red de nodos centrados en las posiciones de los átomos  $C_\alpha$ , se conocen como ANM (del inglés, *Anisotropic Network Model*).

Teniendo en cuenta que los movimientos descritos por los modos de baja frecuencia son colectivos y que ocurren generalmente a nivel de residuos, elementos de estructura secundaria, o incluso dominios, Sanejouand y colaboradores modelaron las estructuras como un conjunto de bloques rígidos cada uno con 6 grados de libertad, *i.e.* 3 de rotación y 3 de traslación (Durand, *et al.*, 1994). Aunque generalmente cada bloque corresponde a un residuo, los bloques pueden

definirse de manera arbitraria para adaptar la resolución del modelo a cada problema particular. Este tipo de modelos CG, denominado RTB (del inglés, *Rotation Translation of Blocks*), hicieron posible llevar a cabo el NMA de proteínas más grandes al conseguir unos modos de baja frecuencia prácticamente equivalentes a los obtenidos con los métodos anteriores (Tama, *et al.*, 2000).

Más recientemente y como caso extremo de modelo CG, el NMA ha sido aplicado con éxito para estimar los movimientos macromoleculares directamente a partir de mapas de baja resolución sin tener en cuenta información detallada de las posiciones atómicas (Chacon, *et al.*, 2003; Ming, *et al.*, 2002).

Los métodos de ANM, GNM y RTB, aunque diferentes, no dejan de ser distintas caras de la misma moneda: el CG-NMA. Además de éstos, existen infinidad de aproximaciones que de forma general son variaciones de las comentadas anteriormente. En la bibliografía existen un buen número de revisiones con distintos enfoques del NMA en las que se comentan con mayor detalle distintos aspectos de esta metodología. En Yang y Chng (Yang y Chng, 2008) se presta especial atención al empleo de representaciones CG y se describen y comparan los modelos más importantes. Tanto Bahar y Rader (Bahar y Rader, 2005) como Ma (Ma, 2004) se centran en las aplicaciones de los modelos de CG, este último con especial énfasis en aquellas aplicaciones relacionadas con el estudio de grandes complejos macromoleculares a partir de datos de baja resolución y de difracción de fibras. Asimismo, en Ma (Ma, 2005) se discuten algunas de las preguntas fundamentales del NMA, tales como por qué el NMA describe tan bien los movimientos colectivos de las macromoléculas o cómo se pueden identificar los modos biológicamente relevantes. Skjaerven y colaboradores (Skjaerven, *et al.*, 2009), introducen brevemente la teoría del NMA y algunas de las aproximaciones necesarias para simplificar los cálculos y hacer tratables los sistemas de gran tamaño. Además, comentan los distintos servidores y bases de datos existentes basados en el análisis vibracional que, o bien permiten realizar el NMA a partir de estructuras biomoléculas, o bien proporcionan información sobre ellas. En dos revisiones actuales del grupo de Bahar se dibuja el panorama actual del NMA: *i*) en las proteínas de membrana (Bahar, *et al.*, 2010a), y *ii*) en el modelado de la respuesta mecánica, en la identificación de residuos alostéricos, sitios catalíticos y de unión a metales, y de su papel en el refinamiento estructural (Eyal, *et al.*, 2011). Recientemente, Dykeman y Sankey (Dykeman y Sankey, 2010) también describieron los principios básicos del NMA con especial interés en los métodos que consideran las estructuras como medios elásticos continuos y en aquellos que permiten estudiar cápsides virales considerando todos sus átomos.

### 3.1.3. Análisis vibracional en coordenadas internas

El sistema de coordenadas utilizado en el NMA tiene un gran impacto en la complejidad y en el coste computacional. Si se emplean CI, tales como los ángulos diedros, los tamaños de las matrices de energía cinética y potencial resultan mucho menores que en CC y los cálculos se aceleran enormemente. Esta reducción en el número de coordenadas supone una gran ventaja cuando el tamaño del sistema es muy grande o los recursos computacionales son limitados. Las bases teóricas del NMA en CI y su aplicación al estudio de las vibraciones moleculares fueron desarrolladas inicialmente por Wilson (Wilson, 1939; Wilson, *et al.*, 1955). Las primeras aplicaciones del NMA en macromoléculas biológicas ya utilizaban CI (Go, *et al.*, 1983; Levitt, *et al.*, 1983; Levy y Karplus, 1979; Noguti y Go, 1982; Tasumi, *et al.*, 1982). En el campo de la biología computacional, existen diversas aproximaciones para obtener los modos normales utilizando un sistema de referencia basado en CI. En 1983, Levitt *et al.* (Levitt, *et al.*, 1983; Levitt, *et al.*, 1985) desarrollaron un procedimiento para obtener las ecuaciones del sistema utilizando cualquier tipo de coordenadas, *i.e.* coordenadas generalizadas. Esta aproximación utiliza diferenciación numérica para obtener la energía potencial en CI y el método de superposición de estructuras de Kabsch (Kabsch, 1976) para eliminar los componentes externos de movimiento de la energía cinética. Paralelamente, Noguti y Go (Noguti y Go, 1983a; Noguti y Go, 1983b) desarrollaron una teoría completa para obtener analíticamente las ecuaciones del sistema utilizando los ángulos diedros como CI que fue generalizada posteriormente para todo tipo de coordenadas (Braun, *et al.*, 1984; Kamiya, *et al.*, 2003). Esta elegante aproximación analítica permite obtener directa y rápidamente la energía del sistema en CI. No obstante, aunque recientemente se ha utilizado la metodología del NMA en CI en el desarrollo de un nuevo método CG (Lu, *et al.*, 2006), en una aplicación para la predicción de interacciones proteína-ligando (Kovacs, *et al.*, 2005) y en el estudio de los cambios conformacionales de las proteínas (Mendez y Bastolla, 2010), su uso se ha visto desplazado por la sencillez y fácil implementación del NMA en CC. Por el contrario, en esta tesis se pretende aprovechar la eficiencia de la metodología del NMA en CI para modelar la flexibilidad de las estructuras macromoleculares manteniendo implícitamente su estructura covalente.

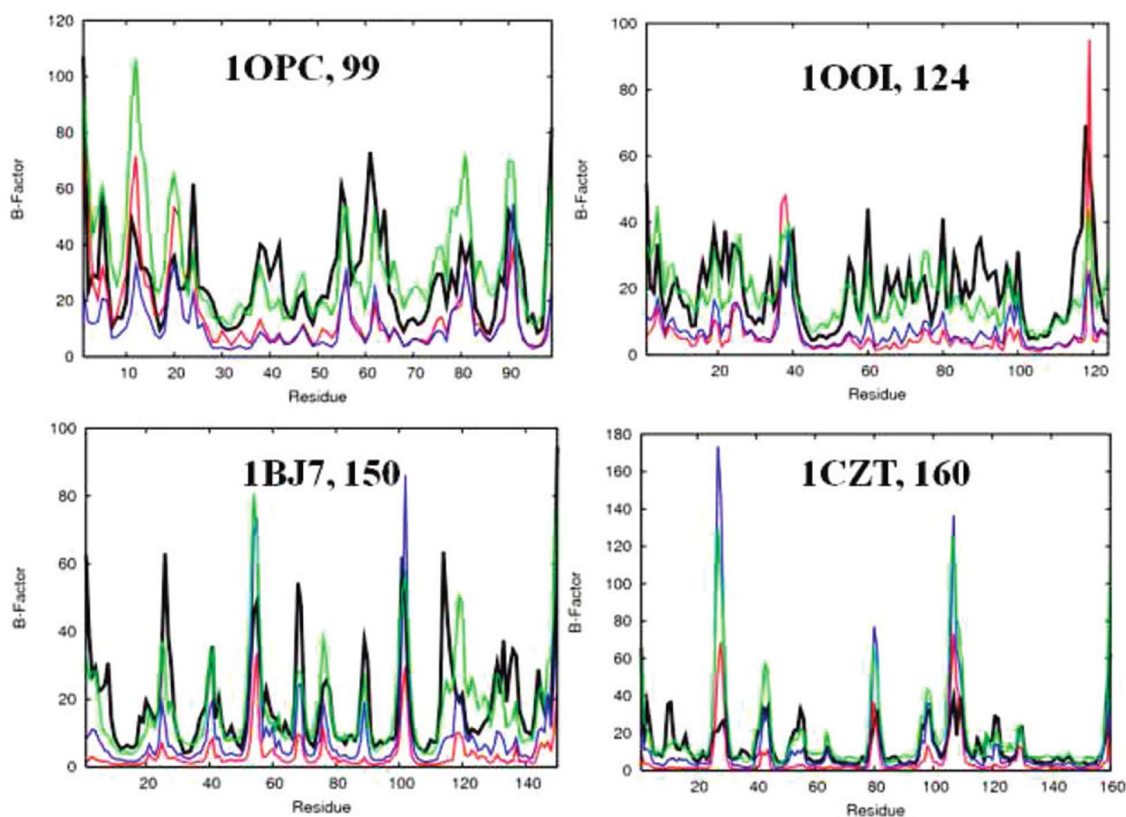
## 3.2. Validación del análisis vibracional

### 3.2.1. Correlación con las fluctuaciones atómicas

A pesar de que es posible obtener información experimental sobre la flexibilidad de las moléculas en disolución, por ejemplo, utilizando medidas de relajación de RMN (Wagner, 1993) o de dispersión de neutrones (Gabel, *et al.*, 2002; Zaccai, 2003), esta información solo



está disponible para unas pocas proteínas de pequeño tamaño, y en ningún caso para grandes macromoléculas. Por el contrario, existe abundante información sobre las fluctuaciones atómicas de las estructuras cristalográficas, *i.e.* los *B-factors*. Por este motivo, la mayoría de los métodos de análisis vibracional se han validado comparando las fluctuaciones predichas teóricamente con los *B-factors* experimentales de CRX. La magnitud de las fluctuaciones atómicas predichas correlaciona bastante bien (0.55-0.59) con estos factores (Brooks y Karplus, 1983; Eyal, *et al.*, 2007; Go, *et al.*, 1983; Yang, *et al.*, 2006). Además, incluso es posible alcanzar la coincidencia perfecta ajustando convenientemente la red elástica en cada caso particular (Erman, 2006). No obstante, el uso de los *B-factors* como validación resulta controvertido porque están condicionados tanto por el empaquetamiento y el desorden cristalino (Halle, 2002; Hinsen, 2008; Soheilifard, *et al.*, 2008) como por los errores de refinamiento (Carugo y Argos, 1999). Por otro lado, las fluctuaciones de los átomos observadas en las trayectorias de MD correlacionan muy bien con los *B-factors* experimentales, aunque las fluctuaciones de los residuos más flexibles son sistemáticamente mayores (Rueda, *et al.*, 2007b). Esto evidencia las restricciones que el entorno cristalino impone en el comportamiento dinámico de las macromoléculas. Notablemente, las magnitudes de las fluctuaciones calculadas con NMA se parecen mucho a

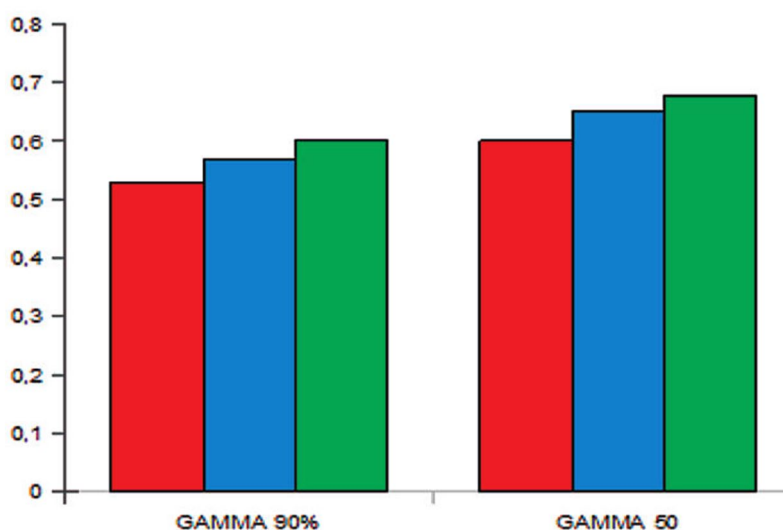


**Figura 9.** Comparación de los *B-factors* de MD y NMA. Se han utilizado tres redes elásticas diferentes para calcular con cuatro proteínas representativas los *B-factors* de NMA: Tirion (rojo), Kovacs (azul) y ed-ENM (verde). En negro se muestran los *B-factors* de MD. Estos perfiles se han tomado de la Figura 4 de Orellana *et al.* (2010).

las de MD. En la Figura 9 se comparan las fluctuaciones obtenidas a partir de las trayectorias de MD (negro) con las de NMA utilizando tres redes elásticas distintas (rojo, azul o verde) y cuatro proteínas representativas. Como puede apreciarse, aunque la amplitud absoluta de las fluctuaciones puede variar, la amplitud relativa entre las fluctuaciones de NMA y las de MD es muy similar en todos los casos. Otra manera de validar las fluctuaciones predichas por el NMA con datos experimentales consiste en comparar las diferencias entre los ángulos diedros de pares de estructuras en distintas conformaciones con las predicciones de deformabilidad (Kovacs, *et al.*, 2004). De este modo es posible verificar que los residuos cuyos ángulos diedros varían en mayor medida durante el cambio conformacional corresponden generalmente a aquellos que se deforman más fácilmente según el NMA.

### 3.2.2. Correspondencia con dinámica esencial

La técnica de dinámica esencial (ED, del inglés *Essential Dynamics*), también conocida como análisis quasi-armónico, permite obtener unos modos de deformación a partir de una trayectoria, generalmente de MD (Amadei, *et al.*, 1993; Brooks, *et al.*, 1995). Para ello, se calcula la matriz de covarianza de la trayectoria y se descompone en una serie de vectores ortogonales utilizando el análisis de componentes principales (PCA, del inglés *Principal Component Analysis*). Estos vectores, o componentes principales, representan las direcciones preferentes a lo largo de las cuales cambia la conformación de la macromolécula durante la simulación. Los componentes principales se obtienen de forma ordenada según la varianza de la trayectoria justificada por cada uno de ellos. De esta forma, los componentes principales de mayor varianza son también los más colectivos, tal y como ocurre con los modos normales de menor frecuencia del NMA. Gracias a la similitud entre componentes principales y modos normales es posible comparar indirectamente las técnicas de MD y NMA. En Rueda *et al.* (Rueda, *et al.*, 2007a) se ha llevado a cabo de manera exhaustiva este tipo de comparativa utilizando 30 estructuras representativas de proteínas. Los componentes principales proceden de trayectorias largas (10-100 ns) de MD obtenidas utilizando campos de fuerza actuales en presencia explícita de disolvente y los modos normales se han obtenido con distintos modelos de red elástica utilizando una representación CG basada en átomos  $C_\alpha$ . Los resultados de la comparativa revelan un gran parecido entre los espacios conformacionales descritos por ambas técnicas. Además, éste es independiente tanto del tipo de plegamiento como del tamaño de la proteína. En Orellana *et al.* (Orellana, *et al.*, 2010) utilizamos como referencia dichas trayectorias junto con otras adicionales para construir un nuevo modelo de red elástica CG a nivel de los átomos  $C_\alpha$  que representa los movimientos de las proteínas mejor que los modelos convencionales. Este nuevo modelo, denominado ed-ENM (del inglés, *essential dynamics*



**Figura 10.** Comparación del solapamiento entre los espacios vectoriales de ED y NMA. Se muestran los solapamientos medios para el conjunto de proteínas estudiado en Orellana *et al.* (2010) obtenidos con las redes elásticas de Tirion (rojo), Kovacs (azul) y ed-ENM (verde). A la izquierda se ha representado el solapamiento utilizando el número de modos necesario para justificar el 90% de la varianza de ED, y a la derecha, utilizando los 50 primeros. Estos datos han sido tomados de la Figura 2 de Orellana *et al.* (2010).

*Elastic Network Model*), endurece las constantes de fuerza de las interacciones que cada átomo establece con los tres que le siguen en la secuencia y utiliza constantes de fuerza que decaen exponencialmente ( $\sim 1/d^6$ ) con la distancia ( $d$ ) para el resto de interacciones. De esta forma, el espacio conformacional descrito por el ed-ENM se parece más al espacio conformacional de ED y, por tanto, a MD. En la Figura 10 se comparan los solapamientos entre el espacio conformacional de ED y los del NMA obtenidos utilizando distintos modelos de red elástica. Como puede apreciarse, cuando se utiliza la red elástica ed-ENM (verde) los solapamientos son mayores que con las redes de Tirion (rojo) y Kovacs (azul). Esta última, a diferencia de la red elástica de Tirion, emplea una constante de fuerza diferente para cada par de átomos que decae exponencialmente ( $\sim 1/d^6$ ) en función de la distancia que los separa (Kovacs, *et al.*, 2004). Véase la sección 4.5.1 de Materiales y métodos para una descripción de la métrica de solapamiento entre espacios vectoriales.

### 3.2.3. Correlación con conjuntos de conformaciones obtenidos por RMN

La técnica de RMN generalmente proporciona un conjunto de estructuras compatibles con los datos experimentales que representan las diferentes conformaciones que coexisten en disolución. Después de alinear las estructuras entre sí, es posible obtener unos perfiles de RMSD que indican la variabilidad estructural para cada residuo. En Yang *et al.* (Yang, *et al.*, 2007) se comparan estos perfiles calculados para 64 proteínas representativas con las fluctuaciones



predichas por los principales métodos de NMA. En principio, todos estos métodos predicen los RMSDs de los conjuntos de estructuras de RMN igual de bien, por ejemplo, el modelo ANM alcanza correlaciones de 0.68 y los modelos GNM y de Tirion llegan a 0.74. Sin embargo, estos resultados han de ser tomados con precaución dado que la incertidumbre experimental puede dar lugar a la existencia de múltiples conformaciones compatibles con los datos.

De manera similar a ED, también es posible realizar el PCA a partir de los conjuntos de estructuras resueltas por RMN y comparar directamente los componentes principales obtenidos con los modos normales de vibración. En la Tabla 1 se muestra el solapamiento entre los espacios vectoriales descritos por los 5 modos de menor frecuencia y los 5 componentes principales de mayor varianza, *i.e.* los más colectivos, de 20 proteínas. Estos datos se han tomado de Orellana *et al.* (Orellana, *et al.*, 2010). Como puede observarse, los solapamientos son mayores al utilizar

**Tabla 1.** Solapamiento entre los espacios vectoriales descritos por NMA y por los componentes principales de conjuntos de estructuras de RMN.

PDB <sup>a</sup>	N <sup>b</sup>	M <sup>c</sup>	$\gamma_5^d$		
			Tirion	Kovacs	ed-ENM
1ro4	58	35	0.56	0.52	0.58
1e9t	59	59	0.51	0.48	0.63
1bw5	66	50	0.69	0.59	0.74
2eot	74	32	0.52	0.42	0.61
1a6x	87	49	0.55	0.44	0.56
1bve	99	28	0.47	0.52	0.49
1q06	101	55	0.57	0.58	0.57
2czn	103	38	0.62	0.55	0.66
1a90	108	31	0.36	0.44	0.49
2bo5	120	44	0.54	0.37	0.58
1e5g	120	50	0.71	0.70	0.69
1cmo	127	43	0.70	0.61	0.70
1iti	133	31	0.53	0.65	0.59
1c89	134	40	0.70	0.76	0.80
1xsb	153	39	0.46	0.43	0.49
1bf8	205	20	0.47	0.55	0.54
1by1	209	20	0.55	0.55	0.56
1n6u	212	22	0.63	0.61	0.59
2jz4	299	20	0.53	0.51	0.61
2d21	370	20	0.60	0.56	0.65
Promedio	142	36	0.56	0.54	0.61

<sup>a-c</sup> Código PDB (<sup>a</sup>), número de aminoácidos (<sup>b</sup>) y número de modelos de RMN (<sup>c</sup>).

<sup>d</sup> Solapamiento entre los componentes principales de ED y los modos de NMA obtenidos a partir de las redes elásticas de Tirion, Kovacs y ed-ENM. En estos casos se han utilizado únicamente los 5 primeros modos y componentes principales.

\* Datos reproducidos de Orellana *et al.* (Orellana, *et al.*, 2010).

la red elástica del ed-NMA. Por tanto, además de aumentar el parecido del NMA con MD, el ed-NMA también describe mejor la flexibilidad de los conjuntos de estructuras de RMN.

### 3.3. Aplicaciones del análisis vibracional

#### 3.3.1. Predicción de flexibilidad

Existen muchos ejemplos que constatan la capacidad del NMA para predecir la flexibilidad de las estructuras macromoleculares. En los últimos años han proliferado una serie de servidores web con bases de datos o herramientas que permiten obtener de manera sencilla información basada en el NMA a partir de las estructuras atómicas de las macromoléculas biológicas (véase la Tabla 2). Entre las bases de datos destacan *Promode* y *MolMovDB*. *Promode* contiene más de 3000 animaciones pre-calculadas de los modos normales obtenidos en CI y considerando todos los átomos. *MolMovDB* recopila una gran cantidad de movimientos de proteínas a partir de las estructuras atómicas en distintas conformaciones depositadas en el PDB. Asimismo, ésta última permite al usuario utilizar sus propias estructuras para calcular los modos normales y predecir qué residuos serán los más flexibles. Con los servidores *MinActionPath* y *PATH-ENM* es posible utilizar los modos normales de vibración para calcular una trayectoria de transición entre dos conformaciones diferentes. El servidor *HingeProt* predice las regiones bisagra, y *DFprot* analiza la deformabilidad y movilidad de cada residuo. Ambos pueden utilizarse con estructuras proporcionadas por el usuario. Con *DFprot* también es posible visualizar y personalizar interactivamente animaciones de los modos vibracionales.

**Tabla 2.** Servidores web basados en NMA.

Servidor	Dirección (http://)	Referencia
<i>AD-ENM</i>	<a href="http://enm.lobos.nih.gov/start.html">enm.lobos.nih.gov/start.html</a>	(Zheng y Doniach, 2003)
<i>DC-ENM</i>	<a href="http://enm.lobos.nih.gov/start_dc.html">enm.lobos.nih.gov/start_dc.html</a>	(Zheng y Brooks, 2005)
<i>DFProt</i>	<a href="http://chaconlab.org/DFprot/index.html">chaconlab.org/DFprot/index.html</a>	(Garzon, <i>et al.</i> , 2007)
<i>ElNémo</i>	<a href="http://www.igs.cnrs-mrs.fr/elnemo/">www.igs.cnrs-mrs.fr/elnemo/</a>	(Suhre y Sanejouand, 2004a)
<i>HingeProt</i>	<a href="http://www.prc.boun.edu.tr/appserv/prc/hingeprot/">www.prc.boun.edu.tr/appserv/prc/hingeprot/</a>	(Emekli, <i>et al.</i> , 2008)
<i>iGNM</i>	<a href="http://ignm.cccb.pitt.edu/">ignm.cccb.pitt.edu/</a>	(Yang, <i>et al.</i> , 2005)
<i>MinActionPath</i>	<a href="http://lorentz.immstr.pasteur.fr/joel/index.php">lorentz.immstr.pasteur.fr/joel/index.php</a>	(Franklin, <i>et al.</i> , 2007)
<i>MolMovDB</i>	<a href="http://www.molmovdb.org/">www.molmovdb.org/</a>	(Gerstein y Krebs, 1998)
<i>MoViES</i>	<a href="http://ang.cz3.nus.edu.sg/cgi-bin/prog/norm.pl">ang.cz3.nus.edu.sg/cgi-bin/prog/norm.pl</a>	(Cao, <i>et al.</i> , 2004)
<i>NOMAD-Ref</i>	<a href="http://lorentz.immstr.pasteur.fr/nomad-ref.php">lorentz.immstr.pasteur.fr/nomad-ref.php</a>	(Lindahl, <i>et al.</i> , 2006)
<i>oGNM</i>	<a href="http://ignm.cccb.pitt.edu/GNM_Online_Calculation.htm">ignm.cccb.pitt.edu/GNM_Online_Calculation.htm</a>	(Yang, <i>et al.</i> , 2006)
<i>Path-ENM</i>	<a href="http://enm.lobos.nih.gov/start_path.html">enm.lobos.nih.gov/start_path.html</a>	(Zheng, <i>et al.</i> , 2007)
<i>ProMode</i>	<a href="http://cube.socs.waseda.ac.jp/pages/jsp/index.jsp">cube.socs.waseda.ac.jp/pages/jsp/index.jsp</a>	(Wako, <i>et al.</i> , 2004)
<i>TMM@</i>	<a href="http://services.cbu.uib.no/tools/tmma">services.cbu.uib.no/tools/tmma</a>	(Skjaerven, <i>et al.</i> , 2007)
<i>WEBnm@</i>	<a href="http://apps.cbu.uib.no/webnma/home">apps.cbu.uib.no/webnma/home</a>	(Hollup, <i>et al.</i> , 2005)

*WEBnm@* es capaz de visualizar los desplazamientos asociados a los modos, identificar los dominios dinámicos y los modos normales que más contribuyen a un cambio conformacional dado. Aunque la mayoría de los servidores utilizan una representación CG basada en los átomos  $C_\alpha$ , el servidor *ElNémo* utiliza la aproximación del RTB por lo que puede obtener los modos normales de grandes estructuras considerando todos sus átomos. Este servidor también permite calcular la colectividad, la amplitud de las fluctuaciones atómicas y la correlación de éstas últimas con los *B-factors* experimentales. Además, es posible perturbar la conformación de una estructura atómica inicial utilizando sus modos normales para generar modelos que ayuden en la obtención de estructuras de CRX mediante la técnica de reemplazamiento molecular (véase la siguiente sección). El servidor *NOMAD-Ref* también permite obtener conjuntos de estructuras de gran tamaño utilizando RTB.

### 3.3.2. Refinamiento estructural

El NMA ha resultado ser de gran utilidad en el refinamiento de datos de CRX. Las primeras aproximaciones permitieron refinar los factores de Debye-Waller (*B-factors*) incluyendo combinaciones lineales de los modos de baja frecuencia en el modelo (Diamond, 1990; Kidera y Go, 1990; Kidera, *et al.*, 1992). El refinamiento basado en los modos normales no solo permite obtener, a partir de una estructura atómica inicial preexistente, estructuras que se ajustan mejor a los datos de difracción (*R-factor*), sino que también mejora los porcentajes de ángulos diedros en las regiones favorecidas y permitidas de Ramachandran resolviendo incluso residuos que no se encontraban en el modelo inicial (Chen, *et al.*, 2007). La técnica cristalográfica denominada reemplazamiento molecular permite resolver rápidamente el problema de las fases a partir de una estructura molde que se parezca a la macromolécula estudiada (Cantor y Schimmel, 1980). En ocasiones, las estructuras molde disponibles se encuentran en conformaciones diferentes, lo que dificulta o impide alcanzar estructuras atómicas válidas. En este contexto, el uso de uno o dos modos de baja frecuencia para generar modelos en conformaciones alternativas a partir del molde ha demostrado ser una aproximación interesante para ayudar a resolver las estructuras atómicas (Suhre y Sanejouand, 2004b). El análisis vibracional también permite mejorar la calidad de estructuras atómicas procedentes de modelado por homología empleando unos pocos modos de baja frecuencia para reducir así el espacio de búsqueda (Leo-Macias, *et al.*, 2005a; Leo-Macias, *et al.*, 2005b). Además, en el refinamiento de este tipo de modelos se ha demostrado utilizando datos simulados que los modos normales se comportan mejor que trayectorias cortas de MD o de MC (Stumpff-Kane, *et al.*, 2008).

### 3.3.3. Transiciones conformacionales

La simulación de la transición entre distintas conformaciones de una misma macromolécula constituye un reto de gran importancia en biología, especialmente en los complejos de gran tamaño. Desafortunadamente, aunque se han desarrollado métodos más o menos sofisticados basados tanto en redes elásticas (Best, *et al.*, 2005; Kim, *et al.*, 2002; Weiss y Levitt, 2009; Zheng, *et al.*, 2007) como en NMA (Franklin, *et al.*, 2007; Lindahl, *et al.*, 2006; Maragakis y Karplus, 2005; Miyashita, *et al.*, 2003; Zheng y Brooks, 2005), no existe una forma fácil y rigurosa de validar los cambios conformacionales debido a las limitaciones de las técnicas experimentales que proporcionan información estructural con resolución temporal. A pesar de ello, ha sido posible utilizar estos métodos para obtener trayectorias factibles que conectan las distintas conformaciones macromoleculares.

### 3.3.4. Predicción de interacciones proteína-proteína

A pesar de que la mayoría de los complejos proteicos todavía no han sido resueltos, existen bastantes estructuras de alta resolución de sus componentes individuales. Por esta razón, la predicción de la estructura atómica de un complejo a partir de la de sus constituyentes representa un reto muy importante en biología. Dentro de este campo, denominado ajuste proteína-proteína (del inglés, *protein-protein docking*) es posible encontrar un buen número de aproximaciones que ya han sido revisadas detalladamente en la bibliografía (Andrusier, *et al.*, 2008; Bonvin, 2006; Camacho y Vajda, 2002; Deremble y Lavery, 2005; Halperin, *et al.*, 2002; Lensink y Mendez, 2008; Pons, *et al.*, 2010; Ritchie, 2008; Sternberg, *et al.*, 1998; Vakser y Kundrotas, 2008). En general, estos métodos utilizan una estrategia de dos pasos: *i*) identificación aproximada de las regiones de interacción, donde se tratan de discernir las posiciones potencialmente cercanas a la correcta, y *ii*) refinamiento de estas soluciones para encontrar la orientación correcta, donde se trata de cribar y mejorar las soluciones encontradas en el paso anterior. El primer paso consiste generalmente en una búsqueda exhaustiva global de posibles soluciones utilizando una función de ajuste basada en la complementariedad de superficies. En un segundo paso, las mejores soluciones encontradas en la búsqueda anterior se evalúan de nuevo para discernir aquellas más cercanas a la interacción real utilizando funciones que describen con mayor precisión la energética de la interacción. En el test ciego internacional, CAPRI (Lensink, *et al.*, 2007), se evalúa la capacidad de predicción actual de las interacciones proteína-proteína a nivel estructural. Generalmente, cuando no se producen cambios conformacionales significativos entre las conformaciones libre y unida, la mayoría de los métodos que participan en dicho test son capaces de predecir adecuadamente la geometría del complejo. Sin embargo, cuando existen cambios conformacionales significativos (incluso

tan pequeños como  $\sim 2\text{\AA}$ ), los mismos métodos fallan en obtener una predicción aceptable. En consecuencia, la consideración o no de la flexibilidad juega un papel decisivo en la solución de los problemas de ajuste proteína-proteína. Aunque la mayoría de los métodos solo consideran la flexibilidad a nivel de cadenas laterales, se han desarrollado distintas aproximaciones para tenerla en cuenta de diferentes formas. Unas lo hacen implícitamente, o bien permitiendo un cierto grado de penetrabilidad entre las superficies de interacción especialmente con las cadenas laterales (*soft docking*) (Fernandez-Recio, *et al.*, 2002; Palma, *et al.*, 2000), o bien utilizando un conjunto de posibles conformaciones tomadas, por ejemplo, de RMN, simulaciones de MD o cualquier otro método de muestreo conformacional (*ensemble docking*) (Dominguez, *et al.*, 2004; Keskin, *et al.*, 2005; Rajamani y Reynolds, 2004). También existen varios métodos que permiten tratar explícitamente la flexibilidad para mejorar los ajustes (Dominguez, *et al.*, 2003; Fernandez-Recio, *et al.*, 2002; Gray, *et al.*, 2003). Dentro de este último grupo, algunos de ellos emplean componentes principales de ED (Smith, *et al.*, 2005) y otros modos normales de vibración (Lindahl y Delarue, 2005; May y Zacharias, 2005; May y Zacharias, 2008a) como variables colectivas adicionales. La utilización de los modos normales de vibración para considerar la flexibilidad en el contexto del ajuste proteína-proteína constituye una manera eficiente de mejorar tanto la geometría como la puntuación de los modelos, especialmente cuando los cambios conformacionales implican movimientos colectivos del esqueleto polipeptídico (Zacharias, 2010).

### 3.3.5. Predicción de interacciones proteína-ligando

Los medicamentos son compuestos químicos que actúan sobre las macromoléculas biológicas modulando su acción. Su uso médico permite combatir y prevenir enfermedades, tratar el dolor y mejorar la calidad de vida de las personas. Una de las estrategias más importantes para descubrir y desarrollar nuevos medicamentos consiste en llevar a cabo el cribado de miles de compuestos, *i.e.* efectuar los ensayos bioquímicos necesarios para seleccionar aquellos que poseen actividad terapéutica. Finalmente, tan solo unos pocos llegan a ser utilizados como fármacos. La realización de estos ensayos es un proceso muy costoso en términos de tiempo, recursos, infraestructura y personal especializado (Mayr y Bojanic, 2009).

El cribado virtual de ligandos (VLS, del inglés *Virtual Ligand Screening*) representa una alternativa computacional eficaz para reducir costes limitando el número de candidatos potencialmente activos que han de ser ensayados experimentalmente (Shoichet, 2004). El objetivo del VLS es seleccionar computacionalmente aquellos compuestos (ligandos) que tienen más probabilidades de interactuar con una diana biológica, generalmente una enzima o un receptor, utilizando sus estructuras atómicas. Para ello, se lleva a cabo el ajuste

proteína-ligando (del inglés, *protein-ligand docking*), *i.e.* la predicción y caracterización de la interacción entre una proteína y sus potenciales ligandos empleando métodos computacionales. Aunque la mayoría de las aproximaciones actuales tienen en cuenta la flexibilidad del ligando, la consideración de la flexibilidad del receptor es la que suele condicionar el éxito de los ajustes (Sousa, *et al.*, 2006). Los avances recientes en este sentido han sido revisados en la bibliografía (Carlson y McCammon, 2000; Cozzini, *et al.*, 2008; Sousa, *et al.*, 2006; Teague, 2003; Teodoro y Kavraki, 2003; Totrov y Abagyan, 2008). Entre los procedimientos propuestos destacan aquellos que utilizan múltiples conformaciones del receptor (MRCs, del inglés *Multiple Receptor Conformations*) (Bottegoni, *et al.*, 2011; Totrov y Abagyan, 2008). Estas conformaciones se pueden obtener a partir de fuentes experimentales (CRX, RMN) y de métodos computacionales basados en simulaciones de MD, ED y NMA. Aunque la mayoría de los estudios utilizan conjuntos de estructuras experimentales (principalmente cristalográficas), para la mayoría de dianas terapéuticas no existe un número suficiente de estructuras que capture adecuadamente su flexibilidad. Por este motivo, el empleo de métodos de simulación resulta muy útil. Probablemente, la mejor manera de obtener MRCs sea utilizar simulaciones de MD en condiciones fisiológicas empleando potenciales atómicos detallados. Por ejemplo, el MD ha sido aplicado con éxito en el estudio de fármacos antivirales contra la neuraminidasa del virus de la gripe aviar (Cheng, *et al.*, 2008) y la integrasa (Carlson y McCammon, 2000) y proteasa (Meagher y Carlson, 2004; Tatsumi, *et al.*, 2004) del virus de inmunodeficiencia humana. También es posible utilizar los componentes principales de una trayectoria de MD, *i.e.* ED, para generar MRCs y predecir con éxito interacciones proteína-ligando (Teodoro, *et al.*, 2003; Zacharias, 2004). No obstante, el elevado coste de los recursos necesarios para llevar a cabo las simulaciones de MD limita seriamente su aplicación en el VLS con receptores de gran tamaño y grandes librerías de compuestos. En este contexto, la utilización de los modos del NMA para generar conformaciones a partir de una única estructura conocida constituye una alternativa muy eficiente (Cavasotto, *et al.*, 2005; Floquet, *et al.*, 2006; Kovacs, *et al.*, 2005; May y Zacharias, 2008b; Rueda, *et al.*, 2009; Sander, *et al.*, 2008; Zacharias y Sklenar, 1999). Aunque la mayoría de las aplicaciones del NMA en el ajuste proteína-ligando emplean CC, en Kovacs *et al.* (Kovacs, *et al.*, 2005) se utilizan CI para representar la flexibilidad del receptor.

### **3.3.6. Respuesta mecánica**

Gracias al desarrollo de las técnicas de microscopía de fuerzas atómicas, actualmente es posible manipular macromoléculas individualmente y ejercer fuerzas minúsculas (del orden de pico-Newtons) sobre ellas y en posiciones específicas (Oberhauser y Carrion-Vazquez, 2008). En este contexto, se puede analizar teóricamente el comportamiento mecánico y la estabilidad



mecánica de las macromoléculas cuando se las somete a estiramientos en distintas direcciones utilizando tanto MD dirigida (Grater, *et al.*, 2005) como ANM (Eyal y Bahar, 2008). Al comparar las predicciones con los datos obtenidos experimentalmente se observa una buena correlación entre ambos, y no solo en las proximidades de la conformación de equilibrio, sino también en el coste efectivo de la desnaturalización mecánica de las proteínas. Por tanto, los métodos de NMA-CG son unas herramientas muy útiles para estimar la respuesta mecánica cuando se ejercen fuerzas externas sobre las estructuras macromoleculares.

### **3.3.7. Ajuste flexible de mapas de microscopía electrónica**

Los modos normales de vibración también han sido utilizados con éxito para tener en cuenta la flexibilidad durante el ajuste de estructuras atómicas en el interior de mapas de densidad de MET. Existen varios métodos que utilizan uno o unos pocos modos de baja frecuencia para generar pequeñas deformaciones de la estructura inicial con las que iterativamente mejorar el ajuste frente al mapa objetivo (Delarue y Dumas, 2004; Hinsen, *et al.*, 2005; Suhre, *et al.*, 2006; Tama, *et al.*, 2004a; Tama, *et al.*, 2004b). Estas técnicas se comentarán con mayor detalle en la sección 3.5.3 dado que la aplicación del NMA en el ajuste flexible de estructuras atómicas en mapas de MET constituye un objetivo principal de esta tesis.

## **3.4. Microscopía electrónica y análisis de partículas individuales**

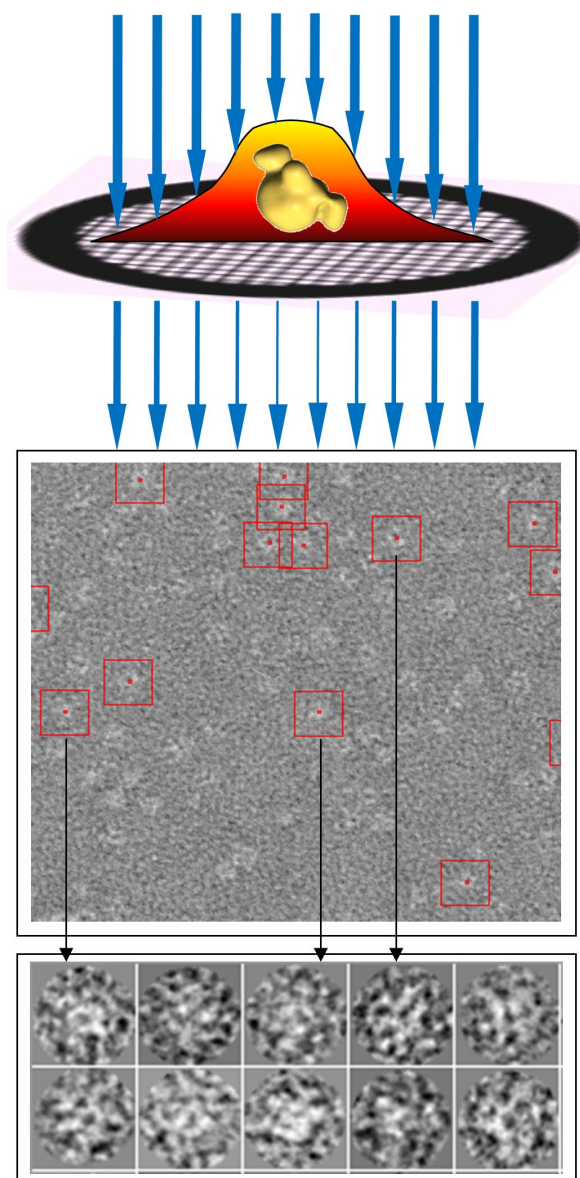
El campo de actuación de esta tesis incluye la utilización y obtención de datos estructurales mediante MET, por esta razón se comentarán brevemente los principios básicos de esta técnica así como los aspectos metodológicos del procesamiento digital y análisis de imagen necesarios para la obtención de reconstrucciones 3D de las macromoléculas (mapas). Aunque existen diversas técnicas basadas en MET para obtener estas reconstrucciones, tales como TE o CRX (véase Motivación), nos centraremos en el análisis de partículas individuales (SP) que es la técnica más utilizada. El SP se ha convertido en una técnica estándar en biología estructural que permite visualizar macromoléculas generalmente a resoluciones de 10 a 40 Å, según la variante empleada (Frank, 2006). En contraste con las técnicas cristalográficas, el SP es capaz de proporcionar mapas de densidad electrónica en una gran variedad de condiciones experimentales cercanas a las fisiológicas a partir de cantidades muy pequeñas de muestra. Una de las principales limitaciones de esta técnica es el tamaño mínimo para que una molécula pueda ser observada. Este tamaño depende principalmente de la variante utilizada: si la muestra se vitrifica en medio acuoso el límite es de ~200 kDa y si se utiliza tinción negativa o criotinción es de ~70 kDa (ver más adelante). La resolución constituye otra limitación importante, aunque en ocasiones sea posible alcanzar resolución atómica (Zhou, 2011). Por otra parte, es posible

reconstruir la estructura atómica a partir de los mapas de baja resolución gracias a las técnicas de ajuste macromolecular (Wriggers y Chacon, 2001). Aún con sus limitaciones, la versatilidad del SP lo convierte en una herramienta clave para avanzar en el conocimiento funcional y dinámico de los principales sistemas macromoleculares que gobiernan los procesos celulares.

En SP la estructura 3D de una molécula o complejo macromolecular se genera a partir de la información contenida en sus imágenes de proyección. Estas imágenes se forman cuando la muestra es atravesada en el interior de un microscopio electrónico de transmisión por un haz coherente de electrones acelerado con un potencial de varios miles de voltios (100-300 kV) (véase la Figura 11). Como la intensidad de dicho haz se ve atenuada proporcionalmente a la cantidad de electrones de los átomos que atraviesa, es posible obtener en el detector (una película fotográfica o un dispositivo digital CCD, del inglés *Charge Coupled Device*) imágenes que integran en 2D toda la información estructural 3D de la muestra. Según el teorema de Radon (Radon, 1986), combinando adecuadamente estas proyecciones es posible reconstruir la estructura 3D original.

#### 3.4.1. Preparación de la muestra

Para obtener las imágenes en proyección, la muestra se deposita sobre una rejilla recubierta con una fina capa de carbón. De este modo, las moléculas se adsorben al carbón, en principio, en todas las orientaciones posibles. Antes de introducir la muestra en el microscopio es necesario prepararla para soportar tanto las condiciones de vacío del interior de la columna como la



**Figura 11.** Imágenes en proyección de partículas individuales de la proteína RepB. Las imágenes se han obtenido en el microscopio electrónico de transmisión utilizando tinción negativa.



radiación electrónica. Existen tres variantes de preparación del espécimen: tinción negativa, vitrificación y criotinción (Frank, 2006; Llorca, 2005). La tinción negativa consiste en cubrir la muestra con un agente de tinción, que suele ser una sal de un metal pesado como el acetato (o formiato) de uranilo. A pesar de que este tipo de tinción incrementa significativamente el contraste de las imágenes obtenidas (véase la Figura 11), la resolución está limitada a unos 20 Å debido a que el tamaño de grano de la tinción hace imposible que ésta penetre en algunas zonas de la macromolécula (Cheng, *et al.*, 2006; Llorca, 2005). La vitrificación consiste en congelar la muestra en solución acuosa a gran velocidad utilizando etano líquido para que el hielo formado sea vítreo y no dañe al espécimen (Cheng y Walz, 2009; Dubochet, *et al.*, 1988; van Heel, *et al.*, 2000). La muestra se mantiene y se visualiza congelada a temperatura de nitrógeno líquido (unos -180 °C), lo que asegura su preservación e hidratación en condiciones cercanas a las fisiológicas. Aunque esta variante es la que permite obtener la mayor resolución, está limitada a macromoléculas de más de ~200 kDa debido al bajo contraste ocasionado por la pequeña diferencia de densidad electrónica entre la muestra y el hielo vítreo circundante. La criotinción es una técnica intermedia entre las dos anteriores que consigue incrementar el contraste vitrificando la muestra en una solución con un agente de tinción negativa (Adrian, *et al.*, 1998). Aunque esta técnica mejora el contraste de la vitrificación y preserva la estructura nativa mejor que con tinción negativa, la resolución alcanzada suele ser del orden de la obtenida con tinción negativa y no se utiliza habitualmente.

### **3.4.2. Toma de micrografías**

Una vez preparada la muestra para soportar las condiciones de vacío, se introduce ésta en el interior de la columna del microscopio. Debido a la alta energía y gran capacidad de interacción del haz de electrones con la materia, durante la toma de micrografías es necesario emplear la técnica de mínima dosis para preservar así la máxima información estructural posible. Esta técnica consiste en utilizar una dosis muy pequeña de electrones ( $\sim 10\text{e}^-/\text{\AA}^2$ ) para evitar un deterioro significativo de la estructura de las macromoléculas. Para ello, se enfoca en una región próxima a la de interés de forma que esta última permanezca intacta hasta la toma de la micrografía. Las micrografías pueden realizarse, o bien digitalmente utilizando una cámara CCD, o bien analógicamente en placas fotográficas. En este último caso es necesario un paso de digitalización posterior. La relación señal-ruido (SNR, del inglés *Signal to Noise Ratio*) de imágenes tomadas en condiciones de mínima dosis es muy baja (Figura 11). Como veremos a continuación, esto no solo condiciona el procesamiento y análisis digital de las imágenes, sino que constituye una de las mayores dificultades del SP. Además, las imágenes tomadas en el microscopio electrónico están afectadas por un efecto conocido como función de

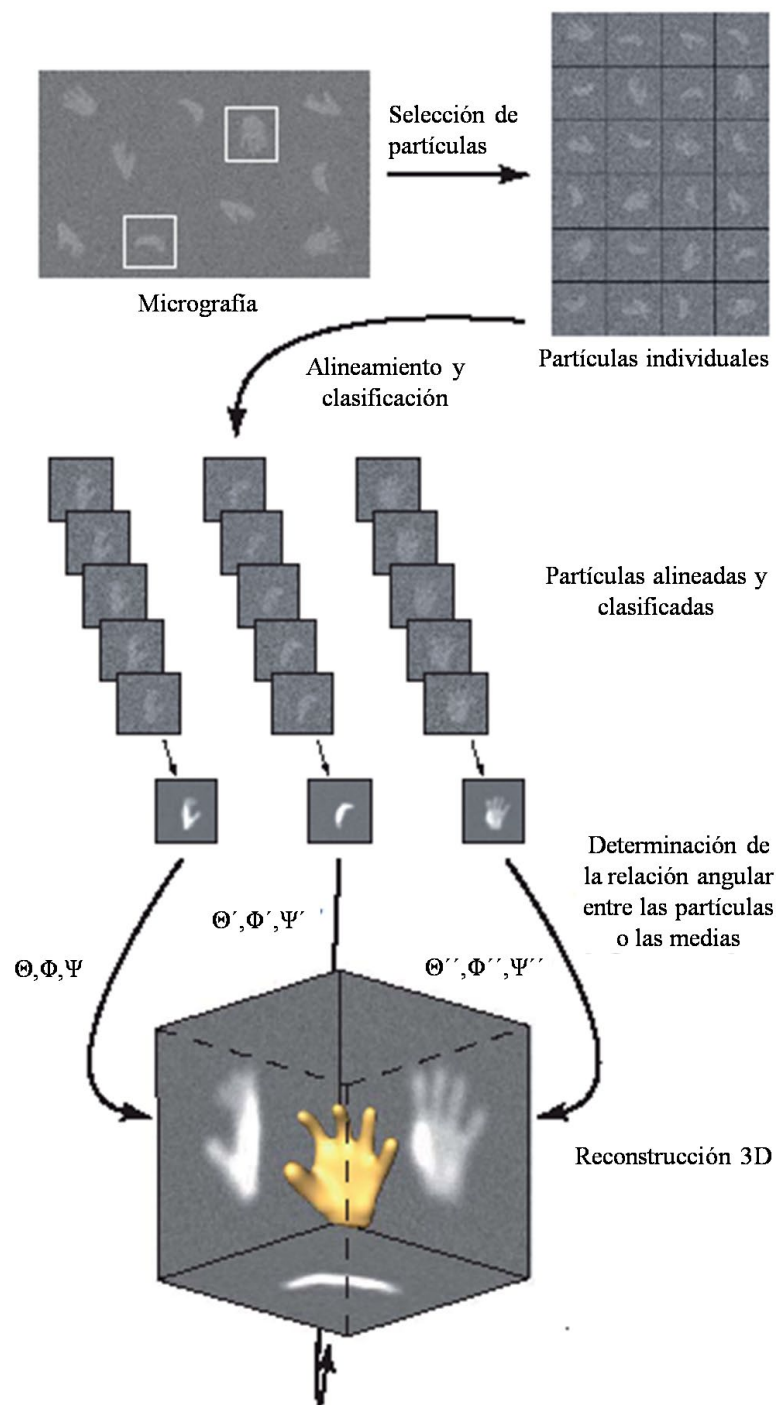
transferencia de contraste (CTF, del inglés *Contrast Transfer Function*). Este efecto proviene de las aberraciones de las lentes del microscopio y está modulado por el desenfoque utilizado para visualizar las imágenes y por la coherencia de los electrones emitidos. Para obtener la estructura de la molécula original a la mayor resolución posible es necesario determinar y corregir la CTF; sin embargo, cuando se emplea tinción negativa sus efectos no son muy importantes por lo que esto no suele ser necesario.

### **3.4.3. Procesamiento de imagen y reconstrucción tridimensional**

El procesamiento y análisis digital de las imágenes puede llevarse a cabo utilizando distintos paquetes informáticos. Entre los más conocidos y usados se encuentran EMAN (Ludtke, 2010; Ludtke, *et al.*, 1999), SPIDER (Frank, *et al.*, 1996), IMAGIC (van Heel, *et al.*, 1996) y XMIPP (Marabini, *et al.*, 1996; Sorzano, *et al.*, 2004). Estos paquetes contienen una serie de herramientas y comandos que automatizan algunas de las tareas necesarias para procesar las imágenes de MET y obtener reconstrucciones 3D (véase la Figura 12), todo el proceso de reconstrucción requiere la interacción y toma de decisiones del usuario. Una vez obtenidas suficientes micrografías en formato digital se procede a extraer de ellas varios miles de imágenes de partículas individuales que representan diferentes orientaciones del espécimen. Como la SNR de las imágenes tomadas en condiciones de mínima dosis es muy baja, es necesario promediar un gran número de ellas para neutralizar el ruido. No obstante, las imágenes que corresponden a la misma orientación han de estar alineadas entre sí previamente. Para ello, las imágenes se centran, clasifican, alinean y promedian iterativamente sin utilizar más información que la contenida en los propios datos (Figura 12). Este proceso se conoce como obtención de imágenes medias libres de patrón. Las diferentes medias obtenidas tras la clasificación representan vistas de las diferentes orientaciones de la molécula, siempre y cuando todas las partículas provengan del mismo espécimen y se encuentren en la misma conformación.

Uno de los pasos críticos del proceso consiste en encontrar la orientación relativa correcta entre las imágenes para obtener la reconstrucción 3D de la macromolécula. Existen dos estrategias principales para determinarla: encaje de proyecciones (del inglés, *projection matching*) y reconstitución angular (del inglés, *angular reconstitution*). Ambas requieren la obtención de un mapa inicial que puede ser generado, o bien a partir de los datos utilizando las técnicas de “líneas comunes” (Crowther, 1971) o RCT (del inglés, *Random Conical Tilt*) (Radermacher, 1988), o bien de manera ajena a ellos, por ejemplo, a partir de ruido, de un “blob” gaussiano o de un mapa independiente. En el encaje de proyecciones, las proyecciones del mapa inicial se emplean como referencias para mejorar el alineamiento y clasificación de las partículas. De esta forma, se asigna una nueva orientación a cada imagen para generar un

mapa mejor que será proyectado de nuevo en la siguiente iteración para refinar el alineamiento y la clasificación de las partículas. En la técnica de reconstitución angular también se usan las proyecciones de los mapas para mejorar el alineamiento y la clasificación. No obstante, no se utilizan las orientaciones de las proyecciones para generar el nuevo mapa, sino que se calculan de nuevo por líneas comunes en cada iteración. En principio, utilizando iterativamente cualquiera de estas dos estrategias es posible obtener un mapa final compatible con los datos.



**Figura 12.** Representación esquemática del proceso de reconstrucción 3D por microscopía electrónica de transmisión y análisis de partículas individuales.

Aunque no existe un criterio unificado para determinar la resolución de la reconstrucción 3D (Rosenthal y Henderson, 2003; van Heel y Schatz, 2005), ésta puede estimarse de distintas maneras. Uno de los criterios más utilizados es el de la FSC 0.5 (del inglés, *Fourier Shell Correlation*) que define la resolución del mapa como la frecuencia espacial (en espacio recíproco o de Fourier) a la cual la correlación entre dos reconstrucciones generadas cada una seleccionando aleatoriamente la mitad de las partículas es igual a 0.5.

Finalmente, la información estructural proporcionada por la reconstrucción debe ser verificada e interpretada utilizando técnicas adicionales. Por ejemplo, se pueden ajustar las estructuras atómicas de algunos de los componentes de la muestra en el interior de los mapas para ver si son compatibles y permiten localizar dominios concretos. Si no existen estructuras atómicas disponibles se puede llevar a cabo la localización de subunidades o dominios mediante ensayos de delección de dominios o de marcaje con distintas sustancias, tales como anticuerpos, oro, proteínas o péptidos de fusión. La integración de la información estructural procedente de las reconstrucciones 3D con la obtenida en estos experimentos adicionales permite interpretar el funcionamiento de las macromoléculas en condiciones cercanas a las fisiológicas y constituye uno de los objetivos principales de esta tesis.

### **3.5. Ajuste macromolecular**

Como se acaba de mencionar, los mapas de densidad obtenidos por microscopía electrónica pueden ser interpretados utilizando las estructuras atómicas disponibles (Wriggers y Chacon, 2001). Al ajustar estas estructuras en el interior de los mapas a media/baja resolución es posible obtener información cuasi-atómica con la que desentrañar el funcionamiento de los complejos macromoleculares. Como se ha visto anteriormente (Figura 3), cuando las estructuras atómicas disponibles y los mapas se encuentran en la misma conformación, basta con encontrar la orientación adecuada entre ambos utilizando una estrategia de ajuste rígido. Sin embargo, en múltiples ocasiones las conformaciones observadas en los mapas difieren significativamente de las que se encuentran en los cristales. En estos casos, además de encontrar una orientación inicial adecuada entre ambas estructuras, es necesario emplear una estrategia de ajuste flexible para tener en cuenta las diferentes conformaciones.

#### **3.5.1. Ajuste rígido**

Durante los últimos años se han desarrollado un buen número de métodos para ajustar de forma rígida estructuras atómicas en el interior de mapas de densidad cuando las conformaciones de ambos son similares. Para ello, existen diversos programas que permiten llevar a cabo esta tarea con éxito: *SITUS* (Wriggers, *et al.*, 1999), *EMFIT* (Rossmann, 2000), *DOCKEM*

(Roseman, 2000), *FOLDHUNTER* (Jiang, *et al.*, 2001), *COLORES* (Chacon y Wriggers, 2002), *URO* (Navaza, *et al.*, 2002), *COAN* (Volkman y Hanein, 2003), *3SOM* (Ceulemans y Russell, 2004) y *ADP\_EM* (Garzon, *et al.*, 2007a). Esencialmente, estas herramientas llevan a cabo una búsqueda automatizada exhaustiva entre todas las rotaciones y traslaciones posibles para encontrar aquellas que maximizan la correlación entre el mapa objetivo y un mapa simulado a partir de la estructura atómica (Fabiola y Chapman, 2005; Wriggers y Chacon, 2001).

### 3.5.2. Ajuste flexible

Para tener en cuenta los cambios conformacionales durante el ajuste de las estructuras atómicas en los mapas de densidad se emplean métodos de ajuste flexible. Aunque los más utilizados son los basados en MD y NMA, también se han empleado con éxito otras aproximaciones alternativas. En las siguientes secciones se comentan brevemente estos métodos junto con sus características más importantes.

### 3.5.3. Métodos de ajuste flexible basados en el análisis vibracional

Actualmente existen dos herramientas automáticas de ajuste flexible que utilizan NMA en CC: *NMFF* (Tama, *et al.*, 2004a) y *NORMA* (Suhre, *et al.*, 2006). La herramienta *NMFF* (<http://mmts.org/software/nmff.html>), emplea una combinación lineal de modos para maximizar iterativamente la correlación entre el modelo de prueba y el mapa objetivo. En ella, los modos se obtienen con la técnica RTB en CC (Durand, *et al.*, 1994; Tama, *et al.*, 2000). Como se ha mencionado anteriormente, esta aproximación considera cada residuo como un bloque rígido con 6 grados de libertad (rotación/traslación), y como potencial la red elástica de Tirion (Tirion, 1996). Las amplitudes y modos utilizados para deformar las estructuras se eligen según el gradiente en coordenadas normales de la función de correlación siguiendo el esquema de optimización de ascenso más pronunciado (del inglés, *steepest ascent*). De esta forma, se generan modelos de prueba a partir de los 5 modos con mayor gradiente que solo son aceptados si aumenta la correlación. Cada vez que se acepta un nuevo modelo los modos normales son actualizados. Para minimizar las distorsiones, la amplitud de los movimientos se limita de forma que el máximo desplazamiento de cualquiera de los átomos nunca excede 1 Å. En las proximidades del máximo, el protocolo de optimización utiliza el método de Newton-Raphson para obtener directamente las amplitudes exactas que maximizan la correlación empleando también los 5 modos con mayor gradiente. La herramienta *NMFF* resulta efectiva para llevar a cabo el ajuste flexible automatizado de estructuras atómicas en mapas de densidad de conformaciones diferentes, como por ejemplo en: el canal conductor de proteínas (PCC, del inglés *Protein-Conducting Channel*) (Mitra, *et al.*, 2005), el complejo GroEL-Glutamina

sintasa (Falke, *et al.*, 2005), el Factor de Elongación G (del inglés, *Elongation Factor G*), la ARN polimerasa y el virus CCMV (*Cowpea Chlorotic Mottle Virus*) (Tama, *et al.*, 2004b).

En cuanto a *NORMA* (<http://www.igs.cnrs-mrs.fr/elnemo/NORMA/>), también emplea RTB y el potencial de Tirion para obtener los modos normales de vibración en CC. No obstante, el esquema de optimización y la función de coste son diferentes. Esta herramienta adopta un algoritmo basado en un símplice (del inglés, *simplex*) multidimensional y, opcionalmente, en recocido simulado (del inglés, *simulated annealing*) para minimizar la diferencia cuadrática entre el mapa objetivo y los mapas simulados obtenidos a partir de los modelos de prueba. Este método tan solo efectúa el análisis vibracional con la estructura inicial y requiere la intervención del usuario para realizar los ajustes en múltiples pasos. A pesar de ello, ha sido utilizado en el estudio de la integración del virus HIV-1 en el genoma humano (Michel, *et al.*, 2009) y es capaz de obtener ajustes comparables a los de *NMFF* con GroEL, la ATPasa de calcio y la proteína viral VP2 del virus IBDV (Suhre, *et al.*, 2006).

Aunque no se encuentra disponible como herramienta, también existe otro método diferente basado en NMA desarrollado por Hinsen (Hinsen, *et al.*, 2005). Este método iterativo utiliza el gradiente de la desviación cuadrática entre el mapa objetivo y un mapa simulado a partir de las coordenadas atómicas como fuerza motriz para deformar, en pequeños incrementos de  $\sim 1\text{\AA}$ , el modelo atómico inicial según sus modos normales de baja frecuencia. Los modos se calculan utilizando los átomos  $C_\alpha$  y un potencial modificado (Hinsen, *et al.*, 2000) para que los resultados sean parecidos a los que se obtendrían con el campo de fuerzas AMBER94. De esta forma se consiguen ajustes flexibles satisfactorios con varias estructuras atómicas de la ATPasa de calcio y un mapa experimental a  $8\text{\AA}$  de resolución (Hinsen, *et al.*, 2005).

En términos generales, estos métodos resultan muy eficientes computacionalmente y justifican la utilización del NMA para modelar la flexibilidad conformacional de las macromoléculas. No obstante, todos utilizan CC durante el análisis vibracional por lo que los modelos generados pueden presentar deformaciones no deseadas que alteran su estructura covalente. Por ello, generalmente es necesario minimizar la energía de los modelos finales (Michel, *et al.*, 2009; Mitra, *et al.*, 2005).

#### 3.5.4. Métodos de ajuste flexible basados en dinámica molecular

El ajuste flexible de estructuras atómicas en mapas de densidad ha sido ampliamente estudiado desde el campo de la MD. Entre los métodos más recientes se encuentran: *Flex-EM* (Topf, *et al.*, 2008), *MDFF* (Trabuco, *et al.*, 2008), *YUPSCX* (Tan, *et al.*, 2008), *MDFIT* (Ratje, *et al.*, 2010) y dos métodos desarrollados por el grupo de Tama: uno basado en todos los átomos (Tama 1) (Orzechowski y Tama, 2008) y otro en  $C_\alpha$  (Tama 2) (Grubisic, *et al.*, 2010). Todos



ellos incorporan la información del mapa de microscopía electrónica en forma de un término de energía potencial adicional que se añade al término energético del campo de fuerzas utilizado en la simulación de MD. Sus diferencias radican principalmente en el tipo de campo de fuerza empleado, el potencial asociado al mapa, el protocolo de optimización y la presencia o ausencia de restricciones para el mantenimiento de la estructura estereoquímica de la macromolécula.

La herramienta de ajuste flexible de *Flex-EM* (Topf, *et al.*, 2008) (<http://salilab.org/Flex-EM/>) considera a la macromolécula como un conjunto de cuerpos rígidos a distintos niveles de resolución. Para dirigir la trayectoria de MD hacia el interior de los mapas se utiliza un término de energía potencial basado en la función de correlación entre el mapa objetivo ( $\rho^{EM}$ ) y el mapa simulado a partir del modelo actual ( $\rho^P$ ):

$$E^{Flex-EM} = \sum_b^B \frac{\sum_{i \in b} \rho_i^{EM} \rho_i^P}{\sqrt{\sum_{i \in b} (\rho_i^{EM})^2 \sum_{i \in b} (\rho_i^P)^2}} \quad (3.5.1)$$

Nótese que el término energético se obtiene a partir de la suma de las correlaciones entre los voxels ( $i$ ) pertenecientes a cada uno de los  $B$  cuerpos rígidos ( $b$ ) y los correspondientes al mapa objetivo. El protocolo consta de tres fases, en cada una de las cuales se va incrementando la resolución del modelo. En la primera, se lleva a cabo un ajuste rígido de los dominios siguiendo un esquema de optimización de MC en el que únicamente se utiliza como función de coste la correlación. Durante esta fase se eliminan temporalmente los residuos que conectan los dominios para poder ajustarlos libremente. En la segunda fase, se escogen aleatoriamente 20 estructuras de la fase anterior para perturbar aleatoriamente la orientación y posición de los dominios. Acto seguido, se minimiza independientemente la energía de estas 20 estructuras por el método del gradiente conjugado teniendo en cuenta la función de correlación, el campo de fuerzas de MD (CHARMM22) y considerando rígidos los dominios. Los modelos resultantes se agrupan según su RMSD relativo. Dentro de cada grupo se elige a la estructura de menor energía como representante. Los 5 representantes de menor energía se optimizan de nuevo considerando rígidos únicamente los elementos de estructura secundaria y aplicando una serie de restricciones para mantener la estructura estereoquímica. Las restricciones contribuyen a la función energética global como una suma de términos armónicos que varían en función de los ángulos diedros y de las longitudes y ángulos de enlace. Sus valores de equilibrio y constantes de fuerza también proceden de CHARMM22. Estas restricciones también incluyen un término específico para las interacciones entre átomos no enlazados covalentemente. Si la estructura inicial solo posee un dominio, únicamente se lleva a cabo la optimización a nivel de estructura secundaria. En la última fase, se refina por MD y recocido simulado el modelo representante

de menor energía de la fase anterior. Finalmente, se relaja la estructura con 200 pasos de minimización por gradiente conjugado teniendo en cuenta tanto la función de correlación como las restricciones estereoquímicas seguidos de otros 200 empleando únicamente las restricciones, *i.e.* sin tener en cuenta la información del mapa objetivo. En todos los casos se consideran rígidos los elementos de estructura secundaria.

El método *MDFF* (Trabuco, *et al.*, 2008) (<http://www.ks.uiuc.edu/Research/mdff/>), en lugar de utilizar la función de correlación como en el caso de *Flex-EM*, emplea un término local de energía potencial basado en la densidad del mapa ( $\rho$ ) para dirigir el ajuste:

$$\begin{aligned}
 E^{MDFF} &= \sum_a^A w_a \xi \left( 1 - \frac{\rho(\mathbf{r}_a) - \rho_{thr}}{\rho_{max} - \rho_{thr}} \right) & \text{si } \rho(\mathbf{r}_a) \geq \rho_{thr} \\
 &= \sum_a^A w_a \xi & \text{si } \rho(\mathbf{r}_a) < \rho_{thr}
 \end{aligned} \tag{3.5.2}$$

Aquí, la suma se extiende sobre los  $A$  átomos de la macromolécula,  $\rho_{max}$  es la densidad máxima del mapa,  $\rho_{thr}$  es el umbral a partir del cual descartar todos los valores inferiores, y  $\mathbf{r}_a$  y  $w_a$  corresponden a la posición y masa del átomo  $a$ , respectivamente. La constante  $\xi$  regula la fuerza con la que los átomos se ven atraídos por el mapa. Este potencial permite que cada átomo perciba una fuerza atractiva hacia las regiones de mayor densidad situadas a su alrededor. La estructura estereoquímica se mantiene gracias al campo de fuerzas CHARMM27. Además del potencial asociado al mapa, se introduce un término extra para preservar la estructura secundaria. Este término impone restricciones armónicas en los ángulos diedros del esqueleto de las proteínas y los ácidos nucleicos. También se impone este tipo de restricciones a las distancias entre los pares de bases para mantener su planaridad.

La herramienta *YUPSCX* (Tan, *et al.*, 2008) (<http://rumour.biology.gatech.edu/>) tampoco utiliza la función de correlación para llevar a cabo los ajustes. En su lugar se emplea un potencial aproximado basado en la función “*SwissCheeseX*” que dirige los átomos del modelo hacia el interior del mapa:

$$\begin{aligned}
 E^{YUPSCX} &= \sum_{a,i} -E_0 w_i, & \text{si } r_{ai} \leq r_x \\
 &= \sum_{a,i} -E_0 w_i \frac{(r_z - r_{ai})^2 (r_z + 2r_{ai} - 3r_x)}{(r_z - r_x)^3}, & \text{si } r_x < r_{ai} < r_z \\
 &= 0, & \text{si } r_{ai} \geq r_z
 \end{aligned} \tag{3.5.3}$$

Según esta función, cada voxel ( $i$ ) lleva asociado un pozo de potencial que atrae a los átomos ( $a$ ) situados en las inmediaciones ( $r_{ai} < r_z$  y  $r_{ai} > r_x$ ) con una fuerza atractiva cuya intensidad



depende tanto de la densidad del voxel ( $w_i$ ) como de la distancia ( $r_{ai}$ ) que los separa. Nótese que  $r_x$  y  $r_z$  definen el rango de actuación de estas fuerzas. El término  $E_0$  define la profundidad máxima de los pozos de potencial. Para encontrar un modelo ajustado *YUPSCX* utiliza una estrategia de optimización de recocido simulado. Para mantener la integridad estereoquímica del modelo durante el ajuste, en lugar de utilizar un campo de fuerzas convencional de MD, este método emplea una red elástica similar a la desarrollada por Tirion. De esta forma es posible acelerar los cálculos, mantener una calidad estereoquímica aceptable y evitar la necesidad de añadir restricciones específicas para mantener la estructura secundaria. No obstante, para prevenir la aproximación excesiva entre los átomos se incluye un término armónico adicional que penaliza las conformaciones con colisiones, *i.e.* aquellas en las que los átomos se acercan a distancias inferiores que la suma de sus radios de van der Waals.

El método de ajuste Tama 1 (Orzechowski y Tama, 2008), al igual que *Flex-EM*, utiliza un término potencial adicional que dirige los átomos hacia el interior del mapa basado en la función de correlación normalizada:

$$E^{Tama} = k \left( 1 - \frac{\sum_i \rho_i^{EM} \rho_i^P}{\sqrt{\sum_i (\rho_i^{EM})^2 \sum_i (\rho_i^P)^2}} \right) \quad (3.5.4)$$

Para mantener la calidad de las estructuras se utiliza el campo de fuerzas AMBER con el conjunto de parámetros *parm99*. La constante  $k$  permite regular la fuerza con la que son atraídos los átomos por el mapa. El modelo atómico empleado tiene en cuenta todos los átomos, y a diferencia de *Flex-EM*, considera que toda la estructura es flexible. El método Tama 2 (Grubisic, *et al.*, 2010) supone una modificación de Tama 1 para reducir el coste computacional a través de la utilización de un modelo de Go basado únicamente en átomos  $C_\alpha$  y un potencial de MD aproximado que tiene en cuenta tanto los ángulos diedros como las distancias y ángulos de enlace. Nótese que este último método, a diferencia de los demás, requiere un programa de modelado externo para obtener estructuras con todos los átomos.

El método *MDFIT* (Ratje, *et al.*, 2010) (<http://smog.ucsd.edu/extension/MDfit.html>) incorpora la información del mapa de forma similar al método Tama 1, *i.e.* añadiendo un término a la energía potencial basado en la función de correlación normalizada:

$$E^{MDFIT} = V^{SB} - W \frac{\sum_i \rho_i^{EM} \rho_i^P}{\sqrt{\sum_i (\rho_i^{EM})^2 \sum_i (\rho_i^P)^2}} \quad (3.5.5)$$

En este caso, para mantener la integridad de la macromolécula y minimizar las distorsiones se tienen en cuenta todos los átomos pesados y se utiliza un potencial basado en la estructura ( $V^{SB}$ ) que define a la estructura inicial como la conformación de equilibrio, *i.e.* de potencial cero. La constante  $W$  se ajusta para que la correlación tenga un valor comparable a  $V^{SB}$ . Este potencial incluye términos específicos para distancias y ángulos de enlace, ángulos diedros y átomos en contacto (a menos de 4Å). Además, también se añade un término repulsivo entre los átomos que no están en contacto. Los ajustes se efectúan utilizando dinámica de Langevin en una versión modificada del programa de MD Gromacs.

Por norma general, los métodos de ajuste flexible basados en MD resultan muy eficaces para obtener modelos finales compatibles con los mapas de densidad. Además, gracias al empleo de campos de fuerzas más o menos detallados, suelen mantener correctamente la estereoquímica de los modelos. Sin embargo, su principal limitación es el elevado coste computacional que normalmente dificulta su utilización en ordenadores corrientes.

### 3.5.5. Métodos de ajuste flexible basados en otras aproximaciones

También existen otro tipo de aproximaciones para resolver el problema del ajuste flexible que no emplean ni modos de baja frecuencia ni simulaciones de MD estándar. Entre ellos, destacan la herramienta *S-flexfit* (Velazquez-Muriel y Carazo, 2007) que utiliza PCA para obtener modelos compatibles con los datos; el método de Jolley *et al.* (Jolley, *et al.*, 2008), basado en simulaciones con restricciones geométricas y un esquema de optimización de MC; el programa *DireX* (Schröder, *et al.*, 2007), basado un modelo de red elástica deformable que también emplea simulaciones con restricciones geométricas; y el método desarrollado recientemente por Zheng (Zheng, 2011) que simula una trayectoria de ajuste minimizando una función de energía elástica modificada. Por conveniencia, en esta sección también se han incluido dos métodos cuyas estrategias de ajuste, a pesar de estar basadas en simulaciones de MD, difieren significativamente de las de los métodos comentados en la sección anterior: *DDFF* (Kovacs, *et al.*, 2008) que utiliza una red elástica de muelles con amortiguadores y *SITUS* (Wriggers, 2010) que emplea representaciones reducidas tanto de la estructura atómica como del mapa objetivo para dirigir los ajustes. Los principios básicos de funcionamiento de estos métodos se describen brevemente a continuación.

La herramienta *S-flexfit* (Velazquez-Muriel, *et al.*, 2006) (<http://biocomp.cnb.csic.es/Sflexfit>) utiliza un reducido número de componentes principales (generalmente los 3 más colectivos) para generar una serie de modelos deformados. Los componentes principales se obtienen del PCA de un conjunto de estructuras conocidas de la misma proteína o de proteínas (o dominios) que comparten el mismo tipo de plegamiento (véase la sección 3.2.2). Una vez

generados dichos modelos, se alinean con el mapa objetivo y se elige el que mejor se ajusta a los datos utilizando una función de correlación local entre dicho mapa y los mapas simulados correspondientes. Para mantener la estructura covalente de los modelos deformados se mantienen rígidos los elementos de estructura secundaria, se reconectan las regiones separadas y se repara su estereoquímica. De esta forma se han obtenido modelos ajustados comparables a los obtenidos con *NMFF* en el ajuste flexible de la estructura atómica del EFG en un mapa experimental (Velazquez-Muriel, *et al.*, 2006). El principal inconveniente de este método consiste en que el conjunto de estructuras ha de ser lo suficientemente numeroso como para caracterizar adecuadamente la variabilidad estructural de la proteína. Más aún, las estructuras no pueden contener huecos en su secuencia. Para minimizar estos inconvenientes e incrementar su aplicabilidad *S-flexfit* ha sido mejorado posteriormente gracias a la implementación de una extensión del PCA convencional que tolera la presencia de estos huecos (Velazquez-Muriel y Carazo, 2007). El método mejorado ha permitido llevar a cabo satisfactoriamente los ajustes flexibles de GroEL y de las proteínas virales VP1 y VP2 del poliovirus 135S en sendos mapas experimentales a 6 y 10 Å de resolución, respectivamente (Velazquez-Muriel y Carazo, 2007).

El método de *Jolley et al.* (Jolley, *et al.*, 2008) considera a las macromoléculas como un conjunto de unidades rígidas articuladas por enlaces giratorios. Estas unidades se definen a partir de la estructura inicial antes de la simulación utilizando una técnica basada en la teoría de grafos. En cada iteración se genera una nueva conformación sin colisiones estereoquímicamente correcta. La nueva conformación se acepta únicamente cuando se cumple el criterio de Metrópolis empleando la función de correlación como energía. Para mantener las restricciones geométricas impuestas por la rigidez de las unidades se utilizan las técnicas de simulación con restricciones geométricas implementadas en *FIRST/FRODA* (Wells, *et al.*, 2005). De este modo, ha sido posible obtener ajustes flexibles satisfactorios con mapas simulados de lactoferrina y adenilato quinasa, y con dos mapas experimentales de GroEL (Jolley, *et al.*, 2008).

El programa *DireX* (Schröder, *et al.*, 2007) (<http://www.schroderlab.org/software/direx/index.html>) combina el muestreo conformacional con restricciones geométricas de *CONCOORD* (de Groot, *et al.*, 1997) con una red elástica deformable para obtener ajustes factibles. Durante las simulaciones, un término de perturbación geométrico dirige iterativamente el muestreo conformacional hacia el interior del mapa objetivo, mientras tanto, la red elástica mantiene la topología de la estructura inicial evitando el sobreajuste. Este método resulta robusto con ruido simulado y ha sido aplicado recientemente en el estudio de la exonucleasa-1 humana (Orans, *et al.*, 2011), el polimorfismo estructural de la actina-F (Galkin, *et al.*, 2010) y el mecanismo de acción de una chaperonina molecular del grupo II (Zhang, *et al.*, 2010).

El método de Zheng (Zheng, 2011) simula una trayectoria de ajuste calculando iterativamente cada conformación según el método de minimización de Newton-Raphson. De esta forma se utilizan implícitamente todos los modos normales para obtener la conformación del mapa objetivo. Este método emplea un modelo atómico reducido de dos pseudo-átomos por residuo, uno de ellos situado en la posición del átomo  $C_\alpha$  y el otro en el centro de masas de los átomos de la cadena lateral. La función de energía a minimizar consta de un potencial armónico modificado con un término no enlazante similar al potencial de Lennard-Jones para mantener la continuidad de la cadena, la estructura secundaria y modelar las interacciones entre los pseudo-átomos no enlazados directamente. Asimismo, la información necesaria para ajustar el modelo inicial en el mapa de densidad se introduce en forma de un término energético adicional basado en la correlación entre el mapa objetivo y los mapas simulados de los modelos de la trayectoria. De esta manera es posible obtener buenos ajustes con datos simulados incluso en presencia de niveles de ruido similares a los observados en los mapas experimentales. También se obtuvieron resultados satisfactorios utilizando mapas experimentales de miosina, EFG y ARN polimerasa (Zheng, 2011).

El método *DDFF* (Kovacs, *et al.*, 2008) está basado en simulaciones de MD que emplean una red elástica de muelles con amortiguadores para mantener la separación entre los átomos cercanos. El ajuste flexible tiene lugar gracias a una fuerza atractiva que dirige los átomos hacia las regiones desocupadas del mapa objetivo de forma proporcional a su densidad. Para simplificar las ecuaciones del movimiento y evitar oscilaciones en las trayectorias, las masas atómicas se consideran nulas. De este modo, el sistema se vuelve totalmente amortiguado y la integración de las ecuaciones es más eficiente. Esta aproximación emplea como modelo atómico una representación reducida de 5 pseudo-átomos por residuo en la cual los átomos situados más allá del  $C_\beta$  se condensan en una única masa puntual. Como coordenadas se utilizan los ángulos diedros  $\varphi$ ,  $\psi$  y  $\chi$  de los aminoácidos más las seis coordenadas que definen la orientación y posición de cada cadena. Para compensar la utilización de una representación reducida y evitar que el ajuste quede bloqueado por las cadenas laterales, la conformación de estas últimas se predice periódicamente durante toda la trayectoria. Este método es capaz de conseguir modelos comparables a los obtenidos anteriormente con el método de Hinsen (Hinsen, *et al.*, 2005) en el ajuste flexible de la estructura atómica de la ATPasa de calcio en un mapa experimental a 8Å de resolución y con la herramienta *NMFF* (Tama, *et al.*, 2004b) en el ajuste flexible del EFG en un mapa experimental a 12Å de resolución (Kovacs, *et al.*, 2008).

El conjunto de herramientas *SITUS* (Wriggers, 2010) (<http://situs.biomachina.org>) incluye distintos métodos basados en MD o en técnicas de interpolación para llevar a cabo el ajuste

flexible de estructuras atómicas en mapas de densidad. Todos ellos utilizan representaciones reducidas tanto de la estructura inicial como del mapa objetivo para dirigir los ajustes. Estas representaciones se obtienen empleando la técnica conocida como *Vector Quantization* (VQ) (Wriggers, *et al.*, 1998). La VQ permite modelar dos conjuntos de datos con el mismo nivel de detalle identificando y representando sus características principales mediante unos pocos puntos denominados *Codebook Vectors* (CV). Por norma general se establecen manualmente restricciones de distancia entre los CV para que los movimientos sean compatibles con la conectividad de la cadena polipeptídica y reducir el sobreajuste (Wriggers, *et al.*, 2004). Generalmente, los ajustes se llevan a cabo añadiendo al potencial de MD un término armónico adicional que penaliza las diferencias entre los CV de la estructura atómica inicial y los del mapa objetivo (Wriggers y Birmanns, 2001; Wriggers y Chacon, 2001). De este modo es posible mantener una buena calidad estereoquímica de los modelos durante toda la trayectoria de ajuste. Asimismo, en lugar de utilizar MD, *SITUS* también puede emplear distintas técnicas de interpolación (Rusu, *et al.*, 2008). Esto simplifica enormemente el proceso, evita tener que configurar la MD y permite efectuar los ajustes flexibles rápidamente, incluso a usuarios no experimentados. No obstante, debe tenerse en cuenta que cuando se emplean estas técnicas de interpolación la calidad estereoquímica de los modelos empeora en comparación con la obtenida utilizando MD. Entre los múltiples ejemplos de la aplicación de *SITUS* al ajuste flexible destacan la ARN polimerasa (Darst, *et al.*, 2002), actina unida a CCT (Wriggers, *et al.*, 2004), GroEL (Rusu, *et al.*, 2008) y miosina (Alamo, *et al.*, 2008).

Según los antecedentes presentados, el NMA es un método válido que permite modelar eficientemente la flexibilidad de los complejos macromoleculares de gran tamaño. Las herramientas de ajuste flexible basadas en NMA en CC permiten integrar la información estructural a distintas resoluciones obtenida por cristalografía y MET a un coste computacional menor que la mayoría de los otros métodos, en especial aquellos basados en MD. En esta tesis doctoral, nos proponemos aplicar la metodología del NMA en CI para desarrollar una nueva herramienta bioinformática de ajuste flexible que permita mejorar las ya existentes. La utilización de CI en la representación de las macromoléculas permitirá minimizar las distorsiones e incrementar la velocidad respecto de los métodos basados en CC. De esta forma será posible extender su rango de aplicabilidad para poder tratar complejos macromoleculares de gran tamaño. Además, también se utilizará esta metodología para desarrollar herramientas de simulación de trayectorias de MC y de transiciones conformacionales entre dos estructuras atómicas de la misma macromolécula.

---

## 4. Materiales y métodos

### 4.1. Análisis vibracional

De acuerdo con la mecánica clásica, el análisis vibracional permite caracterizar el movimiento de un sistema de osciladores armónicos acoplados mediante un conjunto de vectores independientes denominados modos normales de vibración (Goldstein, *et al.*, 2002). En principio, estos modos representan todas las maneras posibles de deformar una estructura alrededor de su conformación de equilibrio.

#### 4.1.1. Análisis vibracional en coordenadas generalizadas

La teoría del análisis vibracional aplicada a estructuras moleculares proporciona una solución analítica completa de las ecuaciones del movimiento del sistema asumiendo que las energías cinética ( $K$ ) y potencial ( $V$ ) varían cuadráticamente con las coordenadas generalizadas ( $\mathbf{q}$ ) y sus velocidades ( $\dot{\mathbf{q}}$ ) (Wilson, *et al.*, 1955):

$$K = \frac{1}{2} \dot{\mathbf{q}} \mathbf{K} \dot{\mathbf{q}}^T \quad V = \frac{1}{2} \mathbf{q} \mathbf{H} \mathbf{q}^T \quad (4.1.1)$$

Donde la matriz de energía cinética  $\mathbf{K}$ , o tensor métrico, corresponde a las segundas derivadas parciales de la energía cinética y la matriz de energía potencial  $\mathbf{H}$ , o Hessiano, a las de la energía potencial. Como se ha visto en la sección 3.1.1, de forma general, los elementos de estas matrices vienen dados por:

$$K_{\alpha\beta} = \frac{\partial^2 K}{\partial \dot{q}_\alpha \partial \dot{q}_\beta} \quad H_{\alpha\beta} = \frac{\partial^2 V}{\partial q_\alpha \partial q_\beta} \quad (4.1.2)$$

Donde las coordenadas  $q$  son generalizadas, *i.e.* éstas pueden ser posiciones atómicas, rotaciones/traslaciones del sólido rígido, ángulos torsionales o cualquier tipo de variable elegida para describir los  $N$  grados de libertad del sistema.

Si se considera a la molécula como un conjunto osciladores armónicos acoplados, *i.e.* átomos conectados por muelles que obedecen la ley de Hooke (véanse las Figuras 6 y 8), la energía potencial puede definirse en CC como:

$$V = \sum_{i < j} \frac{1}{2} f_{ij} r_{ij}^2 \quad (4.1.3)$$

Donde  $f_{ij}$  es la constante de fuerza del muelle que une los átomos  $i$  y  $j$ , y  $r_{ij}$  representa el cambio en el desplazamiento relativo desde la posición de equilibrio.

A partir de las fórmulas 4.1.1 y 4.1.2, y de las ecuaciones del movimiento de Lagrange ( $L = T - V$ ), es posible obtener todas las formas posibles de deformar las estructuras macromoleculares alrededor de su conformación de equilibrio (véase la sección 3.1.1 de la Introducción). Para ello, debe resolverse el siguiente problema generalizado de valores propios (Levitt, *et al.*, 1985):

$$\mathbf{H}\mathbf{V} = \mathbf{A}\mathbf{K}\mathbf{V} \quad (4.1.4)$$

De esta forma, se obtienen los  $N$  modos normales ( $\mathbf{V} = \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N$ ) junto con la matriz diagonal ( $\mathbf{A}$ ) de valores propios ( $\lambda_k$ ). Estos últimos se relacionan con las frecuencias ( $\nu_k$ ) según la expresión:

$$\lambda_k = (2\pi\nu_k)^2 \quad (4.1.5)$$

#### 4.1.2. Análisis vibracional en coordenadas cartesianas

En CC, el cálculo de las matrices  $\mathbf{K}$  y  $\mathbf{H}$  (ec. 4.1.2) se simplifica enormemente. La matriz de energía cinética se reduce a la matriz diagonal de masas atómicas ( $\mathbf{M}$ ):

$$\begin{aligned} M_{xy} &= m_x \quad \text{si } x = y, \\ M_{xy} &= 0 \quad \text{si } x \neq y \end{aligned} \quad (4.1.6)$$

Los índices  $x$  y  $y$  hacen referencia a cada una de las  $3n$  CC de los  $n$  átomos del sistema, siendo  $m_x$  la masa del átomo correspondiente a la coordenada  $x$ . Por su parte, el Hessiano viene definido por las segundas derivadas parciales de la energía potencial:

$$\begin{aligned} H_{xy} &= \frac{f_{xy}}{|\mathbf{r}_{xy}|^2} r_x r_y, \quad \text{si } \mathbf{r}_{xy} \neq \mathbf{0} \\ &= \sum_{z, \mathbf{r}_{xz} \neq \mathbf{0}}^{3n} \frac{f_{xz}}{|\mathbf{r}_{xz}|^2} r_x r_z, \quad \text{si } \mathbf{r}_{xy} = \mathbf{0} \end{aligned} \quad (4.1.7)$$

En este caso,  $f_{xy}$  y  $\mathbf{r}_{xy}$  representan, respectivamente, la constante de fuerza y el vector desplazamiento entre las posiciones de equilibrio de los átomos con coordenadas  $x$  y  $y$ . Los escalares  $r_x$  y  $r_y$  son las coordenadas correspondientes al vector  $\mathbf{r}_{xy}$ . El índice  $z$  hace referencia a cada una de las  $3n$  CC del sistema.

Empleando CC pesadas por las masas es posible simplificar aún más el problema generalizado de valores propios (ec. 4.1.4). El cambio de coordenadas se lleva a cabo utilizando la siguiente transformación:



$$\hat{\mathbf{K}} = \mathbf{M}^{-1/2} \mathbf{K} \mathbf{M}^{-1/2} = \mathbf{I} \quad (4.1.8)$$

$$\hat{\mathbf{H}} = \mathbf{M}^{-1/2} \mathbf{H} \mathbf{M}^{-1/2} \quad (4.1.9)$$

Donde  $\hat{\phantom{x}}$  indica que las matrices correspondientes están pesadas por las masas. De este modo, la obtención de los modos normales de vibración en el nuevo sistema de coordenadas se reduce a resolver el problema estándar de valores propios:

$$\hat{\mathbf{H}} \hat{\mathbf{V}} = \hat{\Lambda} \hat{\mathbf{V}} \quad (4.1.10)$$

Para expresar de nuevo los modos en CC (sin pesar por las masas) simplemente se revierte la transformación de la siguiente manera:

$$\mathbf{V} = \mathbf{M}^{1/2} \hat{\mathbf{V}} \mathbf{M}^{1/2} \quad (4.1.11)$$

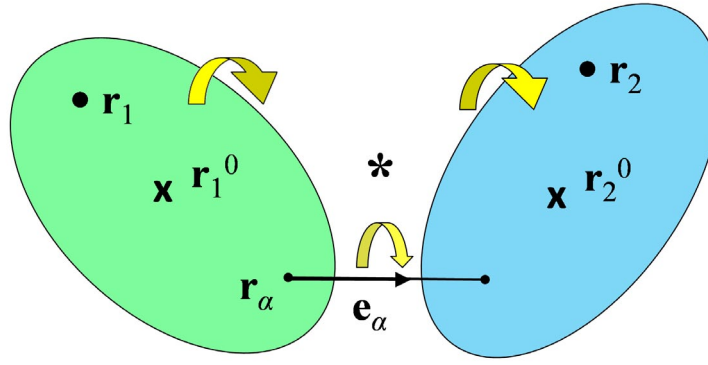
#### 4.1.3. Análisis vibracional en coordenadas internas

La mayoría de las aplicaciones del NMA emplean únicamente CC como variables por su simplicidad y fácil implementación. En su lugar, en esta tesis se ha optado por emplear los ángulos diedros como CI del sistema para: *i*) acelerar los cálculos, *ii*) preservar implícitamente la estructura covalente del sistema, y *iii*) facilitar la reconstrucción de modelos detallados a partir de representaciones más simples.

Las CI son aquellas que únicamente modifican las distancias relativas entre los componentes del sistema. Pueden definirse arbitrariamente siempre y cuando no contengan componentes externos, *i.e.* movimientos que no alteren ni la posición ni la orientación global de la macromolécula. Con el fin de separar los movimientos internos de los externos se emplean las condiciones de Eckart (Eckart, 1935). Estas establecen que cuando un ángulo diedro fluctúa alrededor de su posición de equilibrio, los cuerpos rígidos a ambos lados del enlace correspondiente cambian su posición y orientación relativa para conservar tanto el centro de masas como el momento angular del sistema (véase la Figura 13) (Noguti y Go, 1983a). En nuestro caso, las condiciones de Eckart se aplican implícitamente durante la definición de la matriz de Wilson (Noguti y Go, 1983a; Wilson, 1939):

$$\frac{\partial \mathbf{r}_1}{\partial q_\alpha} = \mathbf{e}_\alpha \times \left\{ \frac{M_2}{M} \mathbf{r}_\alpha + \frac{M_1}{M} \mathbf{r}_1^0 \right\} - \left\{ \left[ M_1 \mathbf{r}_1^0 \times (\mathbf{e}_\alpha \times \mathbf{r}_\alpha) + \mathbf{I}_2 \mathbf{e}_\alpha \right] \mathbf{I}^{-1} \right\} \times \mathbf{r}_1 \quad (4.1.12)$$

$$\frac{\partial \mathbf{r}_2}{\partial q_\alpha} = -\mathbf{e}_\alpha \times \left\{ \frac{M_1}{M} \mathbf{r}_\alpha + \frac{M_2}{M} \mathbf{r}_2^0 \right\} + \left\{ \left[ M_2 \mathbf{r}_2^0 \times (\mathbf{e}_\alpha \times \mathbf{r}_\alpha) + \mathbf{I}_1 \mathbf{e}_\alpha \right] \mathbf{I}^{-1} \right\} \times \mathbf{r}_2 \quad (4.1.13)$$



**Figura 13.** Modelo de dos cuerpos de Go. En ausencia de fuerzas externas, la rotación del cuerpo 1 (verde) respecto del cuerpo 2 (azul) alrededor del enlace  $\alpha$  es compensada por la rotación y traslación de ambos respecto del sistema de referencia de Eckart (representado por el símbolo \*). Los subíndices indican el cuerpo al que corresponden los centros de masa ( $\mathbf{r}^0$ ) y las posiciones de los átomos ( $\mathbf{r}$ ). La posición del enlace y su dirección se representan por  $\mathbf{r}_\alpha$  y  $\mathbf{e}_\alpha$ , respectivamente.

Estas ecuaciones determinan los cambios en las posiciones de los átomos ( $\mathbf{r}$ ) que compensan la aparición de movimientos externos cuando fluctúa el ángulo diedro  $q_\alpha$ . Los subíndices 1 y 2 hacen referencia a los cuerpos rígidos situados a uno y otro lado del enlace  $\alpha$ . El vector  $\mathbf{r}_\alpha$  especifica la posición de dicho enlace y el vector unitario  $\mathbf{e}_\alpha$  su dirección. Los vectores  $\mathbf{r}^0$ , las matrices  $\mathbf{I}$  y las magnitudes escalares  $M$  representan, respectivamente, las posiciones de los centros de masa, los tensores de inercia y las masas de los cuerpos correspondientes. En caso de no llevar subíndice, pertenecen a todo el sistema.

Para obtener la matriz de energía potencial en CI por el método directo, los muelles son proyectados en el nuevo sistema de coordenadas (Kovacs, *et al.*, 2005):

$$H_{\alpha\beta} = \sum_{i < j} \frac{f_{ij}}{|\mathbf{r}_{ij}|^2} \left\langle \mathbf{r}_{ij}, \frac{\partial \mathbf{r}_i - \partial \mathbf{r}_j}{\partial q_\alpha} \right\rangle \cdot \left\langle \mathbf{r}_{ij}, \frac{\partial \mathbf{r}_i - \partial \mathbf{r}_j}{\partial q_\beta} \right\rangle \quad (4.1.14)$$

Donde  $f_{ij}$  es la constante de fuerza del muelle situado entre los átomos  $i$  y  $j$ ,  $\mathbf{r}_{ij}$  representa el vector diferencia entre sus posiciones de equilibrio, y los corchetes angulares ( $\langle \rangle$ ) indican el producto escalar de dos vectores. Las derivadas corresponden a los elementos de la matriz de Wilson (ecuaciones 4.1.12 y 4.1.13).

El tensor métrico expresado en función de las CI también puede calcularse directamente (Levitt, *et al.*, 1985; Noguti y Go, 1983a):

$$K_{\alpha\beta} = \sum_l^n m_l \frac{\partial \mathbf{r}_l}{\partial q_\alpha} \cdot \frac{\partial \mathbf{r}_l}{\partial q_\beta} \quad (4.1.15)$$

Donde  $m_l$  representa la masa de cada uno de los  $n$  átomos del sistema y, al igual que en la ecuación 4.1.14, las derivadas, los elementos de la matriz de Wilson.

#### 4.1.4. Implementación del método rápido de Go – Cálculo del Hessiano

El cálculo del Hessiano a través de la ecuación 4.1.14 (método directo), resulta extremadamente lento. Para evaluar esta ecuación no solo es necesario obtener  $N(N+1)/2$  elementos diferentes, sino que además, cada uno de ellos requiere el cálculo de hasta  $n(n-1)/2$  interacciones distintas. Como el número de átomos ( $n$ ) es aproximadamente proporcional al número de CI ( $N$ ), la cantidad de operaciones necesarias para calcular el Hessiano resulta proporcional a  $\sim N^4$ . Esto dificulta enormemente la aplicación del NMA en CI cuando los sistemas son relativamente grandes. Para obtener la matriz de segundas derivadas del potencial de forma eficiente, Go plantea una aproximación analítica diferente (Noguti y Go, 1983b) cuyos puntos más importantes se detallan a continuación.

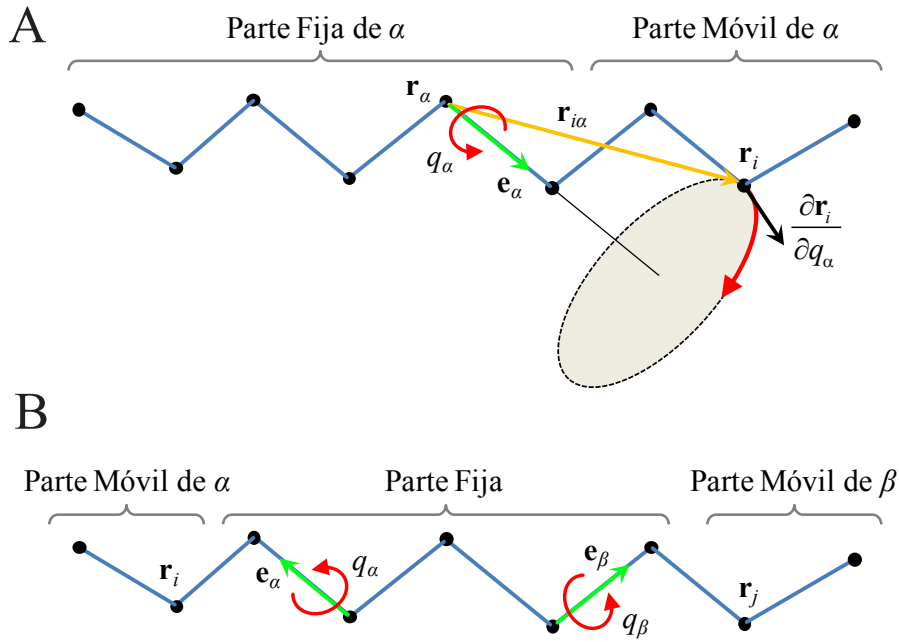
En primer lugar, como la energía potencial depende únicamente de las distancias interatómicas, no cambia con la orientación global de la macromolécula. Por tanto, no es necesario separar los movimientos internos de los externos a través del sistema de referencia de Eckart. En otras palabras, ya no hay que conservar ni el centro de masas y ni el momento angular total del sistema para calcular el Hessiano. Según esto, la relación entre las fluctuaciones en las posiciones atómicas y los cambios en las CI se reduce a la aproximación lineal para la rotación de un punto alrededor de un eje (véase la Figura 14A). Dicha relación viene dada por la siguiente expresión geométrica:

$$\frac{\partial \mathbf{r}_i}{\partial q_\alpha} = \mathbf{e}_\alpha \times \mathbf{r}_{i\alpha} \quad (4.1.16)$$

Aquí  $\mathbf{r}_i$  es la posición del átomo  $i$ , y  $\mathbf{r}_{i\alpha} = \mathbf{r}_i - \mathbf{r}_\alpha$ , donde  $\mathbf{r}_\alpha$  representa la posición de un punto cualquiera localizado en el enlace  $\alpha$ . El vector unitario  $\mathbf{e}_\alpha$  especifica la dirección del enlace correspondiente a la coordenada  $q_\alpha$ . Como se muestra en la Figura 14A, los átomos situados a un lado del enlace  $\alpha$  se mantienen fijos mientras que tan solo los del otro lado se ven afectados por el movimiento. Si se define cada una de estas regiones como parte fija y parte móvil, el sentido del vector  $\mathbf{e}_\alpha$  se define de forma que la parte móvil se sitúa hacia su parte positiva. Los cambios en el ángulo torsional  $q_\alpha$  se consideran positivos cuando la parte móvil del enlace  $\alpha$  gira en el sentido de las agujas del reloj respecto a la parte fija.

Eligiendo convenientemente las partes móviles hacia fuera del par de enlaces  $\alpha$  y  $\beta$ , *i.e.* de forma que no solapen (véase la Figura 14B), las ecuaciones 4.1.12 y 4.1.13 se reducen a la ecuación 4.1.16. De este modo, el cálculo del Hessiano se puede simplificar a:

$$H_{\alpha\beta} = \sum_i \sum_j \frac{f_{ij}}{|\mathbf{r}_{ij}|^2} \cdot \langle \mathbf{r}_{ij}, \mathbf{e}_\alpha \times \mathbf{r}_{i\alpha} \rangle \cdot \langle \mathbf{r}_{ij}, \mathbf{e}_\beta \times \mathbf{r}_{j\beta} \rangle \quad (4.1.17)$$



**Figura 14.** Modelo de movimiento para calcular el Hessiano por el método rápido de Go. En A se definen las partes fijas y móviles del enlace  $\alpha$  a través del vector unitario  $\mathbf{e}_\alpha$ , y el sentido de rotación de los átomos móviles  $\mathbf{r}_i$  alrededor del ángulo  $q_\alpha$ . En B se define la convención seguida durante el cálculo del Hessiano por el método rápido de Go. Nótese que ahora, las partes móviles de los ángulos giratorios  $q_\alpha$  y  $q_\beta$  se eligen por conveniencia hacia los extremos de la cadena. Los subíndices  $\alpha$  o  $\beta$  hacen referencia al enlace correspondiente.

En este caso, las sumas abarcan únicamente los átomos  $i$  y  $j$  situados en las partes móviles de los enlaces  $\alpha$  y  $\beta$ , respectivamente. Al igual que antes,  $f_{ij}$  es la constante de fuerza del muelle situado entre los átomos correspondientes,  $\mathbf{r}_{ij}$  representa el vector diferencia de sus posiciones de equilibrio, y los corchetes angulares ( $\langle \rangle$ ) indican el producto escalar de dos vectores. A pesar de que esta ecuación ya no requiere el cálculo de la matriz de Wilson, el número de operaciones necesario para resolverla sigue siendo proporcional a  $N^4$ . Por lo tanto, sigue siendo poco práctica para obtener el Hessiano de moléculas relativamente grandes.

El paso más importante, y que realmente permite acelerar el cálculo del Hessiano, consiste en la separación de las variables que dependen de las coordenadas (subíndices  $\alpha$  y  $\beta$ ) de las que dependen de los átomos (subíndices  $i$  y  $j$ ). Teniendo en cuenta que  $\mathbf{r}_{i\alpha} = \mathbf{r}_i - \mathbf{r}_\alpha$  y que  $\mathbf{r}_{\alpha\beta} = \mathbf{r}_\alpha - \mathbf{r}_\beta$ , es posible reorganizar la ecuación 4.1.17 para obtener:

$$H_{\alpha\beta} = (\mathbf{e}_\alpha, \mathbf{e}_\alpha \times \mathbf{r}_\alpha) \mathbf{R}_{\alpha\beta} \begin{pmatrix} \mathbf{e}_\beta \\ \mathbf{e}_\beta \times \mathbf{r}_\beta \end{pmatrix} \quad (4.1.18)$$

Donde,

$$\mathbf{R}_{\alpha\beta} = \sum_i \sum_j \mathbf{D}_{ij} \quad (4.1.19)$$

y

$$\mathbf{D}_{ij} = \frac{f_{ij}}{|\mathbf{r}_{ij}|^2} \begin{pmatrix} \mathbf{r}_i \times \mathbf{r}_j \\ \mathbf{r}_{ij} \end{pmatrix} (\mathbf{r}_i \times \mathbf{r}_j, \mathbf{r}_{ij}) \quad (4.1.20)$$

De este modo, las matrices  $\mathbf{R}_{\alpha\beta}$  contienen toda la información de interacción entre los átomos cuyas posiciones se ven afectadas por los cambios en las coordenadas  $q_\alpha$  y  $q_\beta$ . Como los elementos  $\mathbf{D}_{ij}$ , que dependen únicamente de las  $n$  posiciones atómicas, pueden ser precalculados con un coste computacional proporcional a  $n^2$ , y los elementos dependientes de las  $N$  CI ( $\mathbf{R}_{\alpha\beta}$  y  $H_{\alpha\beta}$ ) pueden obtenerse con un número de operaciones proporcional a  $N^2$ , el coste computacional total para calcular el Hessiano resulta proporcional a  $n^2 + N^2$ , *i.e.*  $O(\sim N^2)$ . Nótese que la parte de la derecha de la ecuación 4.1.20 corresponde a la expresión diádica de una matriz de 6x6 y que las sumas en la ecuación 4.1.19 abarcan únicamente los átomos  $i$  y  $j$  situados en las partes móviles de los enlaces  $\alpha$  y  $\beta$ , respectivamente.

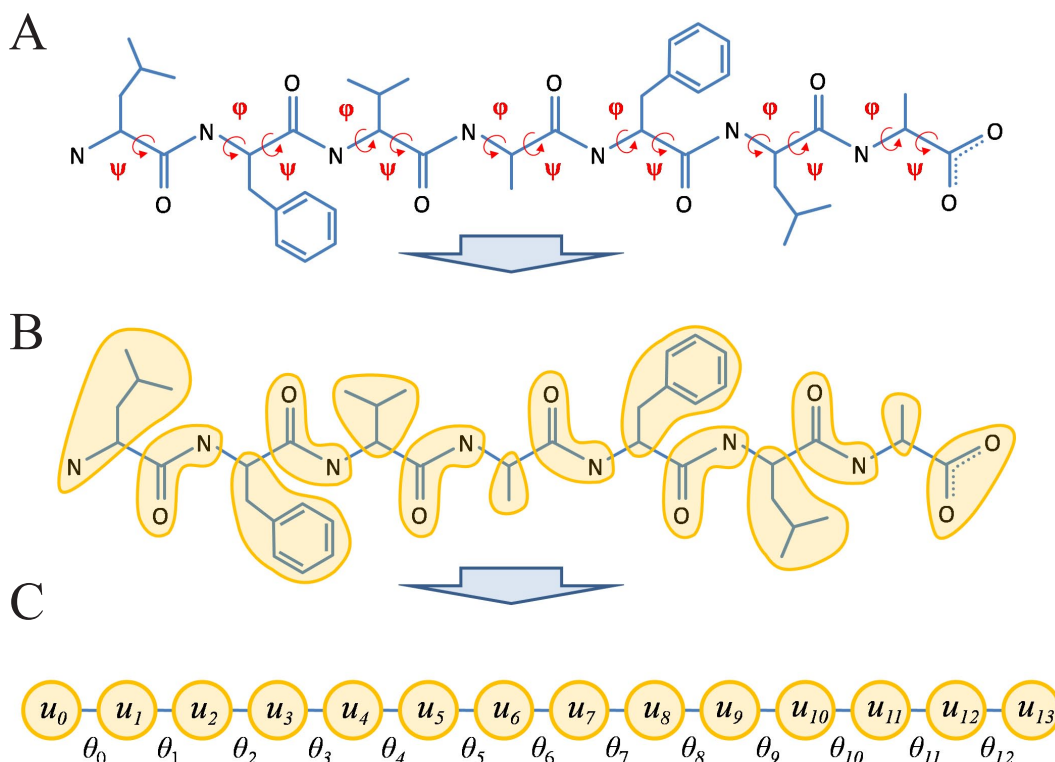
A pesar de haber reducido la complejidad computacional de  $O(\sim N^4)$  a  $O(\sim N^2)$ , la evaluación directa del Hessiano utilizando la ecuación 4.1.18 requiere mantener en memoria los  $n(n-1)/2$  elementos  $\mathbf{D}_{ij}$  con los que calcular cada una de las matrices  $\mathbf{R}_{\alpha\beta}$  correspondientes. Como cada  $\mathbf{D}_{ij}$  tiene 36 componentes y  $n$  suele ser mayor que  $N$ , su almacenamiento puede requerir miles de veces más memoria que las matrices del sistema ( $\mathbf{K}$  y  $\mathbf{H}$  tienen  $N(N-1)/2$  elementos). Por este motivo, la ecuación 4.1.19 todavía es poco práctica para sistemas relativamente grandes.

Para llevar a cabo eficientemente la suma de la ecuación 4.1.19, la macromolécula se representa como una cadena de  $M$  unidades rígidas articuladas por enlaces giratorios. En la Figura 15 se ilustra este modelo para un pequeño péptido. En el ejemplo, cada enlace giratorio corresponde a un ángulo torsional ( $\varphi$  o  $\psi$ ) y sirve de articulación entre las unidades ( $u$ ) (panel A). Estas últimas se consideran cuerpos rígidos de forma que las posiciones relativas de sus átomos no se alteran durante los cambios conformacionales (panel B). A cada unidad se le asigna un número de serie  $a$  o  $b$  que toma valores de 0 a  $M-1$  (panel C). Según este modelo, es posible condensar en una única matriz ( $\mathbf{T}_{ab}$ ) todas las interacciones entre los átomos de cada par de unidades  $ab$ :

$$\mathbf{T}_{ab} = \sum_{i \in a} \sum_{j \in b} \mathbf{D}_{ij} \quad (4.1.21)$$

Nótese que las sumas se llevan a cabo únicamente sobre los átomos  $i$  y  $j$  de las unidades  $a$  y  $b$ . Así, es posible redefinir la ecuación 4.1.19 a partir de los elementos  $\mathbf{T}_{ab}$ :

$$\mathbf{R}_{\alpha\beta} = \sum_a \alpha \sum_b \beta \mathbf{T}_{ab} \quad (4.1.22)$$



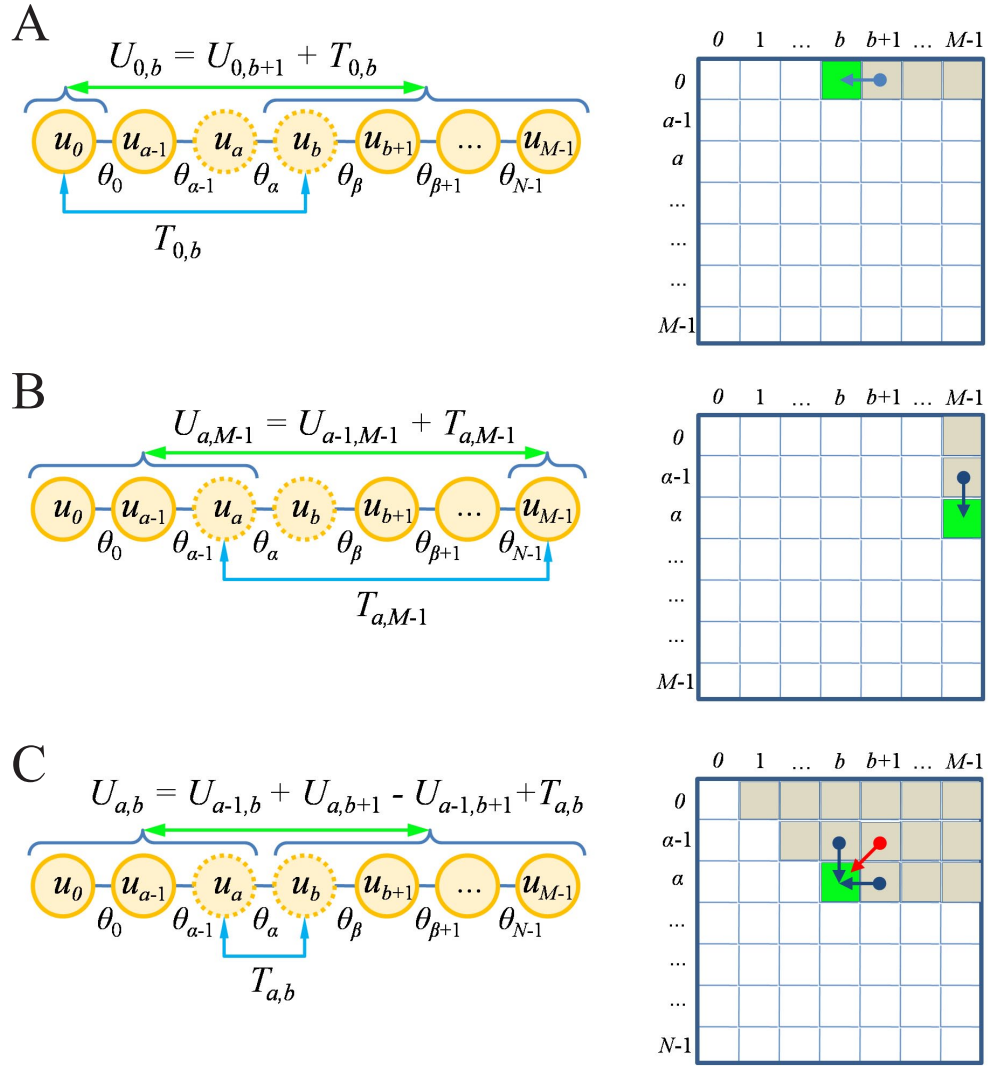
**Figura 15.** Modelo de unidades rígidas articuladas para obtener el Hessiano por el método rápido de Go. En el panel A se representa la estructura atómica de un péptido pequeño con los ángulos diedros canónicos  $\varphi$  y  $\psi$  ilustrados en color rojo. Estos ángulos definen las unidades rígidas sombreadas en amarillo en B. En C las unidades ( $u$ ) se representan esquemáticamente mediante círculos. Nótese que las unidades comienzan a numerarse desde el extremo N-terminal y toman la numeración correspondiente al ángulo giratorio al que preceden ( $\theta$ ).

En este caso, las sumas comprenden todas las unidades  $a$  y  $b$  situadas hacia las partes móviles de los enlaces  $\alpha$  y  $\beta$ , respectivamente. Aunque esta suma necesita unas cien veces menos memoria que la ecuación 4.1.19, los  $M(M-1)/2$  elementos  $\mathbf{T}_{ab}$  todavía ocupan más que las correspondientes matrices del sistema.

Para resolver este problema y calcular eficientemente las matrices  $\mathbf{R}_{\alpha\beta}$ , la suma de la ecuación 4.1.22 puede llevarse a cabo de forma recursiva empleando los elementos  $\mathbf{U}_{ab}$ . Cada elemento  $\mathbf{U}_{ab}$  contiene toda la información de interacción entre las unidades  $a \in [0, a]$  y  $b \in [b, M-1]$ . Para calcularlas se emplea la siguiente relación de recurrencia:

$$\mathbf{U}_{ab} = \mathbf{U}_{a,b+1} + \mathbf{U}_{a-1,b} - \mathbf{U}_{a-1,b+1} + \mathbf{T}_{a,b} \quad (4.1.23)$$

Como se muestra en la Figura 16, cada elemento  $\mathbf{U}_{ab}$  se obtiene sistemáticamente a partir de la ecuación anterior siguiendo un orden determinado. Primero se calcula la primera fila de la matriz  $\mathbf{U}$  comenzando desde la última columna:  $(a,b) = (0,M-1), (0,M-2), \dots, (0,1)$  (panel A), luego se calcula la siguiente fila:  $(a,b) = (1,M-1), (1,M-2), \dots, (1,2)$  (paneles B y C), y así



**Figura 16.** Cálculo de la matriz  $\mathbf{U}$  para obtener el Hessiano por el método rápido de Go. Por simplicidad, únicamente se ilustran los ángulos diedros ( $\theta$ ) como CI.

sucesivamente. Por definición, los elementos  $\mathbf{U}_{ab}$  son 0 si cualquiera de los índices  $a$  o  $b$  están fuera del intervalo  $[0, M-1]$ . De este modo, la obtención de cada  $\mathbf{U}_{ab}$  requiere únicamente la evaluación de un nuevo término de interacción ( $\mathbf{T}_{ab}$ ) y la suma (o resta) de algunos de los elementos calculados previamente. Por tanto, no es necesario almacenar los  $\mathbf{T}_{ab}$  pues solo se utilizan una vez. Además, como la matriz  $\mathbf{U}$  es simétrica, tan solo es necesario obtener los elementos  $\mathbf{U}_{ab}$  con índices  $a < b$ .

Para calcular las matrices  $\mathbf{R}_{\alpha\beta}$  a partir de los elementos  $\mathbf{U}_{ab}$  es necesario relacionar convenientemente los índices de las coordenadas ( $\alpha$  y  $\beta$ ) con los de las unidades ( $a$  y  $b$ ). En el caso más sencillo, las coordenadas del sistema ( $q$ ) corresponden únicamente a los ángulos diedros de la cadena principal ( $\varphi$  y  $\psi$ ). En este caso, según la definición de  $\mathbf{R}_{\alpha\beta}$  (ec. 4.1.22) la interacción entre las unidades de la parte móvil de  $\alpha$  con las unidades de la parte móvil de  $\beta$  viene dada directamente por el elemento  $\mathbf{U}_{ab}$  de subíndices  $a = \alpha$  y  $b = \beta + 1$ :



$$\mathbf{R}_{\alpha\beta} = \mathbf{U}_{\alpha,\beta+1} \quad (4.1.24)$$

Para obtener el Hessiano, los elementos  $\mathbf{R}_{\alpha\beta}$  se van calculando directamente a través de la relación de recursión 4.1.23 y la ecuación anterior. Una vez conocido el  $\mathbf{R}_{\alpha\beta}$ , el Hessiano se evalúa eficientemente empleando la ecuación 4.1.18. Nótese que para calcular cada elemento  $\mathbf{H}_{\alpha\beta}$  siguiendo este esquema, únicamente es necesario mantener en memoria los elementos de la fila anterior de la matriz  $\mathbf{U}$  ( $\mathbf{U}_{\alpha-1,\beta}$ ). De este modo, no solo se ha reducido drásticamente la cantidad de memoria auxiliar necesaria para calcular eficientemente el Hessiano, sino también el número de operaciones.

Cuando alguna de las coordenadas  $q_\alpha$  o  $q_\beta$  (o ambas) no pertenece a la cadena principal, como por ejemplo, los ángulos diedros  $\chi$  de las cadenas laterales, la expresión 4.1.24 que permite obtener las matrices  $\mathbf{R}_{\alpha\beta}$  a partir de los elementos de la matriz  $\mathbf{U}$ , no resulta tan sencilla. En estos casos se emplean las relaciones mostradas en la Tabla I de (Noguti y Go, 1983b).

#### 4.1.5. Implementación del método rápido de Go – Cálculo del tensor métrico

En sistemas grandes, el cálculo del tensor métrico, o matriz de energía cinética ( $\mathbf{K}$ ), con la ecuación 4.1.15 (método directo), también resulta un cuello de botella computacional. Para obtener cada uno de los  $N(N+1)/2$  elementos diferentes de esta matriz simétrica es necesario tener en cuenta la contribución de las masas de los  $n$  átomos del sistema. Por lo tanto, el número de operaciones es proporcional a  $nN^2$ . Considerando que  $N$  depende linealmente de  $n$ , el coste computacional total de la matriz de energía cinética resulta proporcional a  $\sim N^3$ . Con el fin de acelerar los cálculos, es posible substituir las ecuaciones 4.1.12 y 4.1.13 en la ecuación 4.1.15 (Noguti y Go, 1983a) para obtener:

$$K_{\alpha\beta} = \frac{M_1 M_3}{M} \left[ \mathbf{e}_\alpha \times (\mathbf{r}_\alpha - \mathbf{r}_1^0) \right] \cdot \left[ \mathbf{e}_\beta \times (\mathbf{r}_\beta - \mathbf{r}_3^0) \right] + \left[ M_1 \mathbf{r}_1^0 \times (\mathbf{e}_\alpha \times \mathbf{r}_\alpha) - \mathbf{I}_1 \mathbf{e}_\alpha \right] \mathbf{I}^{-1} \left[ M_3 \mathbf{r}_3^0 \times (\mathbf{e}_\beta \times \mathbf{r}_\beta) - \mathbf{I}_3 \mathbf{e}_\beta \right] \quad (4.1.25)$$

Los vectores  $\mathbf{r}^0$ , las matrices  $\mathbf{I}$  y los valores  $M$  corresponden, respectivamente, a las posiciones de los centros de masa, los tensores de inercia y las masas de los cuerpos correspondientes. Los subíndices 1 y 3 hacen referencia, respectivamente, a los cuerpos comprendidos entre el principio de la cadena y el enlace  $\alpha$ , y entre el enlace  $\beta$  y el final de la misma (véase la Figura 14B). En caso de no llevar subíndice pertenecen a todo el sistema. Los vectores  $\mathbf{r}$  y  $\mathbf{e}$  indican, respectivamente, la posición y dirección de los enlaces con subíndices  $\alpha$  y  $\beta$ . Precalculando todos los términos entre corchetes de esta ecuación es posible obtener el tensor métrico con un número total de operaciones proporcional a  $N^2$ .

#### 4.1.6. Mover en coordenadas internas

Para actualizar eficientemente las posiciones atómicas a partir de las CI se ha implementado el método que emplea matrices de rotación descrito en Choi (Choi, 2006). Este método consiste en la actualización secuencial de las posiciones comenzando desde uno de los extremos de la cadena. El extremo inicial se considera fijo para que cada cambio en las CI afecte únicamente a las posiciones de los átomos situados después de ellas, *i.e.* hacia el extremo final de la cadena. Como se ilustra en la Figura 15, al igual que en el cálculo del Hessiano con el método rápido, la macromolécula se representa por una cadena de unidades rígidas ( $u_\alpha$ ) articuladas mediante enlaces giratorios  $\alpha$ . Cada unidad  $u_\alpha$  está formada por todos los átomos comprendidos entre el enlace  $\alpha$  y el anterior ( $\alpha-1$ ). Los dos átomos que definen el enlace  $\alpha$  también pertenecen a la unidad  $u_\alpha$ . El procedimiento de actualización de las posiciones atómicas comienza desde un extremo de la cadena. Primero se calcula la matriz de rotación/traslación asociada a la CI actual ( $q_\alpha$ ), después se acumula ésta a la matriz de rotación/traslación total resultante de todas las contribuciones de las CI anteriores, y finalmente, se utiliza esta última matriz para actualizar las posiciones de los átomos de la siguiente unidad ( $u_{\alpha+1}$ ). De este modo se obtiene la transformación necesaria para colocar los átomos en sus posiciones finales directamente desde la conformación inicial. Este proceso resulta eficiente dado que cada CI y cada posición atómica es considerada solamente una vez. Según esto, teniendo en cuenta que el número de átomos por residuo es aproximadamente constante, el coste computacional escala linealmente con el número de CI. No obstante, este método de actualización de las posiciones atómicas provoca la aparición de cambios no deseados en la orientación global de la macromolécula. En la práctica, estos componentes externos del movimiento se eliminan después de actualizar las posiciones. Para ello, se alinea la nueva conformación con la estructura de partida (Noguti y Go, 1985) empleando el método analítico de superposición de estructuras basado en la minimización del RMSD (Kabsch, 1976; Kabsch, 1978).

Por otro lado, también es posible actualizar las posiciones atómicas ( $\mathbf{r}_i$ ) a partir de los cambios en las CI utilizando la matriz de Wilson (ecuaciones 4.1.12 y 4.1.13):

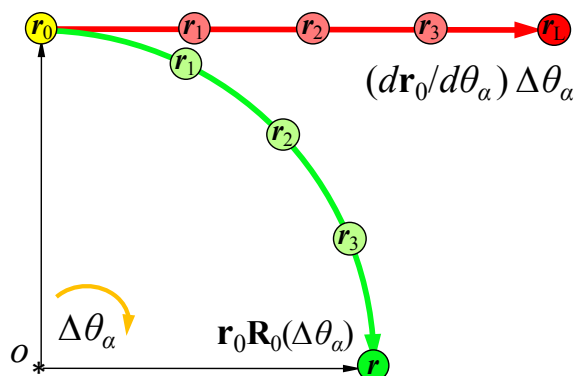
$$\Delta \mathbf{r}_i = \sum_{\alpha}^N \frac{\partial \mathbf{r}_i}{\partial q_{\alpha}} \Delta q_{\alpha} \quad (4.1.26)$$

Según esta ecuación, cada uno de los átomos  $i$  describe una trayectoria lineal proporcional a la magnitud del desplazamiento en la coordenada interna  $\Delta q_{\alpha}$ , exactamente de igual forma que al mover en CC. Sin embargo, las trayectorias descritas por los átomos al girar alrededor de los enlaces son curvas. La actualización de las posiciones atómicas sin tener en cuenta la curvatura de las trayectorias podría alterar la estructura covalente de la macromolécula especialmente

con amplitudes grandes, tal y como sucede en el caso cartesiano. En la Figura 17 se ilustran esquemáticamente las diferencias entre las trayectorias lineales (rojo) y circulares (verde). Como puede observarse, la divergencia entre ambas trayectorias aumenta enormemente cuando los desplazamientos no son pequeños. Sin embargo, como se muestra esquemáticamente en la Figura 18, es posible aproximar una trayectoria curva (verde) mediante la acumulación iterativa de un cierto número de pequeños desplazamientos lineales ( $\Delta \mathbf{r}_i$ ) (trayectoria azul). En cada iteración ( $t$ ) la matriz de Wilson se recalcula a partir de las coordenadas actuales ( $\mathbf{r}_i^{(t)}$ ) para obtener el  $\Delta \mathbf{r}_i^{(t)}$  utilizando la ecuación 4.1.26. En cada una, el  $\Delta q_a^{(t)}$  es constante e igual al  $\Delta q_a$  total dividido entre el número de iteraciones. Así, las posiciones atómicas de la siguiente iteración vienen dadas por:

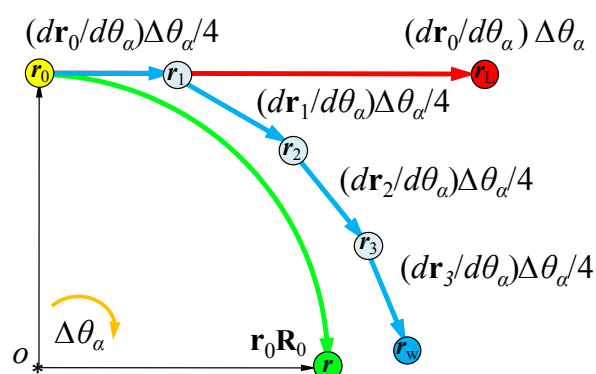
$$\mathbf{r}_i^{(t+1)} = \mathbf{r}_i^{(t)} + \Delta \mathbf{r}_i^{(t)} \quad (4.1.27)$$

De esta forma, es posible descomponer las trayectorias circulares asociadas a los modos en una suma de desplazamientos lineales más pequeños y conseguir resultados similares a los obtenidos con el método de matrices de rotación. Aunque en la práctica tan solo es necesario llevar a cabo 5~10 iteraciones para obtener buenos resultados, el proceso resulta relativamente lento porque el coste computacional de la ecuación 4.1.26 aumenta de forma proporcional a  $\sim N^2$ . Nótese que, en el sistema de referencia de Eckart, la posición de cada uno de los átomos se ve afectada por todas las CI. Por otra parte, el almacenamiento de la matriz de Wilson consume más memoria que las matrices del sistema, pues ocupa  $N \times 3n$  elementos y  $3n$  siempre es mayor que  $N$  en cualquiera de los modelos reduccionistas considerados (véase la siguiente sección). Para evitar este inconveniente, en nuestra implementación, la matriz de Wilson no se mantiene en memoria, sino que cada elemento se recalcula cada vez que es necesario. A pesar de que esto ralentiza ligeramente los cálculos ( $\sim 2x$ ), la disminución drástica del consumo de memoria permite aplicar el método a macromoléculas de gran tamaño.



**Figura 17.** Representación esquemática de la divergencia en la aproximación lineal de una trayectoria curva. La aproximación lineal se muestra en rojo y la solución analítica en verde. El vector  $\mathbf{r}$  representa la posición final resultante de rotar analíticamente un átomo cualquiera alrededor del eje ( $o$ ) un ángulo  $\Delta \theta_a$  desde su posición inicial ( $\mathbf{r}_0$ ) utilizando la matriz de rotación  $\mathbf{R}_0$ . La posición  $\mathbf{r}_L$  es la posición final obtenida aproximando la trayectoria linealmente por la tangente calculada en  $\mathbf{r}_0$ . Las posiciones  $\mathbf{r}_1$ ,  $\mathbf{r}_2$  y  $\mathbf{r}_3$  representan puntos intermedios en ambas trayectorias con los mismos ángulos.

Comparativamente, el método de matrices de rotación no solo es más rápido que el método iterativo basado en la matriz de Wilson, sino que además, el escalado del coste computacional del primero es lineal ( $\sim N$ ) mientras que en el segundo es cuadrático ( $\sim N^2$ ). Sin embargo, al mover con el método de matrices de rotación, los errores en las posiciones atómicas se van acumulando hacia el final de la cadena. A pesar de que estos resultan insignificantes en la mayoría de los casos, al efectuar movimientos de gran amplitud con estructuras grandes y simétricas pueden aparecer distorsiones al final de la cadena. Por el contrario, la actualización de las posiciones aplicando la matriz de Wilson iterativamente resulta muy útil porque los errores en las posiciones no se acumulan, sino que se distribuyen y cancelan homogéneamente sin importar el orden de los átomos en la cadena. Además, este método presenta la ventaja adicional de no introducir componentes externos de movimiento gracias a que las condiciones de Eckart están incluidas implícitamente en dicha matriz. En consecuencia, a diferencia de lo que sucede con el método de matrices de rotación, al mover con la matriz de Wilson no es necesario alinear la nueva conformación con la de la estructura de partida.



**Figura 18.** Representación esquemática de tres aproximaciones distintas para actualizar las posiciones atómicas. Las trayectorias correspondientes a los métodos analítico, lineal y de la matriz de Wilson, se representan en colores verde, rojo y azul, respectivamente. El vector  $\mathbf{r}$  representa la posición final obtenida al rotar analíticamente un átomo cualquiera alrededor del eje ( $o$ ) un ángulo  $\Delta\theta_\alpha$  desde su posición inicial ( $\mathbf{r}_0$ ) utilizando la matriz de rotación  $\mathbf{R}_0$ . La posición  $\mathbf{r}_L$  es la posición final obtenida al mover linealmente desde  $\mathbf{r}_0$ . Las posiciones  $\mathbf{r}_x$  representan cada uno los puntos dónde se actualizan las derivadas  $(d\mathbf{r}_x/d\theta_\alpha)$  para obtener la posición final ( $\mathbf{r}_w$ ) con el método de la matriz de Wilson. En este ejemplo,  $x = [0,3]$ .

## 4.2. Representaciones reduccionistas

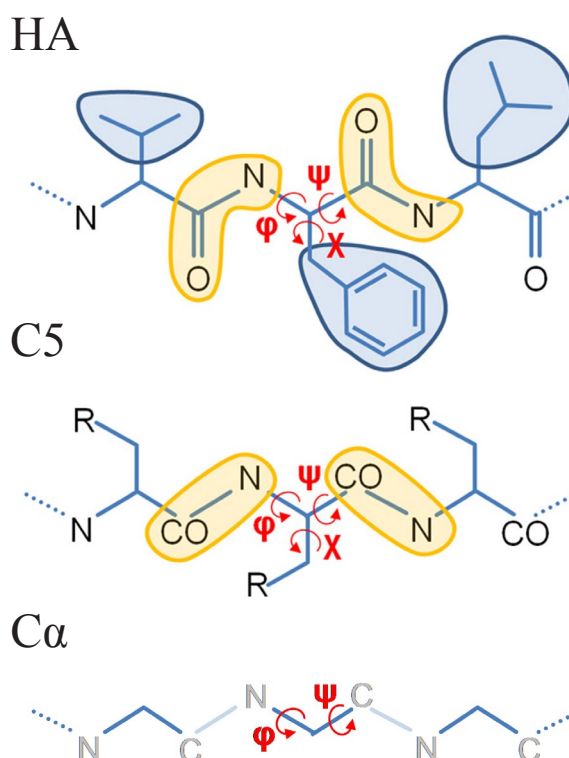
Para adaptar de manera sencilla la resolución del modelo a cada problema particular se han implementado distintas representaciones reduccionistas a tres niveles: CI, modelo atómico y red elástica.

### 4.2.1. Modelos de coordenadas internas

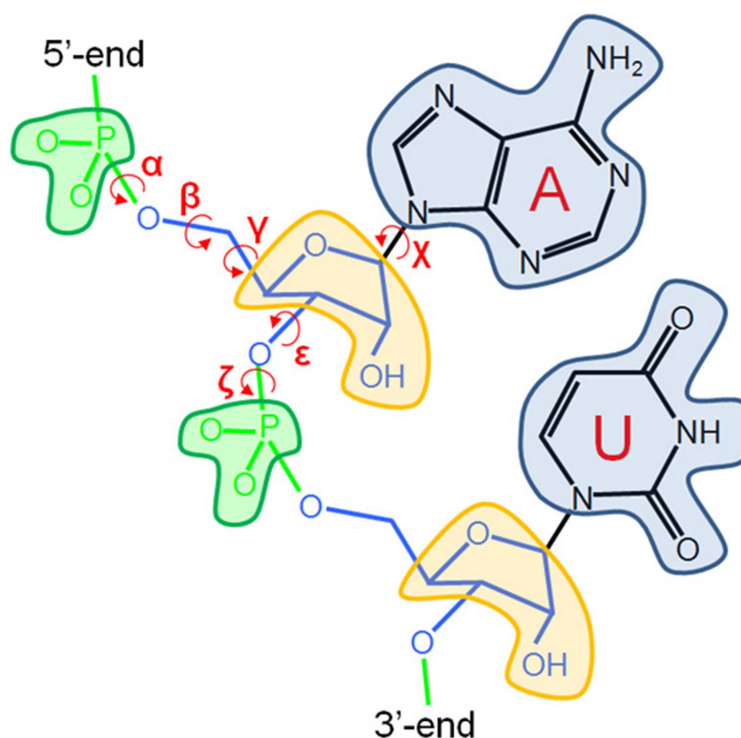
Por defecto, las variables de nuestro sistema vienen definidas por los ángulos diedros canónicos:  $\varphi$  y  $\psi$  en proteínas, y  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\varepsilon$  y  $\zeta$  en ácidos nucleicos. Opcionalmente, puede incluirse el ángulo diedro  $\chi$  de las cadenas laterales excepto en los aminoácidos prolina,

alanina y glicina. En las Figuras 19 y 20 se ilustran esquemáticamente las variables consideradas en proteínas y ácidos nucleicos, respectivamente. Para conservar la estructura covalente del sistema, los restantes diedros junto con las longitudes y ángulos de enlace se mantienen fijos. De esta forma, las cadenas laterales y los átomos del enlace peptídico de las proteínas, y los fosfatos, ribosas, desoxirribosas y bases de los ácidos nucleicos son siempre cuerpos rígidos (véanse las regiones sombreadas de las Figuras 19 y 20). Para tener en cuenta la existencia de múltiples cadenas y sus posibles cambios de orientación relativa se añaden seis grados de libertad que corresponden a la rotación/traslación entre ellas. Nótese que la orientación de la primera cadena no necesita ser especificada al estar implícitamente condicionada por la del resto del sistema a través de las condiciones de Eckart. Cuando las cadenas de proteínas o ácidos nucleicos presentan discontinuidades, *i.e.* cuando los átomos  $C_\alpha$  o P se encuentran separados a más de 6 y 7 Å, respectivamente, éstas se consideran independientes y, de forma general, se añaden seis grados de libertad entre ellas. También se añaden seis grados de libertad en el caso de moléculas no unidas covalentemente a las macromoléculas, tales como antibióticos, nucleótidos, azúcares, metales, moléculas de agua y otros ligandos.

Una gran ventaja de nuestra implementación es la posibilidad de definir arbitrariamente qué partes del sistema se consideran rígidas a través de la fijación de cualquier subconjunto de coordenadas. Así, al excluir simplemente ciertos ángulos diedros del análisis vibracional es posible mantener la rigidez de una región determinada. Como se verá en la sección de Resultados y discusión, esta aproximación resulta una manera efectiva tanto de reducir el coste computacional en sistemas muy grandes como de prevenir deformaciones en regiones que *a priori* se conoce que son rígidas.



**Figura 19.** Representación esquemática de los modelos atómicos y de CI para proteínas. En ellos se consideran: todos los átomos pesados (HA), cinco pseudo-átomos por aminoácido (C5), y un único pseudo-átomo  $C_\alpha$  por residuo ( $C_\alpha$ ). En color rojo se muestran los ángulos diedros y las regiones sombreadas indican los grupos de átomos considerados rígidos. Los átomos grises del modelo  $C_\alpha$  únicamente se emplean para definir los ángulos diedros.



**Figura 20.** Representación esquemática del modelo atómico y de CI para ácidos nucleicos. Se muestran dos nucleótidos de una cadena de ARN. En ADN, el modelo no incluye el oxígeno 2' de la ribosa. Los ángulos diedros correspondientes se ilustran con flechas rojas. Las regiones sombreadas indican los grupos de átomos considerados rígidos. Actualmente solo está disponible el modelo atómico con todos los átomos pesados (HA).

#### 4.2.2. Modelos atómicos

El modelo atómico especifica la distribución de masas del sistema y determina los puntos de interacción de la red elástica (véanse las Figuras 6 y 8). Los tres modelos atómicos reduccionistas de proteínas se muestran esquemáticamente en la Figura 19 y se ilustran para una proteína pequeña en la Figura 21. A continuación se enumeran dichos modelos:

- HA, (del inglés, *Heavy Atoms*) tiene en cuenta todos los átomos pesados (aquellos que no son hidrógeno) con sus respectivas masas atómicas.
- C5, cada aminoácido consta de cinco pseudo-átomos: tres para el esqueleto covalente (NH, HC<sub>α</sub> y CO) y dos para la cadena lateral (C<sub>β</sub> y un pseudo-átomo con la masa del resto de átomos ubicado en su centro de masas) (Kovacs, *et al.*, 2005). Los pseudo-átomos del esqueleto poseen la masa de los átomos que representan.
- Cα, cada residuo está representado por un único pseudo-átomo con la masa total del aminoácido que se sitúa en la posición C<sub>α</sub> (Mendez y Bastolla, 2010). En este caso, los átomos de nitrógeno y carbono carbonílicos del esqueleto son necesarios únicamente para definir los ángulos diedros.



Gracias al empleo de CI es posible reconstruir modelos HA a partir de las representaciones más simples (C $\alpha$  y C5) sin coste computacional adicional. Asimismo, la reducción del número de átomos considerado en el modelo puede mejorar significativamente la eficiencia de los métodos. Por el momento, solo está disponible el modelo HA para ácidos nucleicos. En la Figura 20 se representa esquemáticamente este modelo para ARN. En el caso de ADN, el modelo es idéntico, excepto en el átomo de oxígeno en posición 2' de la ribosa ya que no existe en el ADN.

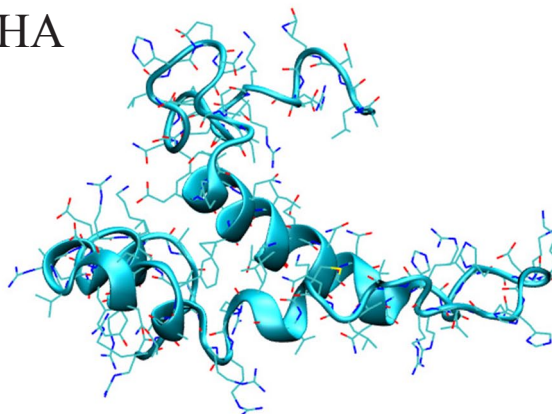
#### 4.2.3. Modelos de red elástica

En función del modelo atómico empleado, los átomos (o pseudo-átomos) se conectan mediante muelles armónicos que forman una red elástica. En la Figura 8, la red elástica de una proteína pequeña se representa mediante segmentos de color rojo. La red elástica define de forma general la energía potencial del sistema según:

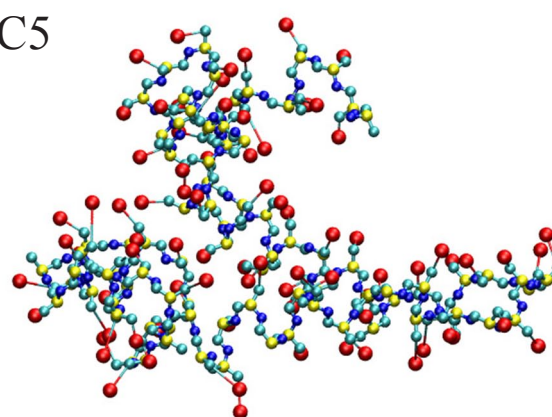
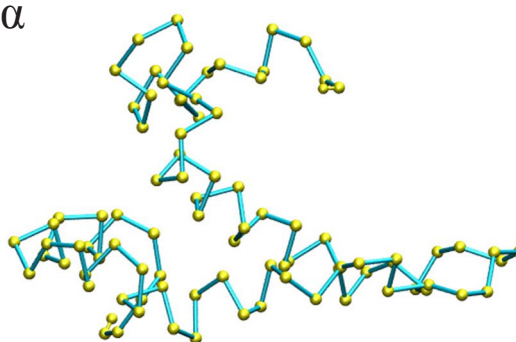
$$V = \sum_{i < j} f_{ij} (r_{ij} - r_{ij}^0)^2 + a \sum_{\alpha} (\theta_{\alpha} - \theta_{\alpha}^0)^2 \quad (4.2.1)$$

Donde el primer término corresponde al potencial entre pares de átomos no enlazados covalentemente,  $r_{ij}$  es su distancia relativa, los elementos  $f_{ij}$  corresponden a las constantes de fuerza de la matriz elástica, y el superíndice 0 indica la conformación inicial de equilibrio. El segundo término aporta rigidez torsional extra ( $a$ ) a cada ángulo diedro ( $\theta_{\alpha}$ ) para prevenir la presencia de modos localizados de baja frecuencia. Este fenómeno, conocido como

HA



C5

C $\alpha$ 

**Figura 21.** Modelos atómicos de proteínas. Se han representado los modelos HA, C5 y C $\alpha$  de la proteína adenilato quinasa. Los átomos C $\alpha$  se muestran en amarillo, los de nitrógeno en azul y los carbonos restantes en cian. Los pseudo-átomos que representan la cadena lateral en el modelo C5 aparecen en rojo. En el modelo HA el esqueleto peptídico se representa por un diagrama de cintas y las cadenas laterales por líneas con el esquema de colores por defecto de *VMD* (Humphrey, *et al.*, 1996). En los restantes modelos se utiliza una representación de bolas y barras.



“*tip effect*”, está causado por la elevada flexibilidad de pequeñas regiones poco conectadas de la estructura. Dichas regiones provocan la aparición de modos locales irracionales que pueden llegar a enmascarar los modos colectivos de la molécula (Lu, *et al.*, 2006).

La matriz elástica puede especificarse de diferentes maneras en nuestra implementación, por ejemplo, una constante de fuerza para todos los pares de átomos situados a distancias por debajo de un cierto umbral (Tirion, 1996), una función exponencial de la distancia (Hinsen, *et al.*, 1999) o un potencial refinado a partir de datos de MD (Orellana, *et al.*, 2010). Además, el usuario también puede introducir su propia función, lo que proporciona una mayor flexibilidad. Por defecto, se emplea la siguiente función sigmoidea cuya aplicabilidad ha sido comprobada tanto para ácidos nucleicos como para las distintas representaciones CG de proteínas (véase la Figura 8) (Lopez-Blanco, *et al.*, 2011):

$$F_{ij} = k / \left( 1 + \left( r_{ij}^0 / r_0 \right)^p \right), \quad \text{si } r_{ij}^0 < r_{cut}, \quad \text{de otro modo } F_{ij} = 0 \quad (4.2.2)$$

Donde  $k$  permite ajustar el valor máximo de la constante de fuerza,  $r_0$  representa el punto de inflexión (donde su valor es la mitad del máximo, *i.e.*  $k/2$ ),  $p$  define la pendiente de la curva, y  $r_{cut}$  define la distancia límite a partir de la cual las interacciones muy débiles no serán consideradas. Así, se eliminan interacciones cuya contribución es despreciable y se incrementa la eficiencia. Los parámetros  $k$ ,  $r_0$ ,  $p$  y  $r_{cut}$  fueron ajustados a 1, 3.8 Å, 6 y 10 Å, respectivamente, para obtener un comportamiento similar al de la función exponencial empleada en la comparación con datos de ED (Rueda, *et al.*, 2007a).

### 4.3. Aplicaciones del análisis vibracional en coordenadas internas

Los métodos de NMA en CI han sido utilizados en las distintas herramientas bioinformáticas que se describen brevemente a continuación.

#### 4.3.1. Análisis vibracional – iMode

El programa *iMode* permite obtener eficientemente los modos normales de vibración en CI a partir de estructuras atómicas de las macromoléculas. Éstas pueden estar compuestas de proteínas, ácidos nucleicos (ADN y ARN), así como también de azúcares, nucleótidos, lípidos, moléculas de agua y otros ligandos. En ordenadores de 32-bit y 4GB de memoria RAM es posible obtener los modos normales de estructuras de hasta ~15000 variables; en cambio, en 64-bit el límite viene dado por la cantidad de memoria disponible. Por ejemplo, en un servidor con 48GB de RAM se han obtenido algunos modos normales de baja frecuencia de la estructura del virus CCMV (*Icwp*) considerando ~75000 variables (Figura 22). Nótese que es posible fijar arbitrariamente cualquier conjunto de CI para adaptar el número de variables a

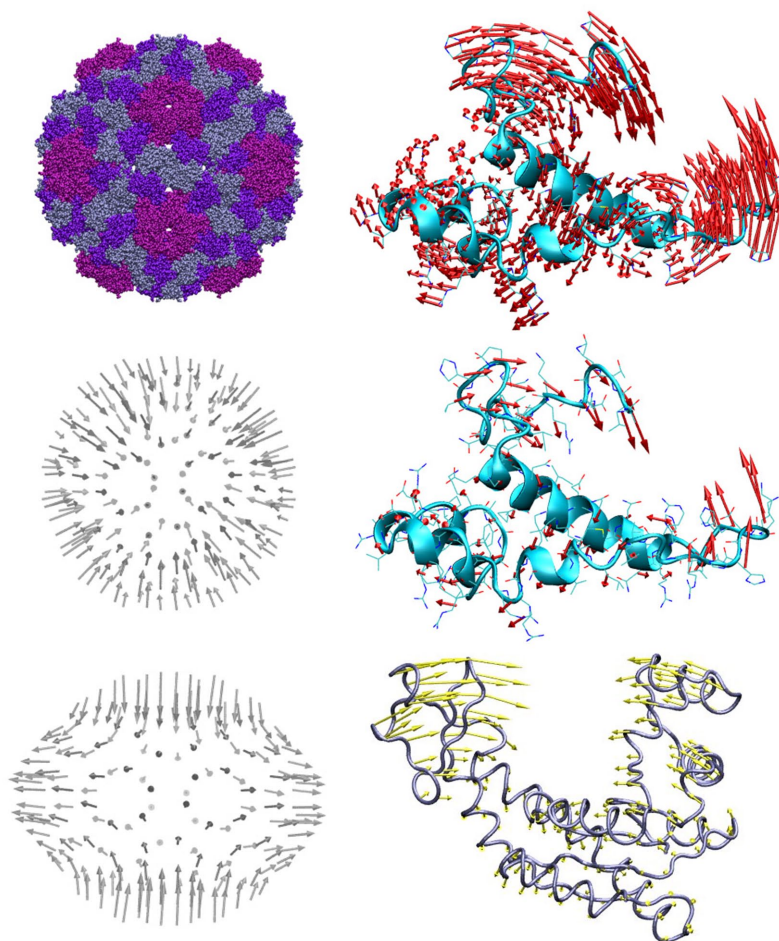
los recursos computacionales disponibles (véase la sección 4.2.1). Con fines comparativos se han implementado los distintos métodos para calcular las matrices descritos en las secciones 4.1.3, 4.1.4 y 4.1.5. Además, es posible utilizar los diferentes modelos atómicos y de potencial expuestos en las secciones 4.2.2 y 4.2.3, respectivamente, así como también cualquier subconjunto de las CI detalladas en la sección 4.2.1.

Para poder visualizar y comparar los modos obtenidos en CI ( $v^I$ ) es necesario expresarlos en CC ( $v^C$ ). Por ello, *iMode* incluye una opción para exportar los modos en CC utilizando la matriz de Wilson:

$$v_{ki}^C = \sum_{\alpha}^N (\partial \mathbf{r}_i / \partial q_{\alpha}) v_{\alpha k}^I \quad (4.3.1)$$

y otra para pesar estos modos por las masas a través de la siguiente ecuación:

$$\hat{\mathbf{V}} = \mathbf{M}^{1/2} \mathbf{V} \quad (4.3.2)$$



**Figura 22.** Representación de modos normales con *iModview*. En la columna de la izquierda se ilustra la estructura atómica del virus CCMV utilizando una representación de bolas (arriba) y sus modos de menor energía: no-degenerado (centro) y degenerado (abajo). En la columna derecha se muestra el primer modo normal de las estructuras: *lab3*, según los modelos HA (arriba) y Ca (centro); y *4ake*, según el modelo Ca (abajo). Estas últimas estructuras se han representado con cintas y tubos, respectivamente.

### 4.3.2. Visualización - iModview

Los desplazamientos infinitesimales codificados en los modos normales de vibración pueden representarse mediante un conjunto de vectores con magnitudes proporcionales al movimiento relativo de los átomos. Con el fin de ilustrar los movimientos codificados en los modos, la herramienta *iModview* permite generar representaciones de flechas a partir de los modos cartesianos que pueden visualizarse directamente en *VMD* (Humphrey, *et al.*, 1996). El tamaño, color y proporciones de las mismas puede ser especificado libremente. De manera opcional, es posible promediar los desplazamientos atómicos a nivel de residuos o cadenas para simplificar y mejorar las representaciones. En la Figura 22 se ilustra la flexibilidad de *iModview* al representar los modos de menor energía de dos estructuras de tamaños y formas muy diferentes. Por otra parte, la red elástica del sistema puede representarse por una malla de segmentos cuyo grosor es proporcional a la constante de fuerza entre cada par de átomos. Estas representaciones también pueden visualizarse directamente en *VMD* como se ha mostrado anteriormente en la Figura 8.

### 4.3.3. Animación de modos - iMove

Las estructuras macromoleculares pueden ser animadas para mostrar de manera intuitiva la información dinámica codificada en los modos vibracionales. Las soluciones de las ecuaciones de Lagrange presentadas en la Introducción (ecs. 3.1.8 y 3.1.9) corresponden a movimientos armónicos en el espacio de coordenadas normales. Así, el movimiento del modo  $k$  se traduce en la oscilación de las coordenadas del sistema alrededor de la conformación de equilibrio según:

$$\Delta \mathbf{q}_k = a_k \mathbf{v}_k \cos(2\pi \nu_k \cdot t) \quad (4.3.3)$$

Dónde la amplitud ( $a_k$ ) del movimiento es una variable que debe ser ajustada en cada caso particular, y  $\nu_k$  es la frecuencia (ec. 4.1.5). Para aplicar esta ecuación a intervalos de tiempo ( $t$ ) constantes se ha desarrollado el programa *iMove*, que es capaz de generar trayectorias representativas del movimiento codificado en los modos normales de vibración. La actualización de las posiciones atómicas puede llevarse a cabo, o bien linealmente, en el caso de modos cartesianos, o bien con el método de las matrices de rotación o con el basado en la matriz de Wilson si los modos son en CI (sección 4.1.6). Este último método está especialmente recomendado para evitar distorsiones cuando el tamaño de la macromolécula y la amplitud del movimiento son muy grandes. Nótese que generalmente es conveniente exagerar dicha amplitud para mejorar la visualización. Al tratarse de un movimiento oscilatorio, únicamente es necesario calcular las conformaciones correspondientes a medio periodo comenzando desde uno de sus extremos para visualizar todo el movimiento. Las trayectorias así producidas pueden

ser representadas en cualquier programa de visualización molecular. En la página web [http://chaconlab.org/imod/tutorial.html#Tuto\\_iMove](http://chaconlab.org/imod/tutorial.html#Tuto_iMove) se incluyen algunas animaciones generadas con *iMove* y *VMD*.

#### 4.3.4. Simulación de Monte Carlo - iMC

Se ha desarrollado una herramienta, *iMC*, que permite explorar el espacio conformacional de las macromoléculas mediante los modos normales de vibración y un muestreo de MC. La exploración se puede llevar a cabo tanto en CI como en CC. El procedimiento, esquematizado en la Figura 23, comienza calculando a partir de la estructura inicial los 5 o 10 modos de menor frecuencia del espectro. En cada paso de MC seleccionamos aleatoriamente un modo ( $k$ ) y una amplitud para generar un nuevo desplazamiento de prueba que será aceptado si se cumple el criterio de Metropolis (Metropolis, *et al.*, 1953), *i.e.* con probabilidad:

$$P_k = \min(1, e^{-\Delta V_k/k_B T}) \quad (4.3.4)$$

Donde,  $k_B$  es la constante de Boltzmann,  $T$  la temperatura absoluta y  $\Delta V_k$  representa la diferencia de energía potencial entre la conformación actual ( $t$ ) y la de prueba ( $p$ ). En coordenadas normales ( $Q$ ), esta diferencia viene dada por:

$$\Delta V_k = \frac{1}{2} \lambda_k \kappa (Q_{k,p}^2 - Q_{k,t}^2) \quad (4.3.5)$$

Esta expresión permite evaluar la energía en las inmediaciones de la estructura de partida de forma mucho más eficiente que la ecuación 4.2.1. Se ha incluido un factor adicional ( $\kappa$ ) para hacer que en promedio, los modelos generados se desvíen un cierto RMSD de la estructura inicial. Este factor se ajusta en cada caso particular antes de la generación de la trayectoria en función del RMSD deseado.

En las ecuaciones anteriores se puede observar que la probabilidad de aceptación del movimiento asociado a cada uno de los modos es inversamente proporcional tanto a su frecuencia (o eigenvalor,  $\lambda_k$ ) como a la amplitud del desplazamiento. En consecuencia, la mayoría de los movimientos asociados a los modos de mayor frecuencia son desechados. Esto influye negativamente en la eficiencia de la simulación. Con el fin de equilibrar las probabilidades de aceptación para aumentar la eficiencia del muestreo, las coordenadas normales son escaladas previamente siguiendo el método de escalado de la variable colectiva (SCV, del inglés *Scaled Collective Variable*) (Noguti y Go, 1985). En el SCV la amplitud de cada coordenada normal ( $Q_k$ ) se pondera por un factor de escala que corresponde a la amplitud térmica asociada al modo seleccionado ( $k$ ):

$$S_k = \sqrt{\frac{k_B T}{\lambda_k}} Q_k \quad (4.3.6)$$

Al utilizar en la ecuación 4.3.5 las coordenadas normales escaladas  $S_k$  en lugar de  $Q_k$  (sin escalar) es posible mantener una tasa de aceptación constante entre los distintos modos incrementando la eficiencia de la simulación. Los desplazamientos en coordenadas normales que son aceptados se van almacenando en un vector  $S$  que especifica la conformación absoluta en el espacio de coordenadas normales. Cada 1000 pasos de MC se aplica el vector  $S$  a la conformación inicial para generar una nueva estructura de la trayectoria. Si los modos están en CI se utiliza el método basado en la matriz de rotación (sección 4.1.6), y si son cartesianos, el movimiento de los átomos es lineal. Siguiendo este procedimiento iterativamente es posible generar eficientemente trayectorias que pueden ser representadas en cualquier programa de visualización molecular. En la página web [http://chaconlab.org/imod/tutorial.html#Tuto\\_iMC](http://chaconlab.org/imod/tutorial.html#Tuto_iMC) se muestra la animación de una trayectoria creada con *iMC* y *VMD*.

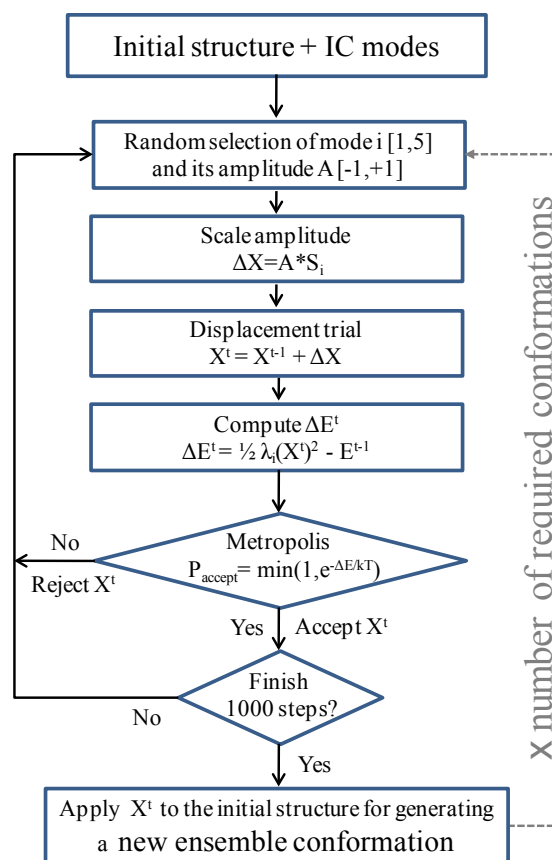


Figura 23. Diagrama de flujo de *iMC*.

#### 4.3.5. Simulación de transiciones conformacionales - iMorph

La herramienta *iMorph* permite simular transiciones conformacionales entre pares de estructuras atómicas empleando los modos normales de vibración en CI. Los eventos más importantes del procedimiento implementado se muestran esquemáticamente en el diagrama de flujo de la Figura 24. Antes de comenzar, ambas estructuras deben estar alineadas entre sí. Para conseguir una mejor representación del movimiento se recomienda superponer únicamente las regiones estáticas de las dos estructuras. Esto último se puede hacer, por ejemplo, con el método de alineamiento iterativo de Damm y Carlson (Damm y Carlson, 2006) que está basado en la minimización del RMSD y emplea una función gaussiana de la distancia para favorecer la superposición de las regiones más similares de ambas estructuras. Tras el alineamiento inicial, el proceso iterativo comienza calculando el 10% de los modos de menor energía a partir de la estructura inicial. A continuación se selecciona el 10% de ellos con una probabilidad

proporcional a su varianza ( $1/\lambda_k$ ). Así, los modos más flexibles y colectivos son elegidos con mayor frecuencia. Estos modos se escalan según el método SCV para evitar distorsiones estructurales y se ponderan aleatoriamente con el fin de incrementar el carácter estocástico del proceso. Después, se combinan en un único vector de deformación para generar una nueva conformación de prueba ( $\mathbf{X}_p$ ) utilizando el método de las matrices de rotación (sección 4.1.6). Para mantener aproximadamente la orientación inicial durante toda la simulación se superpone la conformación de prueba sobre la estructura objetivo minimizando su RMSD, pero ponderando previamente la contribución de cada uno de los átomos de las estructuras por una función gaussiana de su distancia relativa. De esta forma, los átomos situados

a más de 5~10Å no contribuyen significativamente al alineamiento y únicamente se tienen en cuenta de manera efectiva las regiones que inicialmente ya se encontraban superpuestas. Una vez alineado el modelo de prueba, se evalúa su RMSD con la estructura objetivo ( $\mathbf{X}_o$ ):

$$RMSD(\mathbf{X}_p, \mathbf{X}_o) = \sqrt{\frac{1}{n} \sum_i (\mathbf{r}_{i,p} - \mathbf{r}_{i,o})^2} \quad (4.3.7)$$

Donde cada conformación  $\mathbf{X}$  viene definida por el conjunto de todas las CC de los  $n$  átomos correspondientes ( $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n$ ). Únicamente serán aceptados conformeros que se aproximen a la estructura objetivo, en caso contrario, la conformación se desecha. Todo el proceso se repite iterativamente hasta alcanzar la convergencia. Se considera que la simulación ha convergido cuando el promedio de los RMSD de las 1000 últimas estructuras no cambia más de 0.01Å.

Los modos normales se recalculan cuando la conformación actual se desvía más de 0.1Å (RMSD) de la última estructura empleada en el NMA. Así, el sistema siempre se encuentra donde la aproximación del NMA es válida, *i.e.* en las proximidades de la conformación considerada de mínima energía. De este modo, la calidad de las estructuras apenas disminuye en comparación con las estructuras cristalográficas de partida según los criterios del programa de validación estructural *Molprobit* (Chen, *et al.*, 2010).

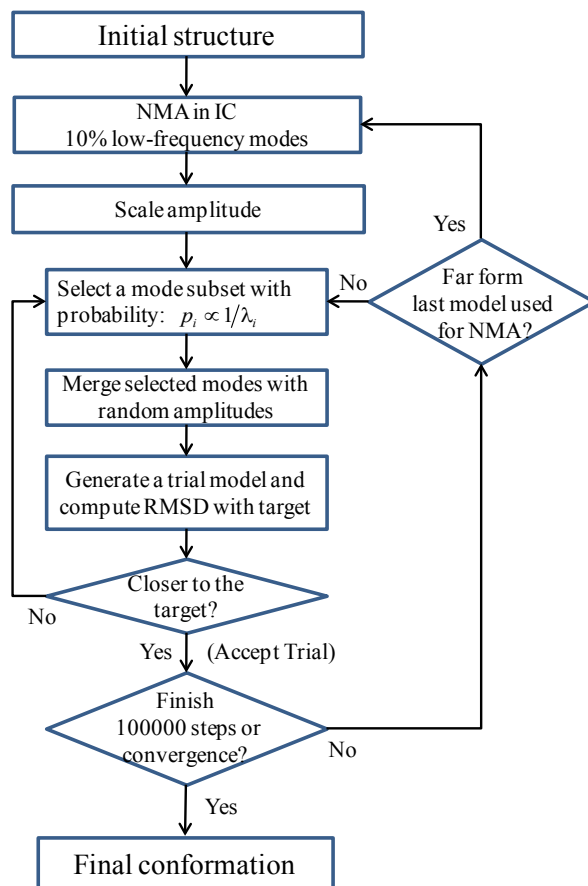


Figura 24. Diagrama de flujo de *iMorph*.



#### 4.3.6. Ajuste flexible de estructuras atómicas en mapas de densidad de microscopía electrónica - iModfit

Empleando únicamente la región de menor frecuencia del espectro de modos normales la aplicación *iModfit* es capaz de ajustar estructuras atómicas macromoleculares en el interior de mapas de densidad de media/baja resolución. Estos mapas corresponden a reconstrucciones 3D de macromoléculas obtenidas mediante la técnica experimental de microscopía electrónica y SP (véase la sección 3.4 de la Introducción).

El proceso de ajuste iterativo implementado se basa, al igual que en el caso de la simulación de trayectorias de transiciones conformacionales, en la perturbación aleatoria de una estructura de partida utilizando unos pocos modos normales de baja frecuencia para obtener un modelo atómico de prueba. Éste será aceptado únicamente si mejora la correlación normalizada entre el mapa objetivo y un mapa simulado a partir del modelo de prueba. Aplicando este procedimiento de forma iterativa es posible ajustar gradualmente la estructura atómica inicial en el interior del mapa de densidad objetivo. En la Figura 25 se muestran esquemáticamente los pasos más importantes del proceso. Antes de iniciar el refinamiento, la estructura atómica de partida debe estar colocada aproximadamente en el interior del mapa objetivo. Esto debe hacerse de forma automatizada mediante algún programa de ajuste rígido como *ADP\_EM* (Garzon, *et al.*, 2007a) o *SITUS* (Wriggers, 2010). Tras el alineamiento inicial se obtiene un conjunto de modos de baja frecuencia (el 5%) a partir de la estructura de partida. Una parte de ellos (el 10%) son seleccionados con probabilidad proporcional a la varianza ( $1/\lambda_k$ ) para generar un vector desplazamiento. La aplicación de este vector sobre el modelo inicial dará lugar al modelo de prueba. En *iModfit*, la función de coste es la correlación normalizada. Por ello, en cada iteración el modelo de prueba se convierte en un mapa de densidad ( $\rho^p$ ) que se filtra a una resolución semejante a la del mapa objetivo ( $\rho^f$ ). La función de correlación normalizada se define de la siguiente manera:

$$C(\rho^p, \rho^f) = \sum_m \frac{(\rho_m^p - \langle \rho^p \rangle)(\rho_m^f - \langle \rho^f \rangle)}{\sigma^p \sigma^f} \quad (4.3.8)$$

Dónde el subíndice  $m$  representa cada uno de los  $M$  voxels,  $\langle \rho \rangle$  la densidad promedio y  $\sigma$  la desviación estándar de los mapas correspondientes. La conversión del modelo atómico en un mapa de densidad de resolución comparable constituye una de las tareas más costosas computacionalmente. El filtrado puede realizarse principalmente por dos métodos: *i*) mediante la convolución del mapa con una función gaussiana, o *ii*) por filtrado en espacio de Fourier. Aunque los dos obtienen ajustes equivalentes, el primero es más rápido con mapas relativamente pequeños (o altas resoluciones), y el segundo con los más grandes (o bajas resoluciones). Su

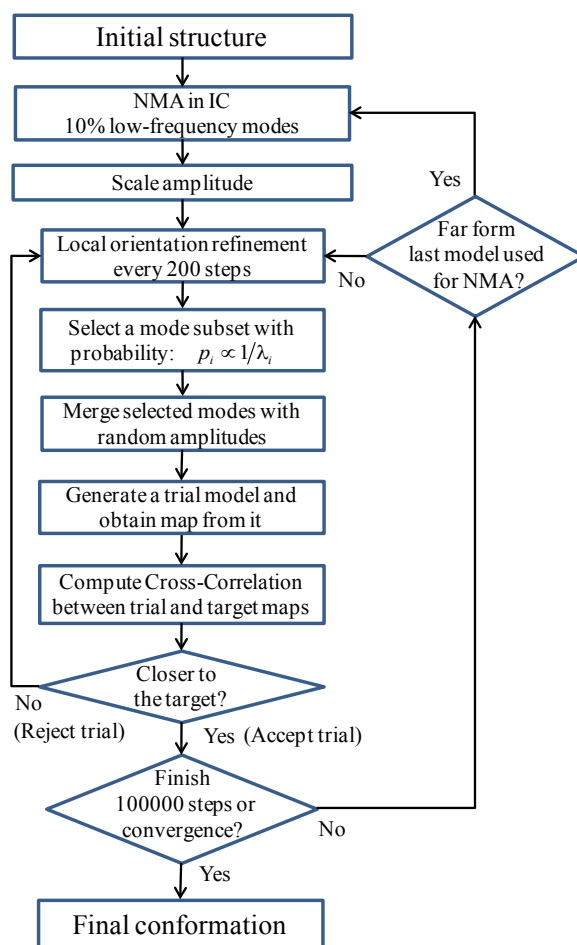


rapidez está condicionada fundamentalmente por las características del mapa (tamaño, muestreo y resolución), aunque también influyen otros factores relacionados con la arquitectura del computador. Para emplear el método óptimo en cada caso particular se hace una estimación previa filtrando la estructura atómica inicial 10 veces con cada uno y se elige el más rápido.

Durante el proceso de ajuste flexible el modelo atómico cambia continuamente de forma pudiendo afectar a su alineación con el mapa. Para minimizar la deriva del modelo y reducir las distorsiones que esto puede provocar es conveniente mantener una orientación adecuada durante el todo el proceso. Para ello, se realiza un alineamiento local entre el modelo atómico y el mapa. Como los modos normales en CI resultan, por definición, inefectivos para alterar la

orientación global de la macromolécula, no pueden emplearse para mantener una orientación adecuada. Por ello, se ha adaptado el esquema de Interpolación Parabólica Inversa (IPI) (Press, *et al.*, 1992) para refinar localmente la orientación del modelo a través de la maximización iterativa de la función de correlación sin recurrir a los modos normales. Una ronda de cuatro ciclos de IPI cada 200 pasos de ajuste flexible son suficientes para mantener una orientación adecuada durante todo el proceso. Nótese que aquí también es necesario convertir varias veces el modelo atómico en un mapa de resolución comparable a la del mapa objetivo.

Además de la IPI, también se lleva a cabo el alineamiento (basado en la minimización del RMSD) entre cada nuevo modelo de prueba y el último modelo aceptado. En lugar de utilizar toda la estructura, solamente se consideran los átomos que están en el interior del mapa objetivo. Esto se consigue ponderando la contribución que tienen en el RMSD los átomos del último modelo aceptado por los valores de densidad de los voxels que ocupan. Así, solo se tienen en cuenta significativamente las regiones colocadas dentro del mapa objetivo y se mantiene la orientación relativa entre cada ronda de IPI.



**Figura 25.** Diagrama de flujo de *iModfit*.

#### 4.4. Reconstrucción tridimensional de RepB y su complejo con ADN

##### 4.4.1. Preparación de las muestras

La proteína RepB de *Streptococcus pneumoniae* fue purificada por el grupo de G. del Solar, según el procedimiento descrito en Ruiz-Masó *et al.* (Ruiz-Masó, *et al.*, 2004). Se nos suministró una alícuota de la proteína purificada 59  $\mu$ M (8.6 mg/ml) en tampón H (Tris/HCl 20mM pH 8.0, EDTA 1mM, DTT 0.1mM y KCl 450mM). Todas las concentraciones molares se refieren al hexámero. Inmediatamente después de diluir la muestra 1000 veces hasta aproximadamente 60 nM (9  $\mu$ g/ml) en el mismo tampón, ésta se adsorbió a rejillas de Cu/Pd (Gilder, G400-CP3) revestidas con una fina película de carbono que previamente habían sido sometidas a descarga iónica para incrementar su carácter hidrofílico. Acto seguido, fueron teñidas negativamente utilizando una solución acuosa del 2% (en peso) de acetato de uranilo siguiendo procedimientos estándar de microscopía electrónica. La unión del ADN a RepB<sub>6</sub> ha sido efectuada incubando la proteína durante 20 minutos a 25°C en una solución de tampón B (Tris/HCl 20mM pH 8.0, EDTA 1mM, DTT 5mM y KCl 300mM) con 10 veces de exceso molar del fragmento de ADN de 42 pares de bases de la región *bind*. Al igual que en el caso anterior, una vez unida la proteína al ADN se prepararon las rejillas inmediatamente después de diluir la muestra.

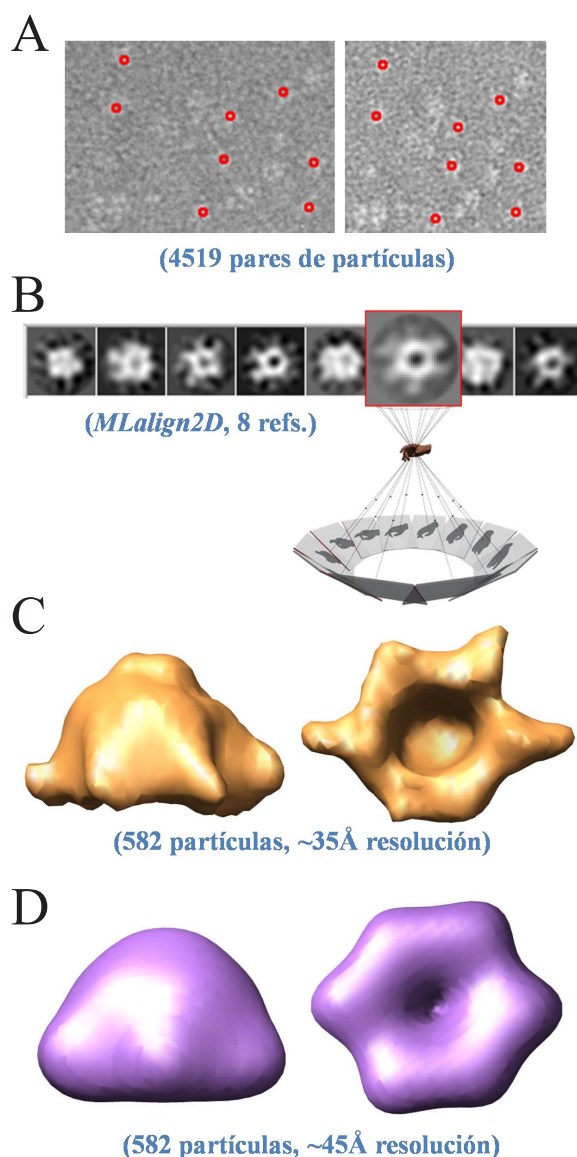
##### 4.4.2. Adquisición de datos

Las rejillas se han observado en el microscopio electrónico de transmisión JEOL 1230 del CIB operando a 100 kV. Las micrografías se han tomado con una magnificación de 50000x inclinando el soporte de la rejilla 0, 20 y 40° bajo condiciones de mínima dosis ( $\sim 10e^-/\text{\AA}^2$ ). En un ensayo independiente se han tomado micrografías de la misma región de la rejilla en dos orientaciones distintas 0 y 40° según el procedimiento del RCT (Figura 26A) (Radermacher, 1988). Todas las micrografías se han digitalizado en un escáner Minolta Dimage Scan Multi PRO a 10.6  $\mu$ m y luego se han promediado hasta una resolución de 4.0 $\text{\AA}$ /pixel.

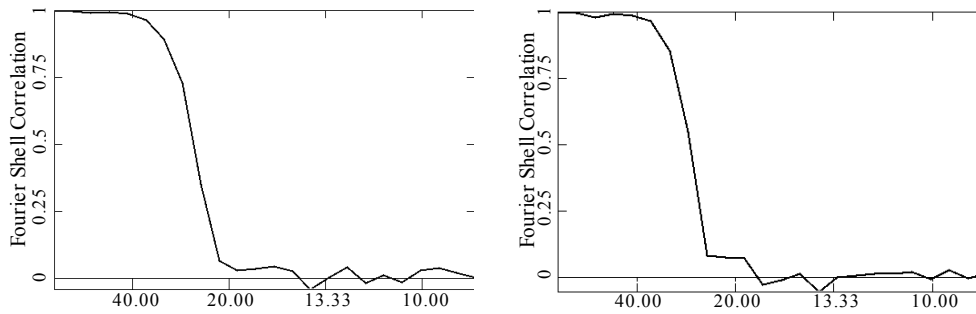
##### 4.4.3. Procesamiento de imagen y reconstrucción tridimensional

Para resolver la estructura 3D de la proteína RepB y la ambigüedad de la simetría observada en las estructuras cristalográficas se ha empleado la metodología de reconstrucción 3D por microscopía electrónica y SP (véase la sección 3.4 de la Introducción). Dado el pequeño tamaño de RepB<sub>6</sub> (145 kDa), se ha utilizado tinción negativa para poder observarla en el microscopio y obtener micrografías. El procesamiento de imagen para obtener las reconstrucciones tridimensionales finales de RepB<sub>6</sub> tanto en ausencia como en presencia de ADN se ha llevado a cabo utilizando el programa EMAN v1.7 (Ludtke, *et al.*, 1999). A partir de

las micrografías correspondientes tomadas a 0, 20 y 40° de inclinación se han extraído semi-automáticamente con el programa *autoboxer* de EMAN un total de 11628 partículas individuales sin ADN y otras 8659 con ADN. Después de enmascarar, filtrar, centrar y normalizar cada conjunto de datos independientemente, se ha llevado a cabo el refinamiento angular por encaje de proyecciones de cada uno de ellos utilizando como modelo inicial una reconstrucción de RCT filtrada a 45Å (Figura 26D). Esta reconstrucción se ha llevado a cabo con el programa *XMIPP* (Sorzano, *et al.*, 2004) a partir de 4519 pares de imágenes de partículas individuales extraídas manualmente de las micrografías de la proteína sin ADN. Las imágenes no-inclinadas fueron alineadas y clasificadas utilizando el método libre de referencia basado en la maximización de la probabilidad (Scheres, *et al.*, 2005) implementado en el comando *MLalign2D* de *XMIPP* (Figura 26B). Finalmente, se ha generado un modelo 3D (Figura 26C) utilizando las 582 imágenes inclinadas a 40° de una de las clases de referencia con una vista frontal característica de la proteína (Figura 26B). En la Figura 27 se muestran la FSC para las reconstrucciones obtenidas en ausencia (izquierda) y en presencia de ADN (derecha). Según el criterio de FSC=0.5 las resoluciones alcanzadas son 22 y 24Å, respectivamente. La mano de las reconstrucciones se ha determinado por comparación directa con la estructura cristalográfica de simetría C2. Los umbrales de representación se han elegido para justificar aproximadamente la masa de la proteína.



**Figura 26.** Modelo inicial de RepB obtenido por RCT. En el panel A se muestran dos vistas de la misma región tomadas a 0 (izquierda) y 40° (derecha) de inclinación. Los puntos rojos indican la posición de algunas de las partículas. En B, pueden verse las imágenes medias obtenidas al clasificar las vistas a 0°. Se ha resaltado con un recuadro rojo la clase utilizada para generar el modelo 3D (C). En D, se ilustra este modelo pero filtrado a ~45Å.



**Figura 27.** Perfiles de FSC de las reconstrucciones 3D de RepB. Para la reconstrucción en ausencia (izquierda) y en presencia (derecha) de ADN, el FSC=0.5 indica resoluciones de 22 y 24Å, respectivamente. Nótese que la escala del eje de abscisas es logarítmica.

## 4.5. Herramientas auxiliares

Con el fin de validar la metodología del análisis vibracional en CI se han utilizado y desarrollado una serie de herramientas de comparación auxiliares cuyos principios básicos se detallan a continuación:

### 4.5.1. Métricas de comparación de modos normales

Para evaluar el solapamiento entre los movimientos predichos por los modos normales de vibración y las transiciones conformacionales observadas experimentalmente se emplea el siguiente producto escalar:

$$\alpha_k = \left| \mathbf{v}_k^C \cdot \Delta \mathbf{X} \right| \quad (4.5.1)$$

Dónde  $\mathbf{v}_k^C$  es el  $k$ -ésimo modo expresado en CC y  $\Delta \mathbf{X}$  representa el vector de transición entre las conformaciones inicial ( $\mathbf{X}^0$ ) y final ( $\mathbf{X}^f$ ). Este último vector se define así:

$$\Delta \mathbf{X} = \mathbf{X}^0 - \mathbf{X}^f / \left\| \mathbf{X}^0 - \mathbf{X}^f \right\| \quad (4.5.2)$$

Cuanto mayor sea el valor de  $\alpha$ , tanto mayor será el parecido del modo con el vector de transición. Nótese que  $\alpha$  únicamente puede tomar valores entre 0 y 1 si los modos normales y el vector de transición están normalizados, *i.e.*  $\mathbf{v}_k^C \cdot \mathbf{v}_k^C = 1$  y  $\Delta \mathbf{X} \cdot \Delta \mathbf{X} = 1$ .

Para evaluar cuantitativamente lo bien que los  $n$  primeros modos normales describen una cierta transición conformacional se ha empleado el solapamiento acumulado  $\delta$ :

$$\delta_n = \sqrt{\sum_{k=1}^n \alpha_k^2} \quad (4.5.3)$$

Su valor puede variar entre 0 y 1. Cuanto mayor sea éste, tanto mejor será representada la transición por el conjunto de modos.

Cuando es necesario comparar dos conjuntos de modos normales  $\mathbf{u}$  y  $\mathbf{v}$  obtenidos en distintas condiciones se emplea el solapamiento entre espacios vectoriales  $\gamma$ :

$$\gamma_n = \frac{1}{n} \sum_{k,l}^n (\mathbf{u}_k \cdot \mathbf{v}_l)^2 \quad (4.5.4)$$

Esta métrica condensa en un único valor el grado de solapamiento entre los vectores de ambos espacios. Al igual que las dos métricas anteriores, sus valores pueden variar entre 0 y 1, y cuanto mayor sea este, tanto mayor será la similitud entre los dos conjuntos de vectores. Para que los valores de  $\gamma$  estén acotados en el intervalo [0,1] cada conjunto de modos debe ser totalmente ortogonal entre sí. Esto solo es posible si los modos vienen expresados en CC pesadas por las masas (ec. 4.3.2) (Brooks, *et al.*, 1995).

#### 4.5.2. Actualización del programa de análisis vibracional – DefProt

Para comparar el NMA en CI con el NMA en CC se ha actualizado el programa de análisis vibracional en CC *DefProt* (Garzon, *et al.*, 2007b). Se han incluido los modelos atómicos y de potencial descritos en los apartados 4.2.2 y 4.2.3, respectivamente. Además, se han implementado rutinas de diagonalización más eficientes que extienden el rango de aplicación de *DefProt* a sistemas de hasta 5000 átomos en arquitecturas de 32-bit.

#### 4.5.3. Dinámica esencial y análisis de componentes principales – iPca

Se ha implementado una herramienta de dinámica esencial (ED, del inglés *Essential Dynamics*) denominada *iPca* para analizar la información contenida en las trayectorias generadas con nuestras aplicaciones. Esta herramienta lleva a cabo el PCA de las matrices de covarianza obtenidas a partir de trayectorias de MD para obtener unos modos esenciales de movimiento denominados componentes principales (Amadei, *et al.*, 1993). Aunque se obtienen de manera diferente a los modos de NMA, poseen un significado similar, *i.e.* representan las principales maneras de deformar el sistema y son ortogonales entre sí (Rueda, *et al.*, 2007a).

En CC, la matriz de covarianza ( $\mathbf{C}$ ) puede calcularse a partir del vector de coordenadas ( $\mathbf{X}$ ) de cada conformación de la trayectoria según la expresión:

$$\mathbf{C} = \left\langle (\mathbf{X} - \langle \mathbf{X} \rangle)(\mathbf{X} - \langle \mathbf{X} \rangle)^T \right\rangle \quad (4.5.5)$$

Donde los corchetes angulares ( $\langle \rangle$ ) representan el promedio sobre todas las conformaciones. La matriz  $\mathbf{C}$  puede descomponerse en un conjunto de componentes principales ( $\mathbf{P}$ ) a través del problema estándar de valores propios:

$$\mathbf{C} = \mathbf{P}\mathbf{\Sigma}\mathbf{P}^T \quad (4.5.6)$$

Donde  $\Sigma$  es la matriz diagonal de valores propios. De este modo, es posible validar las trayectorias comparando sus componentes principales con los modos normales utilizando las métricas de la sección 4.5.1.

Para resolver el problema que surge al comparar modos normales (o componentes principales) obtenidos a partir de distintos modelos atómicos se ha seguido el siguiente procedimiento. Primero se genera una matriz de covarianza empleando únicamente las coordenadas comunes de los modos normales (o componentes principales) a comparar. Y luego, al igual que en ED (Amadei, *et al.*, 1993), se lleva a cabo el PCA de dicha matriz para obtener un conjunto totalmente ortogonal de vectores con las coordenadas seleccionadas.

Para calcular analíticamente los elementos de la matriz de covarianza ( $C_{xy}$ ) a partir de los modos normales puede emplearse la siguiente expresión (Brooks, *et al.*, 1995):

$$C_{xy} = \sum_k^N \frac{k_B T}{\lambda_k} v_{xk}^C v_{yk}^C \quad (4.5.7)$$

Donde  $T$  es la temperatura absoluta (300 K) y  $k_B$  la constante de Boltzmann.

## 4.6. Datos de validación

Con el fin de validar las aplicaciones desarrolladas y ajustar sus parámetros se han escogido distintos conjuntos de estructuras de proteínas y ácidos nucleicos. Sus detalles se comentan en las siguientes secciones.

### 4.6.1. Conjuntos de pruebas de estructuras atómicas

El conjunto de pruebas de proteínas está formado por 23 pares de estructuras abiertas y cerradas con desplazamientos mayores de 2 Å RMSD seleccionados de la base de datos de movimientos moleculares *MolMovDB* (<http://molmovdb.org>) (Flores, *et al.*, 2006). Estos 46 movimientos (de abierto a cerrado y *viceversa*) abarcan diferentes tamaños (de 186 a 994 aminoácidos) y una amplia variedad de cambios conformacionales de gran amplitud (hasta 18 Å RMSD), siendo la desviación promedio de  $7.57 \pm 4.39$  Å. Las estructuras son relativamente grandes y se ajustan a los estándares de calidad del programa de validación estructural *Molprobit* (Chen, *et al.*, 2010). Aunque predominan los movimientos bisagra (*hinge*), también se incluyen movimientos complejos y de cizalla (*shear*). La descripción detallada de estas estructuras se encuentra en la Tabla 3.

Con la intención de complementar los ensayos con movimientos de ácidos nucleicos, hemos cribado el PDB (<http://www.rcsb.org>) (Berman, *et al.*, 2000) en busca de cambios conformacionales colectivos de ARN. Para ello, de las 1762 estructuras del PDB que contenían



ARN el 11/5/2010, se seleccionaron únicamente las cadenas de ARN. Después de descartar las estructuras con el nucleótido inosina, quedaron 9088 estructuras diferentes de cadena única. De éstas, se seleccionaron las cadenas que tenían más de 40 nucleótidos, *i.e.* 1562. Para buscar

**Tabla 3.** Conjuntos de pruebas de transiciones conformacionales.

	Abierta		Cerrada		N <sup>c</sup>	Rmsd <sup>d</sup>		Tipo <sup>e</sup>	Nombre <sup>f</sup>
	PDB <sup>a</sup>	Rg <sup>b</sup>	PDB	Rg		MolMov	Kabsch		
MolMovDB	1ex6	A - 17.09	1ex7	A - 16.26	186	4.58	3.64	D-h-2	Guanilato quinasa (GDK)
	4ake	A - 19.46	1ake	A - 16.37	214	8.26	7.13	D-h-2	Adenilato quinasa (ADK)
	1ggg	A - 19.03	1wdn	A - 17.47	220	10.26	5.34	D-h-2	Proteína de unión a glutamina
	2lao	A - 19.05	1lst	A - 17.70	238	8.67	4.70	D-h-2	Proteína de unión a Lys/Arg/Ornitina (LAO)
	1urp	A - 20.08	2dri	A - 19.08	271	7.69	4.06	D-h-2	Proteína de unión a ribosa
	1ram	B - 24.56	1lei	A - 23.11	273	4.96	3.07	D-?-2	NF-κB
	5at1	A - 19.67	8atc	A - 19.18	310	2.41	2.35	S-a-2	Aspartato carbamoiltransferasa
	1ckm	A - 20.89	1ckm	B - 19.63	317	4.35	3.49	D-h-2	ARNm capping enzyme
	3dap	A - 21.86	1dap	B - 20.56	317	5.83	4.18	D-h-2	Deshidrogenasa del ácido diaminopimélico
	1bp5	A - 21.31	1a8e	A - 19.32	329	12.16	6.71	D-h-2	Lóbulo N-terminal de la transferrina
	1jql	A - 25.63	2pol	A - 25.16	366	2.55	1.99	D-s-2	ADN polimerasa III-β
	1omp	A - 21.63	1anf	A - 20.73	370	7.29	3.77	D-h-2	Proteína de unión a maltodextrina (MBP)
	8adh	A - 20.59	2jhf	A - 20.44	374	2.19	1.28	D-s-2	Alcohol deshidrogenasa (ADH)
	9aat	A - 21.14	1ama	A - 20.89	387	2.19	1.54	D-s-2	Aspartato amino transferasa (AAT)
	1bnc	A - 23.46	1dv2	A - 21.75	452	5.72	4.96	D-h-2	Biotin carboxilasa
	1dpe	A - 24.66	1dpp	A - 22.77	507	12.43	6.53	D-h-2	Proteína de unión a dipéptidos (DppA)
	1rkm	A - 23.81	2rkm	A - 23.24	517	5.76	3.08	D-h-2	Proteína de unión a oligopéptidos
	1sx4	A - 29.23	1oel	A - 25.36	524	15.83	12.20	C----	Monómero de GroEL
	1i7d	A - 29.71	1d6m	A - 29.65	609	4.20	3.47	D-?-2	ADN topoisomerasa III
	1lfg	A - 29.41	1lfh	A - 28.33	691	8.08	6.43	D-h-2	Lactoferrina (LF)
	1oao	D - 28.44	1oao	C - 27.03	728	13.60	7.05	D-h-2	Acetyl-CoA sintasa/CO deshidrogenasa
ARN	1ih7	A - 31.34	1ig9	A - 30.66	890	7.20	6.50	C----	ADN polimerasa de Rb69
	1su4	A - 38.48	1t5s	A - 37.56	994	17.99	13.55	C----	Bomba ATPasa de calcio (SERCA)
	1a51	A 3 20.95	1a51	A 9 20.08	40	8.72	5.08	-	Fragmento del ARNr 5S de E.Coli
	2i7z	B 17 21.93	2jyf	A 1 19.61	42	10.79	8.13	-	Receptor de ARN del bucle GAAA (sintético)
	1z2j	A 13 21.06	1z2j	A 2 20.61	44	10.49	5.04	-	Inductor de desplazamiento de marco HIV-1
	1u63	D - 19.86	2vpl	A - 19.73	48	4.65	3.38	-	ARNm unido a la proteína ribosomal L1
	3knj	W - 23.90	1gts	A - 23.51	74	6.34	5.21	-	ARNt-Gln de E.Coli
	3fih	Y - 23.96	1pns	A - 23.36	75	7.95	5.81	-	Sitio A/T del ARNt-Phe (sintético)
	1p5o	A 6 29.28	1p5o	A 12 29.24	76	8.31	5.74	-	Dominio II del sitio IRES del HCV (sintético)
	3cul	D - 25.57	3cun	B - 24.47	91	3.64	3.14	-	Aminoacilasa de ARNt (ribozima sintética)
	1mfq	A - 30.75	2go5	B - 30.13	126	5.23	3.38	-	ARN 7S de la SRP humana
	2wdj	A - 63.75	3hux	A - 62.27	721	5.27	5.00	-	Fragmento del ARNr 23S de T.Thermophilus
	3ela	1 - 67.93	2aw7	A - 66.88	1529	8.12	6.73	-	ARNr 16S de E.Coli

<sup>a</sup> Código de PDB, cadena y modelo (si la estructura es de RMN).

<sup>b</sup> Radio de giro:  $Rg = \left( \frac{1}{N} \sum_i (r_i - r_c)^2 \right)^{1/2}$  donde  $r_i$  es el vector de posición del átomo  $i$ , y  $r_c$  el del centro geométrico.

<sup>c</sup> Número de aminoácidos o nucleótidos.

<sup>d</sup> Desviación RMSD basada en los átomos  $C_\alpha$  entre las estructuras abierta y cerrada para los dos métodos de alieneamiento utilizados: minimización por mínimos cuadrados (Kabsch, 1976; Kabsch, 1978) y superposición de regiones estáticas del servidor MolMovDB (Echols, *et al.*, 2003). En ARN, en lugar de este último método, se ha empleado una implementación propia del método de Damm y Carlson (Damm y Carlson, 2006).

<sup>e</sup> Tipo de movimiento de proteínas según la clasificación de MolMovDB. I) Movimientos de Dominio: de cizalladura (D-s-2), de bisagra (D-h-2) y aquellos que no pueden ser clasificados actualmente (D-?-2). II) Movimientos de subunidades (más grandes que los de dominios): aquellos que consisten en una transición alostérica (S-a-2) y movimientos complejos (C----).

<sup>f</sup> Nombre de la macromolécula.



pares de estructuras con la misma secuencia se llevó a cabo el alineamiento entre todas ellas con el comando *megablast* basado en BLAST (Camacho, *et al.*, 2009). Todas las secuencias excepto 42, presentaban alineamientos significativos. Entre las 1520 secuencias resultantes se encontraron 93508 regiones con alineamientos individuales significativos. De éstos, solo se seleccionaron aquellos con el 100% de identidad. En los casos con más de un marco de alineamiento se tomaron aquellos de mayor tamaño. A partir de los 57831 alineamientos resultantes se seleccionaron los pares de estructuras con el mismo número de átomos y con RMSD superior a 2Å. En los casos en los que existen múltiples cadenas con la misma secuencia, se elige aquellos de mayor RMSD. Así, se obtuvieron 2189 pares de estructuras no redundantes. Debido al gran número de pares de estructuras diferentes, éstas se clasificaron por homología de secuencia usando el programa *blastclust* de BLAST. De este modo se obtuvieron 80 familias cuyos miembros comparten el 95% de homología. En cada una de ellas se seleccionó el par de estructuras con mayor RMSD como representante. Después de alinear entre sí cada par de representantes con el método de Damm y Carlson (Damm y Carlson, 2006) se visualizaron sus cambios conformacionales. Como muchos de ellos representaban un mismo tipo de movimiento, especialmente los de menor tamaño, se tomaron aquellos de mayor RMSD. De este modo se seleccionaron manualmente 25 pares de estructuras con movimientos diferentes a partir de los 80 iniciales. Posteriormente, se eliminaron los casos más grandes para acelerar los ensayos. Asimismo, ha resultado imposible alinear correctamente varias estructuras debido a que presentaban conformaciones muy diferentes ( $>20\text{\AA}$ ). Además, la mayoría de las estructuras de mayor tamaño que proceden de MET tienen aspecto distorsionado y baja calidad estructural por lo que los casos correspondientes también se han eliminado. También se descartaron aquellos casos con movimientos demasiado localizados por quedar fuera del alcance de los modos normales de baja frecuencia. Finalmente, nos quedamos con 11 pares de estructuras con movimientos colectivos mayores de 3 Å y que superan los criterios de calidad de *Molprobit*. En estos casos, la desviación promedio es  $7.23 \pm 2.37$  Å. Desafortunadamente, la mayoría de estas estructuras son relativamente pequeñas (40-126 nucleótidos) a excepción de dos grandes ARNr de 721 y 1529 residuos. La descripción de estas estructuras de prueba también se detalla en la Tabla 3.

Tanto en la simulación de transiciones conformacionales como en el ajuste de estructuras atómicas en mapas de densidad las estructuras iniciales deben situarse aproximadamente sobre las estructuras objetivo. En los casos de proteínas se empleó el servidor *MolMovDB* (Echols, *et al.*, 2003) que determina la superposición óptima de las regiones invariantes de ambas estructuras y favorece la visualización del cambio conformacional. Para obtener resultados

equivalentes con ARN se ha implementado y utilizado el método alineamiento de Damm y Carlson (Damm y Carlson, 2006) que está basado en la minimización del RMSD y emplea una función gaussiana de la distancia para favorecer iterativamente la superposición de las regiones más similares.

#### 4.6.2. Casos de ajuste flexible con mapas simulados

Se han generado problemas de ajuste flexible a distintas resoluciones tanto de proteínas como de ARN a partir de las estructuras en conformaciones abiertas y cerradas obtenidas en la sección anterior. Para ello, se han simulado mapas a 8, 10, 15, 20 y 25 Å de resolución con un muestreo de 2 Å/voxel utilizando el método de filtrado por convolución. De esta forma, es posible obtener 46 problemas de ajuste diferentes para proteínas y otros 22 para ARN, cuya solución final es conocida. En el caso de proteínas, se han considerado dos orientaciones iniciales diferentes: *i*) la misma que en la sección anterior, y *ii*) la obtenida tras el ajuste rígido con *ADP-EM* de las estructuras iniciales en los mapas objetivo simulados (~2.8° de muestreo angular y 2 Å de muestreo traslacional). En ARN, únicamente se ha empleado la primera orientación.

Con el fin de evaluar el comportamiento de *iModfit* con otras herramientas de ajuste flexible basadas en MD (*MDFF*, Tama 1, Tama 2) se han simulado algunos de los casos de ajuste con los que fueron validados estos métodos. Después de obtener del PDB los pares de estructuras atómicas correspondientes se ha empleado el comando *profix* del programa *Jackal* ([http://wiki.c2b2.columbia.edu/honiglab\\_public/index.php/Software:Jackal](http://wiki.c2b2.columbia.edu/honiglab_public/index.php/Software:Jackal)) para completar los átomos pesados (no-hidrógeno) que les faltan y *clustalW2* (Larkin, *et al.*, 2007) para seleccionar únicamente las secuencias comunes entre ambas. Antes de generar los mapas, estas estructuras han sido alineadas minimizando su RMSD (Kabsch, 1978). Se han empleado resoluciones, muestreos y orientaciones iniciales equivalentes a las utilizadas por sus autores.

#### 4.6.3. Casos de ajuste flexible con ruido simulado

Para estudiar cómo afecta la presencia de ruido a los ajustes realizados con *iModfit* en condiciones controladas y comparar con el método de ajuste descrito por Zheng (Zheng, 2011) se han generado mapas simulados con distintas relaciones de señal ruido (SNR) utilizando el conjunto de pruebas con el que fue validado dicho método. El ruido se ha modelado como una distribución gaussiana de  $\sigma$  proporcional a la raíz cuadrada de la densidad de cada voxel (Jolley, *et al.*, 2008). Así, es posible añadir al mapa simulado el nivel de ruido necesario para conseguir la SNR deseada. La SNR se define según la convención de procesamiento de imagen:

$$SNR = \frac{\mu_{señal}}{\sigma_{ruido}} \quad (4.6.1)$$

Dónde la  $\mu_{señal}$  representa el valor medio de los voxels con densidades positivas del mapa simulado y  $\sigma_{ruido}$  es la desviación estándar del ruido correspondiente. Los mapas han sido generados a 10, 15 y 20 Å de resolución con un muestreo de 3, 5 y 6 Å, respectivamente, para parecerse a los mapas simulados de Zheng (Zheng, 2011).

#### 4.6.4. Casos de ajuste flexible basados en modelado por homología

Para completar la comparación de *iModfit* con otras herramientas de ajuste basadas en MD se ha utilizado el conjunto de estructuras de prueba creado por el grupo de Sali (Topf, *et al.*, 2008). Este conjunto de pruebas representa problemas de ajuste flexible en los cuales la estructura atómica de partida se ha obtenido por modelado comparativo y el mapa objetivo se ha simulado a partir de una estructura atómica experimental. Esto supone una dificultad adicional para el método de ajuste puesto que las estructuras modeladas presentan más distorsiones que las obtenidas experimentalmente. A continuación se describe brevemente cómo fue obtenido este conjunto de pruebas. Primero se seleccionaron pares de estructuras en distintas conformaciones de las bases de datos *MolMovDB* y de *DBAli* (Marti-Renom, *et al.*, 2007). En cada uno de los pares se eligió una de las estructuras como objetivo y la otra como molde. Los modelos iniciales fueron generados por modelado comparativo a partir de dichos moldes con el programa *MODELLER* (Eswar, *et al.*, 2007). Los mapas de densidad se generaron a partir de las estructuras objetivo con el comando *pdb2vol* de *SITUS* a 10 Å de resolución y con un muestreo 1 Å/voxel (Wriggers, 2010). Los ajustes entre las estructuras *1blb* y *1a45* y *1u0s* y *1ffg* tienen los dominios separados y en orientaciones muy diferentes en las estructuras molde y objetivo. Por tanto, como estos casos incluyen la dificultad adicional de encontrar la orientación inicial relativa entre los dominios, no han sido utilizados en ningún ensayo. Los 11 problemas de ajuste restantes poseen una amplitud promedio de  $6.21 \pm 1.94$  Å y se detallan en la Tabla 4.

#### 4.7. Ajuste flexible con otras herramientas

Para comparar *iModfit* con otros métodos automáticos de ajuste flexible basados tanto en NMA como en MD se han utilizado las herramientas: *NMFF* (Tama, *et al.*, 2004a), *NORMA* (Suhre, *et al.*, 2006) y *YUPSCX* (Tan, *et al.*, 2008), ya comentadas en la sección 3.5.3 de la Introducción. A continuación se describen las opciones empleadas con estas herramientas.

Los principales parámetros de *NMFF* son: el número de modos, el umbral de distancia que define la red elástica, la distancia máxima a la que pueden moverse los átomos en cada iteración, y el factor de convergencia. Por defecto, los autores utilizan solamente 20 modos, aunque recomiendan no emplear más de 30 para evitar problemas de sobreajuste. En cuanto al resto de opciones, el umbral que define la red elástica es 5 Å cuando se utilizan todos los átomos

**Tabla 4.** Casos de ajuste flexible basados en estructuras de modelado por homología.

Molde		Objetivo		N <sup>b</sup>	RMSD <sup>c</sup>		Nombre <sup>d</sup>
PDB <sup>a</sup>		PDB			D&C	MODELLER	
1a03	B	1ozo	A	84	5.21	4.68	<i>S100P humana unida a Calcio</i>
1k9p	A	1uwo	A	90	4.49	4.73	<i>S100B humana unida a Calcio</i>
2ggm	B	1c1l	A	144	4.97	4.90	<i>Calmodulina humana</i>
1hnf	A	1ccz	A	170	6.71	5.58	<i>Dominio de unión CD2 de CD58</i>
1ex7	A	1jxm	A	181	5.90	5.34	<i>Guanilato quinasa unida a GMP</i>
3erd	A	1g5y	D	212	5.44	5.17	<i>Receptor del ácido retinoico RXR-<math>\alpha</math></i>
1dvr	B	1ake	A	213	3.89	4.50	<i>Adenilato quinasa de E.coli</i>
1svc	A	1ikn	A	253	8.57	10.40	<i>NF-<math>\kappa</math>B</i>
1p16	A	1ckm	A	260	8.89	8.32	<i>mARN capping enzyme</i>
1gtm	A	1clx	A	338	5.74	6.51	<i>L-fenilalanina deshidrogenasa de Rhodococcus</i>
1hwz	A	1hrd	C	425	6.14	8.20	<i>Glutamato deshidrogenasa de Pyrococcus furiosus</i>

<sup>a</sup> Código del PDB y de la cadena de las estructuras empleadas como Molde y Objetivo para construir las estructuras iniciales por modelado comparativo (véase Materiales y Métodos).

<sup>b</sup> Número de aminoácidos.

<sup>c</sup> RMSD entre los modelos iniciales y las estructuras empleadas para generar los mapas objetivo. Estas estructuras se han alineado empleando una implementación propia del método de ajuste de Damm y Carlson (D&C) (Damm y Carlson, 2006). También se ha incluido el RMSD obtenido tras superponer las estructuras con el comando *superpose* de *MODELLER*. En todos los casos únicamente se han tenido en cuenta los átomos C $_{\alpha}$ .

<sup>d</sup> Nombre de las proteínas.

pesados, la distancia máxima permitida a la que pueden moverse los átomos es 1Å, y el factor de convergencia es 1. Al utilizar estas opciones, los ajustes muestran grandes desviaciones entre los modelos finales y las estructuras de los mapas, lo que evidencia una posible falta de convergencia. Para aprovechar mejor el potencial de esta herramienta se han optimizado los parámetros. Con este fin, se han llevado a cabo los ajustes empleando distintas combinaciones de los parámetros más influyentes, *i.e.* de 10 a 30 modos y factores de convergencia de 1 a 0.001. Además, para minimizar la influencia de los diferentes criterios de resolución a la hora de generar los mapas simulados, los ajustes con *NMFF* se han llevado a cabo empleando resoluciones ligeramente mayores (6Å) que las correspondientes a los mapas (10Å). De esta forma, se ha determinado que es posible obtener mejores ajustes cambiando los valores por defecto del número de modos de 20 a 30 y del factor de convergencia de 1 a 0.001. Por ejemplo, estos parámetros son capaces de reducir de 2.5 hasta 1.4Å la desviación promedio entre los modelos finales y las estructuras utilizadas para generar los mapas de los ajustes del conjunto de pruebas de proteínas (a 10Å de resolución).

El método *NORMA* de ajuste flexible posee multitud de parámetros: el número de modos, la amplitud del movimiento estimada, el número de iteraciones de cada paso de minimización, el número de rondas de minimización y el número de pasos (y temperatura) del recocido simulado. Como los autores ajustan estos valores en cada problema particular, se han ensayado

distintas combinaciones de parámetros para determinar un conjunto que sea válido en todos los casos. Se han realizado distintos ajustes probando de 5 a 30 modos, de 1 a 15 rondas de minimización y de 0 a 10 pasos de recocido. Desafortunadamente, las ejecuciones de *NORMA* finalizan inesperadamente en algunos casos al emplear ciertas combinaciones de parámetros. A pesar de ello, ha sido posible encontrar un conjunto que funciona relativamente bien. Estos parámetros son: 20 modos, amplitud estimada de 500, 10 pasos de recocido simulado, y 15 rondas de minimización con 25 pasos cada una. Además, cabe destacar que *NORMA* es un conjunto de *scripts* que sirve para comunicar los programas *URO* y *elNEMO* a través del algoritmo de minimización, en consecuencia, la utilización de los recursos computacionales resulta muy ineficiente (bajos porcentajes de uso de CPU) por lo que los tiempos de ejecución son generalmente muy largos en algunos casos (días e incluso semanas). A pesar de que los autores recomiendan llevar a cabo los ajustes de mayor amplitud de forma manual y en múltiples pasos para minimizar posibles sobreajustes, en el presente trabajo los ajustes se han efectuado automáticamente y en un único paso.

Los ajustes flexibles con el método *YUPSCX*, basado en MD y recocido simulado, se han efectuado utilizando sus parámetros por defecto.

#### **4.8. Detalles técnicos**

Las aplicaciones han sido desarrolladas íntegramente en los lenguajes de programación C/C++. Se han elaborado distintas distribuciones para arquitecturas de 32 y 64 bits utilizando los compiladores GNU GCC (v 4.0.3) e Intel® (v12.0.0) para proporcionar una difusión adecuada y minimizar los problemas de compatibilidad en los sistemas de los usuarios finales. Con el fin de facilitar el uso de las herramientas y eliminar posibles problemas de dependencias con librerías dinámicas también incluimos distribuciones enlazadas estáticamente que pueden ser utilizadas directamente. Tanto en la resolución del problema generalizado de valores propios como en el filtrado de mapas de densidad se utilizaron las rutinas de álgebra lineal LAPACK (Anderson, *et al.*, 1999) y de transformada rápida de Fourier (FFT, del inglés *Fast Fourier Transform*), respectivamente, optimizadas para procesadores Intel® a través de la implementación no-comercial Intel® MKL (v10.3). En el presente trabajo únicamente se han utilizado las versiones mono-hilo (sin paralelizar) de esta librería utilizando la opción de enlazado denominada *sequential*. Para el cálculo de las métricas entre espacios vectoriales y el análisis de los datos se elaboraron varios *scripts* en los lenguajes interpretados Perl y Bash.

A no ser que se especifique lo contrario, los cálculos cronometrados se llevaron a cabo en un PC Linux de sobremesa con procesador Intel® Core™2 Quad Q6600 a 2.44GHz y 4GB de

RAM. Los cálculos con las macromoléculas de mayor tamaño (virus y ribosomas) se efectuaron en un servidor Linux equipado con dos procesadores Intel® Xeon™ Quad de 48GB de RAM debido al gran número de variables que poseen (>15000). Los restantes cálculos se efectuaron en los recursos computacionales (*clusters*) del Centro de Investigaciones Biológicas y del Instituto de Química Física Rocasolano (ambos del CSIC) utilizando los nodos equipados con procesadores Intel® Xeon™ E5410 a 2.33GHz y 8GB de RAM.

La visualización de todas las estructuras atómicas y mapas de densidad mostrados en este trabajo se han obtenido íntegramente con los programas *VMD* (Humphrey, *et al.*, 1996) y *Chimera* (Pettersen *et al.*, 2004). Para el alineamiento entre estructuras atómicas se ha utilizado tanto *Chimera* como una implementación propia, denominada *emt\_rmsd*, de los métodos basados en la minimización del RMSD de Kabsch (Kabsch, 1976; Kabsch, 1978) y de Damm y Carlson (Damm y Carlson, 2006). Todas las desviaciones entre estructuras atómicas se han calculado utilizando la ecuación 4.3.7.

## 5. Resultados y discusión

En esta sección se describen los resultados obtenidos con las distintas herramientas bioinformáticas desarrolladas. Estas herramientas, cuyos detalles técnicos se describen en la sección de Materiales y métodos, se utilizarán en los siguientes contextos:

- i) Cálculo de los modos normales de vibración en CI – *iMode*.
- ii) Ilustración de modos vibracionales y redes elásticas – *iModview*.
- iii) Animación de modos – *iMove*.
- iv) Exploración conformacional basada en el muestreo de MC – *iMC*.
- v) Simulación de trayectorias de transiciones conformacionales entre pares de estructuras atómicas – *iMorph*.
- vi) Ajuste flexible de estructuras atómicas en el interior de mapas de densidad de microscopía electrónica – *iModfit*.

En primer lugar, se valorarán los detalles técnicos de diseño e implementación de los algoritmos básicos con los que se obtienen los modos normales de vibración. En segundo lugar, se llevará a cabo una extensa validación de las distintas herramientas bioinformáticas con datos simulados y experimentales. Se pondrá especial énfasis en *iModfit*, comparándolo con otras herramientas de ajuste flexible actuales. Finalmente, se estudiará el sistema RepB/ADN mediante microscopía electrónica y SP, lo que incluye la aplicación práctica de la herramienta de ajuste flexible desarrollada.

### 5.1. Diseño e implementación

En el diseño e implementación de la metodología del NMA en CI se ha prestado especial atención en la eficiencia de las tareas computacionalmente más exigentes: el análisis vibracional, el movimiento en CI y la conversión de estructuras atómicas en mapas de densidad. A continuación se comentan brevemente los esfuerzos encaminados a reducir el coste computacional del NMA.

#### 5.1.1. Implementación del análisis vibracional en coordenadas internas

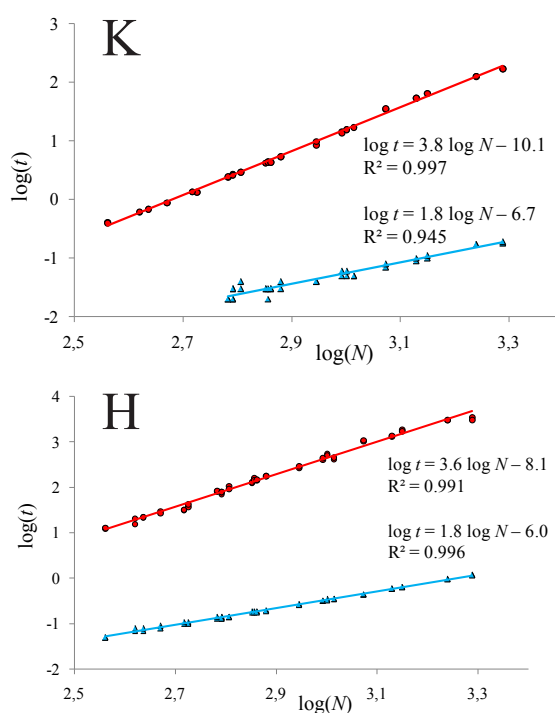
La tarea más costosa es el cálculo de los modos normales de vibración. Esencialmente, el análisis vibracional se divide en dos fases: *i*) el cálculo de las matrices de energía cinética (**K**) y de energía potencial (**H**, o Hessiano), y *ii*) la diagonalización de ambas para resolver el problema generalizado de valores propios ( $\mathbf{H}\mathbf{V}=\mathbf{A}\mathbf{K}\mathbf{V}$ ). Para llevar a cabo la primera fase se



han implementado los algoritmos rápidos de Go, y para la segunda, se han empleado las rutinas optimizadas de la librería estándar LAPACK (Anderson, *et al.*, 1999).

Mis esfuerzos se han centrado fundamentalmente en la implementación y desarrollo de algoritmos eficientes para la obtención de las matrices del sistema (**K** y **H**). El método directo para calcular estas matrices resulta muy costoso computacionalmente y consume mucha memoria, incluso con macromoléculas pequeñas (~300 residuos). Por esta razón se ha implementado un método rápido basado en los algoritmos rápidos de Go (secciones 4.1.4 y 4.1.5 de Materiales y métodos) que permite acelerar los cálculos y reducir el gasto de memoria. Para evaluar su eficiencia se ha determinado el tiempo de cálculo dedicado a la obtención de las matrices con ambos métodos (directo y rápido) en el análisis vibracional de cada una de las estructuras del conjunto de pruebas de proteínas. En estos ensayos se ha utilizado el programa *iMode* con los parámetros por defecto. En la Figura 28 se ilustran los tiempos de cálculo que necesita cada método para obtener las matrices

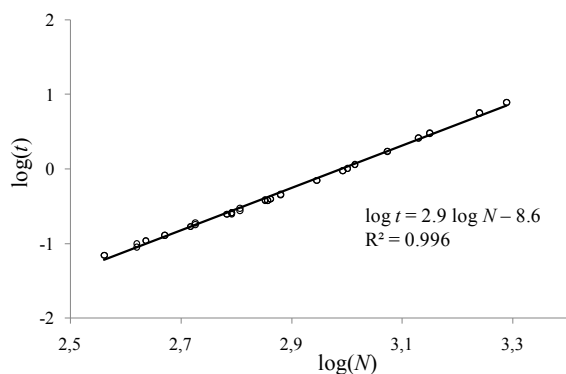
**K** y **H** en función del número de variables ( $N$ ). Como puede apreciarse, la velocidad del método rápido (azul) es muy superior a la del directo (rojo) en todo el rango de  $N$  y con ambas matrices. Debido a que los tiempos de cálculo de la matriz **K** por el método rápido son muy cortos ( $<0.01s$ ) con los casos más pequeños, únicamente se han incluido los datos correspondientes a proteínas de más de 300 aminoácidos. Al emplear una representación logarítmica, la distancia vertical entre las rectas determina el coste computacional relativo expresado en órdenes de magnitud. Como era de esperar, el coste computacional disminuye drásticamente al emplear el método rápido. Por ejemplo, el cálculo de ambas matrices llega a ser entre 3 y 4 órdenes de magnitud más rápido en el caso de mayor tamaño (*Isu4*, ~1000 residuos y ~2000 variables). Nótese además que las diferencias crecen con el tamaño del sistema.



**Figura 28.** Coste computacional del cálculo de las matrices del sistema por los métodos rápido y directo. Se ha muestra la representación logarítmica de los tiempos ( $t$ , [s]) de cálculo del tensor métrico (K) y del Hessiano (H) en función del número de variables ( $N$ ). Los resultados del método directo aparecen en rojo, y en azul los del rápido. También se incluyen las regresiones lineales con sus valores de  $R^2$ .

El coste computacional en función del tamaño del sistema se ha estimado aproximando el tiempo de cálculo a una función exponencial dependiente del número de variables ( $t = k \cdot N^g$ ), siendo  $k$  una constante y el exponente  $g$  el factor de escala del coste computacional. Para determinar este último se transforma esta expresión en una recta tomando el logaritmo a ambos lados de la ecuación, *i.e.*  $\log t = g \cdot \log N + \log k$ . De este modo, al representar el logaritmo del tiempo ( $\log t$ ) en función del logaritmo del número de variables ( $\log N$ ) obtiene una recta cuya pendiente es el factor de escala del coste computacional. En la Figura 28 también se han incluido las regresiones lineales obtenidas. Claramente, todos los datos se ajustan a esta aproximación lineal, con coeficientes de correlación  $R^2$  muy altos ( $>0.99$ ). Como puede observarse en dicha figura, los exponentes del coste computacional para el cálculo de la energía cinética por los métodos directo y rápido son 3.8 y 1.8, respectivamente, por lo que se aproximan a los valores ideales considerados en la sección 4.1.5 de Materiales y métodos, *i.e.* 3 y 2, respectivamente. En cuanto al coste computacional para obtener el Hessiano, se obtienen valores de 3.6 para el método directo y 1.8 para el rápido que resultan ser algo inferiores a los ideales, *i.e.* 4 y 2, respectivamente. Estas desviaciones pueden deberse tanto a los detalles de la implementación como a la paralelización de algunas tareas que efectúa automáticamente el compilador. En el caso del método directo existe un factor adicional para explicar la desviación respecto del valor ideal de 4. Como se detalla en la sección 4.1.4 de Materiales y métodos, el coste computacional de este método para calcular el Hessiano escala de forma proporcional a  $\sim N^2 n^2$ , donde  $n$  es el número de átomos considerado en el modelo. En la práctica, el factor  $n^2$ , que proviene del número total de interacciones, es menor debido a que el número de átomos adyacentes con los que interacciona cada átomo (aquellos a menos de 10Å) es independiente del tamaño de la macromolécula. En consecuencia, el coste computacional efectivo debe escalar de forma proporcional a  $\sim N^2 n$ , lo que justifica aproximadamente la desviación observada. En resumen, puede concluirse que nuestra implementación del método rápido de Go permite reducir el número de operaciones respecto del método directo lo suficiente como para obtener las matrices del sistema en tiempos proporcionales a  $N^2$ .

En cuanto a la fase de diagonalización, los requerimientos computacionales son proporcionales a  $N^2$  en memoria y a  $N^3$  en coste computacional. Por tanto, una vez reducido el coste computacional a  $\sim N^2$  para el cálculo de las matrices  $\mathbf{K}$  y  $\mathbf{H}$ , el cuello de botella del análisis vibracional debe corresponder únicamente a la fase de diagonalización, *i.e.* la obtención de los modos. Para comprobar que esta fase es efectivamente el cuello de botella, en la Figura 29 se ha ilustrado la representación logarítmica del tiempo dedicado a la diagonalización en función de  $N$ . A partir de la regresión lineal correspondiente se ha determinado que el factor de escala del coste

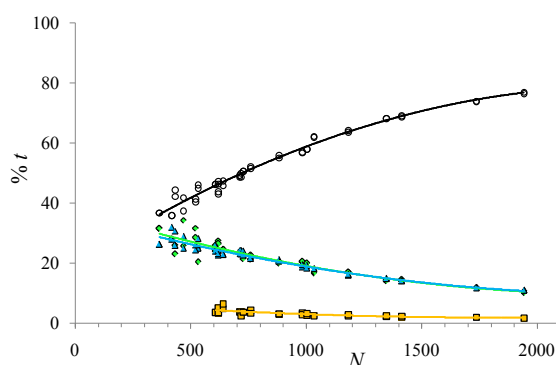


**Figura 29.** Coste computacional de la fase de diagonalización. Se muestra la representación logarítmica del tiempo de cálculo ( $t$ , [s]) empleado en diagonalizar las matrices del sistema en función del número de variables ( $N$ ). También se muestra la regresión lineal con el coeficiente de correlación  $R^2$ . La diagonalización se ha llevado a cabo utilizando la rutina DSPGVX de la librería estándar LAPACK.

la diagonalización siempre consume más tiempo que las demás fases del NMA. Por ejemplo, en la proteína de mayor tamaño, la diagonalización llega a representar el 77% del total, mientras que las fracciones correspondientes al Hessiano, la matriz de energía cinética y las tareas restantes únicamente alcanzan el 11%, 2%, y 10%, respectivamente. Además, se observa claramente que la diagonalización constituye la mayor parte del análisis vibracional incluso para las proteínas más pequeñas (<300 aminoácidos).

Con el fin de comparar el consumo de memoria del NMA al utilizar los métodos directo y rápido se ha repetido el ensayo anterior empleando el programa de análisis de memoria *valgrind* (<http://valgrind.org>). Al igual que para determinar el coste computacional, es posible asumir que el consumo de memoria ( $m$ ) varía según una función exponencial del número de variables ( $m = k \cdot N^h$ ). De este modo, la regresión lineal de la representación logarítmica correspondiente ( $\log m = h \cdot \log N + \log k$ ) también es válida para caracterizar el consumo de memoria.

computacional de la fase de diagonalización en nuestra implementación es 2.9, muy próximo al valor ideal de 3. En consecuencia, para valores muy altos de  $N$  el paso limitante será siempre la diagonalización de las matrices del sistema. En la Figura 30 se han representado, en función del número de variables, las fracciones de tiempo dedicadas a las distintas fases del NMA, *i.e.* diagonalización, cálculo de las matrices y resto de tareas. En ella, puede apreciarse que la fracción de tiempo dedicada a la diagonalización aumenta progresivamente con el número de variables, mientras que la fracción dedicada a la obtención de matrices y a las demás tareas tiende a disminuir. Además,



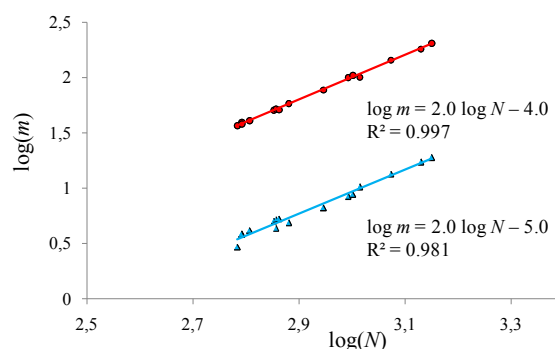
**Figura 30.** Porcentajes de tiempo requeridos para efectuar las distintas tareas del NMA. Se han representado los porcentajes del tiempo total (%t) que se requieren para llevar a cabo las siguientes tareas del análisis vibracional en función de las  $N$  variables del sistema: cálculo del Hessiano (triángulos azules), cálculo del tensor métrico (cuadrados amarillos), diagonalización (círculos blancos), y otras tareas (rombos verdes).

Aunque en este caso es necesario restar a la memoria total, la memoria inicial utilizada por el programa cuando  $N$  es  $\sim 0$ . En la Figura 31 se han ilustrado las representaciones logarítmicas y los ajustes lineales correspondientes a la cantidad de memoria total consumida por *iMode* cuando utiliza los métodos directo (rojo) o rápido (azul) para calcular las matrices del sistema. En estos casos tampoco se han incluido los valores correspondientes a las proteínas de menos de 300 aminoácidos. Las representaciones obtenidas para los dos métodos se ajustan muy bien a líneas rectas alcanzando coeficientes  $R^2$  superiores a 0.98. De las pendientes de estas rectas se obtienen

exponentes ( $h$ ) de 2.0, por lo que el consumo total de memoria del análisis vibracional escala cuadráticamente en ambas aproximaciones. Sin embargo, el método rápido requiere mucha menos memoria que el directo. Para determinar el número de veces que resulta más económico un método sobre el otro, basta con observar la distancia vertical entre ambas rectas. Como evidencian las ordenadas en el origen, la distancia es de  $\sim 1$  unidad en la representación logarítmica. Por tanto, nuestra implementación del método rápido siempre requiere  $\sim 10$  veces menos memoria que la del método directo.

### 5.1.2. Eficiencia del análisis vibracional en coordenadas internas y cartesianas

Como se muestra en la sección anterior, el número de grados de libertad considerado durante el análisis vibracional influye directamente tanto en el coste computacional como en el consumo de memoria, condicionando la aplicabilidad de los métodos en los sistemas de gran tamaño. En CC, el número de variables de una macromolécula es  $3n$ , donde  $n$  es el número de átomos. Sin embargo, en CI se han considerado dos ángulos diedros por residuo ( $r$ ) para proteínas ( $\varphi$  y  $\psi$ ) y cinco para ácidos nucleicos ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\varepsilon$  y  $\zeta$ ). La eficiencia relativa del análisis vibracional al utilizar CI o CC varía considerablemente en función del modelo atómico empleado. En proteínas, con el modelo  $C\alpha$  únicamente se tiene en cuenta un átomo por aminoácido, con el  $C5$  este número siempre es menor o igual a cinco, y con HA es de unos 8 (en promedio). En el modelo HA de ácidos nucleicos el promedio del número de átomos por nucleótido asciende a  $\sim 21$ . Según estas estimaciones, en CC el número de variables necesario para describir los



**Figura 31.** Consumo de memoria del análisis vibracional. Se muestra la representación logarítmica del consumo de memoria ( $m$ , [MB]) en función del número de variables ( $N$ ). Este consumo se ha obtenido restandole a la memoria total utilizada en cada NMA la memoria correspondiente a la ejecución del programa con un péptido pequeño ( $N \approx 0$ ). También se incluyen las regresiones lineales correspondientes y sus valores de  $R^2$ .

modelos de proteínas es  $3r$ ,  $15r$  y  $24r$  para los modelos  $C\alpha$ ,  $C5$  y  $HA$ , respectivamente, mientras que en ácidos nucleicos alcanza  $\sim 63r$ . Sin embargo, el número de CI en proteínas y ácidos nucleicos es bastante menor, *i.e.*  $2r$  y  $5r$ , respectivamente. Según estos cálculos, en proteínas, es posible reducir el número de variables un factor de  $3/2$  con el modelo atómico  $C\alpha$ ,  $15/2$  con  $C5$  y hasta 12 veces con  $HA$ . En ácidos nucleicos, la disminución puede ser de hasta 13 veces. Estas reducciones se traducen en grandes ahorros en el consumo de recursos computacionales.

Como el consumo de memoria escala cuadráticamente con el número de variables, es posible reducir los requisitos de memoria unas 2, 50 y 150 veces con los modelos  $C\alpha$ ,  $C5$  y  $HA$  de proteínas, respectivamente. En ácidos nucleicos la reducción es de unas 170 veces debido a la mayor proporción de átomos por coordenada que presentan los nucleótidos. Las ventajas del empleo de CI se hacen aún más evidentes en cuanto al tiempo de cálculo ya que el coste computacional total del análisis vibracional escala de forma proporcional a  $N^3$ . Si consideramos estas disminuciones de forma conjunta, los tiempos de cálculo deben reducirse unas 3, 400 y 1700 veces para los modelos de proteínas mencionados anteriormente. En ácidos nucleicos la reducción debe alcanzar las 2200 veces.

La comprobación experimental de estas mejoras se ha efectuado mediante el análisis vibracional en CI y CC de todas las estructuras del conjunto de pruebas de *MolMovDB* empleando el modelo atómico de mayor resolución ( $HA$ ). En la Tabla 5 se muestran los tiempos totales de ejecución correspondientes a estos casos. Al comparar los tiempos obtenidos en CC con los correspondientes a CI, las mejoras son evidentes. Las aceleraciones

**Tabla 5.** Comparación entre los tiempos de cálculo del NMA en CI y CC empleando estructuras de proteínas.

PDB <sup>a</sup>	N <sup>b</sup>	n <sup>c</sup>	Internas		Cartesianas	
			DoF <sup>d</sup>	t <sup>e</sup> [s]	DoF	t [s]
1ex6	186	1446	364	0.2	4338	45.5
4ake	214	1656	417	0.3	4968	63.2
1ggg	220	1714	433	0.3	5142	72.7
2lao	238	1833	468	0.3	5499	91.4
1urp	271	2002	532	0.4	6006	116.4
1ram	273	2176	521	0.4	6528	142.6
5at1	310	2415	607	0.5	7245	195.6
1ckm	317	2561	619	0.6	7362	205.1
3dap	317	2454	619	0.6	7683	232.0
1bp5	329	2550	640	0.6	7650	234.0
1jql	366	2837	711	0.8	8511	331.6
1omp	370	2877	718	0.8	8631	326.7
8adh	374	2785	727	0.8	8355	299.8
9aat	387	3048	759	0.9	9144	378.2
1bnc	452	3480	882	1.3	10440	555.5
1dpe	507	4048	982	1.7	12144	892.5
1rkm	517	4165	1002	1.8	12495	982.1
1sx4	524	3842	1033	1.9	11526	790.9
1i7d	609	4846	1182	2.8	14538	1528.9
1lfg	691	5332	1346	3.9	15996	2002.2
1oao	728	5739	1412	4.5	17217	2571.9
1ih7	890	7259	1736	8.1	21777	5352.9*
1su4	994	7671	1942	10.7	23013	6194.3*

<sup>a</sup> Código PDB. El NMA de los 23 pares de estructuras del conjunto de pruebas de proteínas se ha efectuado tanto con *iMode* (CI) como con *DFprot* (CC) utilizando sus parámetros por defecto y el modelo atómico  $HA$ .

<sup>b-d</sup> Número de residuos (<sup>b</sup>), átomos (<sup>c</sup>) y coordenadas (<sup>d</sup>).

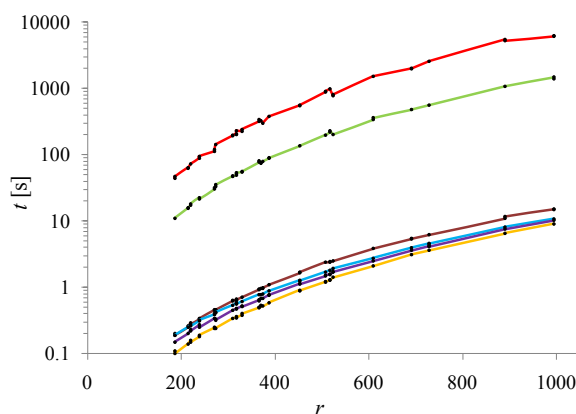
<sup>e</sup> Tiempo total de ejecución. Aunque el código del PDB mostrado es el de la estructura abierta, los valores mostrados corresponden al promedio de cada par.

\* Estos casos han sido obtenidos en 64-bit.

aumentan desde las  $\sim 200$  veces en los casos de menor tamaño hasta las  $\sim 600$  veces en los más grandes. Estas aceleraciones resultan ligeramente inferiores a las esperadas para tiempos de cálculo proporcionales a  $N^3$ . Esto se debe a que los tamaños de las estructuras de prueba no son lo suficientemente grandes como para conseguir que el coste computacional del cálculo de las matrices del sistema ( $O[N^2]$ ), sea despreciable en comparación con el coste de la diagonalización ( $O[N^3]$ ). A pesar de esto, se estima que los incrementos de velocidad serán mayores con macromoléculas más grandes, pues como se muestra en la Figura 30 la diagonalización tiende a ocupar todo el tiempo de ejecución cuando  $N$  es suficientemente grande.

Para determinar las mejoras relativas que se obtienen al utilizar los distintos modelos atómicos, en la Figura 32 se han representado los tiempos totales de cálculo en función del número de residuos. Las líneas de color rojo, verde y marrón corresponden a los tiempos obtenidos en CC para los modelos HA, C5 y Ca, respectivamente, y siguiendo el mismo orden, las líneas de color azul, violeta y amarillo representan los tiempos correspondientes obtenidos en CI. Nótese que se ha empleado una escala logarítmica en el eje de ordenadas. En dicha figura puede apreciarse claramente que el NMA en CI es más rápido que en CC con cualquiera de los modelos atómicos ensayados. Las diferencias de velocidad entre CI y CC cobran mayor importancia en los modelos con mayor número de átomos, *i.e.* C5 y HA, dónde se alcanzan incrementos de  $\sim 150$  y  $\sim 600$  veces, respectivamente. Con el modelo Ca, en CI se alcanza aproximadamente el doble de velocidad que con su homólogo cartesiano.

La comparación experimental del consumo de memoria entre el NMA en CI y el NMA en CC se han llevado a cabo con los distintos modelos atómicos. Para monitorizar el gasto de memoria se ha utilizado el programa *valgrind*. En la Figura 33 se muestra la cantidad de memoria necesaria para realizar el análisis vibracional del conjunto de pruebas de proteínas en función del número de residuos. No se han incluido los dos casos de mayor tamaño porque *valgrind* ralentiza enormemente los cálculos de forma que cada uno puede llegar a tardar más de 1 día. Al igual que en la Figura 32, las líneas de color rojo, verde y marrón corresponden a los resultados

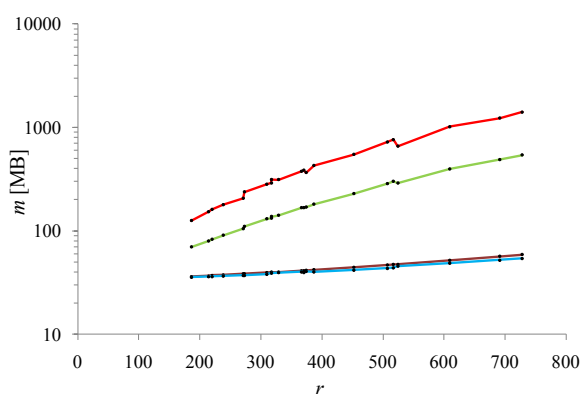


**Figura 32.** Comparación de los tiempos de cálculo del NMA en CI y CC. Se han representado, en función del número de residuos ( $r$ ), los tiempos de cálculo totales ( $t$ ) necesarios para obtener los modos tanto en CC con los modelos HA (rojo), C5 (verde) y Ca (marrón), como en CI con los modelos HA (azul), C5 (morado) y Ca (amarillo). Por claridad, se ha utilizado la escala logarítmica en el eje de ordenadas.



obtenidos en CC para los modelos HA, C5 y C $\alpha$ , respectivamente. El gasto de memoria en CI se ha representado únicamente para HA (azul) porque es prácticamente independiente del modelo atómico ( $\approx N$ ). Como puede apreciarse en la Figura 33, el NMA en CC también es el más costoso en términos de memoria, especialmente al emplear los modelos C5 y HA. Por ejemplo, en el caso más grande, los modelos atómicos C5 y HA en CC consumen 10 y 30 veces más memoria, respectivamente, que los modelos correspondientes en CI. Además, estas diferencias tienden a aumentar con el tamaño del sistema. Por otra parte, aunque el modelo C $\alpha$  cartesiano presenta un

consumo bastante parecido al obtenido con cualquiera de los modelos atómicos en CI, aún puede distinguirse que requiere más memoria ( $\sim 10\%$ ) y que ésta diferencia también tiende a aumentar con el tamaño del sistema. Nótese que los resultados del NMA en CC con los dos casos más grandes (890 y 994 aminoácidos) y el modelo atómico HA se han obtenido en 64-bit por requerir más de los 2GB de memoria disponibles para 32-bit.



**Figura 33.** Comparación del consumo de memoria del NMA en CI y CC. Se ha representado el consumo de memoria ( $m$ ) en función del número de residuos ( $r$ ) durante el NMA en CC utilizando los modelos atómicos HA (rojo), C5 (verde) y C $\alpha$  (marrón). En CI el consumo de memoria con cualquiera de los modelos es equivalente, por lo que únicamente se muestra para HA (azul). Nótese que el eje de ordenadas emplea escala logarítmica.

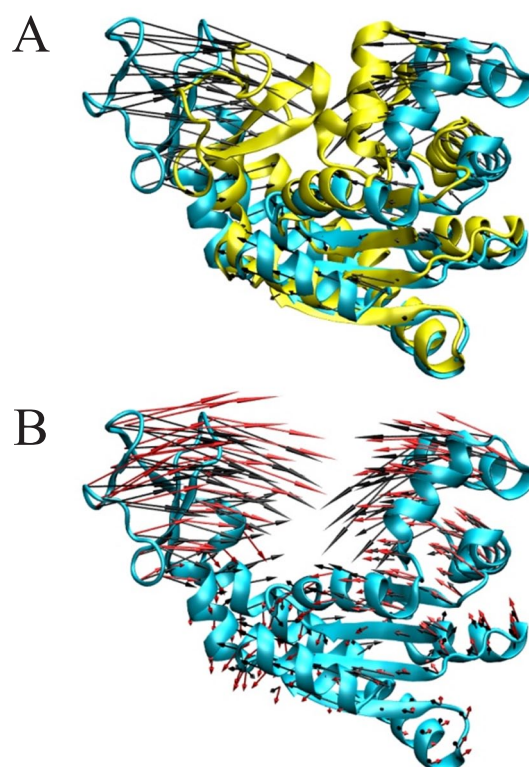
## 5.2. Validación del análisis vibracional en coordenadas internas

Con el fin de validar el análisis vibracional en CI se han comparado los 23 movimientos experimentales extraídos de *MolMovDB* (véase la sección 4.6.1 de Materiales y métodos) con los modos normales obtenidos a partir de cada uno de los confórmeros. Los modos se han obtenido tanto en CI como en CC utilizando los programas *iMode* y *DFprot*, respectivamente. Por simplicidad, el análisis se ha restringido únicamente al modelo atómico C $\alpha$  y se ha empleado el potencial optimizado frente a un conjunto representativo de trayectorias de MD obtenidas con campos de fuerza detallados. Previamente, hemos comprobado que este potencial se comporta mejor que otros al compararlo con datos de ED, RMN y transiciones conformacionales (Orellana, *et al.*, 2010). Además, facilita la comparación pues no requiere ajustar ningún parámetro, como por ejemplo, el umbral de distancia por debajo del cual tener en cuenta las interacciones o el factor de escala de las constantes de fuerza. En el análisis vibracional en CI se tendrán en cuenta únicamente los ángulos diedros  $\varphi$  y  $\psi$  como variables.



La comparativa entre los movimientos observados experimentalmente y los modos normales de vibración se ha llevado a cabo empleando las métricas  $\alpha$  y  $\delta$  descritas en la sección 4.5.1 de Materiales y métodos. Éstas se basan en el producto escalar entre los modos y el vector de transición obtenido directamente del movimiento. Para obtener cada vector, el par de conformaciones de cada movimiento se alinea previamente entre sí (véase la Figura 34A) minimizando su RMSD (Kabsch, 1978). La métrica  $\alpha$  mide el solapamiento de cada uno de los modos con el vector de transición correspondiente y la métrica  $\delta$ , o solapamiento acumulado, evalúa la eficacia con la que un conjunto de modos es capaz de describir dicho vector. En la Tabla 6 puede apreciarse que el solapamiento promedio del modo que mejor describe cada movimiento ( $\alpha_1$ ) es 0.70, lo que indica una excelente coincidencia con los movimientos

experimentales. El solapamiento acumulado de los tres primeros modos también es excelente ( $\delta_3=0.77$ ). Si se incluye hasta el 5° o 10° modo, los valores de  $\delta_5$  y  $\delta_{10}$  aumentan hasta 0.83 y 0.90, respectivamente. Estos elevados solapamientos indican que solo son necesarios unos pocos modos de baja frecuencia para describir la mayor parte de los movimientos experimentales. Por ejemplo, solo hacen falta unos 10 modos para describir el 90% de estos movimientos. Además, el modo de mayor solapamiento con el vector de transición normalmente corresponde al primer o segundo modo de menor energía ( $N\alpha_1=1.7$ ). Los valores individuales se encuentran detallados en la Tabla A1 (véase el Apéndice). En la Figura 34B puede comprobarse el gran parecido entre el vector de transición (flechas negras) y el primer modo obtenido en CI (flechas rojas) a partir de la estructura abierta de un caso representativo. Estos resultados corroboran la buena correlación entre los modos de baja frecuencia y los movimientos biológicamente relevantes, coincidiendo además con los solapamientos obtenidos por el grupo de Gerstein a partir de proteínas de la base de datos de movimientos moleculares *MolMovDB* (Krebs, *et al.*, 2002).



**Figura 34.** Comparación visual del vector de transición con el primer modo normal en CI. Las estructuras abierta (*4ake*) y cerrada (*1ake*) de la adenilato quinasa se muestran en azul y amarillo, respectivamente. El vector de transición (flechas negras) y el primer modo (flechas rojas) se han obtenido a partir del modelo atómico Ca.

**Tabla 6.** Solapamientos de los vectores de transición de proteínas con los modos normales obtenidos en CI y CC

	$\alpha_1^b$	$\alpha_2$	$\alpha_3$	$\delta_3^c$	$\delta_5$	$\delta_{10}$	$N\alpha_1^d$	$N\sigma_{90\%}^e$	$\gamma_3^f$	$\gamma_{10}$	$\gamma_{50}$
Promedio CI <sup>a</sup>	0.70	0.34	0.25	0.78	0.84	0.89	1.7	107	0.98	0.93	0.90
Abierto a cerrado	0.77	0.30	0.23	0.86	0.89	0.92	1.3	90	0.98	0.94	0.90
Cerrado a abierto	0.63	0.38	0.28	0.71	0.80	0.86	2.2	125	0.97	0.93	0.89
Promedio CC	0.70	0.34	0.24	0.77	0.83	0.89	1.8	155	1.00	1.00	1.00

<sup>a</sup> El NMA en CI y en CC se ha llevado a cabo con los programas *iMode* y *DFprot* (Garzon, *et al.*, 2007), respectivamente. En ambos casos se ha empleado el modelo atómico *C $\alpha$*  con el potencial refinado con datos de ED (Orellana, *et al.*, 2010). Los promedios incluyen todas las transiciones de proteínas de la Tabla 3.

<sup>b</sup> Solapamiento entre los vectores de transición  $\Delta r$  y el 1<sup>er</sup>, 2<sup>o</sup> y 3<sup>er</sup> modos que más solapan ( $v_n$ ). El solapamiento se calcula como:  $\alpha_{1,2,3} = |\Delta r \cdot v_n| / [\|\Delta r\| \|v_n\|]$

<sup>c</sup> Solapamiento acumulado de los 3, 5 y 10 modos de menor energía:  $\delta_n = (\sum_{k=1}^n \alpha_k^2)^{1/2}$

<sup>d</sup> Rango del modo con mayor solapamiento.

<sup>e</sup> Número de modos necesario para justificar el 90% de la varianza.

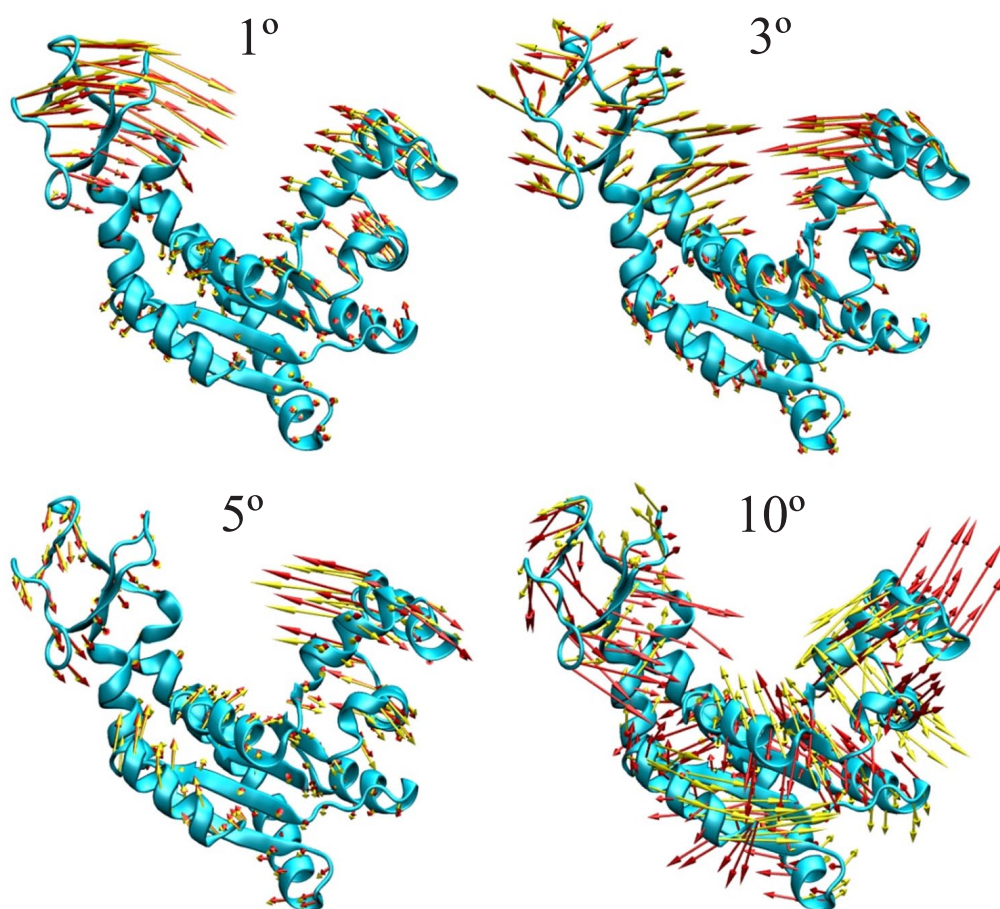
<sup>f</sup> El solapamiento entre los espacios conformacionales  $u$  y  $v$  se ha calculado según (Noy, *et al.*, 2006):

$\gamma_n = 1/n \sum_{i,j}^n (u_i \cdot v_j)^2$  donde  $n$  es el número de modos considerado.

No obstante, se observan promedios ligeramente diferentes cuando el NMA se lleva a cabo a partir de las conformaciones abiertas o cerradas. Al comparar los solapamientos de los modos procedentes de estructuras abiertas con los correspondientes a las estructuras cerradas, el solapamiento del modo que mejor describe el movimiento ( $\alpha_1$ ) se reduce de 0.77 a 0.63 (~8%) y el solapamiento acumulado de los 10 primeros modos disminuye de 0.92 a 0.86 (~7%). Esto indica que el NMA se comporta mejor al emplear estructuras abiertas, hecho que ha sido observado anteriormente por otros autores (Tama y Sanejouand, 2001). Sin embargo, los valores del solapamiento acumulado ( $\delta$ ) obtenidos con las estructuras cerradas siguen siendo lo suficientemente elevados como para describir aceptablemente los vectores de transición. De hecho, tan solo los 5 primeros modos obtenidos a partir de estas estructuras son capaces de representar el 80% del cambio conformacional. Además, aunque los valores de  $N\alpha_1$  al utilizar las estructuras cerradas (2.2) suelen ser algo mayores que con las estructuras abiertas (1.3), en casi todos los casos el modo con mayor solapamiento es uno de los tres primeros (véase la Tabla A1).

Con fines comparativos, en la Tabla 6 también se han incluido los solapamientos obtenidos a partir de los modos en CC. Como se muestra en esta tabla, en CC los niveles de solapamiento acumulado ( $\delta$ ) y de los modos que mejor representan cada movimiento ( $\alpha$ ) son prácticamente idénticos a los obtenidos en CI. Este hecho indica que los primeros modos de baja frecuencia procedentes de ambas aproximaciones son capaces de describir los movimientos experimentales de manera muy similar. También se ha incluido la métrica  $\gamma$  (véase la sección 4.5.1 de Materiales y métodos) para comparar el solapamiento entre los espacios de deformación correspondientes a

los modos normales obtenidos en CC y en CI. Los valores  $\gamma$  obtenidos al emplear los 3 primeros modos son muy próximos a la unidad ( $\gamma_3=0.98$ ), revelando que estos 3 modos representan direcciones de movimiento muy similares independientemente del sistema de coordenadas utilizado. Conforme se van incluyendo más modos, la divergencia entre ambos espacios aumenta progresivamente ( $\gamma_{10}=0.93$ ), aunque todavía el solapamiento entre los 50 primeros ( $\gamma_{50}$ ) sigue siendo muy alto alcanzando valores próximos al 90%. El elevado solapamiento entre los espacios de deformación obtenidos en CC y CI prueba que los modos de baja frecuencia corresponden principalmente a movimientos de los ángulos diedros. Estos resultados están en consonancia con lo observado por el grupo de Go (Kitao, *et al.*, 1994) al comparar los modos cartesianos de una proteína globular pequeña con los obtenidos considerando únicamente los ángulos diedros como CI del sistema. La divergencia observada entre los modos obtenidos en ambos sistemas de coordenadas al ir añadiendo modos de mayor frecuencia, probablemente sea debida a que las fluctuaciones en las longitudes y ángulos de enlace no están contempladas en nuestro modelo de CI. En la Figura 35 se han representado algunos de los primeros modos



**Figura 35.** Comparación visual de los primeros modos normales de una proteína en CI y CC. Se han representado los modos normales de la forma abierta de la adenilato quinasa (*4ake*) obtenidos en CI (flechas rojas) y en CC (flechas amarillas) utilizando el modelo atómico Ca.

obtenidos en CI (rojo) y CC (amarillo) de un caso representativo para ilustrar visualmente su gran parecido. Como puede apreciarse, los modos del 1 al ~5 son prácticamente idénticos, mientras que hacia el ~10 se hace evidente la divergencia. Por otro lado, se han detectado diferencias en el número de modos necesario para justificar el mismo porcentaje de varianza (Tabla 6). En promedio, en CI se requiere un 30% menos modos que en CC para expresar el 90% de la varianza. Estos resultados sugieren que el espacio conformacional descrito por los modos obtenidos en CI es más compacto que su homólogo cartesiano.

### 5.3. Validación de las representaciones reduccionistas

#### 5.3.1. Validación de los modelos atómicos

Para validar los modelos atómicos se ha calculado el solapamiento entre los vectores de transición del conjunto de pruebas de proteínas y los modos normales obtenidos a partir de los modelos atómicos propuestos: C $\alpha$ , C5 y HA. Por simplicidad, en este análisis vibracional se ha empleado la función sigmoidea para definir la energía potencial (véase Materiales y métodos). En la Tabla 7 se muestran los promedios de los solapamientos  $\alpha$  y  $\delta$  entre los vectores de transición y los modos procedentes de los distintos modelos atómicos. Los valores de solapamiento de los modos que mejor representan el movimiento ( $\alpha$ ) y del solapamiento acumulado ( $\delta$ ) son muy altos. Por ejemplo,  $\alpha_1$  está comprendido entre 0.68 y 0.70, y  $\delta_{10}$  entre 0.86 y 0.88. Esto nos indica que los primeros modos describen muy bien los cambios conformacionales independientemente del modelo atómico utilizado. La robustez de estos modos frente a cambios en el detalle de las representaciones ya ha sido observada anteriormente. Lu y Ma (Lu y Ma, 2005) comprobaron

**Tabla 7.** Solapamientos de los vectores de transición de proteínas y ARN con los modos normales de representaciones reduccionistas.

Modelo <sup>a</sup>	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\delta_3$	$\delta_5$	$\delta_{10}$	$N\alpha_1$	$N\sigma_{90\%}$	$\gamma_3^b$	$\gamma_5$	$\gamma_{10}$
C $\alpha$	0.70	0.34	0.23	0.77	0.83	0.88	1.7	118	1.00	1.00	1.00
C5	0.68	0.33	0.23	0.75	0.81	0.86	1.7	226	0.94	0.89	0.87
HA	0.70	0.33	0.22	0.76	0.82	0.87	1.8	368	0.91	0.87	0.85
C $\alpha$ -50%	0.69	0.37	0.24	0.78	0.84	0.89	1.6	85	0.98	0.95	0.94
C $\alpha$ -90%	0.63	0.40	0.25	0.74	0.80	0.88	2.0	30	0.75	0.69	0.66
ARN	0.66	0.38	0.25	0.77	0.82	0.86	1.6	45	-	-	-

<sup>a</sup> Todos los cálculos se han llevado a cabo como en la Tabla 5, pero empleando la función sigmoidea como potencial y las siguientes representaciones reduccionistas (véase Materiales y métodos): C $\alpha$ , un átomo C $\alpha$  por residuo; C5, 3 átomos para el esqueleto peptídico y 2 para la cadena lateral; C $\alpha$ -50%, como C $\alpha$  pero fijando aleatoriamente el 50% de los ángulos diedros; C $\alpha$ -90%, lo mismo pero fijando el 90%; y HA, considerando todos los átomos pesados. En ARN solamente se ha usado HA.

<sup>b</sup> Todos los solapamientos  $\gamma$  se han obtenido utilizando los átomos C $\alpha$ . Para ello, se han generado vectores comparables a partir de los modelos C5 y HA diagonalizando las correspondientes matrices de covarianza (véase Materiales y métodos).



que los modos de menor frecuencia permanecen prácticamente intactos ante cambios en la resolución del modelo y perturbaciones aleatorias en la fortaleza de los muelles siempre y cuando no se altere significativamente la topología de las interacciones. Incluso es posible conseguir modos de baja frecuencia similares a los obtenidos con modelos de resolución atómica partiendo de mapas de densidad a media/baja resolución (Chacon, *et al.*, 2003; Ming, *et al.*, 2002; Tama, *et al.*, 2002).

Para extender el análisis y comparar directamente entre sí los modos obtenidos a partir de las distintas representaciones atómicas se ha empleado la métrica de solapamiento entre espacios vectoriales  $\gamma$ . Como el número y la naturaleza de los átomos (o pseudo-átomos) de los distintos modelos atómicos es diferente, los modos no pueden ser comparados directamente. Por este motivo, la comparación se lleva a cabo con un conjunto de vectores ortogonales obtenidos a partir de la matriz de covarianza de los átomos  $C_\alpha$  tal y como se describe en la sección 4.5.3 de Materiales y métodos. En la Tabla 7 se muestran los promedios de los factores  $\gamma$  obtenidos entre los 3, 5 y 10 primeros modos del modelo  $C_\alpha$  y los correspondientes a los tres modelos atómicos. A pesar del gran parecido entre los vectores de transición y los modos obtenidos con los distintos modelos atómicos (elevados valores de solapamientos  $\alpha$  y  $\delta$ ), los espacios de deformación divergen significativamente a medida que se consideran más modos. Las diferencias con los modos de referencia (obtenidos con el modelo  $C_\alpha$ ) van aumentando progresivamente según se incluyen más átomos, alcanzándose factores  $\gamma_{10}$  de 0.87 y 0.85 para los modelos C5 y HA, respectivamente. Nótese que, aunque el número de átomos considerado en el modelo no altera la forma global de la macromolécula, la inclusión de más átomos provoca la aparición de nuevas interacciones que alteran ligeramente la dirección de los modos. La divergencia entre los modos obtenidos con los modelos más detallados y los correspondientes a  $C_\alpha$  también aumenta progresivamente al incluir en la comparación modos de mayor frecuencia. Por ejemplo, al comparar el modelo  $C_\alpha$  con HA se obtienen valores de  $\gamma$  de 0.91, 0.87 y 0.85 al incluir los 3, 5 y 10 primeros modos, respectivamente. Lu y Ma (Lu y Ma, 2005) también observaron que las diferencias entre los modos procedentes de representaciones reduccionistas basadas en  $C_\alpha$  y los obtenidos a partir de un potencial atómico detallado se acentúan en los modos de mayor frecuencia. Estos resultados evidencian la robustez de los primeros modos normales ante la utilización de distintos modelos atómicos.

### 5.3.2. Validación de la reducción del número de coordenadas internas

La versatilidad de *iMode* ha permitido estudiar cómo se ven afectados los modos normales de vibración ante la reducción del número de grados de libertad. Para ello, se han obtenido los modos fijando aleatoriamente diferentes porcentajes de los ángulos diedros  $\phi$  y  $\psi$ .

Por simplicidad, se ha empleado el modelo atómico Ca con los parámetros por defecto. En la Tabla 7 se muestran los valores promedio del solapamiento entre espacios vectoriales ( $\gamma$ ) obtenidos fijando y sin fijar. Fijando el 50%, los promedios de  $\gamma$  se mantienen muy próximos a 1, tomando valores de 0.98, 0.95 y 0.94 para los 3, 5 y 10 primeros modos, respectivamente. Esto indica que al eliminar aleatoriamente la mitad de las variables no se producen diferencias significativas en el espacio conformacional descrito por estos modos. En cambio, al prescindir del 90% de las variables, el grado de solapamiento disminuye en mayor medida, hasta alcanzar valores  $\gamma$  de 0.75, 0.69 y 0.66 para los mismos modos, respectivamente. A pesar de que la disminución del número de variables no altera la forma de la macromolécula, la eliminación de un porcentaje elevado de ellas acaba afectando a los modos. Del descenso de los valores de  $\gamma$  conforme se incluyen más modos se puede deducir que la fijación de variables afecta sobre todo a los de mayor frecuencia. Como estos modos corresponden a cambios localizados de unas pocas variables, la eliminación de alguna de ellas puede afectar drásticamente a su movimiento. Por el contrario, los modos de baja frecuencia representan el cambio concertado de un gran número de variables, *i.e.* son colectivos. De este modo, la restricción impuesta por la ausencia de alguna de las variables puede ser compensada efectivamente por las demás.

En la Tabla 7 también se incluyen los promedios de los solapamientos entre los vectores de transición y los modos obtenidos fijando aleatoriamente el 50 y 90% de los ángulos diedros. Al fijar el 50%, estos solapamientos son muy similares a los obtenidos sin fijar en todos los casos. Como se mostró anteriormente, los modos de baja frecuencia apenas se ven afectados por la eliminación aleatoria de la mitad de las variables. Aunque los espacios vectoriales correspondientes a estos modos divergen significativamente entre sí ( $\gamma_{10} = 0.66$ ), el vector de transición puede ser representado de igual manera por ambos. Nótese que fijando el 90% de los ángulos  $\varphi$  y  $\psi$ , el solapamiento acumulado los 10 primeros modos con el vector de transición es idéntico al obtenido sin fijar ( $\delta_{10} = 0.88$ ). No obstante, para evitar el deterioro excesivo de los modos de menor frecuencia, no se recomiendan porcentajes de fijación superiores al 90-95%. Con todo esto es posible concluir que el espacio de deformación descrito por estos modos también resulta robusto ante la fijación aleatoria de los ángulos diedros.

#### 5.4. Validación del análisis vibracional en ARN

Aunque son varios los estudios que contemplan el NMA de ácidos nucleicos (Feig y Burton, 2010; Fulle y Gohlke, 2010; Orozco, *et al.*, 2008; Van Wynsberghe y Cui, 2005; Yang, *et al.*, 2006), esta es la primera validación sistemática del NMA de ARN con un conjunto representativo de 11 transiciones conformacionales (véase Materiales y métodos).

La validación se ha realizado comparando los solapamientos de los vectores de transición con los modos normales en CI empleando el modelo atómico HA para ácidos nucleicos y los parámetros por defecto. En la Tabla 7 se han incluido los valores promedio de los solapamientos  $\alpha$  y  $\delta$  obtenidos con los modos de las estructuras de ARN. El solapamiento de los modos que mejor describen el movimiento también es elevado en ARN. No obstante, aunque el valor promedio de  $\alpha_1$  es un poco más bajo (0.63) en comparación con el de proteínas (0.70), los promedios de  $\alpha_2$  y  $\alpha_3$  son ligeramente superiores, por lo que compensan efectivamente el valor de  $\alpha_1$ . El solapamiento acumulado también es muy parecido, con valores de  $\delta_{10}$  de 0.86. Como el rango promedio del modo con mayor solapamiento ( $N\alpha_1=1.6$ ) es tan bajo como en proteínas (1.8), generalmente uno de los 2 o 3 primeros modos es el que mejor representa los cambios conformacionales del ARN. Los resultados detallados correspondientes a estos promedios se muestran en la Tabla A2 (véase el Apéndice).

### 5.5. Validación del análisis vibracional en sistemas muy grandes

Nuestra implementación de NMA resulta muy rápida incluso en los sistemas de mayor tamaño. Por ejemplo, *iMode* puede completar el análisis vibracional de una proteína de ~1000 aminoácidos en un PC estándar de 32-bit empleando el modelo atómico de máxima resolución (HA) en tan solo 11 segundos (véase la Tabla 5). Entre todas las estructuras de los conjuntos de pruebas, este tiempo tan solo se ve superado en los dos casos más grandes, dos ARNr de 721 y 1529 nucleótidos que tardan 1 y 9 minutos, respectivamente (véase la Tabla 8). Cabe destacar que el NMA en CC de estos casos requiere tanta memoria que resulta imposible realizarlo en ordenadores corrientes sin recurrir a simplificaciones adicionales.

La eficiencia y versatilidad de *iMode* se hacen evidentes al tratar sistemas de gran tamaño en ordenadores de sobremesa. Por ejemplo, el NMA de la cápside del virus CCMV (del inglés, *Cowpea Chlorotic Mottle Virus*), que contiene 180 cadenas de proteína y 28620 aminoácidos, requiere alrededor de un día de cálculo y más de 26 GB de RAM al considerar únicamente sus ~56500 ángulos diedros  $\varphi$  y  $\psi$ . También es posible efectuar el NMA del ribosoma procariota completo, un complejo macromolecular de unos 150000 átomos, 11500 residuos y 36500 variables, en unas 9 horas utilizando unos 15 GB de RAM. Este tipo de cálculos no es posible en ordenadores convencionales que suelen tener actualmente entre 4 y 8 GB de RAM. Sin embargo, mediante la fijación de un porcentaje de los ángulos diedros se puede ajustar fácilmente el número de variables a los recursos computacionales disponibles. Por ejemplo, la fijación aleatoria del 75% de los ángulos diedros del CCMV reduce el tamaño del sistema lo suficiente (hasta ~15000 variables) como para obtener en apenas una hora la parte de baja frecuencia de



su espectro vibracional en un PC corriente de 32-bit y 4GB de RAM.

Utilizando la herramienta *iMove* se han animado dos modos característicos del CCMV. Las animaciones se muestran en la página web ([http://chaconlab.org/imod/faq.html#FAQ\\_CCMV\\_NMA](http://chaconlab.org/imod/faq.html#FAQ_CCMV_NMA)). La primera, corresponde al modo no degenerado de menor energía que ha sido propuesto para explicar la ruta de maduración de la forma nativa al estado expandido (Tama y Brooks, 2002). La segunda, ilustra el modo de menor energía del grupo de simetría icosaédrico. Como puede apreciarse en las animaciones, y a pesar de haber eliminado aleatoriamente el 75% de las variables, estos modos resultan prácticamente idénticos a los obtenidos a partir de los análisis vibracionales más detallados (van Vlijmen y Karplus, 2005).

**Tabla 8.** Comparación entre los tiempos de cálculo del NMA en CI y CC con ARN.

PDB <sup>a</sup>	N <sup>b</sup>	n <sup>c</sup>	Internas		Cartesianas	
			DoF <sup>d</sup>	t <sup>e</sup> [s]	DoF	t [s]
1a51	40	901	198	0.1	2703	12.2
2i7z	42	941	208	0.1	2823	13.9
1z2j	44	986	218	0.1	2958	15.6
1u63	48	1031	238	0.1	3093	17.6
3knj	74	1573	368	0.2	4719	61.1
3fih	75	1602	373	0.2	4806	63.7
1p5o	76	1695	378	0.2	5085	75.3
3cul	91	1941	453	0.3	5823	113.6
1mfq	126	2708	628	0.6	8124	295.6
2wdj	721	15505	3603	60.9	46515	*
3e1a	1529	32811	7641	531.9	98433	*

<sup>a</sup> Código PDB. El NMA de los 11 pares de estructuras del conjunto de pruebas de ARN se ha efectuado tanto con *iMode* (CI) como con *DFprot* (CC) utilizando sus parámetros por defecto y el modelo atómico HA.

<sup>b-d</sup> Número de residuos (<sup>b</sup>), átomos (<sup>c</sup>) y coordenadas (<sup>d</sup>).

<sup>e</sup> Tiempo total de ejecución. Aunque el código del PDB mostrado es el de la estructura abierta, los valores mostrados corresponden al promedio de cada par.

\* En estos casos no es posible realizar el NMA en CC debido al elevado número de coordenadas.

## 5.6. Simulación de trayectorias de transiciones conformacionales

La simulación de la transición estructural entre dos conformaciones conocidas de una proteína se ha llevado a cabo con éxito mediante aproximaciones basadas en distintos modelos de red elástica o técnicas de interpolación (Franklin, *et al.*, 2007b; Hollup, *et al.*, 2005; Kim, *et al.*, 2002; Krebs y Gerstein, 2000; Lindahl, *et al.*, 2006; Miyashita, *et al.*, 2003; Vonnrhein, *et al.*, 1995; Weiss y Levitt, 2009). De forma similar, en esta tesis se ha desarrollado e implementado *iMorph*, una herramienta sencilla para la generación de trayectorias factibles basada en el análisis vibracional en CI (véase la sección 4.3.5 de Materiales y métodos).

### 5.6.1. Simulación de trayectorias de transiciones conformacionales en proteínas

En primer lugar, se han empleado los 23 pares de estructuras de proteínas procedentes de la base de datos de movimientos moleculares *MolMovDB* para validar *iMorph* con datos experimentales. En la Tabla 9 se muestran los resultados obtenidos al simular todas las transiciones conformacionales posibles empleando el 10% de los modos, el modelo atómico HA y los ángulos  $\varphi$  y  $\psi$  como variables. En todos los casos, el modelo inicial se deforma

**Tabla 9.** Resultados promedio de las simulaciones de transiciones conformacionales.

Modelo <sup>a</sup>	RMSD [Å] <sup>b</sup>		Colisiones <sup>c</sup>			%R <sup>d</sup>			Molprobit <sup>e</sup>			t <sup>f</sup>
	I	F	I	P	F	I	P	F	I	P	F	
HA	7.57	0.77	21	27	35	0.6	0.7	0.8	2.7	2.8	2.9	6.9
HA-cierre	7.57	0.74	23	27	38	0.7	0.7	0.9	2.8	2.9	3.1	6.2
HA-apertura	7.57	0.80	19	26	32	0.6	0.6	0.7	2.6	2.8	2.8	7.6
HA <sup>+</sup>	7.57	0.78	21	30	40	0.6	0.7	0.8	2.7	2.9	3.1	1.5
C5	7.57	0.75	21	29	37	0.6	0.6	0.8	2.7	2.9	3.0	6.1
Cα	7.57	0.79	21	36	48	0.6	0.6	0.7	2.7	2.9	3.1	5.1
ARN	7.23	1.34	32	33	37	-	-	-	-	-	-	13.0
ARN-cierre	7.23	1.30	33	34	38	-	-	-	-	-	-	13.5
ARN-apertura	7.23	1.37	31	32	36	-	-	-	-	-	-	12.4
HA-50%	7.57	0.80	21	27	35	0.6	0.6	0.7	2.7	2.8	2.9	2.4
HA-90%	7.57	1.06	21	33	50	0.6	0.6	0.6	2.7	2.9	3.1	2.1
Cα-50%	7.57	0.86	21	34	45	0.6	0.6	0.6	2.7	2.9	3.1	1.2
Cα-90%	7.57	1.06	21	37	53	0.6	0.6	0.6	2.7	2.9	3.1	0.6
ARN-50%	7.23	1.39	32	33	37	-	-	-	-	-	-	5.1
ARN-90%	7.23	1.58	32	41	54	-	-	-	-	-	-	3.9

<sup>a</sup> Promedios resultantes de las simulaciones de transiciones conformacionales efectuadas *iMorph* para ir de la conformación abierta a la cerrada (movimiento de cierre) y *viceversa* (movimiento de apertura). En todos los casos se ha utilizado la función sigmoidea como potencial. Se muestran los valores correspondientes a los modelos iniciales (*I*), finales (*F*) y los promedios de las conformaciones intermedias (*P*). Las representaciones reduccionistas se indican de la misma manera que en la Tabla 7.

<sup>b</sup> Desviación calculada teniendo en cuenta únicamente los átomos C<sub>α</sub> o P.

<sup>c-e</sup> Métricas de validación estructural de *Molprobit* (Chen, *et al.*, 2010): número de colisiones importantes cada 1000 átomos (<sup>c</sup>), porcentaje de ángulos diedros  $\varphi$  y  $\psi$  fuera de las regiones permitidas de Ramachandran (<sup>d</sup>), y resolución cristalográfica equivalente a la cual cabría esperar aquellos valores de las métricas de *Molprobit* (<sup>e</sup>). Nótese que estas métricas han sido obtenidas después de añadir todos los átomos de hidrógeno a las estructuras correspondientes.

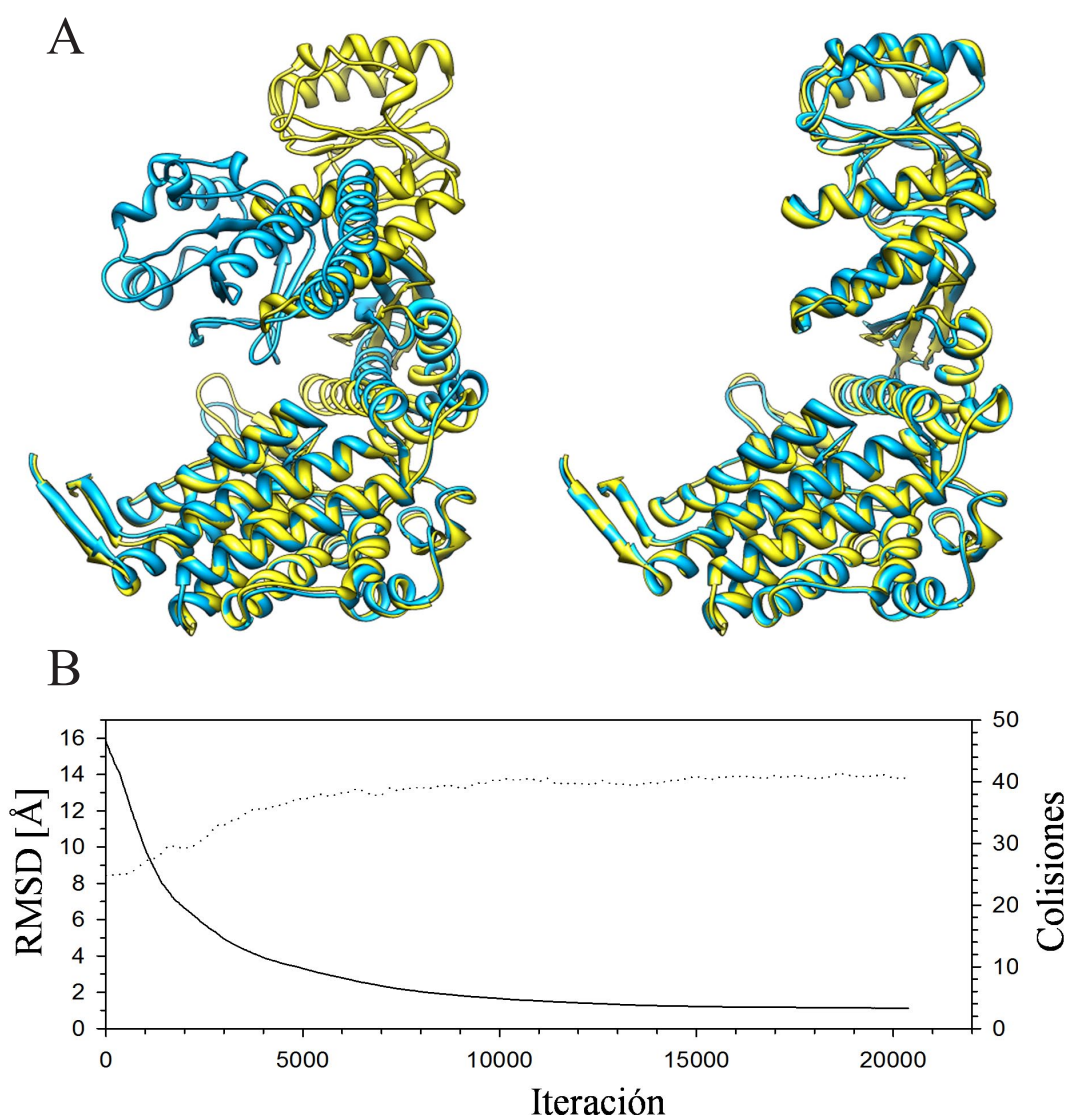
<sup>f</sup> Tiempo de ejecución medio.

\* Por eficiencia, en las simulaciones de las dos estructuras de ARN de mayor tamaño y en las de proteínas HA<sup>+</sup> el umbral para recalcular los modos se ha cambiado de 0.1 a 1Å.

rápida y gradualmente hasta alcanzar la conformación de la estructura objetivo. Como puede apreciarse en dicha tabla, las desviaciones entre la estructura objetivo y la conformación final llegan a los 0.77 Å partiendo de una desviación inicial de 7.57 Å. Al separar los resultados según correspondan a movimientos de cierre o de apertura, aparecen sutiles diferencias, *i.e.* los movimientos de cierre alcanzan desviaciones ligeramente menores (0.74 Å) que las de apertura (0.80 Å). Estas discrepancias no se consideran significativas pues al observar estos resultados individualmente las diferencias son menores de ~0.3Å en todos los casos (datos mostrados en la Tabla A3 del Apéndice).

Para evaluar la calidad de las trayectorias se ha empleado el programa de validación estructural *Molprobit* (Chen, *et al.*, 2010). Este programa requiere que las estructuras posean todos los átomos. Por este motivo se añaden previamente todos los átomos de hidrógeno a los modelos HA utilizando el comando *reduce-nobuild* de *Molprobit*. En las Tablas 9 y A3 (véase

el Apéndice) se han incluido varias métricas para valorar la calidad de las estructuras: *i*) número de choques importantes cada 1000 átomos, *ii*) porcentaje de ángulos  $\varphi$  y  $\psi$  fuera de las regiones permitidas de Ramachandran, y *iii*) resolución cristalográfica equivalente para una estructura de calidad similar. La conformación final mantiene una resolución cristalográfica equivalente a la de las estructuras iniciales, pues apenas aumenta de 2.7 a 2.9 Å. Tan solo se observan pequeños incrementos en el porcentaje de ángulos diedros fuera de las regiones permitidas de Ramachandran y en el número de colisiones, *i.e.* del 0.6 al 0.8% y de 21 a 35, respectivamente. A lo largo de las simulaciones estas métricas se mantienen en valores intermedios, *i.e.* del 0.7% en el porcentaje de ángulos fuera de las regiones permitidas y de 27 en las colisiones. Estos



**Figura 36.** Simulación de una trayectoria de transición conformacional. En el panel A se han superpuesto las conformaciones inicial (*1oel*, izquierda) y final (derecha) de GroEL, en azul, sobre la estructura abierta (*1sx4*) en amarillo. En B, se ha representado la variación del RMSD-Cα (trazo continuo) y las colisiones de *Molprobity* (trazo punteado) a lo largo de la transición. El modelo final alcanza una desviación de tan solo 1.12Å y 41 colisiones empleando el modelo HA.

resultados reflejan la excelente calidad tanto de las estructuras finales como de las estructuras intermedias e indican que los modelos generados por *iMorph* pueden ser factibles. Téngase en cuenta además que no se ha seguido ninguna estrategia adicional para evitar las distorsiones o mantener la estructura secundaria de los modelos durante la toda la simulación. En la Figura 36 se ilustra un ejemplo representativo, la simulación de la transición de la conformación cerrada (azul) del monómero de GroEL hacia su conformación abierta (amarillo). En el panel A puede apreciarse que tanto el grado de ajuste como el mantenimiento de los elementos de estructura secundaria son excelentes. En el panel B se muestra como aumenta progresivamente el número de colisiones conforme disminuye la desviación con la estructura objetivo.

Gracias a la eficiencia de las implementaciones realizadas, la obtención de los modos normales en CI y su animación es un proceso muy rápido que acelera enormemente las simulaciones. En la Tabla 9 se incluyen los promedios de los tiempos de simulación requeridos en estos casos. Con el modelo atómico de mayor resolución (HA) la duración promedio es de ~7 minutos. Las simulaciones no duran más de 20 minutos excepto en el caso de mayor tamaño y amplitud de movimiento (*Isu4*) que tarda entre una y dos horas para completar la trayectoria según del sentido de la transición (datos mostrados en la Tabla A3 del Apéndice). No obstante, es posible acelerar los cálculos fácilmente aumentando el umbral para recalcular los modos (véase la sección 4.3.5 de Materiales y métodos). En la Tabla 9 se han repetido las simulaciones incrementando este umbral desde 0.1Å (por defecto) hasta 1Å (véase el caso HA<sup>+</sup>). De esta forma, no solo se consigue acelerar los cálculos 4-5 veces, sino que además se obtienen trayectorias y modelos finales muy similares a los obtenidos con los parámetros por defecto. Las desviaciones son prácticamente idénticas (0.78Å) y el número de colisiones tan solo aumenta hasta 40 en las estructuras finales y 30 en las promedio.

En consonancia con las observaciones de la sección 5.2, estos resultados no solo demuestran que una pequeña parte del espectro de modos (~10%) es capaz de describir los movimientos colectivos de las macromoléculas, sino que además permite ir de una conformación a otra (RMSD <1Å) manteniendo la calidad de las estructuras.

### 5.6.2. Simulación de trayectorias de transiciones conformacionales con distintos modelos atómicos de proteínas

Las trayectorias de transiciones conformacionales con los restantes modelos atómicos de proteínas se han obtenido repitiendo las simulaciones de la sección anterior. Gracias a la representación en CI de *iMorph*, es posible reconstruir modelos HA directamente a partir de los modelos atómicos C5 y Cα con los que evaluar la calidad de las estructuras. Los resultados de las desviaciones y métricas de calidad estructural de *Molprobit* se muestran también en la Tabla 9.

Tanto las trayectorias como las desviaciones finales obtenidas con los modelos C5 (0.75Å) y Cα (0.79Å) son muy similares a las obtenidas con el modelo HA (0.77Å), lo que valida el empleo de estos modelos más reduccionistas. Únicamente se aprecia un incremento en el número de colisiones. Este número tan solo aumenta hasta 37 con el modelo C5 en comparación con las 35 del modelo HA, y hasta 48 con el modelo Cα. Lógicamente, los modos no son capaces de evitar colisiones entre átomos que no han sido tenidos en cuenta durante el análisis vibracional. A pesar del aumento en las colisiones, los modelos atómicos más reduccionistas (C5 y Cα) son interesantes pues permiten acelerar significativamente las simulaciones respecto del modelo HA. Como se muestra en la Tabla 9, los tiempos medios para los modelos C5 y Cα resultan un 12 y un 26% menores, respectivamente, aunque si lo que se desea es minimizar las colisiones siempre es preferible el modelo HA.

### 5.6.3. Simulación de trayectorias de transiciones conformacionales en ARN

Se han simulado las 22 transiciones conformacionales disponibles en el conjunto de pruebas de ARN empleando el modelo atómico HA. Al igual que en los casos de proteínas, el modelo inicial se deforma rápida y gradualmente hasta alcanzar la conformación de la estructura objetivo (Tabla 9). Partiendo de una desviación inicial de 7.23Å se alcanza un promedio final de 1.34Å. Aunque esta desviación es algo superior a la observada en proteínas (0.77Å), la diferencia resulta irrelevante teniendo en cuenta la distinta naturaleza de los datos. Con ARN, los movimientos de cierre también alcanzan desviaciones promedio ligeramente menores (1.30Å) que los de apertura (1.37Å), aunque, al igual que en proteínas, las diferencias son tan pequeñas que es posible considerar que la simulación resulta igualmente satisfactoria en ambos sentidos. Desafortunadamente, con estructuras de ARN, *Molprobit* únicamente permite determinar el número de choques importantes cada 1000 átomos. En la Tabla 9 puede comprobarse que el número de colisiones inicial (32) apenas aumenta hasta 37 en las estructuras finales y hasta 33 en las estructuras intermedias de la trayectoria. Como sucede en proteínas, los movimientos son graduales y conservan adecuadamente la estructura secundaria a lo largo de las simulaciones. Estos resultados validan el uso de los modos normales de baja frecuencia para generar trayectorias factibles y estudiar los cambios conformacionales del ARN. Los resultados individuales con ARN se muestran en la Tabla A4 (véase el Apéndice).

### 5.6.4. Simulación de trayectorias de transiciones conformacionales reduciendo el número de coordenadas en proteínas y ARN

Para comprobar el efecto de la reducción del número de grados de libertad en las trayectorias se han simulado las transiciones conformacionales de todos los casos de proteínas y ARN fijando



aleatoriamente distintos porcentajes de ángulos diedros. En la Tabla 9 se muestran los resultados correspondientes a la fijación del 50 y 90% de estos ángulos empleando los distintos modelos atómicos. Sorprendentemente, al fijar el 50%, tanto las desviaciones finales como las métricas de calidad estructural resultan prácticamente idénticas a las obtenidas sin fijar. Más aún, al quitar el 90% de las variables, todavía se obtienen resultados aceptables con desviaciones finales muy próximas a 1 Å, aunque con más colisiones (~50). Nótese que esta reducción del número de variables no es permanente, pues el conjunto de ángulos diedros considerado fijo durante la trayectoria cambia aleatoriamente cada vez que se lleva a cabo el análisis vibracional (~100 veces por simulación). De este modo, todos estos ángulos tienen la oportunidad de moverse a lo largo de la trayectoria. No obstante, aunque el cálculo de transiciones factibles empleando los modelos más detallados ya es muy rápido, es posible aumentar aún más la velocidad al reducir aleatoriamente el número de variables. Esto resulta muy interesante sobre todo en los sistemas de mayor tamaño. Según el modelo atómico empleado, los incrementos obtenidos fijando el 50 y 90% de las variables son, respectivamente, de 2.9 y 3.3 veces en HA, de 4.3 y 8.5 en Ca, y de 2.5 y 3.3 en ARN. De este modo, es posible obtener aceleraciones de 2 a 4 veces fijando aleatoriamente el 50% de los ángulos diedros sin afectar significativamente a la calidad de las estructuras, y de 3 a 8 veces fijando el 90%, aunque con más colisiones. Nótese que las aceleraciones serán tanto mayores cuanto más grandes sean las macromoléculas.

### 5.7. Validación de las simulaciones de Monte Carlo

Para explorar el espacio conformacional de las macromoléculas se ha desarrollado *iMC*, una herramienta que permite generar rápidamente conformaciones factibles a partir de una única estructura inicial y de sus modos normales de vibración (véase la sección 4.3.4 de Materiales y métodos). El proceso iterativo se basa en la activación de los modos de baja frecuencia para generar conformaciones que solo serán aceptadas cuando se cumpla el criterio de Metropolis. Gracias al empleo de los modos normales en CI la evaluación de la energía es muy rápida y las distorsiones estructurales son pequeñas, lo que permite obtener eficientemente un gran número de modelos con elevada calidad estructural. En la validación de este método se ha empleado un conjunto de 10 proteínas que cierran sus dominios tras la unión de sus ligandos. Aunque este conjunto de proteínas procede de un estudio reciente para predecir estructuras holo a partir de conformaciones apo (Seeliger y de Groot, 2010), aquí se emplea únicamente para ilustrar la eficacia del muestreo conformacional de *iMC*. En la Tabla 10 se recogen los resultados de estas simulaciones. A partir de cada una de las 10 estructuras apo se genera un conjunto de 1000 conformaciones empleando tan solo los 5 modos normales de menor frecuencia. Estos modos

son capaces de representar la mayor parte del cambio conformacional entre las estructuras apo y holo ( $\delta_5=0.94$ ). Al analizar estos conjuntos de conformaciones se ha observado que la estructura más cercana a la conformación holo correspondiente resulta ser tan solo de 1.28Å en promedio. Además, un 1.6 y un 6.6% de las conformaciones se encuentran a menos de 1.5 y 2.0Å de dicha conformación, respectivamente. Estos resultados son muy interesantes teniendo en cuenta que el promedio de las desviaciones entre las estructuras apo y holo es relativamente grande (4.04Å). Por tanto, *iMC* es capaz de obtener conformaciones muy próximas a la forma holo empleando únicamente la información codificada en los 5 primeros modos normales de la estructura apo. En la Figura 37A se muestra la mejor estructura del conjunto de conformaciones (azul) superpuesta sobre la estructura holo (amarillo) para un caso representativo (*Isw5/Isw2*). Respecto a la calidad de las estructuras, el porcentaje de ángulos diedros fuera de las regiones permitidas permanece constante, únicamente aumentan respecto de las conformaciones iniciales el número de choques (de 12 a 40) y la resolución equivalente (de 2.0 a 2.5Å). Además, estos métodos son muy rápidos, la generación de 1000 estructuras solo tarda unos pocos segundos.

**Tabla 10.** Exploración conformacional basada en los modos normales en CI y muestreo de MC.

Apo		Holo					Libre <sup>f</sup>			Restringido <sup>h</sup>			Col. <sup>i</sup>		%R		Molprobability	
Id.	Rg <sup>a</sup>	Id.	Rg	N <sup>b</sup>	$\delta_5^c$	Rmsd <sup>d</sup>	Rmin <sup>e</sup>	<1.5Å <sup>g</sup>	<2Å	Rmin	<1.5Å	<2Å	I	P	I	P	I	P
1ex6	17.09	1ex7	16.26	186	0.92	3.64	1.42	0.3	3.6	1.20	3.1	18.8	31	50	0.0	0.0	3.2	3.4
1fto	19.11	1ftm	18.23	257	0.96	2.26	0.57	7.8	22.9	0.49	70.1	91.4	13	22	0.4	0.4	2.1	2.3
1sw5	19.46	1sw2	18.52	270	0.94	5.04	1.16	0.2	1.2	1.24	1.5	7.5	3	75	0.0	0.0	1.6	2.8
1urpC	20.14	2dri	19.08	271	0.96	4.29	1.32	0.3	1.3	0.90	3.7	10.4	8	38	0.7	0.7	2.2	2.7
1gud	21.01	1rpj	19.34	288	0.92	4.45	1.32	0.1	1.0	1.15	3.5	19.7	3	32	0.7	0.7	1.1	1.9
2e2n	20.66	2e2o	19.69	298	0.93	2.99	1.13	0.5	5.2	0.95	14.4	47.7	12	26	0.0	0.0	1.7	2.0
1usg	22.13	1usi	20.03	345	0.95	7.14	1.85	0.0	0.1	1.25	0.3	2.2	6	55	0.3	0.3	1.3	2.1
1jej	21.92	1jg6	20.91	351	0.91	2.10	0.85	7.1	28.3	0.83	58.6	90.2	19	25	0.3	0.3	2.4	2.5
1rf5	22.62	1rf4	21.08	427	0.92	3.69	1.55	0.0	1.6	1.28	6.4	39.0	11	22	0.9	0.9	2.7	3.0
1y3q	24.62	1y3n	23.18	490	0.95	4.79	1.60	0.0	0.7	1.47	0.3	10.3	12	56	0.0	0.0	2.0	2.6
Promedio				318	0.94	4.04	1.28	1.6	6.6	1.08	16.2	33.7	12	40	0.3	0.3	2.0	2.5
Desv.Est.				88	0.02	1.48	0.37	3.1	10.2	0.28	25.8	33.2	8	18	0.4	0.4	0.7	0.5

<sup>a</sup> Radio de giro (calculado según Tabla 3).

<sup>b</sup> Número de aminoácidos.

<sup>c</sup> Solapamiento acumulado de los 5 modos de menor energía (calculado según Tabla 6).

<sup>d</sup> Desviación entre las conformaciones apo y holo.

<sup>e</sup> Desviación mínima entre la estructura holo y todas las conformaciones muestreadas.

<sup>f</sup> Muestreo libre desde la conformación apo empleando los parámetros por defecto de *iMC* con el modelo Ca y el potencial refinado a partir de datos de MD.

<sup>g</sup> Porcentaje de modelos que se desvían a menos de 1.5 y 2Å de la estructura holo.

<sup>h</sup> Muestreo restringido para obtener conformaciones cuyo Rg se desvía un 1% del de la estructura holo.

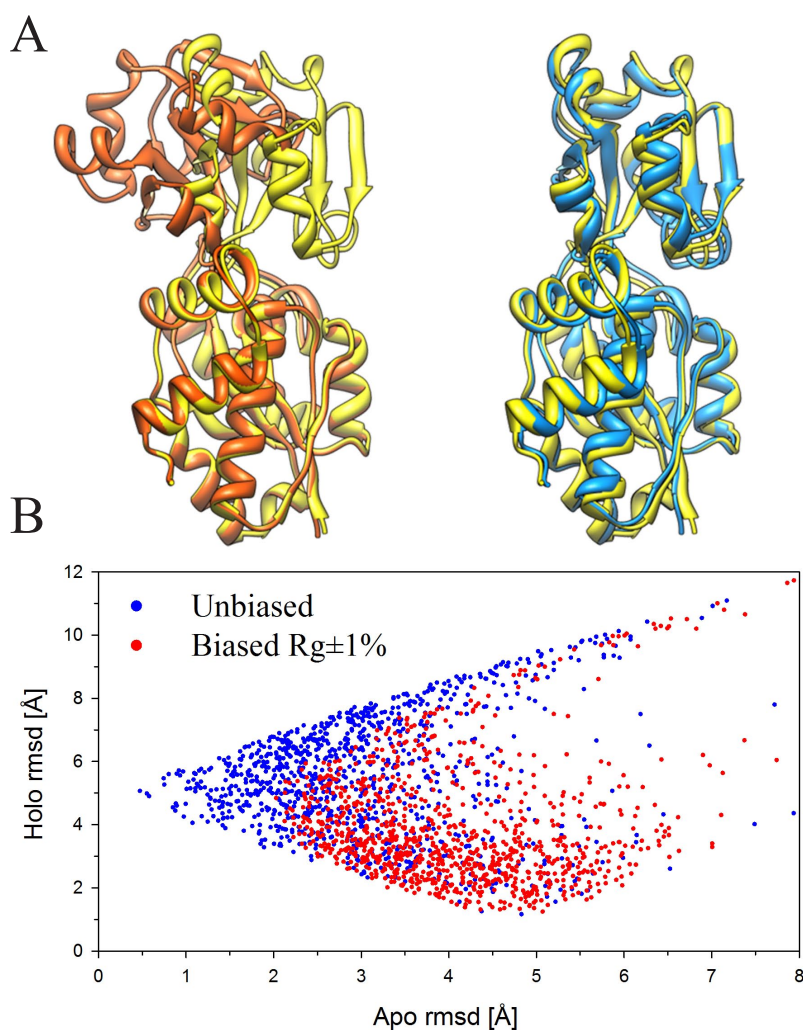
<sup>i</sup> Métricas de validación estructural de *Molprobability* (véase descripción en Tabla 9). Se muestran los valores iniciales (*I*) junto con los promedios correspondientes a los 1000 modelos del muestreo libre (*P*). También se obtienen valores similares a partir de los modelos del muestreo restringido (datos no mostrados).

\* En todos los casos las desviaciones se ha calculado empleando únicamente los átomos C <sub>$\alpha$</sub> .



### 5.7.1. Muestreo restringido

Para enriquecer el muestreo en conformaciones próximas a las estructuras holo, es posible imponer restricciones en el radio de giro ( $R_g$ ) de las conformaciones generadas. Por ejemplo, si únicamente se generan estructuras con un  $R_g$  parecido al de la estructura holo (con una tolerancia del 1%), el 16% de ellas se encuentra a menos de 1.5Å de RMSD de las estructuras holo y el 34% a menos de 2.0Å. Esto supone un enriquecimiento de 10 y 6 veces, respectivamente, respecto del muestreo libre. Las diferencias se ilustran en la Figura 37B para el mismo caso representativo. En ella puede apreciarse cómo el muestreo restringido (rojo) se localiza más cerca a la conformación holo que el muestreo sin restricciones (azul). Al aplicar las restricciones los tiempos de ejecución resultan un poco mayores, aunque siempre están por debajo de los 4 minutos.



**Figura 37.** Muestreo conformacional de MC. Superposición de la estructura holo (amarillo) con la estructura apo (naranja) a la izquierda, y con la mejor estructura (azul) a la derecha para el caso *Isw5/Isw2* (A). Comparación del muestreo obtenido al restringir el radio de giro (puntos rojos) con el obtenido sin restricciones (puntos azules) (B).

## 5.8. Validación de la herramienta de ajuste flexible – iModfit

Se ha desarrollado e implementado *iModfit*, una herramienta basada en NMA en CI para el ajuste flexible de estructuras atómicas en el interior de mapas de densidad a media/baja resolución procedentes de microscopía electrónica. Al igual que con los demás métodos se ha procedido a su validación sistemática tanto con datos simulados como experimentales.

### 5.8.1. Validación de iModfit en casos simulados de proteínas

Para obtener problemas de ajuste flexible con los que validar la herramienta *iModfit* se han generado mapas simulados a 8, 10, 15, 20 y 25 Å de resolución a partir de los 23 pares de estructuras presentes en el conjunto de pruebas de proteínas de *MolMovDB*. Con cada par de estructuras es posible obtener dos ajustes diferentes, que corresponden a la adaptación de la estructura atómica de la conformación cerrada en el interior del mapa de densidad de la conformación abierta, y *viceversa*. Como las estructuras atómicas de las dos conformaciones son conocidas, es posible evaluar el error cometido durante los ajustes mediante la comparación de los modelos finales con las estructuras atómicas utilizadas en la generación de los mapas. Para validar la herramienta con estos casos simulados se han efectuado los ajustes empleando el 5% de los modos, el modelo atómico HA y los ángulos  $\varphi$  y  $\psi$  como variables del sistema. Para evaluar la dispersión de los resultados se ha repetido cada ajuste en 10 ocasiones. La orientación relativa entre la estructura atómica inicial y el mapa objetivo es la misma que la empleada en las simulaciones de transiciones conformacionales (Sección 5.6).

En la Tabla 11 se muestra el RMSD promedio entre las estructuras finales y las estructuras atómicas utilizadas para simular los correspondientes mapas objetivo en función de la resolución. Como puede apreciarse en esta tabla, los bajos valores de RMSD obtenidos en todo el rango de resoluciones ( $\sim 1$  Å) indican que se han caracterizado correctamente las conformaciones codificadas en los mapas. Recuérdese que la amplitud inicial de los cambios conformacionales es de 7.57 Å. En cuanto a los cambios observados en función de la resolución, las desviaciones promedio aumentan gradualmente desde 1.06 Å al utilizar mapas de 8 Å hasta 1.33 Å en los casos de menor resolución (25 Å). Nótese que las diferencias son muy pequeñas ( $\sim 0.3$  Å) teniendo en cuenta el amplio rango de resoluciones ensayado. Estos resultados confirman que es posible obtener buenos ajustes a partir de mapas a media/baja resolución en ausencia de ruido. En la Figura 38, se muestran los resultados detallados para cada uno de los ajustes. Cada barra representa el promedio de los 10 ajustes realizados en cada caso. En casi todos ellos, los RMSDs son muy bajos ( $\sim 1$  Å) y aumentan ligeramente conforme disminuye la resolución. Además, la dispersión de los resultados es muy baja y no supera en ningún caso los  $\sim 0.1$  Å (datos no mostrados). Únicamente en los ajustes correspondientes a las estructuras *IgggA/IwdnA* y

**Tabla 11.** Resultados promedio de los ajustes flexibles de proteínas realizados con el modelo atómico HA y el 5% de los modos en función de la resolución.

Resolución <sup>a</sup> [Å]	RMSD <sup>b</sup> [Å]	Colisiones <sup>c</sup> [%]	Rama. <sup>d</sup> [%]	Enlaces <sup>e</sup> [%]	Angulos <sup>f</sup> [%]	Molprobability <sup>g</sup> [Å]
8	1.06 ± 0.66	35 ± 17	0.9 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
10	1.07 ± 0.68	35 ± 17	0.9 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
15	1.14 ± 0.72	36 ± 17	0.8 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
20	1.22 ± 0.74	35 ± 17	0.8 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
25	1.33 ± 0.77	35 ± 15	0.8 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
Inicial	7.57 ± 4.34	21 ± 11	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	2.7 ± 0.6
<i>iMorph</i>	0.85 ± 0.44	33 ± 14	0.8 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5

<sup>a</sup> Resolución de los mapas objetivo. Se han incluido los valores correspondientes a las estructuras iniciales (Inicial), junto con los resultados procedentes de las simulaciones de las trayectorias conformacionales entre las estructuras atómicas iniciales y las estructuras utilizadas para simular los mapas (*iMorph*).

<sup>b</sup> Promedio de las desviaciones entre los modelos finales y las estructuras empleadas para generar los mapas objetivo. En el cálculo del RMSD solo se han tenido en cuenta los átomos C<sub>α</sub>.

<sup>c-g</sup> Después de añadir todos los átomos de hidrógeno a los modelos finales, se han obtenido los promedios correspondientes a las siguientes métricas de validación estructural del programa *Molprobability*: (<sup>e</sup>) número de colisiones importantes por cada 1000 átomos, (<sup>d</sup>) porcentaje de ángulos diedros fuera de las regiones permitidas de Ramachandran, (<sup>e</sup>) porcentaje de longitudes de enlace incorrectas, (<sup>f</sup>) porcentaje de ángulos de enlace incorrectos, y (<sup>g</sup>) resolución cristalográfica equivalente para una estructura de calidad similar.

*Isu4A/It5sA* las desviaciones finales no bajan de ~3Å a pesar de haber convergido. En estos casos, o bien el sistema queda atrapado en un máximo local de la función de correlación, o bien los movimientos necesarios para ajustar las estructuras no están representados en los modos de baja frecuencia. En los ajustes de *IgggA/IwdnA*, el empleo del RMSD como función de coste en la simulación de las transiciones conformacionales permite alcanzar desviaciones menores de 1Å con el 5% de los modos (véase *iMorph* en la Figura 38). Por tanto, en este caso parece más razonable pensar que los ajustes quedan atrapados en un máximo local de la función de correlación dado que el 5% de los modos es suficiente para representar el cambio conformacional. Sin embargo, en los ajustes de *Isu4A/It5sA* el problema es diferente ya que ni la correlación ni el RMSD son capaces de conducir los ajustes más allá de los ~3Å. En este caso, probablemente sea necesario incluir modos de mayor frecuencia para reducir las desviaciones.

Al igual que en secciones anteriores, para evaluar la calidad de los modelos finales se han utilizado las métricas de validación estructural del programa *Molprobability*. En la Tabla 11 se han incluido los valores de estas métricas en función de la resolución. Como puede observarse, el número de colisiones importantes por cada 1000 átomos tan solo aumenta de 21 en las estructuras iniciales hasta 35 en las finales y, además, su número no depende de la resolución de los mapas. En cuanto al porcentaje de ángulos diedros con valores no permitidos, tan solo aumenta desde

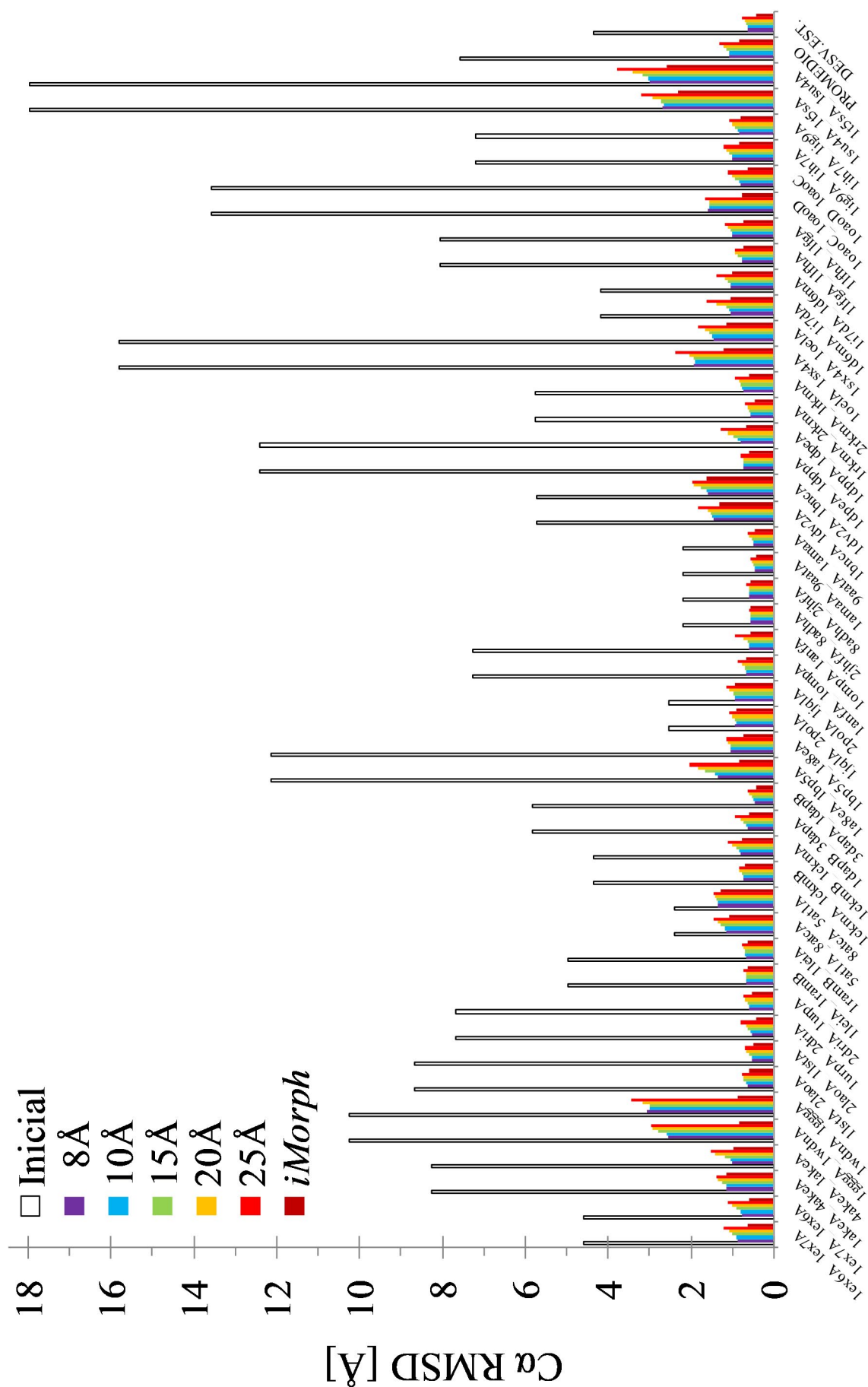
**Figura 38.** Desviaciones individuales de los ajustes flexibles con el modelo HA y mapas simulados de proteínas. En el eje de abscisas se indica el problema de ajuste mediante el código del PDB de la estructura atómica inicial, seguido del código de la estructura empleada en la simulación del mapa. Por ejemplo, *Ilei\_IramB* representa el ajuste de la estructura *Ilei* (cadena A) en el mapa simulado de la estructura *Iram* (cadena B). En la leyenda se muestran los colores correspondientes a cada resolución. Nótese que cada barra representa el promedio de 10 ajustes diferentes y que su desviación estándar no se ha representado porque en ningún caso supera los 0.1 Å. También se han incluido los resultados promedio correspondientes a los modelos finales obtenidos de las simulaciones de las trayectorias conformacionales entre las estructuras atómicas iniciales y las utilizadas para simular los mapas (*iMorph*).

el 0.6% inicial hasta el 0.8 o 0.9% en todo el rango de resoluciones. Asimismo, la resolución cristalográfica equivalente inicial (2.7 Å) apenas empeora en las estructuras finales (3.0 Å). Cabe destacar que ni las distancias ni los ángulos de enlace se ven afectados en ningún caso gracias al empleo de los ángulos diedros  $\varphi$  y  $\psi$  como CI del sistema. Por tanto, la calidad de las conformaciones finales es excelente y similar a la obtenida en las simulaciones de transiciones conformacionales (*iMorph*). Todas estas observaciones validan el uso de los modos normales de baja frecuencia para obtener buenos modelos atómicos de proteínas a partir de mapas de densidad simulados de microscopía electrónica.

### 5.8.2. Validación de los modelos atómicos reduccionistas de proteínas

En la Tabla 12 se han representado los promedios de RMSD entre las estructuras finales y las estructuras atómicas correspondientes a los mapas objetivo para los tres modelos atómicos de proteínas (C $\alpha$ , C5 y HA). En ella puede apreciarse que todas las desviaciones son pequeñas, *i.e.* entre ~1.1 y ~1.6 Å. Al igual que con el modelo HA, con C $\alpha$  y C5 la desviación también es inversamente proporcional a la resolución de los mapas. Al comparar el modelo C5 con HA, los promedios resultan prácticamente idénticos en todo el rango de resoluciones ensayadas. Esto no es de extrañar pues el número de átomos por residuo de ambos modelos es muy similar, *i.e.* 5 átomos en el C5 y 7~8 en el HA. Sin embargo, con el modelo C $\alpha$  las desviaciones son ligeramente superiores (~0.2-0.3 Å) a las de C5 y HA para todas las resoluciones. Cuando se comparan individualmente los ajustes procedentes del modelo C $\alpha$  con los demás, se observa que las desviaciones tan solo empeoran significativamente en dos de los 46 casos: *IakeA\_4akeA* y *IwdnA\_IgggA* (datos no mostrados). Al ajustar la estructura de *IakeA* en los mapas de *4akeA*, las desviaciones aumentan hasta ~4 Å con todas las resoluciones, mientras que el ajuste de *IwdnA* en los mapas de *IgggA*, únicamente empeora con el mapa de 8 Å hasta alcanzar ~6 Å de desviación. Estas diferencias puntuales ponen de manifiesto algunos de los beneficios que supone el empleo de los modelos atómicos más detallados.

En la Tabla 12 puede apreciarse que con el modelo C5 se alcanzan unas 38 colisiones para las distintas resoluciones, lo que supone un aumento relativamente pequeño en comparación



con las colisiones de las estructuras iniciales (21). Además, estos valores resultan muy similares a los obtenidos con el modelo HA (~35). En cambio, con el modelo Ca el número de colisiones triplica al de las estructuras iniciales para todo el rango de resoluciones ensayadas, *i.e.* unas 60~70. Esto se debe a que no es posible evitar los choques entre los átomos que no se han tenido en cuenta explícitamente en el NMA. En cuanto al porcentaje de ángulos fuera de las regiones permitidas de Ramachandran, puede observarse que apenas aumenta al utilizar los modelos Ca y C5. En ambos casos se obtienen valores de ~0.7 para todo el rango de resoluciones, lo que supone un incremento insignificante respecto del valor de la conformación inicial (0.6). No obstante, con el modelo HA estos valores son ligeramente superiores, *i.e.* entre el 0.8 y 0.9%. En la Tabla 12 también se han incluido los porcentajes de longitudes y ángulos de enlace inválidos, así como también, del *molprobability score* que indica la resolución cristalográfica equivalente a estructuras cristalográficas de calidad similar. Estos valores tampoco varían significativamente en ninguna de las condiciones ensayadas, aunque son algo superiores (0.2-0.3Å) con el modelo Ca debido a que aparecen más colisiones. Cabe destacar que la calidad de las estructuras finales

**Tabla 12.** Comparación entre los ajustes flexibles de proteínas realizados con los distintos modelos atómicos y el 5% de los modos en función de la resolución.

Modelo <sup>a</sup> Atómico	Resolución <sup>b</sup> [Å]	RMSD [Å]	Colisiones [%]	Rama. [%]	Enlaces [%]	Ángulos [%]	Molprobability [Å]
HA	8	1.06 ± 0.65	35 ± 17	0.9 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	10	1.07 ± 0.64	35 ± 17	0.9 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	15	1.14 ± 0.66	36 ± 17	0.8 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	20	1.22 ± 0.70	35 ± 17	0.8 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	25	1.33 ± 0.76	35 ± 15	0.8 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	<i>iMorph</i>	0.85 ± 0.44	33 ± 14	0.8 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
C5	8	1.08 ± 0.66	38 ± 17	0.7 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	10	1.09 ± 0.65	38 ± 17	0.7 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	15	1.15 ± 0.66	36 ± 15	0.7 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	20	1.24 ± 0.70	39 ± 17	0.7 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	25	1.33 ± 0.75	39 ± 16	0.7 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	<i>iMorph</i>	0.85 ± 0.43	35 ± 14	0.7 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
Ca	8	1.36 ± 0.95	64 ± 22	0.7 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.2 ± 0.5
	10	1.31 ± 0.68	70 ± 24	0.7 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.3 ± 0.5
	15	1.39 ± 0.68	67 ± 25	0.7 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.3 ± 0.5
	20	1.50 ± 0.73	65 ± 25	0.7 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.2 ± 0.5
	25	1.61 ± 0.79	61 ± 24	0.7 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.2 ± 0.5
	<i>iMorph</i>	0.97 ± 0.57	48 ± 18	0.7 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.5

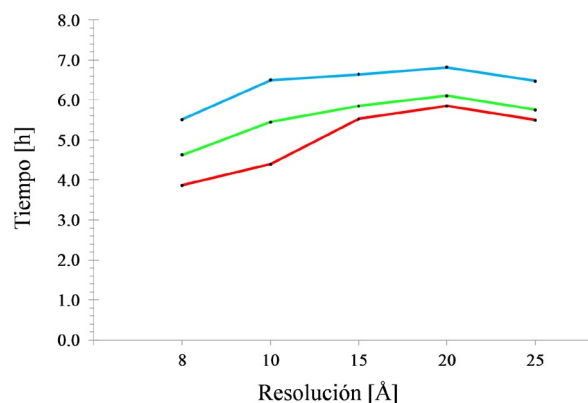
<sup>a</sup> Modelo atómico utilizado en los ajustes.

<sup>b</sup> La descripción de estas métricas se detalla en el pie de la Tabla 11.



no depende significativamente de la resolución de los mapas ni de de los modelos atómicos empleados. Estos hechos evidencian la robustez de los modos normales de vibración en CI a la hora de representar los cambios conformacionales colectivos de las macromoléculas.

En la Figura 39 se han cronometrado los tiempos necesarios para llevar a cabo los 46 ajustes con los mapas de 8 a 25 Å para los tres modelos atómicos de proteínas. Con el modelo de mayor resolución (HA) es posible realizar los ajustes en unas 6.5 horas. Como es esperable, los ajustes son tanto más rápidos cuanto menor es el número de átomos del modelo atómico en todo el rango de resoluciones ensayado, aunque las mejoras son moderadas. Al emplear los mapas de mayor resolución (8 y 10 Å) se obtienen aceleraciones del 30 y del 15% con los modelos Ca y C5, respectivamente, mientras que, con los de menor resolución (15, 20 y 25 Å) las mejoras son más modestas, *i.e.* del 15% con Ca y del 10% con C5. La dependencia de estas mejoras con el modelo atómico y la resolución puede explicarse a partir de los tiempos parciales empleados en completar cada una de las tareas más costosas computacionalmente, *i.e.* el NMA, el movimiento en CI y el filtrado de los modelos de prueba para convertirlos en mapas de densidad (datos no mostrados). El tiempo dedicado al NMA, que supone aproximadamente un 50% del total, es independiente de la resolución y disminuye ligeramente con el número de átomos del modelo, llegando a ser un 10% menor que el correspondiente a HA con el modelo C5 y un 20% con el Ca. Según esto, el NMA es capaz de justificar mejoras del ~5% con el modelo C5 y del ~10% con Ca. El tiempo asociado al movimiento en CI resulta dos y tres veces menor al emplear los modelos C5 y Ca, respectivamente, independientemente de la resolución, aunque tan solo supone entre el 7 y el 11% del tiempo total. Por tanto, las aceleraciones observadas en los casos de menor resolución se pueden justificar únicamente a partir de las mejoras procedentes del NMA y del movimiento en CI. En cambio, para entender las aceleraciones obtenidas en los casos de mayor resolución hay que tener en cuenta como varía el tiempo necesario para obtener los mapas a partir del modelo de prueba. Esta tarea, consume aproximadamente entre el 20 y el 45% del tiempo total y depende de múltiples factores, como la resolución, el muestreo, el tamaño de



**Figura 39.** Tiempo de cálculo necesario para efectuar los ajustes flexibles de proteínas en función del modelo atómico y la resolución. En colores rojo, verde y azul se muestran los tiempos correspondientes a los modelos atómicos Ca, C5 y HA, respectivamente. Cada uno de los 46 ajustes de *MolMovDB* se han efectuado una sola vez (en lugar de 10) utilizando un PC Intel® Core<sup>2</sup>duo Q6600 a 2.4 GHz.



los mapas y el número de átomos del modelo. Para llevarla a cabo de la forma más eficiente posible, *iModfit* selecciona automáticamente el método más rápido para cada problema particular (véase la sección 4.3.6 de Materiales y métodos). Con los mapas ensayados de 2Å/voxel y resoluciones más altas ( $\leq 10\text{\AA}$ ) el filtrado por convolución suele ser más rápido, mientras que a resoluciones más bajas ( $>10\text{\AA}$ ) lo es el filtrado de Fourier. Al observar los tiempos de filtrado, puede verse que el método de Fourier es prácticamente independiente de la resolución y del modelo atómico; sin embargo, el método de convolución depende de ambos (datos no mostrados). Así, el filtrado a 8Å resulta el doble de rápido que a 10Å y el filtrado a partir de un modelo Ca también es el doble de rápido en comparación con el de HA. Estas observaciones justifican las mejores aceleraciones obtenidas en los casos de mayor resolución. De todas formas, al utilizar muestreos más gruesos (3~4Å/voxel) en los mapas de menor resolución es posible alcanzar aceleraciones similares a las de los casos de mayor resolución (datos no mostrados).

En resumen, a pesar de que con Ca se obtiene un mayor número de colisiones y que con HA el porcentaje de ángulos fuera de las regiones permitidas de Ramachandran es ligeramente

**Tabla 13.** Comparación entre los ajustes flexibles de proteínas obtenidos fijando aleatoriamente distintos porcentajes de ángulos diedros con el modelo atómico HA en función de la resolución.

Fijación <sup>a</sup> [%]	Resolución <sup>b</sup> [Å]	RMSD [Å]	Colisiones [%]	Rama. [%]	Enlaces [%]	Ángulos [%]	Molprobability [Å]
0%	8	1.06 ± 0.65	35 ± 17	0.9 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	10	1.07 ± 0.64	35 ± 17	0.9 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	15	1.14 ± 0.66	36 ± 17	0.8 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	20	1.22 ± 0.70	35 ± 17	0.8 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	25	1.33 ± 0.76	35 ± 15	0.8 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	<i>iMorph</i>	0.85 ± 0.44	33 ± 14	0.8 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
50%	8	1.09 ± 0.67	35 ± 16	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	10	1.10 ± 0.66	35 ± 16	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	15	1.16 ± 0.69	35 ± 16	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	20	1.24 ± 0.74	35 ± 15	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	25	1.34 ± 0.81	34 ± 14	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	<i>iMorph</i>	0.90 ± 0.49	33 ± 14	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	2.9 ± 0.5
90%	8	1.45 ± 1.21	50 ± 23	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.5
	10	1.45 ± 1.13	49 ± 22	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.5
	15	1.47 ± 1.00	50 ± 22	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.5
	20	1.55 ± 1.05	49 ± 21	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.5
	25	1.69 ± 1.14	46 ± 18	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.5
	<i>iMorph</i>	1.18 ± 0.76	48 ± 21	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.5

<sup>a</sup> Porcentaje fijado de ángulos diedros  $\varphi$  y  $\psi$ . Los ajustes se han llevado a cabo empleando el 5% de los modos.

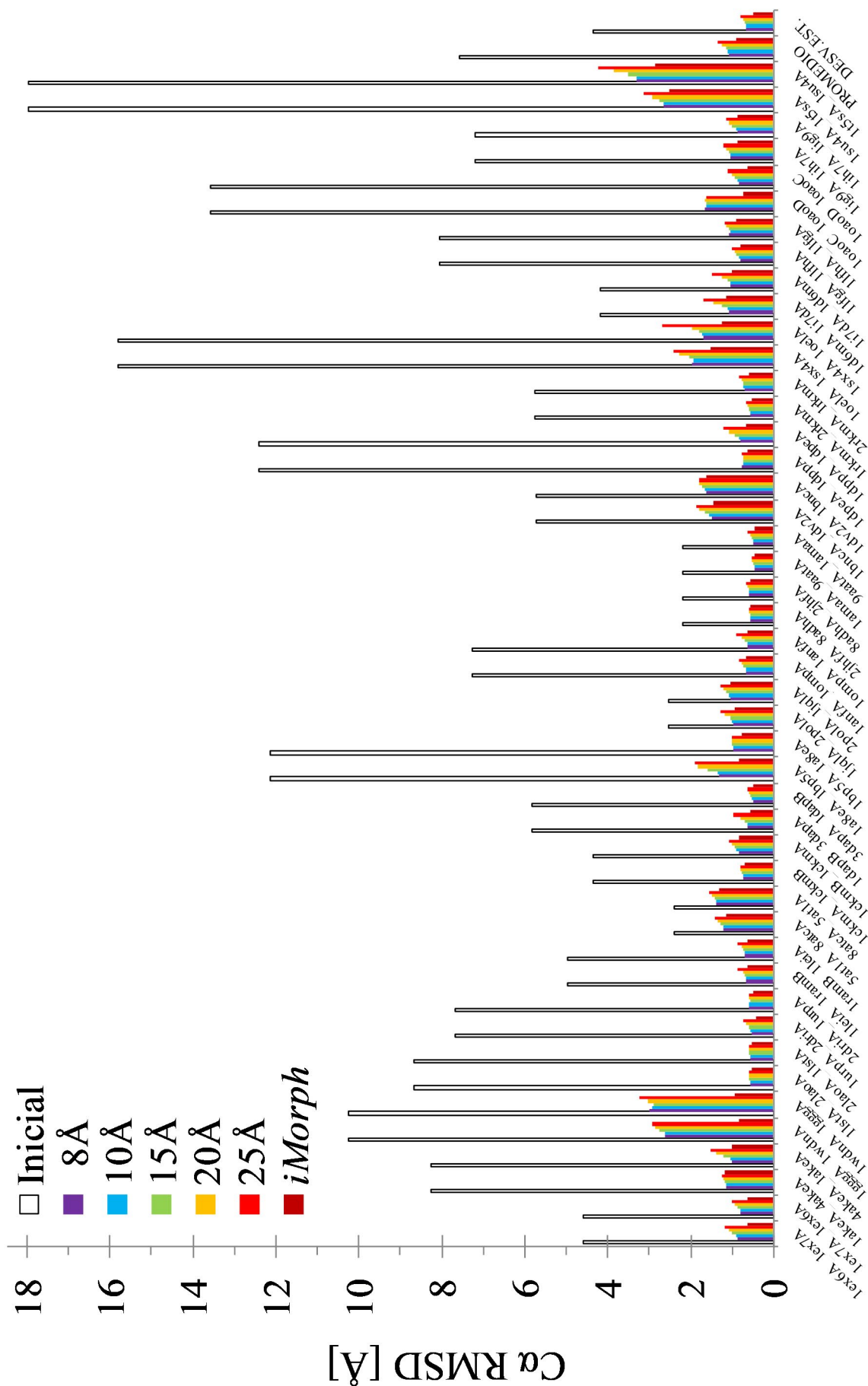
<sup>b</sup> La descripción de estas métricas se detalla en el pie de la Tabla 11.

mayor, la eficacia y calidad de los ajustes resulta excelente con cualquiera de los modelos atómicos utilizados, especialmente con C5 y HA. No obstante, el empleo de los modelos más reduccionistas C5 y Ca permite acelerar los tiempos de ajuste entre un 10 y un 30%, respectivamente, en comparación con los correspondientes al modelo HA.

### 5.8.3. Validación de la reducción del número de coordenadas internas en proteínas

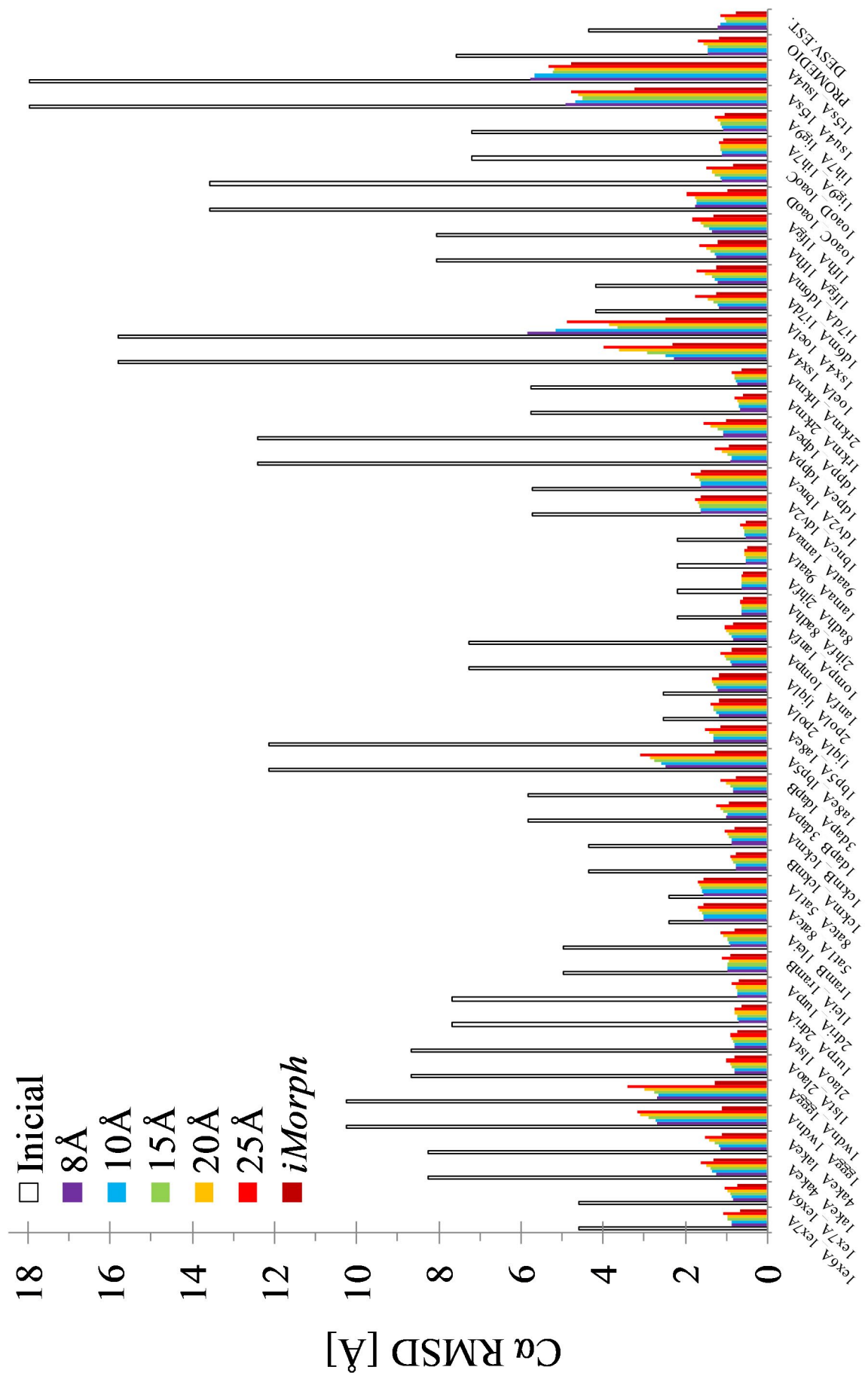
Para comprobar la eficacia de la fijación aleatoria de los ángulos diedros ( $\varphi$  y  $\psi$ ) en el ajuste flexible se han repetido las pruebas de la sección 5.8.1 (5% de los modos y modelo atómico HA) eliminando aleatoriamente el 50 y 90% de las variables (Tabla 13). En general, las desviaciones entre los modelos finales y las estructuras con las que se simulaban los mapas objetivo aumentan suavemente según disminuye la resolución. Al fijar el 50% de los ángulos diedros las desviaciones son muy bajas, *i.e.* entre 1.09 y 1.35 Å. Sorprendentemente, las diferencias con los valores obtenidos empleando todos los ángulos son mínimas y no superan los 0.03Å a ninguna resolución. Al analizar las desviaciones individualmente en la Figura 40, los resultados también son prácticamente idénticos a los obtenidos sin fijar (Figura 38). En ningún caso se obtiene un RMSD mayor de 2Å, a excepción de los ya señalados en la sección 5.8.1. Pese a eliminar el 90% de las variables todavía se consiguen desviaciones entre 1.45 y 1.69Å que tan solo son ~0.4Å superiores a las obtenidas sin fijar. Con el fin de comprobar si el aumento de estas desviaciones ocurre de manera generalizada o si, por el contrario, se debe al empeoramiento de algunos casos puntuales, en la Figura 41 se han representado estos resultados individualmente. En ella, se observa que algunos de los ajustes con cambios conformacionales de más de 12Å de amplitud presentan desviaciones significativamente superiores de entre ~3 y ~5Å: *la8eA\_lbp5A*, *loelA\_lsx4A*, *lsx4A\_loelA*, *lsu4A\_lt5sA* y *lt5sA\_lsu4A*. Cuando se calculan de nuevo las desviaciones sin incluir estos 5 casos, los RMSD al fijar el 90% mejoran ~0.3Å hasta valores de 1.11-1.36Å. En consecuencia, cuando se fija el 90% de los ángulos diedros las desviaciones son prácticamente iguales en la mayoría de los casos.

En la Tabla 13 se muestra el número de colisiones en función del porcentaje de ángulos fijado y la resolución de los mapas. Es importante destacar que al eliminar el 50% de las variables el número de colisiones es prácticamente idéntico al obtenido sin fijar (~35). En cambio, quitando el 90% de las variables las colisiones aumentan hasta alcanzar valores de ~50. Este incremento se debe probablemente a que la excesiva rigidez del sistema dificulta significativamente los cambios conformacionales. En cuanto a la calidad de los ángulos diedros, los porcentajes fuera de las regiones permitidas se mantienen en torno al ~0.6% tanto al eliminar el 50 como el 90% de las variables. Al igual que en casos anteriores, la calidad estructural se mantiene ante los cambios en la resolución de los mapas.



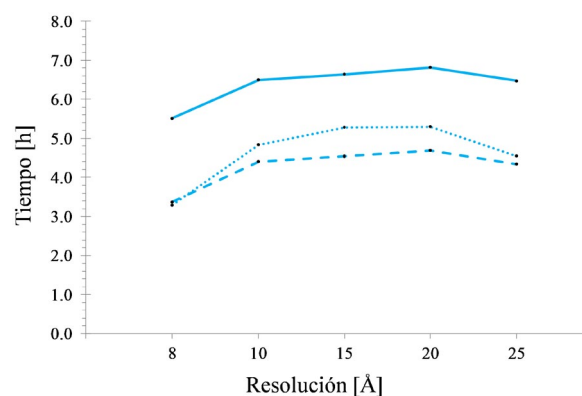
◀ **Figura 40.** Desviaciones individuales de los ajustes flexibles de proteínas obtenidos eliminando aleatoriamente el 50% de las variables con el modelo HA en función de la resolución. Véase el pie de la Figura 38.

Como cabría esperar, los tiempos necesarios para llevar a cabo los 46 casos se reducen al eliminar el 50% de las variables en comparación con los obtenidos sin fijar (Figura 42). Esta reducción de poco más de 2 horas se mantiene constante para todo el rango de resoluciones, y permite completar, por ejemplo, los ajustes en unas 4.4 horas con los mapas a 10Å de resolución. Al examinar los tiempos parciales de las fases más costosas computacionalmente, puede comprobarse que la reducción en los tiempos globales se debe principalmente a que disminuye unas 4 veces el tiempo dedicado al NMA, *i.e.* de 3.1 horas pasa a tardar tan solo 0.8 (datos no mostrados). Por tanto, la aceleración del NMA justifica por sí misma las mejoras en los tiempos totales obtenidos al fijar el 50% de los ángulos. Sin embargo, al eliminar el 90% de las variables, los ajustes con los mapas de 10 a 20Å son más lentos que al fijar el 50%. Estos hechos, aparentemente contradictorios, pueden deberse a dos factores principales. El primero está relacionado con el aumento de rigidez del sistema debido a la eliminación de variables. Esta rigidez adicional restringe los movimientos, por lo que es necesario llevar a cabo un mayor número de iteraciones para alcanzar la convergencia. Por ejemplo, con el modelo HA el número de iteraciones promedio tan solo aumenta entre el 2 y el 8% con el 50% de las variables, mientras que al fijar el 90% de los ángulos los incrementos son mayores, *i.e.* entre el 13 y el 35% en función de la resolución (datos no mostrados). El segundo factor tiene que ver con el escalado del coste computacional del NMA con el número de átomos ( $n$ ) y de variables ( $N$ ). Cuando  $N$  es significativamente menor que  $n$ , el coste computacional del NMA ya no es proporcional a  $\sim N^3$  (fase de diagonalización) y se vuelve proporcional a  $\sim n^2$  (cálculo del Hessiano). Por este motivo, el NMA con el 10% de las variables no es  $\sim 1000$  veces más rápido que al emplear todas ellas, como cabría esperar si el coste computacional fuese proporcional a  $\sim N^3$ , sino que la aceleración es de tan solo  $\sim 6$  veces. Por estos factores, la fijación del 90% de los ángulos no merece la pena, dado que el aumento en la velocidad del NMA ocasionado por la disminución del número de variables no compensa el incremento en el número de iteraciones necesario para llevar a cabo los ajustes. Por el contrario, la fijación aleatoria del 50% de los ángulos  $\varphi$  y  $\psi$  constituye un método eficaz para reducir el número de grados de libertad, acelerar los cálculos un 30-40% y obtener modelos finales de mayor calidad estructural. Únicamente se recomienda fijar porcentajes superiores al 50% para reducir los requerimientos computacionales cuando el tamaño de la macromolécula es tan grande que la memoria necesaria para almacenar las matrices del sistema supera la cantidad máxima disponible.



◀ **Figura 41.** Desviaciones individuales de los ajustes flexibles de proteínas obtenidos eliminando aleatoriamente el 90% de las variables con el modelo HA en función de la resolución. Véase el pie de la Figura 38.

Con el fin de completar el estudio de los efectos de la reducción del número de variables, en las Tablas 14 y 15 se muestran los resultados de los ajustes obtenidos fijando aleatoriamente los ángulos diedros con los modelos atómicos C5 y C $\alpha$ , respectivamente. La eliminación del 50% de las variables utilizando estos modelos produce resultados muy similares a los obtenidos sin fijar. Los promedios de las desviaciones entre estos modelos y las estructuras de los mapas tan solo aumentan entre 0.02 y 0.04 Å. Al examinar individualmente estos valores, los resultados también son muy similares (datos no mostrados). Únicamente hay dos excepciones con el modelo C $\alpha$  en las que se alcanzan desviaciones significativamente diferentes. En *Iso4A\_IoelA* los RMSD empeoran hasta unos 3-4 Å al utilizar mapas de 8, 20 y 25 Å de resolución. En cambio, *IwdnA\_IgggA* mejora hasta los 3 Å con el mapa de 8 Å, alcanzando valores similares a los obtenidos con las otras resoluciones. Nótese que este último caso, presenta desviaciones de ~6 Å cuando se utilizan todos los ángulos diedros (véase la sección 5.8.2). Esto demuestra que, en ocasiones, la consideración de estructuras más rígidas puede mejorar algunos ajustes. En cuanto a la calidad estructural, mientras que con el modelo C5 el número de colisiones ya es muy parecido (~38) al obtenido con todos los ángulos, con C $\alpha$  se obtienen valores algo mejores (~55). Además, el porcentaje de ángulos diedros en regiones no permitidas disminuye ligeramente del 0.7% al 0.6% con ambos modelos. Por tanto, la eliminación del 50% de las variables con los modelos C5 y C $\alpha$  no solo mantiene la eficacia de los ajustes, sino que además mejora ligeramente la calidad de las estructuras. Al fijar el 90% de los ángulos, las desviaciones promedio empeoran respecto de las obtenidas sin fijar en ~0.4 Å con el modelo C5 y en ~0.2 Å con C $\alpha$  para todo el rango de resoluciones. Al analizar estas desviaciones individualmente se observa que la mayor parte de las diferencias son debidas al deterioro de los mismos 5 casos señalados



**Figura 42.** Tiempo de cálculo necesario para efectuar ajustes flexibles de proteínas con el modelo HA en función de la resolución y del porcentaje de ángulos fijado. Se han representado los resultados obtenidos fijando aleatoriamente el 0, 50 y 90% de los ángulos diedros ( $\phi$  y  $\psi$ ) mediante líneas azules de trazo continuo, discontinuo y punteado, respectivamente. Cada uno de los 46 ajustes de *MolMovDB* se han efectuado una sola vez (en lugar de 10) utilizando un PC Intel® Core<sup>2</sup>duo Q6600 a 2.4 GHz.

**Tabla 14.** Comparación entre los ajustes flexibles de proteínas obtenidos al fijar aleatoriamente distintos porcentajes de ángulos diedros con el modelo atómico C5 en función de la resolución.

Fijación <sup>a</sup> [%]	Resolución <sup>b</sup> [Å]	RMSD [Å]	Colisiones [%]	Rama. [%]	Enlaces [%]	Ángulos [%]	Molprobability [Å]
0%	8	1.08 ± 0.66	38 ± 17	0.7 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	10	1.09 ± 0.65	38 ± 17	0.7 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	15	1.15 ± 0.66	36 ± 15	0.7 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	20	1.24 ± 0.70	39 ± 17	0.7 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	25	1.33 ± 0.75	39 ± 16	0.7 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	<i>iMorph</i>	0.85 ± 0.43	35 ± 14	0.7 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
50%	8	1.12 ± 0.68	37 ± 16	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	10	1.13 ± 0.68	37 ± 15	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	15	1.18 ± 0.67	38 ± 15	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	20	1.26 ± 0.71	36 ± 13	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	25	1.37 ± 0.79	37 ± 14	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	<i>iMorph</i>	0.91 ± 0.48	36 ± 13	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
90%	8	1.47 ± 1.26	49 ± 21	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.4
	10	1.42 ± 1.03	49 ± 21	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.4
	15	1.47 ± 1.00	49 ± 20	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.4
	20	1.57 ± 1.06	48 ± 20	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.4
	25	1.69 ± 1.14	47 ± 20	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.5
	<i>iMorph</i>	1.22 ± 0.92	47 ± 20	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.4

<sup>a</sup> Porcentaje fijado de ángulos diedros  $\phi$  y  $\psi$ . Los ajustes se han llevado a cabo empleando el 5% de los modos.

<sup>b</sup> La descripción de estas métricas se detalla en el pie de la Tabla 11.

en el análisis de los resultados correspondientes a la eliminación del 90% de las variables con modelo HA (datos no mostrados). En el caso del modelo C5, si se calculan las desviaciones promedio sin tenerlos en cuenta se consiguen valores de 1.12 a 1.36Å que tan solo superan en ~0.2Å a los obtenidos sin fijar (0.96-1.18Å). En cuanto al modelo C $\alpha$ , el caso *IakeA\_4akeA* mejora hasta conseguir los ~1.5Å para todas las resoluciones. Nótese que tanto al fijar el 50% de los ángulos como sin fijarlos tan solo se alcanzan desviaciones de ~4Å con este modelo atómico. Además, el caso *IgggA\_IwdnA* con el mapa de 8Å también mejora respecto del ajuste sin fijar. Sin embargo, la calidad de las estructuras se ve afectada ligeramente al eliminar el 90% de las variables con los modelos atómicos más reduccionistas, pues las colisiones aumentan hasta ~49 con C5 y ~60 con C $\alpha$ , y el porcentaje de diedros fuera de las regiones permitidas disminuye levemente del 0.7 al 0.6% con ambos modelos. Por consiguiente, la eliminación del 90% de las variables con estos modelos afecta poco a las desviaciones de la mayoría de los ajustes y la calidad de las estructuras apenas disminuye.



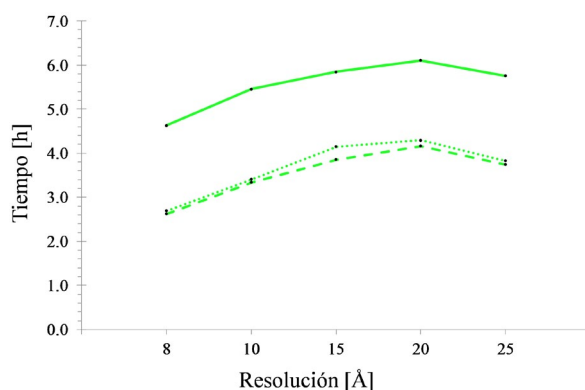
**Tabla 15.** Comparación entre los ajustes flexibles de proteínas obtenidos al fijar aleatoriamente distintos porcentajes de ángulos diedros con el modelo atómico C $\alpha$  en función de la resolución.

Fijación <sup>a</sup> [%]	Resolución <sup>b</sup> [Å]	RMSD [Å]	Colisiones [%]	Rama. [%]	Enlaces [%]	Ángulos [%]	Molprobability [Å]
0%	8	1.36 ± 0.95	64 ± 22	0.7 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.2 ± 0.5
	10	1.31 ± 0.68	70 ± 24	0.7 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.3 ± 0.5
	15	1.39 ± 0.68	67 ± 25	0.7 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.3 ± 0.5
	20	1.50 ± 0.73	65 ± 25	0.7 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.2 ± 0.5
	25	1.61 ± 0.79	61 ± 24	0.7 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.2 ± 0.5
	<i>iMorph</i>	0.97 ± 0.57	48 ± 18	0.7 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.5
50%	8	1.32 ± 0.81	55 ± 21	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.4
	10	1.31 ± 0.71	58 ± 23	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.2 ± 0.4
	15	1.37 ± 0.69	58 ± 22	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.2 ± 0.4
	20	1.48 ± 0.74	55 ± 21	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.4
	25	1.59 ± 0.81	54 ± 20	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.4
	<i>iMorph</i>	1.02 ± 0.55	45 ± 17	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.4
90%	8	1.56 ± 1.33	61 ± 26	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.2 ± 0.4
	10	1.54 ± 1.18	62 ± 26	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.2 ± 0.4
	15	1.57 ± 1.00	59 ± 24	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.2 ± 0.4
	20	1.67 ± 1.04	57 ± 23	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.4
	25	1.79 ± 1.10	55 ± 22	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.4
	<i>iMorph</i>	1.26 ± 1.04	51 ± 20	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.4

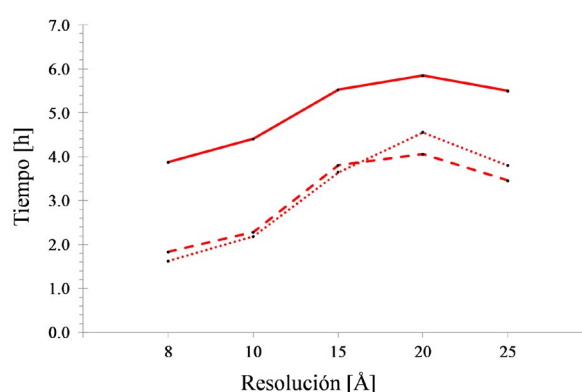
<sup>a</sup> Porcentaje fijado de ángulos diedros  $\phi$  y  $\psi$ . Los ajustes se han llevado a cabo empleando el 5% de los modos.

<sup>b</sup> La descripción de estas métricas se detalla en el pie de la Tabla 11.

Los tiempos necesarios para llevar a cabo los ajustes con los modelos C5 y C $\alpha$  fijando distintos porcentajes de ángulos diedros se muestran en las Figuras 43 y 44, respectivamente. En general, cuando se utiliza tanto el 10 como el 50% de las variables con ambos modelos se observan aceleraciones del 30 al 40% respecto de los correspondientes ajustes obtenidos sin fijar; sin embargo, en los casos de mayor resolución (8 y 10 Å) con el modelo C $\alpha$  los ajustes llegan a ser el doble de rápidos. Cuando se comparan los tiempos con los obtenidos a partir del modelo de mayor resolución (HA sin fijar), al emplear C5 las aceleraciones son del ~50% con los mapas de mayor (8 y 10 Å) y del ~40% con los de menor resolución ( $\geq 15$  Å). Si se utiliza C $\alpha$ , los tiempos de cálculo disminuyen 3 y 1.5 veces con los mapas de mayor y menor resolución, respectivamente. En consecuencia, el empleo del modelo atómico C $\alpha$  con la fijación aleatoria del 50% de los ángulos es la opción más rápida para obtener modelos ajustados con buena calidad estructural. Al igual que sucede con el modelo HA, las aceleraciones son debidas principalmente a la reducción en el tiempo dedicado al NMA y la velocidad es prácticamente la misma al fijar el 50 o el 90%. Por tanto, mientras que la fijación del 50% de los ángulos diedros



**Figura 43.** Tiempo de cálculo necesario para efectuar los ajustes de proteínas con el modelo C5 en función de la resolución y del porcentaje de ángulos fijado. Véase el pie de la Figura 42.



**Figura 44.** Tiempo de cálculo necesario para efectuar los ajustes de proteínas con el modelo Cα en función de la resolución y del porcentaje de ángulos fijado. Véase el pie de la Figura 42.

también permite acelerar los cálculos con los modelos atómicos C5 y Cα, la fijación del 90% tampoco merece la pena a no ser que, como se ha mencionado anteriormente, la macromolécula sea muy grande y el tamaño de las matrices del sistema suponga un problema.

#### 5.8.4. Dependencia de los ajustes con el número de modos

En la Tabla 16 se muestran los RMSD promedio entre los modelos finales y las estructuras con las que se generaron los mapas objetivo para los ajustes obtenidos con el 2, el 5 y el 10% de los modos de baja frecuencia y el modelo atómico HA. En todos los casos, las desviaciones disminuyen progresivamente conforme aumenta el número de modos empleado y aumentan suavemente al disminuir la resolución de los mapas. Por ejemplo, al utilizar el 10% de los modos, la desviación promedio alcanza valores de 1.00Å con los mapas de 8Å de resolución, y de 1.28Å con los de 25Å. Incluso empleando tan solo el 2% todavía es posible alcanzar desviaciones muy bajas, *i.e.* entre 1.17 y 1.49Å para todo el rango de resoluciones ensayado. Esto supone una disminución de únicamente unos 0.2-0.3Å respecto de las desviaciones obtenidas con el 10%. En consonancia con lo observado durante la validación del análisis vibracional (sección 5.2), estos resultados muestran que es posible utilizar una pequeña parte del espectro de modos (~2%) para obtener modelos atómicos compatibles con los mapas objetivo. En la Tabla 16 también se ha evaluado la calidad de las estructuras finales con *Molprobability*. Al emplear el 2% de los modos se obtienen las estructuras de mayor calidad, pues únicamente aparecen 33 colisiones y los ángulos fuera de las regiones permitidas de Ramachandran tan solo aumentan del 0.6% al ~0.8% para todas las resoluciones. Además, estos valores son prácticamente idénticos a los obtenidos en las simulaciones de transiciones conformacionales con *iMorph*. Sin embargo, con el 10% de los modos la calidad estructural disminuye ligeramente. En estos casos se alcanzan 37-39 colisiones y los porcentajes de ángulos en regiones no permitidas alcanzan del 0.8 al

1.0%. Estas distorsiones probablemente sean debidas a que los mapas a media/baja resolución no contienen suficiente información para distinguir qué conformaciones deben ser elegidas o rechazadas cuando éstas corresponden a los movimientos más localizados asociados a los modos de mayor frecuencia.

Se han extendido estos ensayos para comprobar si la utilización de distinto porcentaje de modos afecta de igual manera a los ajustes llevados a cabo con los modelos C $\alpha$  y C5 (datos no mostrados). Al emplear tanto el modelo C $\alpha$  como el C5, los RMSD finales también disminuyen progresivamente conforme se incluyen más modos, aunque las diferencias observadas entre las desviaciones obtenidas con el 2% y el 10% de los modos son únicamente de 0.1~0.2Å cuando se comparan las mismas resoluciones. En cuanto a la calidad estructural de los ajustes con C $\alpha$ , al reducir el número de modos del 10 al 2% el número de colisiones mejora de 80~90 hasta ~55. Con C5 las colisiones también siguen una tendencia descendente al considerar menos modos, alcanzando valores de ~41 y ~36 para el 10 y el 2% de los modos, respectivamente. Para estos modelos atómicos el porcentaje de ángulos fuera de las regiones permitidas también es similar (~0.7%) y no depende de la resolución de los mapas.

**Tabla 16.** Comparación entre los ajustes flexibles de proteínas obtenidos empleando distintos porcentajes de modos con el modelo atómico HA en función de la resolución.

Modos <sup>a</sup> [%]	Resolución <sup>b</sup> [Å]	RMSD [Å]	Colisiones [%]	Rama. [%]	Enlaces [%]	Ángulos [%]	Molprobability [Å]
2%	8	1.17 ± 0.71	33 ± 14	0.8 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	10	1.19 ± 0.70	33 ± 14	0.8 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	15	1.25 ± 0.72	33 ± 14	0.8 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	20	1.36 ± 0.78	33 ± 13	0.8 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	25	1.49 ± 0.86	33 ± 13	0.8 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	<i>iMorph</i>	1.00 ± 0.56	32 ± 12	0.8 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
5%	8	1.06 ± 0.65	35 ± 17	0.9 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	10	1.07 ± 0.64	35 ± 17	0.9 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	15	1.14 ± 0.66	36 ± 17	0.8 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	20	1.22 ± 0.70	35 ± 17	0.8 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	25	1.33 ± 0.76	35 ± 15	0.8 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	<i>iMorph</i>	0.85 ± 0.44	33 ± 14	0.8 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
10%	8	1.00 ± 0.62	39 ± 22	1.0 ± 0.9	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	10	1.01 ± 0.61	39 ± 22	1.0 ± 0.9	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	15	1.08 ± 0.62	39 ± 21	1.0 ± 0.9	0.2 ± 0.6	2.6 ± 4.7	3.1 ± 0.5
	20	1.17 ± 0.67	38 ± 20	0.9 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	25	1.28 ± 0.73	37 ± 18	0.8 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	<i>iMorph</i>	0.77 ± 0.38	35 ± 16	0.9 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5

<sup>a</sup> Porcentaje de modos.

<sup>b</sup> La descripción de estas métricas se detalla en el pie de la Tabla 11.

Independientemente del modelo atómico empleado, la utilización de un mayor número de modos mejora las desviaciones y deteriora ligeramente la calidad de los modelos finales. No obstante, las diferencias son pequeñas, *i.e.* 0.1~0.2Å al utilizar entre el 2 y el 10% de los modos. En cuanto a la calidad de las estructuras, las diferencias también son pequeñas. Al comparar las colisiones obtenidas a partir del 10% de los modos con las obtenidas utilizando el 2%, tan solo aparecen 5 colisiones más con el modelo C5 y unas 30 con Ca. El porcentaje de ángulos fuera de las regiones permitidas no depende significativamente ni del modelo atómico ni del número de modos empleado. En consecuencia, el ajuste flexible resulta muy robusto y eficaz también al utilizar distintos porcentajes de modos de baja frecuencia.

### 5.8.5. Dependencia de los ajustes con la orientación inicial

Para determinar el grado de dependencia de *iModfit* con la orientación entre el modelo atómico inicial y el mapa de densidad se han repetido los ajustes a partir de dos orientaciones iniciales diferentes: *MolMovDB* y *ADP\_EM*. Estas orientaciones generalmente corresponden a dos extremos distintos donde, o bien se superponen las regiones de las estructuras atómicas que permanecen estáticas durante el movimiento (*MolMovDB*), o bien se maximiza la correlación entre las densidades de las estructuras completas (*ADP\_EM*). En consecuencia, en el primer caso las desviaciones iniciales con las estructuras objetivo son mayores (7.57Å) que en el segundo (5.73Å). En la Tabla 17 se han representado los resultados de estos ajustes.

**Tabla 17.** Comparación del efecto de la orientación inicial en los ajustes flexibles de proteínas con el 5% de los modos y el modelo atómico HA.

Orientación Inicial <sup>a</sup>	Resolución <sup>b</sup> [Å]	RMSD [Å]	Colisiones [%]	Rama. [%]	Enlaces [%]	Angulos [%]	Molprobability [Å]
<i>MolMovDB</i>	8	1.06 ± 0.66	35 ± 17	0.9 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	10	1.07 ± 0.68	35 ± 17	0.9 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	15	1.14 ± 0.72	36 ± 17	0.8 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	20	1.22 ± 0.74	35 ± 17	0.8 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	25	1.33 ± 0.77	35 ± 15	0.8 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	Inicial	7.57 ± 4.34	21 ± 11	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	2.7 ± 0.6
<i>ADP_EM</i>	8	1.09 ± 0.99	34 ± 15	0.8 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	10	1.01 ± 0.55	33 ± 14	0.8 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	15	1.15 ± 0.61	35 ± 16	0.8 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	20	1.32 ± 0.75	36 ± 17	0.8 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	25	1.50 ± 0.86	35 ± 16	0.8 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	Inicial	5.73 ± 3.32*	21 ± 11	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	2.7 ± 0.6

<sup>a</sup> Método empleado para obtener la orientación inicial relativa.

<sup>b</sup> La descripción de estos valores se detalla en el pie de la Tabla 11.

\* Como al efectuar el ajuste rígido con *ADP\_EM* los valores iniciales varían ligeramente con la resolución de los mapas, únicamente se muestran los correspondientes a 10Å (el resto de resoluciones alcanzan valores similares).

Las desviaciones promedio de los modelos finales con las estructuras empleadas para simular los mapas son bajas en ambos casos. Al utilizar la orientación de *ADP\_EM* se obtienen desviaciones promedio de 1.09Å con los mapas de 8Å que aumentan progresivamente hasta 1.50Å a 25Å de resolución. Estos valores son muy similares a los obtenidos a partir de la orientación de *MolMovDB*, *i.e.* de 1.06 a 1.33Å con los mapas de 8 a 25Å, respectivamente. En dicha tabla, también se han representado los promedios de las métricas de validación estructural de *Molprobability* correspondientes. Como puede verse, la calidad de los modelos finales es muy buena independientemente de la orientación inicial utilizada. Con todo esto puede concluirse que *iModfit* es muy robusto en cuanto a la orientación inicial.

### 5.8.6. Validación de iModfit en casos simulados de ARN

Para validar el ajuste flexible de las estructuras atómicas de ARN se han realizado los ajustes de algunas de las secciones anteriores en condiciones similares, pero utilizando el conjunto de pruebas de ARN descrito en la sección 4.6.2 de Materiales y métodos. Cada caso de ajuste se ha repetido en 10 ocasiones utilizando *iModfit* con el 5% de los modos, el modelo atómico HA de ácidos nucleicos y los ángulos  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\epsilon$  y  $\zeta$  como variables.

En la Tabla 18 se muestran las desviaciones promedio entre los modelos finales y las estructuras con las que se generaron los mapas en función de la resolución. A pesar de la gran amplitud de las desviaciones iniciales ( $7.23 \pm 2.31$  Å), los valores finales son bastante bajos ( $\sim 2.5$ Å) en todo el rango de resoluciones y aumentan suavemente desde los 2.33Å con los mapas de 8Å de resolución hasta los 2.68Å con los de 25Å. Como sucede en proteínas, la resolución de los mapas afecta relativamente poco a las desviaciones finales ( $\sim 0.4$ Å). Sin embargo, las desviaciones obtenidas con ARN son significativamente mayores que con proteínas ( $\sim 1$ Å). En la Figura 45 se examinan estos casos individualmente. Aunque en la mayoría de ellos es posible alcanzar los  $\sim 2$ Å, en algunos las desviaciones resultan peores (3~4Å): *1a51A09\_1a51A03*, *1pnsW\_3fihY*, *3fihY\_1pnsW*, *2aw7A\_3elal* y *3elal\_2aw7A*, y en: *1p5oA06\_1p5oA12* y *1p5oA12\_1p5oA06* las desviaciones apenas disminuyen en comparación con sus valores iniciales. En los dos últimos casos, a pesar de que la posición final de los átomos no es

**Tabla 18.** Resultados promedio de los ajustes flexibles de ARN realizados con el 5% de los modos en función de la resolución.

Resolución <sup>a</sup> [Å]	RMSD [Å]	Colisiones [%]
8	$2.33 \pm 1.46$	$36 \pm 21$
10	$2.36 \pm 1.50$	$36 \pm 20$
15	$2.48 \pm 1.66$	$36 \pm 20$
20	$2.60 \pm 1.80$	$35 \pm 20$
25	$2.68 \pm 1.86$	$35 \pm 20$
Inicial	$7.23 \pm 2.31$	$32 \pm 20$
<i>iMorph</i>	$1.46 \pm 0.79$	$35 \pm 19$

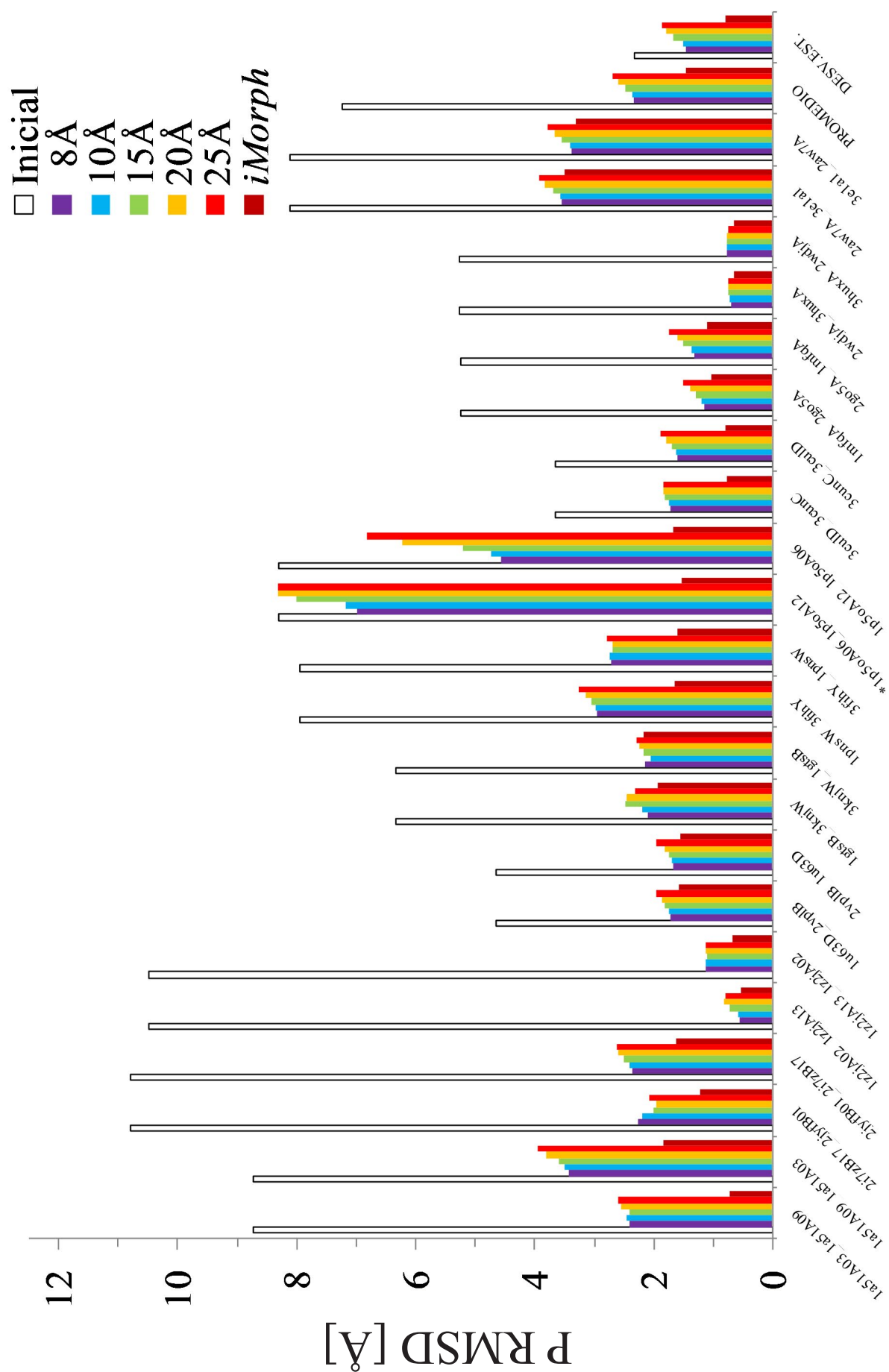
<sup>a</sup> La descripción de estos valores se detalla en el pie de la Tabla 11. En este caso, para el cálculo del RMSD solo se han tenido en cuenta los átomos P. En ARN no está disponible la resolución cristalográfica equivalente (*molprobability score*).

**Figura 45.** Desviaciones individuales de los ajustes flexibles con mapas simulados de ARN. En el caso marcado con (\*), como las desviaciones finales no mejoran a 20 y 25 Å de resolución, se han representado en su lugar los valores iniciales. Véase el pie de la Figura 38.

la adecuada, el modelo atómico se encuentra totalmente cubierto por el mapa de densidad (datos no mostrados). Todo esto puede deberse principalmente a dos factores, o bien el 5% de los modos no contiene los movimientos necesarios para alcanzar la estructura final, o bien la función de correlación no es capaz de conducir el ajuste adecuadamente. Para comprobar qué factor es el dominante, en dicha figura también se han incluido las desviaciones finales obtenidas con *iMorph* utilizando el mismo número de modos que en los ajustes. Al utilizar *iMorph*, sí es posible obtener desviaciones de 1 a 2 Å. Por lo tanto, lo más probable parece ser que estos ajustes quedan atrapados en un máximo local de la función de correlación que les impide adquirir las conformaciones finales. Este efecto también se aprecia, aunque en menor medida en otros casos: *1a51A03\_1a51A09*, *2i7zB17\_2jyfB01*, *2jyfB01\_2i7zB17*, *3culD\_3cunC* y *3cunC\_3culD*. Cabe destacar que las desviaciones finales obtenidas con proteínas y ARN no son totalmente comparables entre sí debido a la diferente naturaleza de las estructuras y al empleo de átomos distintos en el cálculo de las desviaciones (en ARN se emplean los átomos de P en lugar de los C<sub>α</sub>). En cuanto a la calidad estructural, las colisiones apenas aumentan de las 32 iniciales hasta 35~36 en todo el rango de resoluciones y, además, son prácticamente iguales a las obtenidas en la simulación de transiciones conformacionales (*iMorph*). Por tanto, también es posible conseguir con ácidos nucleicos modelos ajustados de calidad utilizando los modos de baja frecuencia en CI.

También se han cronometrado estos ajustes para evaluar su velocidad. Debido al pequeño tamaño de la mayoría de los casos del conjunto de pruebas de ARN, la mayoría de ellos tardan menos de 6 minutos. Sin embargo, los cuatro casos de mayor tamaño son mucho más lentos, *i.e.* los ajustes de *2wdjA/3huxA* (721 nucleótidos) tardan unas 2 horas y los de *2aw7A/3e1a1* (1529 nucleótidos) unas 20. En estos casos la mayor parte del tiempo de cálculo se invierte en el NMA (70~80%). Esta lentitud se debe principalmente al gran número de variables considerado en estas estructuras, *i.e.* ~3600 en *2wdjA/3huxA* y ~7600 en *2aw7A/3e1a1*. No obstante, en estos casos es posible acelerar los cálculos sin afectar mucho a la calidad de los ajustes incrementando el umbral para recalcular los modos de 0.1 Å (por defecto) hasta 1 Å. De este modo se consiguen tiempos de 15-45 minutos con *2wdjA/3huxA* y 4-8 horas con *2aw7A/3e1a1* según el sentido del ajuste. En cuanto a los casos más pequeños, la velocidad apenas se ve afectada debido a que el tiempo dedicado al NMA representa menos del 3% del total. En la Tabla 19 se recogen los resultados obtenidos en estas condiciones. Como puede apreciarse en esta tabla, las desviaciones no se ven afectadas y el número de colisiones aumenta tan solo de 35~36 a 38~42.







La reducción del número de variables durante los ajustes de ácidos nucleicos también permite obtener buenos modelos finales. Los resultados promedio obtenidos para los 22 casos ensayados se muestran en la Tabla 20. Al fijar aleatoriamente el 50% de los ángulos diedros, tanto las desviaciones como el número de colisiones son idénticos a los obtenidos sin fijar, incluso en los casos individuales (datos no mostrados). Sin embargo, al fijar el 90%, las colisiones aumentan significativamente hasta ~57, mientras que las desviaciones tan solo son ~0.2Å mayores. Por tanto, el incremento de la rigidez del sistema afecta negativamente a la calidad de los modelos de ARN tan solo cuando se elimina más del ~50% de las variables. Estos resultados son equivalentes a los obtenidos en proteínas (véase la Tabla 13).

**Tabla 19.** Resultados promedio de los ajustes flexibles de ARN realizados con el 5% de los modos y 1Å de umbral de re-diagonalización en función de la resolución.

Resolución <sup>a</sup> [Å]	RMSD [Å]	Colisiones [%]
8	2.29 ± 1.40	42 ± 27
10	2.32 ± 1.42	41 ± 26
15	2.35 ± 1.59	38 ± 23
20	2.53 ± 1.73	40 ± 24
25	2.67 ± 1.81	40 ± 23
Inicial	7.23 ± 2.31	32 ± 20
<i>iMorph</i>	1.46 ± 0.79	39 ± 23

<sup>a</sup> Véase el pie de la Tabla 11.

**Tabla 20.** Comparación entre los ajustes flexibles de ARN obtenidos al fijar aleatoriamente distintos porcentajes de ángulos diedros en función de la resolución.

Fijación <sup>a</sup> [%]	Resolución <sup>b</sup> [Å]	RMSD [Å]	Colisiones [%]
0%	8	2.33 ± 1.46	36 ± 21
	10	2.36 ± 1.50	36 ± 20
	15	2.48 ± 1.66	36 ± 20
	20	2.60 ± 1.80	35 ± 20
	25	2.68 ± 1.86	35 ± 20
	<i>iMorph</i>	1.46 ± 0.79	35 ± 19
50%	8	2.33 ± 1.42	36 ± 22
	10	2.36 ± 1.42	36 ± 22
	15	2.47 ± 1.58	36 ± 23
	20	2.58 ± 1.72	35 ± 22
	25	2.67 ± 1.77	35 ± 23
	<i>iMorph</i>	1.52 ± 0.82	35 ± 23
90%	8	2.51 ± 1.26	57 ± 33
	10	2.54 ± 1.30	57 ± 33
	15	2.64 ± 1.42	57 ± 34
	20	2.78 ± 1.53	57 ± 34
	25	2.91 ± 1.66	52 ± 30
	<i>iMorph</i>	1.67 ± 0.89	52 ± 31

<sup>a</sup> Porcentaje fijado de ángulos diedros  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\epsilon$  y  $\zeta$ . Los ajustes se han llevado a cabo empleando el 5% de los modos.

<sup>b</sup> La descripción de estos valores se detalla en el pie de la Tabla 11.

Para estudiar el aumento en la velocidad asociado a la reducción del número de variables se han cronometrado los 22 ajustes de ARN (datos no mostrados). Igual que antes, al reducir el número de las variables, los cuatro casos de mayor tamaño tardan mucho más que el resto. Al utilizar el 50% de las variables, los tiempos se reducen entre 2 y 4 veces, *i.e.* los casos de *2wdjA* y *3huxA* tardan de 25 a 45 minutos y los de *2aw7A* y *3elal* entre 4 y 8 horas según el sentido del ajuste. Si se reduce aún más el porcentaje de variables hasta el 90% apenas se producen mejoras. Esto se debe fundamentalmente a que cuando el número de átomos es muy grande en comparación con el número de variables (véase la sección 5.8.3), el coste computacional del NMA solo depende del número de átomos. En cambio, los tiempos del resto de casos apenas varían al reducir el número de variables porque el tiempo dedicado al NMA representa menos del 15% del total (datos no mostrados). Según estos resultados, la fijación aleatoria del 50% de los ángulos diedros constituye un método eficaz para acelerar los ajustes de las estructuras de ARN de gran tamaño sin afectar a la calidad de los modelos.

Para comprobar si los ácidos nucleicos se comportan de forma similar a las proteínas cuando se emplean distintos porcentajes de modos se han repetido los ensayos de la sección 5.8.4 con los casos de ARN. En la Tabla 21 se incluyen los resultados obtenidos con el 2, 5

**Tabla 21.** Comparación entre los ajustes flexibles de ARN obtenidos con distintos porcentajes de modos en función de la resolución.

Modos <sup>a</sup> [%]	Resolución <sup>b</sup> [Å]	RMSD [Å]	Colisiones [‰]
2%	8	2.38 ± 1.40	39 ± 23
	10	2.41 ± 1.43	39 ± 23
	15	2.52 ± 1.54	39 ± 22
	20	2.64 ± 1.72	38 ± 22
	25	2.73 ± 1.78	40 ± 23
	<i>iMorph</i>	1.64 ± 0.86	38 ± 22
5%	8	2.33 ± 1.46	36 ± 21
	10	2.36 ± 1.50	36 ± 20
	15	2.48 ± 1.66	36 ± 20
	20	2.60 ± 1.80	35 ± 20
	25	2.68 ± 1.86	35 ± 20
	<i>iMorph</i>	1.46 ± 0.79	35 ± 19
10%	8	2.25 ± 1.41	44 ± 26
	10	2.29 ± 1.45	44 ± 26
	15	2.39 ± 1.62	43 ± 26
	20	2.52 ± 1.78	41 ± 24
	25	2.64 ± 1.85	41 ± 24
	<i>iMorph</i>	1.35 ± 0.73	40 ± 24

<sup>a</sup> Porcentaje de modos.

<sup>b</sup> La descripción de estos valores se detalla en el pie de la Tabla 11.

y 10% de los modos de baja frecuencia. Aunque las desviaciones son menores conforme se emplean más modos, las mejoras son muy pequeñas ( $\sim 0.1 \text{ \AA}$ ) en todo el rango de resoluciones. Respecto a la calidad de los modelos, se alcanzan de 39 a 41 colisiones por cada 1000 átomos al utilizar el 2% de los modos y hasta 41-44 con el 10%, lo que supone un ligero incremento en comparación con los valores obtenidos con el 5% de los modos ( $\sim 36$ ). De igual modo que en los casos de proteínas (sección 5.8.4), las distorsiones observadas con el 10% de los modos probablemente sean debidas a que se aceptan como válidos algunos movimientos locales a pesar de que los mapas a media/baja resolución no son tan detallados como para discriminar si estos movimientos son adecuados. Sin embargo, a diferencia de lo que sucede en proteínas, con el 2% de los modos las colisiones aumentan ligeramente, aunque estos incrementos son muy pequeños y están dentro de la incertidumbre experimental.

### 5.8.7. Comparación con otros métodos basados en análisis vibracional

Con el fin de comparar *iModfit* con otros métodos automáticos de ajuste flexible basados en NMA y CC se han efectuado los 46 ajustes de proteínas de *MolMovDB* utilizando los programas *NMFF* (Tama, *et al.*, 2004a) y *NORMA* (Suhre, *et al.*, 2006). En la Tabla 22 se comparan los resultados obtenidos con estas herramientas utilizando los mapas de  $10 \text{ \AA}$  y el modelo atómico HA. Para llevar a cabo la comparación se han calculado las desviaciones de los modelos finales con las estructuras empleadas en la simulación de los mapas y se han obtenido las métricas de *Molprobability* correspondientes.

Al utilizar *NMFF* con los parámetros por defecto (*NMFF\**) se observan desviaciones relativamente elevadas, por lo que se ha procedido a optimizarlos. Después de probar sistemáticamente distintas combinaciones de parámetros, se ha determinado que las mejores

**Tabla 22.** Comparación de *iModfit* con otros métodos de ajuste flexible basados en NMA y CC.

Programa <sup>a</sup>	RMSD <sup>b</sup> [ $\text{\AA}$ ]	Colisiones <sup>c</sup> [%]	Rama. [%]	Enlaces [%]	Angulos [%]	Molprobability [ $\text{\AA}$ ]
<i>NMFF*</i>	$2.52 \pm 2.93$	$27 \pm 12$	$1.0 \pm 0.8$	$11.7 \pm 11.8$	$13.4 \pm 11.6$	$2.9 \pm 0.5$
<i>NMFF</i>	$1.38 \pm 2.16$	$31 \pm 21$	$1.4 \pm 1.2$	$17.1 \pm 11.1$	$18.6 \pm 10.8$	$3.0 \pm 0.5$
<i>NORMA</i>	$2.60 \pm 2.50$	$82 \pm 72$	$1.9 \pm 1.6$	$43.2 \pm 23.9$	$20.7 \pm 10.5$	$3.4 \pm 0.5$
<i>iModfit</i>	$1.07 \pm 0.64$	$35 \pm 17$	$0.9 \pm 0.8$	$0.2 \pm 0.6$	$2.6 \pm 4.7$	$3.0 \pm 0.5$
<i>Inicial</i>	$7.57 \pm 4.34$	$21 \pm 11$	$0.6 \pm 0.6$	$0.2 \pm 0.6$	$2.6 \pm 4.7$	$2.7 \pm 0.6$

<sup>a</sup> Programa utilizado en los 46 problemas de ajuste flexible de proteínas de *MolMovDB*. Los ajustes se han llevado a cabo con los mapas de  $10 \text{ \AA}$  de resolución y el modelo atómico HA. Tanto con *NMFF* como con *NORMA*, se han empleado los parámetros optimizados (véase texto), y con *iModfit* los parámetros por defecto. En el caso de *NMFF* también se han obtenido los resultados con los parámetros por defecto (\*).

<sup>b</sup> RMSD promedio entre los modelos finales de los ajustes y las estructuras empleadas para generar los mapas objetivo. Únicamente se han tenido en cuenta los átomos  $C_{\alpha}$ .

<sup>c</sup> Estas métricas se describen en el pie de la Tabla 11.

desviaciones se obtienen incrementando el número de modos de 20 a 30 y reduciendo el factor de convergencia de 1 a 0.001 (véase la sección 4.7 de Materiales y métodos). En el caso de *NORMA*, aunque los parámetros deben ajustarse para cada problema particular, se han ensayado distintas condiciones para determinar un conjunto válido en todos los casos. Estos parámetros corresponden al empleo de 20 modos, una amplitud estimada de 500, 10 pasos de recocido simulado y 15 rondas de minimización de 25 pasos cada una. Con *iModfit* se utilizan los parámetros por defecto.

En la Tabla 22 se observa que con *NMFF* también es posible conseguir desviaciones promedio muy bajas ( $\sim 1.4\text{\AA}$ ), aunque resultan algo más altas que las de *iModfit* ( $\sim 1.1\text{\AA}$ ). Sin embargo, *NORMA* obtiene resultados significativamente peores ( $\sim 2.6\text{\AA}$ ). En la Figura 46 se examinan estos resultados individualmente. En ella, puede apreciarse que las desviaciones de *NORMA* (rojo) son más grandes que las obtenidas con los otros métodos en prácticamente todos los casos. Además, éstas tienden a ser mucho mayores en los ajustes de gran amplitud, como por ejemplo, los correspondientes a las estructuras *la8eA/lbp5A*, *loelA/lx4A* y *lsu4A/lt5sA*. Estos resultados sugieren que la actualización de los modos puede ser determinante en el éxito de los ajustes. Por otra parte, todas las diferencias individuales de RMSD entre *iModfit* (azul) y *NMFF* (verde) son pequeñas excepto en tres casos: *lgggA\_lwdnA*, *lwdnA\_lgggA* y *lsx4A\_loelA*. En los dos primeros, las desviaciones resultan  $1\text{--}2\text{\AA}$  mayores con *iModfit*. En estos casos, los modelos finales obtenidos con *iModfit* quedan atrapados en mínimos locales de la función de correlación, dado que el control de *iMorph* sí es capaz de alcanzar desviaciones menores de  $1\text{\AA}$  al emplear las estructuras atómicas y el RMSD como función de coste. Sin embargo, en el tercer caso, mientras que *NMFF* no consigue disminuir prácticamente la desviación inicial, *iModfit* obtiene  $\sim 1.5\text{\AA}$ .

En cuanto a la calidad de las estructuras finales, en la Tabla 22 también se han incluido las métricas de *Molprobability*. Aunque con *NORMA* se obtiene un número de colisiones bastante elevado (82), con *NMFF* este número es muy bajo (31), mejorando incluso el obtenido con *iMorph* (35). Sin embargo, el porcentaje de longitudes de enlace incorrectos crece enormemente hasta alcanzar valores superiores al 17% con *NMFF* y al 43% con *NORMA*. El porcentaje de ángulos de enlace incorrectos también aumenta mucho hasta alcanzar el  $\sim 20\%$  con estos métodos. Por tanto, al utilizar estas aproximaciones basadas en CC la estructura covalente de los modelos finales se ve seriamente distorsionada. Sin embargo, la estructura secundaria se mantiene bastante bien dado que el porcentaje de ángulos diedros fuera de las regiones permitidas de Ramachandran tan solo aumenta con moderación, *i.e.* hasta el 1.4% y el 1.9% para *NMFF* y *NORMA*, respectivamente.

**Figura 46.** Desviaciones individuales obtenidas por las herramientas de ajuste flexible basadas en NMA en los casos de ajuste con mapas simulados de proteínas. En el eje de abscisas se indica el problema de ajuste mediante el código del PDB de la estructura atómica inicial seguido del código de la estructura empleada en la simulación del mapa. En la leyenda se muestran los colores correspondientes a cada herramienta. En el caso de *NMFF* también se muestran las desviaciones obtenidas con los parámetros por defecto (*NMFF\**).

Estos resultados evidencian que con las aproximaciones basadas en NMA en CC es posible obtener modelos similares a los empleados para generar los mapas, aunque la baja calidad de las estructuras compromete su posterior utilización. Sin embargo, *iModfit* permite obtener, gracias al NMA en CI, los mejores modelos finales tanto en calidad estructural como en similitud con las estructuras a partir de las cuales se generaron los mapas.

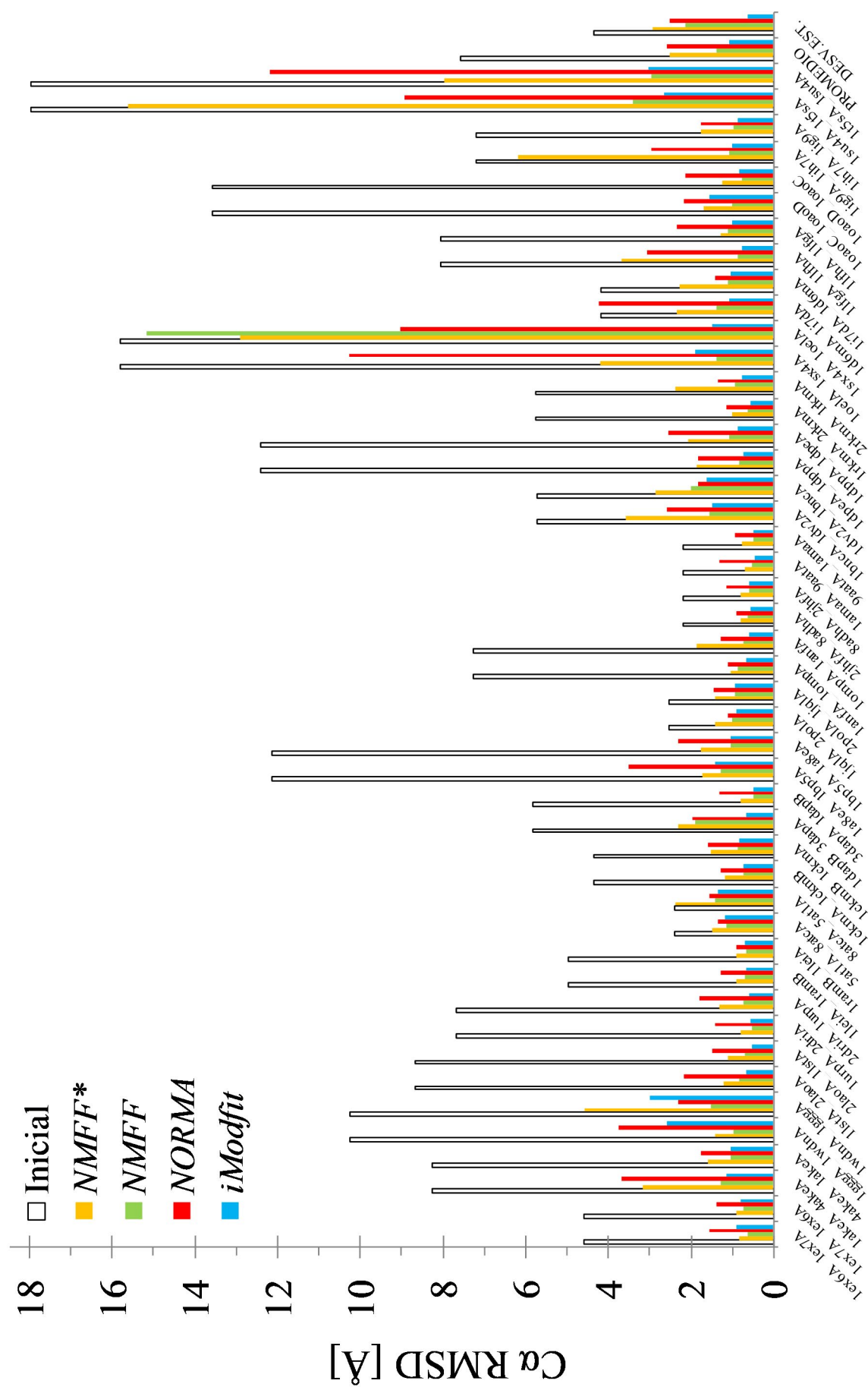
Para estudiar cómo afecta la resolución de los mapas a la calidad de los ajustes al utilizar un método basado en NMA y CC, en la Tabla 23 se recogen los resultados de *NMFF* con los mapas simulados a 8, 10, 15, 20 y 25 Å de resolución. Con *NMFF* las desviaciones aumentan gradualmente de 1.20 a 1.91 Å con los mapas de 8 a 25 Å, respectivamente. La calidad de las estructuras se mantiene constante en todo el rango de resoluciones excepto a 25 Å, donde los promedios empeoran debido únicamente al caso de *lakeA\_4akeA* (datos no mostrados). Estos resultados son similares a los obtenidos anteriormente con *iModfit* (véase la sección 5.8.1), aunque las desviaciones son de 0.2 a 0.6 Å más elevadas y la estructura covalente se ha deteriorado significativamente. Por tanto, puede concluirse que los métodos de ajuste flexible basados en el análisis vibracional resultan bastante robustos en un amplio rango de resoluciones, independientemente del sistema de coordenadas utilizado.

**Tabla 23.** Comparación entre los ajustes flexibles de proteínas llevados a cabo con *NMFF* a distintas resoluciones.

Método <sup>a</sup>	Resolución <sup>b</sup> [Å]	RMSD [Å]	Colisiones [%]	Rama. [%]	Enlaces [%]	Angulos [%]	Molprobability [Å]
<i>NMFF</i>	8	1.20 ± 0.96	29 ± 13	1.3 ± 0.9	16.6 ± 10.2	18.4 ± 9.7	3.0 ± 0.5
	10	1.38 ± 2.16	31 ± 21	1.4 ± 1.2	17.1 ± 11.1	18.6 ± 10.8	3.0 ± 0.5
	15	1.50 ± 1.98	33 ± 20	1.5 ± 1.2	19.7 ± 13.1	21.3 ± 13.4	3.0 ± 0.5
	20	1.57 ± 1.92	30 ± 17	1.3 ± 1.0	17.4 ± 10.8	19.4 ± 10.7	3.0 ± 0.5
	25	1.91 ± 2.36	42 ± 55	1.6 ± 1.7	22.6 ± 18.2	22.9 ± 14.9	3.1 ± 0.5
Inicial	-	7.57 ± 4.34	21 ± 11	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	2.7 ± 0.6

<sup>a</sup> Método de ajuste. Estos ajustes se han efectuado a partir de la orientación inicial de *MolMovDB* utilizando *NMFF* con los parámetros optimizados. También se incluyen los valores correspondientes a los modelos iniciales.

<sup>b</sup> La descripción de estas métricas se detalla en el pie de la Tabla 11.





### 5.8.8. Comparación con métodos basados en dinámica molecular

Se ha comparado *iModfit* con los siguientes cinco métodos de ajuste flexible basados en MD: *YUPSCX* (Tan, *et al.*, 2008), *MDFF* (Trabuco, *et al.*, 2008), Tama 1 (Orzechowski y Tama, 2008), Tama 2 (Grubisic, *et al.*, 2010) y *Flex-EM* (Topf, *et al.*, 2008). La comparación con este último, al tratarse de una aplicación específica para el ajuste de estructuras procedentes de modelado por homología, se realizará en la siguiente sección. Los detalles de estos métodos se han descrito brevemente en la sección 3.5.4 de la Introducción.

Desafortunadamente, algunas de estas aproximaciones de ajuste flexible, o bien no se encuentran disponibles públicamente como herramientas automatizadas, o bien son demasiado costosas computacionalmente como para ser utilizadas sistemáticamente con nuestros conjuntos de pruebas. Los métodos *Flex-EM*, *MDFF* y Tama 1 emplean tanto campos de fuerza como representaciones detalladas por lo que resultan demasiado lentos. Por ejemplo, se tardan unas 100 horas en completar el ajuste de una proteína de ~200 aminoácidos en un mapa de 10Å de resolución utilizando *Flex-EM* con una única CPU (Topf, *et al.*, 2008). Por otra parte, el método Tama 2 (basado en un modelo reduccionista), aunque consigue tiempos de ejecución aceptables, *i.e.* ~1 h con proteínas de ~1000 aminoácidos, no se encuentra disponible públicamente. El único método basado en MD que es lo suficientemente rápido para ser utilizado sistemáticamente y a la vez está disponible como herramienta es *YUPSCX*. Por todo ello, se han adoptado diferentes estrategias para comparar cada uno de estos métodos con *iModfit*.

La comparación sistemática de *YUPSCX* con *iModfit* se ha llevado a cabo utilizando los 46 casos de ajuste de proteínas de *MolMovDB* con mapas simulados a distintas resoluciones. En esta ocasión se han utilizado las orientaciones iniciales procedentes del ajuste rígido de las estructuras atómicas en los mapas objetivo con *ADP\_EM*. En la Tabla 24 se recogen los resultados obtenidos empleando *YUPSCX* con sus parámetros por defecto. En cuanto a las desviaciones entre los modelos finales y las estructuras objetivo utilizadas para simular los mapas, se obtienen valores promedio de 1.27 a 2.64Å con *YUPSCX* y de 1.01 a 1.50Å con *iModfit* en el rango de resoluciones de 8 a 25Å. En la Figura 47 se muestran los resultados de *YUPSCX* para cada caso individual. Como era de esperar, conforme disminuye la resolución las desviaciones aumentan progresivamente. La única excepción es el caso *Isx4A\_IoelA*, en el cual tanto a 15 como a 25Å se obtienen desviaciones mucho mayores que con el resto de resoluciones. Por todo ello, aunque *YUPSCX* es capaz de obtener modelos con desviaciones pequeñas respecto de las estructuras objetivo, los resultados son mejores con *iModfit*, especialmente en los casos de menor resolución. Para evaluar la calidad de los modelos en la Tabla 24 se incluyen las métricas de validación estructural de *Molprobability*. En los modelos producidos por *YUPSCX* las colisiones

**Tabla 24.** Comparación entre los ajustes flexibles de proteínas llevados a cabo con *YUPSCX* e *iModfit* a distintas resoluciones.

Método <sup>a</sup>	Resolución <sup>b</sup> [Å]	RMSD [Å]	Colisiones [%]	Rama. [%]	Enlaces [%]	Angulos [%]	Molprobability [Å]
<i>YUPSCX</i>	8	1.27 ± 0.60	37 ± 6	0.6 ± 0.6	0.6 ± 0.7	7.6 ± 6.1	3.1 ± 0.4
	10	1.33 ± 0.58	41 ± 6	0.6 ± 0.6	0.6 ± 0.8	8.0 ± 6.3	3.1 ± 0.4
	15	1.68 ± 0.83	46 ± 7	0.7 ± 0.6	0.6 ± 0.7	8.6 ± 6.4	3.2 ± 0.4
	20	1.99 ± 0.65	49 ± 7	0.7 ± 0.7	0.7 ± 0.7	8.7 ± 6.3	3.2 ± 0.4
	25	2.64 ± 1.26	50 ± 8	0.7 ± 0.7	0.7 ± 0.9	8.6 ± 6.4	3.2 ± 0.4
<i>iModfit</i>	8	1.09 ± 0.99	34 ± 15	0.8 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	10	1.01 ± 0.55	33 ± 14	0.8 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	15	1.15 ± 0.61	35 ± 16	0.8 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	20	1.32 ± 0.75	36 ± 17	0.8 ± 0.8	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
	25	1.50 ± 0.86	35 ± 16	0.8 ± 0.7	0.2 ± 0.6	2.6 ± 4.7	3.0 ± 0.5
Inicial	-	5.73 ± 3.32*	21 ± 11	0.6 ± 0.6	0.2 ± 0.6	2.6 ± 4.7	2.7 ± 0.6

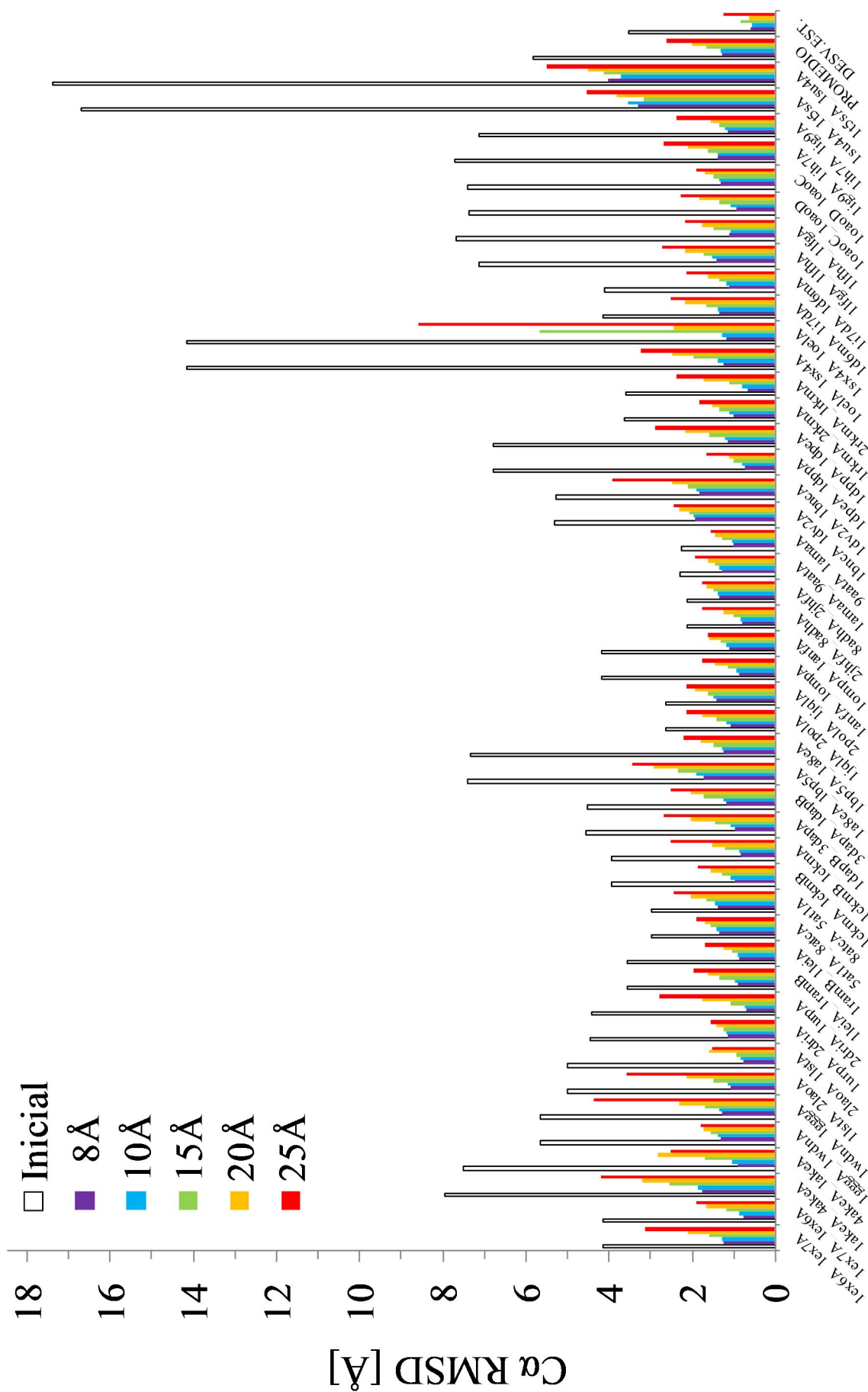
<sup>a</sup> Método de ajuste. Estos ajustes se han realizado utilizando *YUPSCX* e *iModfit* con las opciones por defecto después de ajustar rígidamente las estructuras atómicas en el interior de los mapas de densidad con *ADP\_EM*. También se incluyen los valores correspondientes a los modelos iniciales.

<sup>b</sup> La descripción de estas métricas se detalla en el pie de la Tabla 10.

\* Como al efectuar el ajuste rígido con *ADP\_EM* los valores iniciales varían ligeramente con la resolución de los mapas, únicamente se muestran los correspondientes a 10Å (el resto obtienen valores similares).

aumentan progresivamente de 37 con los mapas de 8Å hasta las 50 en los de 25Å. Por el contrario, con *iModfit* el número de colisiones permanece constante y es sensiblemente menor (34~36) en todo el rango de resoluciones. En ambas aproximaciones los porcentajes de ángulos fuera de las regiones permitidas de Ramachandran permanecen bajos independientemente de la resolución con valores del 0.6-0.7% para *YUPSCX* y del 0.8% para *iModfit*. Sin embargo, al utilizar *YUPSCX* se triplican los porcentajes iniciales de longitudes (0.2%) y ángulos (2.6%) de enlace incorrectos, alcanzándose valores del 0.6~0.7% y del 7.6-8.7%, respectivamente. En cambio, con *iModfit* se mantienen los valores basales gracias a la utilización de los ángulos diedros como CI. Además, la resolución cristalográfica equivalente es ligeramente mejor con *iModfit*, *i.e.* 3.0Å frente a los 3.1-3.2Å de *YUPSCX*. Teniendo en cuenta las desviaciones y todas las métricas de validación estructural, es posible afirmar que en los ajustes con mapas simulados los modelos finales producidos con *iModfit* son ligeramente mejores que los obtenidos con *YUPSCX* tanto por su parecido con los modelos utilizados para generar los mapas como por la calidad de las estructuras.

La comparación con *MDFF* se lleva a cabo utilizando el conjunto de pruebas que se muestra en la Tabla S1 de Trabuco *et al.* (Trabuco, *et al.*, 2008). Para ello, se han efectuado los ajustes con *iModfit* a distintas resoluciones empleando los parámetros por defecto (modelo HA con el 5% de modos). Para reproducir las condiciones iniciales en las que fue validado



◀ **Figura 47.** Desviaciones individuales finales obtenidas por *YUPSCX* en los casos de ajuste flexible con mapas simulados de proteínas. Véase el pie de la Figura 38.

dicho método de MD se han ajustado las estructuras iniciales en el interior de los mapas con *ADP\_EM* (véase la sección 4.6.2 de Materiales y métodos). En la Tabla 25 se recogen las desviaciones de estos ajustes al utilizar mapas de 10 y 15 Å de resolución y un muestreo de 2 Å por voxel. No se han incluido en la comparativa los resultados obtenidos a 5 Å por estar fuera del rango accesible por SP. Los valores procedentes de *MDFF* fueron tomados de la Tabla S1 del material suplementario de Trabuco *et al.* y corresponden al RMSD promedio de los últimos 200 ps de las trayectorias. Sin embargo, las desviaciones obtenidas con *iModfit* corresponden al promedio de los modelos finales de 10 ejecuciones distintas. Dado que para *MDFF* no se indica el sentido del ajuste, *i.e.* qué estructura es la inicial y cuál la objetivo, con *iModfit* se han mostrado los resultados de los ajustes en ambos sentidos. Como puede apreciarse

**Tabla 25.** Comparación entre *iModfit* y *MDFF* utilizando ajustes con mapas simulados.

PDB <sup>a</sup>		N <sup>b</sup>	Res. <sup>c</sup>	<i>MDFF</i> <sup>d</sup>		<i>iModfit</i> (X→Y) <sup>e</sup>		<i>iModfit</i> (Y→X)	
X	Y		[Å]	I [Å]	F [Å]	I [Å]	F [Å]	I [Å]	F [Å]
1pvi A	1pvu A	154	10	3.54	0.96	3.47	1.14	3.47	1.20
			15	3.73	1.49	3.47	1.27	3.46	1.33
1f3y* A	1jkn* A	165	10	3.82	2.53	4.34	1.97	4.39	1.85
			15	3.96	2.80	4.78	2.30	4.81	2.08
1g6n A	1g6n B	200	10	2.79	1.17	3.21	0.62	3.21	0.57
			15	2.76	1.86	3.22	0.66	3.22	0.66
1ake A	4ake A	214	10	8.19	1.24	8.00	1.18	7.45	1.28
			15	7.64	1.79	8.89	1.76	8.98	2.27
1ckn A	1ckn B	316	10	3.71	1.26	3.98	0.76	3.93	0.82
			15	3.58	2.04	3.92	1.00	3.87	0.90
1bnc A	1dv2 A	452	10	4.59	1.62	5.40	1.51	5.36	1.64
			15	4.10	2.06	5.33	1.69	5.30	1.92
1dpe A	1dpp A	507	10	12.22	1.22	6.78	0.61	6.79	0.78
			15	6.97	1.65	6.77	0.67	6.78	0.90
Promedios			10	5.55	1.43	5.03	1.11	4.94	1.16
			15	4.68	1.96	5.20	1.33	5.20	1.44

<sup>a-c</sup> Código PDB y cadena de las estructuras empleadas (<sup>a</sup>), número de aminoácidos (<sup>b</sup>) y resolución (<sup>c</sup>).

<sup>d</sup> RMSD de los modelos iniciales (I) y finales (F) con las estructuras empleadas para generar los mapas. Estos valores proceden de la Tabla S1 de Trabuco *et al.* (Trabuco, *et al.*, 2008). Los valores finales corresponden al promedio de las desviaciones obtenidas con las estructuras de los últimos 200ps de las trayectorias de *MDFF*. La orientación inicial fue obtenida con el programa de ajuste rígido *SITUS* utilizando los parámetros por defecto. En los cálculos solo se han tenido en cuenta los átomos del esqueleto.

<sup>e</sup> Desviaciones obtenidas con *iModfit* al ajustar la conformación X en el mapa de la conformación Y, y *viceversa*. En estos casos, la orientación inicial se ha obtenido con el programa de ajuste rígido *ADP\_EM* (véase Materiales y Métodos) y las desviaciones se han calculado a partir de los átomos C<sub>α</sub>.

\* Estas estructuras proceden de RMN (solo se ha utilizado el primer modelo).

en la Tabla 25, los dos métodos alcanzan desviaciones promedio bastante bajas. Con *MDFF* es posible obtener promedios de 1.43Å con los mapas de 10Å y de 1.96Å con los de 15Å. Con *iModfit*, estos valores son algo más bajos, *i.e.* de ~1.1 y ~1.4Å al emplear los mapas de 10 y 15Å, respectivamente. Por tanto, ambas aproximaciones permiten obtener modelos muy parecidos a las estructuras con las que se generaron los mapas, aunque con *iModfit* las desviaciones son ligeramente mejores.

Se ha comparado con *iModfit* el método Tama 1 utilizando los 5 pares de estructuras de proteínas en distintas conformaciones con los que fue validado este último. Para ello, se han efectuado dichos ajustes empleando *iModfit* con los parámetros por defecto. En la Tabla 26 se muestran las desviaciones obtenidas con ambos métodos. Los datos correspondientes al método Tama 1 fueron tomados de la Tabla 4 de Orzechowski y Tama (Orzechowski y Tama, 2008). Solo se han utilizado los resultados obtenidos a 8Å por ser los únicos en el rango accesible por SP. Como en esta referencia tampoco se indica el sentido del ajuste, también se han incluido los resultados de *iModfit* en ambos sentidos. Para reproducir la orientación inicial utilizada en estos casos se ha minimizado previamente el RMSD entre las estructuras atómicas iniciales y las correspondientes a los mapas objetivo. Como puede observarse en la Tabla 26, las desviaciones promedio de *iModfit* (0.72Å) son significativamente menores que las correspondientes al método Tama 1 (1.50Å). Además, con *iModfit* no solo los promedios son más bajos, sino que también lo son las desviaciones individuales en ambos sentidos.

También se ha comparado *iModfit* con la aproximación Tama 2 que está basada en un modelo reducido de Go. Para ello, se han llevado a cabo con *iModfit* los mejores ajustes

**Tabla 26.** Comparación entre *iModfit* y el método Tama 1 utilizando ajustes con mapas simulados.

PDB <sup>a</sup>		N	Res. [Å]	Tama 1 <sup>b</sup>		<i>iModfit</i> (X→Y) <sup>c</sup>		<i>iModfit</i> (Y→X)	
X	Y			I [Å]	F [Å]	I [Å]	F [Å]	I [Å]	F [Å]
1ake A	4ake A	214	8	7.10	2.16	7.13	1.17	7.13	1.05
1lst A	2lao A	238	8	5.10	1.16	4.70	0.63	4.70	0.51
3tgl A	4tgl A	265	8	2.80	1.39	1.59	0.82	1.59	0.95
1anf A	1omp A	370	8	3.80	1.22	3.77	0.65	3.77	0.60
1ama A	9aat A	387	8	1.80	1.59	1.54	0.46	1.54	0.49
Promedios			8	4.12	1.50	3.75	0.75	3.75	0.72

<sup>a</sup> La descripción de estos valores se detalla en el pie de la Tabla 25.

<sup>b</sup> Desviaciones de los modelos iniciales (I) o finales (F) con las estructuras empleadas para generar los mapas obtenidas con el método Tama 1. Estos valores corresponden a los resultados a 8Å de resolución de la Tabla 4 de Orzechowski y Tama (Orzechowski y Tama, 2008).

<sup>c</sup> Desviaciones entre los modelos obtenidos con *iModfit* y las estructuras de los mapas al ajustar la estructura X en el mapa de la conformación Y, y *viceversa*. En estos casos, la orientación inicial se ha obtenido minimizando el RMSD entre X e Y con una implementación propia del método de Kabsch (Kabsch, 1978).

\* Todas las desviaciones se han calculado utilizando únicamente los átomos C<sub>α</sub>.

mostrados en la Tabla 5 de Grubisic *et al.* (Grubisic, *et al.*, 2010) utilizando mapas simulados de 8, 10, 15 y 20 Å de resolución y las mismas condiciones iniciales. En la Tabla 27 se recogen las desviaciones obtenidas. Los dos métodos son capaces de conseguir modelos muy similares a las estructuras utilizadas para generar los mapas en todo el rango de resoluciones ensayado, *i.e.* con desviaciones finales de 1.4 a 2.2 Å con Tama 2 y de 1.6 a 1.8 Å con *iModfit*. Individualmente, este último obtiene mejores desviaciones en 3 de los 5 ajustes, aunque las diferencias entre ambos son muy pequeñas y no superan generalmente los ~0.5 Å. No obstante, en los ajustes de la estructura atómica de *Iiwo* en los mapas de *Isu4* a 8 y 10 Å resolución, *iModfit* obtiene desviaciones ~1 Å

**Tabla 27.** Comparación entre *iModfit* y el método Tama 2 utilizando ajustes con mapas simulados.

PDB <sup>a</sup>		N	Res.	Tama 2 <sup>b</sup>		<i>iModfit</i>	
Inicial	Objetivo		[Å]	I [Å]	F [Å]	I [Å]	F [Å]
1ake A	4ake A	214	8	7.2	1.5	7.1	1.2
			10	7.2	1.8	7.1	1.2
			15	7.2	1.7	7.1	1.3
			20	7.2	1.9	7.1	1.4
1lfg A	1lfh A	691	8	6.4	1.1	6.4	0.8
			10	6.4	1.3	6.4	0.8
			15	6.4	1.4	6.4	0.9
			20	6.4	1.7	6.4	0.9
1oao C	1oao D	728	8	7.0	1.0	7.1	0.8
			10	7.0	1.2	7.1	0.8
			15	7.0	1.3	7.1	0.8
			20	7.0	1.4	7.1	0.9
1n0v D	1n0u A	818	8	14.5	1.2	14.4	1.9
			10	14.5	1.3	14.4	1.9
			15	14.5	1.5	14.4	2.0
			20	14.5	1.9	14.4	2.1
1iwo A	1su4 A	994	8	14.0	2.2	14.0	3.5*
			10	14.0	2.8	14.0	3.5*
			15	14.0	3.0	14.0	3.4*
			20	14.0	3.9	14.0	3.6*
Promedios			8	9.8	1.4	9.8	1.6
			10	9.8	1.7	9.8	1.6
			15	9.8	1.8	9.8	1.7
			20	9.8	2.2	9.8	1.8

<sup>a</sup> La descripción de estos valores se detalla en el pie de la Tabla 25.

<sup>b</sup> RMSD C<sub>α</sub> de los modelos iniciales (I) y finales (F) con las estructuras empleadas para generar los mapas. Los datos correspondientes al método Tama 2 proceden de la Tabla 5 de Grubisic *et al.* (Grubisic, *et al.*, 2010). *iModfit* se ha utilizado con las opciones por defecto y la orientación inicial se ha obtenido como en la Tabla 26.

\* En estos casos se generaron 100 modelos a partir de la estructura inicial (ΔRMSD promedio 4 Å) utilizando *iMC* y los 5 primeros modos obtenidos con *iMode*. Después, cada uno de ellos se ajustó al mapa correspondiente con *iModfit*. Se usaron las opciones por defecto. La desviación mostrada es la del modelo con mayor correlación.



mayores. Esto puede deberse a que las estructuras *Liwo* y *Isu4* presentan diferencias localizadas que impiden la obtención de desviaciones menores de  $\sim 3\text{\AA}$  utilizando únicamente los modos de menor frecuencia. Recuérdese que en la sección 5.8.1 ocurre algo similar en los casos de las estructuras *It5s* y *Isu4*. No obstante, para poder llevar a cabo los ajustes partiendo de *Liwo*, fue necesario perturbar ligeramente esta estructura utilizando sus 5 primeros modos con *iMC*. En caso contrario, la rotación de los dominios necesaria para alcanzar la conformación objetivo no se efectúa correctamente (véase el pie de la Tabla 27). A pesar de esto, cabe destacar que los ajustes llevados a cabo utilizando la estructura *It5s*, una conformación alternativa a *Liwo*, se han efectuado con éxito directamente en secciones anteriores empleando diferentes modelos atómicos, grados de libertad, número de modos y orientaciones iniciales.

### 5.8.9. Comparación con métodos basados en dinámica molecular utilizando estructuras modeladas por homología

Cuando no existe estructura atómica de la misma proteína con la que se obtuvo el mapa de microscopía, la alternativa consiste en generar un modelo atómico a partir de otras estructuras homólogas conocidas. Para evaluar el comportamiento de *iModfit* en este tipo de ajustes y compararlo con los métodos de ajuste flexible basadas en MD *Flex-EM*, *YUPSCX* y *MDFF* se ha empleado el conjunto de pruebas utilizado en Topf *et al.* (Topf, *et al.*, 2008). Este conjunto de pruebas consiste en 12 problemas de ajuste de proteínas cuyas estructuras iniciales han sido obtenidas mediante modelado por homología con el programa *MODELLER* (Eswar, *et al.*, 2007) (véase la sección 4.6.4 de Materiales y métodos). Para completar la comparativa con todos los métodos posibles se han llevado a cabo los ajustes tanto con *iModfit* como con *YUPSCX* utilizando sus parámetros por defecto. Las estructuras de partida fueron obtenidas mediante el ajuste rígido con *ADP\_EM*. En la Tabla 28 se recogen los resultados promedio correspondientes a estos métodos. Los valores de *Flex-EM* se han tomado de las Tablas 1 y 2 de Topf *et al.* (Topf, *et al.*, 2008) y los de *MDFF* de la Tabla S2 del material suplementario de Trabuco *et al.* (Trabuco, *et al.*, 2008). Las desviaciones promedio entre los modelos finales y las estructuras utilizadas en la obtención de los mapas son similares para los cuatro métodos ( $\sim 4\text{\AA}$ ), aunque con *iModfit* resultan ligeramente mejores ( $3.73\text{\AA}$ ). La Figura 48 permite analizar las desviaciones individualmente. En general, *iModfit* alcanza desviaciones similares a las obtenidas con las aproximaciones basadas en MD, y en dos casos es capaz de mejorarlas significativamente. En el caso de *IsvaA\_Iikna*, mientras que *MDFF* y *YUPSCX* tan solo alcanzan  $\sim 6\text{\AA}$  y *Flex-EM* unos  $4\text{\AA}$ , *iModfit* es capaz de reducir la desviación hasta los  $\sim 2\text{\AA}$ . Cabe destacar que este problema de ajuste corresponde al cambio conformacional de mayor amplitud de este conjunto de pruebas ( $\sim 10\text{\AA}$ ). Asimismo, cuando en el caso *3erda\_Ig5yd* ni *Flex-EM* ni *MDFF* consiguen reducir

**Tabla 28.** Comparación de *iModfit* con otros métodos basados en MD utilizando el conjunto de pruebas de modelado por homología.

Programa <sup>a</sup>	RMSD <sup>b</sup> [Å]	NO3.5 <sup>c</sup> [%]	Colisiones <sup>d</sup> [%]	Rama. [%]	Enlaces [%]	Angulos [%]	Molprobability [Å]
<i>Flex-EM</i>	4.20 ± 2.32	70 ± 18	249 ± 37	1.1 ± 0.8	0.0 ± 0.0	1.9 ± 0.8	4.0 ± 0.2
<i>MDFF</i>	4.00 ± 1.64	-	-	-	-	-	-
<i>YUPSCX</i>	4.02 ± 1.62	76 ± 11	77 ± 12	2.6 ± 1.6	5.8 ± 3.5	14.4 ± 6.7	3.5 ± 0.3
<i>iModfit</i>	3.73 ± 1.43	74 ± 13	157 ± 38	2.8 ± 1.5	0.2 ± 0.4	1.3 ± 0.9	3.6 ± 0.5
<i>iModfit*</i>	3.23 ± 1.58	81 ± 14	190 ± 50	5.0 ± 2.5	0.2 ± 0.4	1.3 ± 0.9	3.8 ± 0.4
Inicial	6.21 ± 1.94	60 ± 18	110 ± 23	2.0 ± 1.1	0.2 ± 0.4	1.3 ± 0.9	3.4 ± 0.3
Objetivo	0.00 ± 0.00	100 ± 0	36 ± 43	2.2 ± 3.2	0.0 ± 0.0	1.3 ± 1.1	2.8 ± 0.9

<sup>a</sup> Programa con el que se han llevado a cabo los ajustes flexibles del conjunto de pruebas de Sali (véase Materiales y Métodos). Con *iModfit* se han utilizado los parámetros por defecto. También se incluyen los valores promedio correspondientes: a los modelos finales de *iModfit* obtenidos con el 20% de los modos (\*), a las estructuras iniciales y a las estructuras objetivo empleadas para generar los mapas. En todos los casos se ha utilizado el modelo atómico HA, el 5% de los modos y la orientación inicial correspondiente al ajuste rígido los modelos iniciales en los mapas objetivo (*ADP\_EM*). Los problemas de ajuste *Iblb\_1a45A* y *Iu0sAY\_IffgAB* no han sido incluidos en los promedios ya que incluyen la dificultad adicional de encontrar la orientación inicial relativa entre sus dominios (véase Materiales y Métodos).

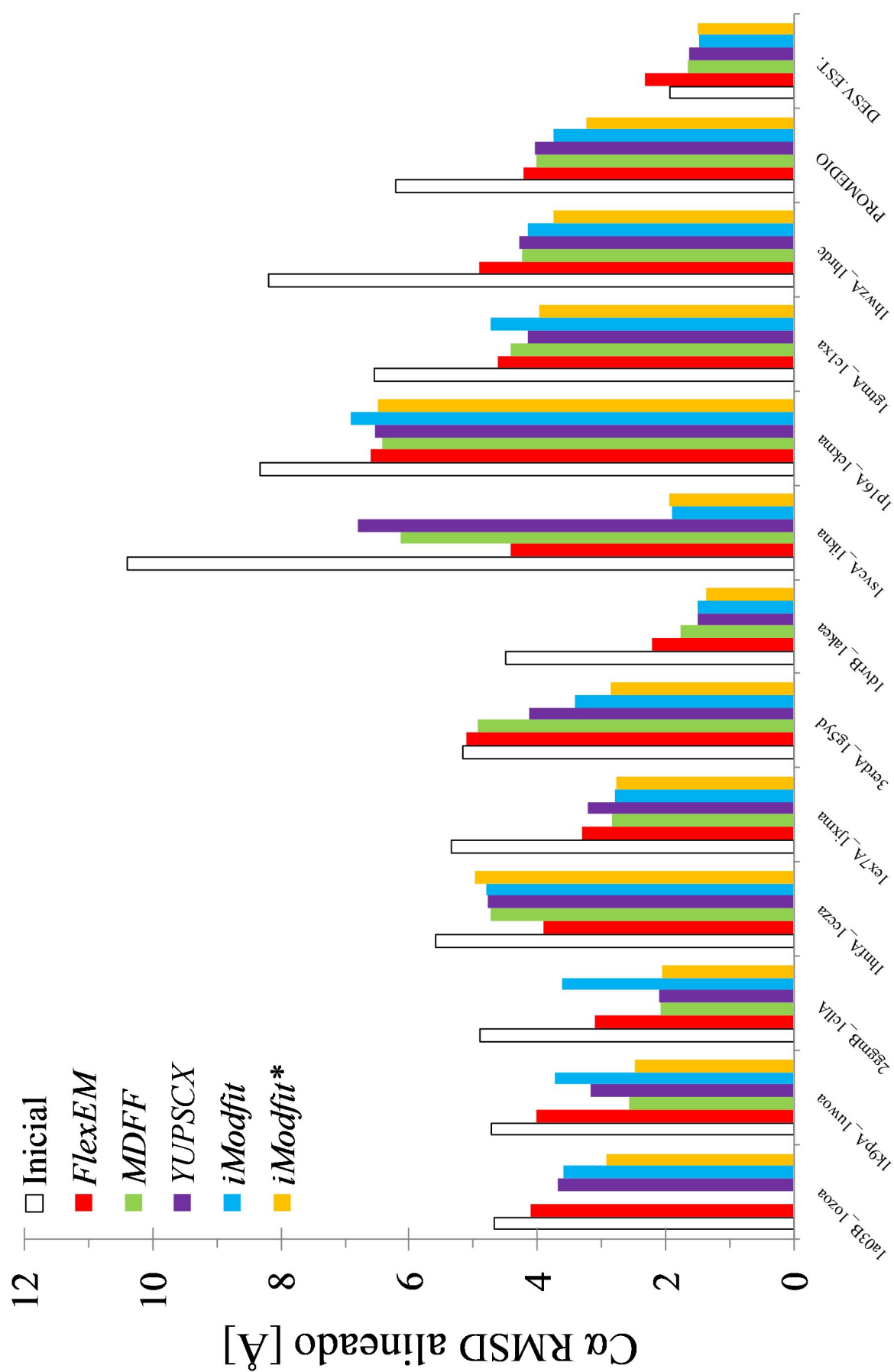
<sup>b,c</sup> RMSD promedio entre los modelos finales de los ajustes y las estructuras empleadas para generar los mapas objetivo (<sup>b</sup>). Porcentaje promedio de átomos de los modelos finales situados a menos de 3.5Å de los átomos correspondientes en las estructuras utilizadas para generar los mapas (<sup>c</sup>). Estas métricas se han obtenido después de minimizar el RMSD entre las estructuras con el comando *superpose* de *MODELLER* considerando únicamente los átomos C<sub>α</sub>. Al no estar disponibles públicamente los modelos finales de *MDFF*, no se han podido obtener los valores correspondientes.

<sup>d</sup> Las métricas de validación estructural ya han sido descritas en el pie de la Tabla 11.

apreciablemente los ~5Å de desviación inicial, *iModfit* permite alcanzar unos 3Å. Con *YUPSCX* se obtienen mejoras intermedias en este caso (~4Å). A pesar de que *iModfit* obtiene desviaciones similares a las alcanzadas con *Flex-EM* en los ajustes de *1k9pA\_luwoa* y *2ggmB\_1c1la*, éstas no son tan bajas como con *MDFF*. Esto puede deberse a la necesidad de utilizar más modos para representar estos cambios conformacionales. Para comprobar si un mayor número de modos es capaz de mejorar los resultados de *iModfit*, tanto en la Tabla 28 como en la Figura 48 se han incluido los resultados correspondientes a los ajustes obtenidos con el 20% de los modos. Como puede observarse en esta figura, al incluir más modos es posible alcanzar las desviaciones de *MDFF* en estos ajustes. Estas mejoras pueden generalizarse al resto de casos, pues al emplear el 20% de los modos se alcanzan desviaciones promedio de 3.23Å, *i.e.* ~0.5Å mejores que las obtenidas con el 5%. Al comparar estos ajustes individualmente, los valores, o bien mejoran, o bien son muy similares a los obtenidos con el 5%. La inclusión de más del ~20% de los modos no produce mejoras significativas en las desviaciones. Por ejemplo, con el 50% de los modos se obtienen valores promedio de 3.20Å (datos no mostrados). En la Tabla 28 también se ha incluido el porcentaje promedio de átomos C<sub>α</sub> situados a menos de 3.5Å de los átomos

**Figura 48.** Desviaciones individuales obtenidas con las herramientas de ajuste flexible basadas en MD y el conjunto de pruebas de modelado por homología. En el eje de abscisas se indica el problema de ajuste mediante el código del PDB de la estructura atómica inicial, seguido del código de la estructura empleada en la simulación del mapa. La leyenda muestra los colores asociados a cada herramienta. En *iModfit\** se ilustran las desviaciones alcanzadas con el 20% de los modos. El RMSD se ha obtenido después de superponer las estructuras con el comando *superpose* de *MODELLER* considerando únicamente los átomos C $\alpha$ . Estos resultados corresponden a los promedios mostrados en la Tabla 28.

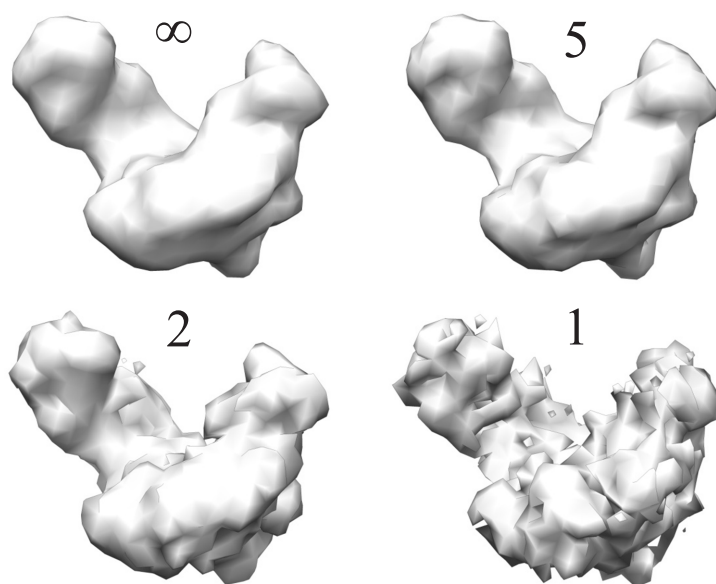
correspondientes en las estructuras utilizadas para generar los mapas (NO3.5, del inglés, *Native Overlap*). Desafortunadamente, como *MDFE* es un método computacionalmente muy costoso y los modelos finales no se encuentran disponibles públicamente, únicamente es posible la comparación entre *iModfit*, *Flex-EM* y *YUPSCX*. Según se aprecia en la Tabla 28, *Flex-EM* consigue un promedio de NO3.5 del 70%, mientras que *iModfit* obtiene valores superiores tanto con el 5 como con el 20% de los modos, *i.e.* 74 y 81%, respectivamente. En cambio, *YUPSCX* alcanza un NO3.5 del 76% superando tanto a *Flex-EM* como a *iModfit* utilizando sus parámetros por defecto. No obstante, cabe destacar que con el 20% de los modos *iModfit* obtiene los mejores valores, *i.e.* 81%. Respecto a la calidad de las estructuras, en la Tabla 28 también se han incluido los promedios de las métricas de *Molprobability* obtenidas a partir de los modelos finales disponibles. Lo primero que llama la atención es el gran número de colisiones (>150) observado en los modelos procedentes tanto de *Flex-EM* como de *iModfit*. Esto se debe en gran medida a que, como se mencionó en la sección 4.6.4 de Materiales y métodos, las estructuras iniciales procedentes de modelado por homología presentan más distorsiones que las estructuras objetivo (experimentales). Para estudiar estas distorsiones, en la Tabla 28 se han calculado las métricas de validación estructural con las estructuras iniciales y las estructuras objetivo utilizadas en la simulación de los mapas. De este modo, se comprueba que las 110 colisiones observadas en los modelos de homología triplican las de las estructuras experimentales (36) y que las longitudes y ángulos de enlace incorrectos permanecen prácticamente inalterados con *Flex-EM* e *iModfit*. A pesar de las distorsiones iniciales, *iModfit* es capaz de obtener menos colisiones y porcentaje de ángulos de enlace incorrectos que *Flex-EM*, *i.e.* 157 y 1.3% frente al 249 y 1.9%, respectivamente. Sin embargo, aunque con *iModfit* el porcentaje de ángulos diedros fuera de las regiones permitidas de Ramachandran tan solo aumenta del 2.0% inicial hasta el 2.6%, *Flex-EM* no solo es capaz de reducir el porcentaje inicial a la mitad, *i.e.* hasta el 1.1%, sino que además puede reparar las distancias de enlace incorrectas (0.0%). En cambio, la resolución cristalográfica equivalente (*molprobability score*) para los modelos finales de *Flex-EM* es de unos 4.0 Å, *i.e.* significativamente peor que la obtenida con *iModfit* (3.6 Å). Al emplear el 20% de los modos con *iModfit*, el número de colisiones, el porcentaje de ángulos fuera de las regiones permitidas de Ramachandran y el *molprobability score* empeoran hasta alcanzar valores



de 190, 5.0% y 3.8Å, respectivamente. En cambio, *YUPSCX* no solo reduce drásticamente el número de colisiones inicial (110) hasta alcanzar valores de 77, sino que también es capaz de obtener porcentajes de ángulos fuera de las regiones permitidas de Ramachandran (2.6%) y resoluciones cristalográficas equivalentes (3.5Å) ligeramente menores que los obtenidos con *iModfit*. Esto se debe a la presencia de un término armónico adicional que evita el acercamiento excesivo entre los átomos no enlazados (véase la sección 3.5.4 de la Introducción). No obstante, estas mejoras tienen lugar a costa de un incremento importante en el porcentaje de longitudes y ángulos de enlace incorrectos, que alcanzan valores mucho mayores que con los demás métodos *i.e.* del 5.8 y del 14.4%, respectivamente. En resumen, aunque generalmente los modelos obtenidos con los distintos métodos a partir de estructuras modeladas por homología son bastante similares, *iModfit* es el método más equilibrado teniendo en cuenta tanto la calidad estructural de los modelos finales como su desviación de las estructuras objetivo.

#### 5.8.10. Comparación con métodos basados en otras aproximaciones

También se ha comparado *iModfit* con dos métodos de ajuste flexible que no emplean ni modos de baja frecuencia ni simulaciones de MD. Éstos son: el método Zheng (Zheng, 2011), basado en el algoritmo de Newton-Raphson para minimizar una función de energía elástica modificada, y el método Jolley (Jolley, *et al.*, 2008) que utiliza simulaciones de MC con restricciones geométricas. Sus principios básicos de funcionamiento se han descrito brevemente en la sección 3.5.5 de la Introducción. Desafortunadamente, ninguno de estos métodos se encuentra disponible como herramienta. Por esta razón, se han repetido con *iModfit* los ajustes con mapas simulados que sus autores utilizan para validarlos. En concreto, para comparar con el método Zheng se han empleado 10 casos de ajuste con mapas simulados a 10, 15 y 20Å de resolución y distintas relaciones señal/ruido (SNR= 1, 2, 5 e  $\infty$ ). Como se indica en Zheng (Zheng, 2011), el ruido se ha generado según Jolley *et al.* (Jolley, *et al.*, 2008), *i.e.* se añade a cada voxel una variable aleatoria que sigue una distribución gaussiana cuya  $\sigma$  es proporcional a la raíz cuadrada de la densidad (véase la sección 4.6.3 de Materiales y métodos). En la Figura 49 se muestra la apariencia de estos mapas con un caso representativo. El tamaño de voxel utilizado por Zheng corresponde a la mitad de la resolución nominal, *i.e.* de 5, 7.5 y 10Å por voxel para los mapas de 10, 15 y 20Å de resolución, respectivamente. Estos muestreos resultan insuficientes y provocan una disminución efectiva de la resolución (Frank, 2006). Nosotros hemos empleado muestreos menores acordes con la resolución, *i.e.* de 3, 5 y 6Å por voxel en los mapas de 10, 15 y 20Å de resolución, respectivamente. No se han realizado los ajustes a resoluciones de 5Å por encontrarse fuera del rango de SP. En la Tabla 29 se comparan las desviaciones entre los modelos finales y las estructuras empleadas



**Figura 49.** Mapas simulados a 10Å de resolución de la adenilato quinasa con distintos niveles de ruido.

para simular los mapas obtenidos con *iModfit* y el método Zheng. Los valores correspondientes a este último se han tomado de la Tabla 1 del artículo de Zheng (Zheng, 2011). Para facilitar la comparación se han subrayado las desviaciones más bajas. Al utilizar mapas libres de ruido, el método Zheng consigue mejores desviaciones promedio que *iModfit*, es decir, de 0.71, 0.76 y 0.86Å con los mapas de 10, 15 y 20Å, respectivamente. No obstante, con *iModfit* también resultan muy bajas, *i.e.* 1.08, 1.17 y 1.28Å. En cuanto a los casos individuales, la aproximación Zheng obtiene las menores desviaciones en todos los casos excepto en 4. Por el contrario, al añadir distintas cantidades de ruido a los mapas esta tendencia se invierte. Mientras que a un nivel de ruido moderado (SNR=5) y 10Å de resolución los métodos Zheng e *iModfit* obtienen desviaciones bastante similares, *i.e.* de 1.16 y 1.09Å, respectivamente, a resoluciones de 15 y 20Å las diferencias se acentúan. En estos casos, aunque el método Zheng alcanza valores de 2.51Å con los mapas de 15Å y de 2.93Å con los de 20Å, *iModfit* consigue 1.23 y 1.35Å a 15 y 20Å, respectivamente. No obstante, 3 de los casos a 10Å de resolución son mejores con el método Zheng. Al aumentar la cantidad de ruido hasta niveles típicos de los mapas experimentales (SNR~2), las desviaciones de *iModfit* tan solo aumentan ligeramente hasta 1.12, 1.48 y 1.82Å a 10, 15 y 20Å de resolución, respectivamente. Por el contrario, con Zheng se alcanzan valores significativamente mayores, *i.e.* de 1.46, 2.95 y 3.72Å. Nótese que en estos casos las diferencias también se acentúan con los mapas de menor resolución. A pesar de ello, 5 de las 30 desviaciones individuales resultan ligeramente mejores al utilizar Zheng. Con el máximo nivel de ruido ensayado (SNR=1), *iModfit* también permite obtener mejores ajustes en 24 de los 30 casos (véase la Tabla 29). Por ejemplo, en presencia de ruido, el método



**Tabla 29.** Comparación entre *iModfit* y el método de Zheng utilizando ajustes con mapas simulados y distintos niveles de ruido.

PDB <sup>a</sup>		N	Res.		Zheng SNR <sup>b</sup>				<i>iModfit</i> SNR			
Inicial	Objetivo		[Å]	I [Å]	∞	5	2	1	∞	5	2	1
1ake A	4ake A	214	10	7.13	<u>0.74</u>	1.25	1.74	2.26	1.21	<u>1.24</u>	<u>1.26</u>	<u>1.38</u>
			15	7.13	<u>0.86</u>	2.27	2.41	3.00	1.39	<u>1.48</u>	<u>1.42</u>	<u>2.10</u>
			20	7.13	<u>0.99</u>	2.85	4.45	4.66	1.42	<u>1.46</u>	<u>1.54</u>	<u>2.71</u>
1l1st A	2lao A	238	10	4.70	<u>0.51</u>	0.71	0.74	1.02	0.64	<u>0.63</u>	<u>0.67</u>	<u>0.90</u>
			15	4.70	<u>0.58</u>	1.46	2.32	3.82	0.73	<u>0.76</u>	<u>1.09</u>	<u>1.71</u>
			20	4.70	<u>0.56</u>	1.93	4.54	4.69	0.78	<u>0.86</u>	<u>1.46</u>	<u>2.64</u>
3tgl A	4tgl A	265	10	1.59	1.02	1.40	1.57	1.58	<u>0.83</u>	<u>0.86</u>	<u>0.85</u>	<u>1.17</u>
			15	1.59	<u>0.70</u>	1.59	<u>1.59</u>	<u>1.59</u>	0.91	<u>1.08</u>	1.64	3.45
			20	1.59	<u>0.68</u>	1.59	<u>1.59</u>	<u>1.59</u>	1.00	<u>1.28</u>	1.70	5.94
1i5s A	1vfv A	313	10	1.81	<u>0.73</u>	1.64	1.79	1.81	1.39	<u>1.40</u>	<u>1.44</u>	<u>1.54</u>
			15	1.81	<u>0.92</u>	1.81	1.81	<u>1.81</u>	1.41	<u>1.48</u>	<u>1.71</u>	2.35
			20	1.81	<u>1.09</u>	1.81	<u>1.82</u>	<u>1.81</u>	1.48	<u>1.53</u>	2.83	2.41
1omp A	1anf A	370	10	3.77	<u>0.52</u>	1.08	1.28	1.51	0.63	<u>0.63</u>	<u>0.80</u>	<u>1.04</u>
			15	3.77	<u>0.60</u>	2.77	3.57	3.71	0.71	<u>0.80</u>	<u>1.29</u>	<u>2.74</u>
			20	3.77	<u>0.66</u>	3.74	3.76	3.78	1.02	<u>0.91</u>	<u>1.45</u>	<u>3.68</u>
9aat A	1ama A	387	10	1.66	0.59	0.95	1.07	1.33	<u>0.51</u>	<u>0.51</u>	<u>0.57</u>	<u>0.77</u>
			15	1.66	0.76	1.87	1.91	1.96	<u>0.65</u>	<u>0.72</u>	<u>1.16</u>	<u>1.81</u>
			20	1.66	0.95	1.90	2.01	<u>2.14</u>	<u>0.71</u>	<u>0.69</u>	<u>1.89</u>	3.33
1aon H	1aon A	524	10	12.25	<u>1.02</u>	<u>1.28</u>	<u>1.35</u>	<u>1.48</u>	2.05	2.05	2.02	2.08
			15	12.25	<u>1.08</u>	7.84	8.22	7.41	2.13	<u>2.09</u>	<u>2.23</u>	<u>2.45</u>
			20	12.25	<u>1.13</u>	8.72	9.36	9.45	2.32	<u>2.40</u>	<u>2.39</u>	<u>2.23</u>
1lfg A	1lfh A	691	10	6.44	<u>0.64</u>	1.06	2.24	1.53	0.81	<u>0.81</u>	<u>0.82</u>	<u>0.94</u>
			15	6.44	<u>0.61</u>	1.75	1.35	3.39	0.89	<u>0.96</u>	<u>1.15</u>	<u>1.52</u>
			20	6.44	<u>0.80</u>	1.92	2.34	4.34	0.93	<u>1.05</u>	<u>1.20</u>	<u>2.45</u>
1oao C	1oao D	728	10	7.05	<u>0.47</u>	<u>0.80</u>	0.88	1.08	0.78	<u>0.80</u>	<u>0.81</u>	<u>0.87</u>
			15	7.05	<u>0.51</u>	1.26	3.02	2.29	0.87	<u>0.89</u>	<u>1.03</u>	<u>1.49</u>
			20	7.05	<u>0.57</u>	2.50	2.24	4.16	0.97	<u>1.02</u>	<u>1.28</u>	<u>2.49</u>
1n0v D	1n0u A	818	10	14.44	<u>0.89</u>	<u>1.41</u>	<u>1.90</u>	2.69	1.95	1.93	1.97	<u>2.05</u>
			15	14.44	<u>1.01</u>	2.50	3.30	3.75	2.06	<u>2.05</u>	<u>2.10</u>	<u>2.34</u>
			20	14.44	<u>1.16</u>	2.34	5.07	7.52	2.21	<u>2.27</u>	<u>2.42</u>	<u>2.64</u>
Promedios			10	6.08	<u>0.71</u>	1.16	1.46	1.63	1.08	<u>1.09</u>	<u>1.12</u>	<u>1.27</u>
			15	6.08	<u>0.76</u>	2.51	2.95	3.27	1.17	<u>1.23</u>	<u>1.48</u>	<u>2.20</u>
			20	6.08	<u>0.86</u>	2.93	3.72	4.41	1.28	<u>1.35</u>	<u>1.82</u>	<u>3.05</u>

<sup>a</sup> La descripción de estos valores se detalla en el pie de la Tabla 27.

<sup>b</sup> Desviaciones finales (RMSD  $C_{\alpha}$ , [Å]) obtenidas utilizando mapas de distinta relación señal ruido ( $SNR = \mu_{\text{señal}} / \sigma_{\text{ruido}}$ , donde  $\mu$  es el valor medio y  $\sigma$  la desviación estándar). Los resultados de *iModfit* se han efectuado utilizando los parámetros por defecto para ajustar las estructuras iniciales en los mapas de las estructuras objetivo. Los datos de Zheng provienen de la Tabla 1 de Zheng (Zheng, 2011). La orientación inicial se ha determinado minimizando el RMSD entre las estructuras iniciales y objetivo. Se han subrayado los casos donde la desviación es menor.

Zheng no es capaz de mejorar significativamente las desviaciones de algunos de los ajustes de mayor amplitud cuando se utilizan los mapas de menor resolución: *lakeA\_4akeA*, *1lstA\_2laoA*, *laonH\_laonA*, *1lfgA\_1lfhA* y *1n0vD\_1n0uA*. Según estos resultados, aunque el método Zheng obtiene generalmente las mejores desviaciones con mapas simulados libres de ruido, cuando aumenta éste (SNR=1~5) es *iModfit* el que consigue los mejores resultados.

Para comparar *iModfit* con el método Jolley se han empleado los dos casos de ajuste con mapas simulados con los que fue validado este último, *i.e.* la adenilato quinasa (de *lake* a *4ake*) y la lactoferrina (de *1lfg* a *1lfh*). Los RMSD basados en todos los átomos (hidrógenos incluidos) obtenidos en estos casos a 8 y 10Å de resolución y muestreo de 2Å por voxel se han tomado de la Tabla 4 de Jolley *et al.* (Jolley, *et al.*, 2008). Para llevar a cabo la comparación con *iModfit* en las mismas condiciones y utilizando los parámetros por defecto se han calculado las desviaciones (también con todos los átomos) a partir de los ajustes correspondientes utilizados en la sección anterior para comparar con el método de Tama 2 (véase la Tabla 27). Las desviaciones obtenidas por el método Jolley al ajustar la estructura *lake* en el mapa simulado de *4ake*, son de 1.5 y 1.8Å con los mapas de 8 y 10Å de resolución, respectivamente. Asimismo, *iModfit* permite obtener desviaciones parecidas, *i.e.* de 1.8 y 1.9Å con los mismos mapas, respectivamente. En el caso de ajuste de *1lfg*, el método Jolley alcanza valores de 1.8Å con el mapa de 8Å y de 2.1Å con el de 10Å, mientras que con *iModfit* las desviaciones son algo mayores, *i.e.* 2.8Å a las dos resoluciones. En la Figura 5 de Jolley *et al.* también se estudia el efecto del ruido en los ajustes para el caso de la adenilato quinasa; no obstante, únicamente se muestran los resultados obtenidos a partir de mapas de 3.3Å de resolución, *i.e.* resoluciones fuera del rango de SP. En dicha figura se muestra que el método Jolley permite obtener desviaciones de ~1Å cuando la SNR es mayor o igual a 2, mientras que los valores ascienden hasta los ~2Å cuando la SNR es 1. Sin embargo, y a pesar de que no es posible la comparación directa, cabe destacar que a 10Å *iModfit* es capaz de obtener desviaciones de 2.0 y 1.9Å cuando la SNR es 1 y 2, respectivamente. Nótese que en estos casos las desviaciones se han calculado teniendo en cuenta todos los átomos. Aunque esta comparativa no es tan exhaustiva como las anteriores, permite mostrar que los dos métodos son capaces de obtener modelos finales similares incluso en presencia de ruido.

#### **5.8.11. Efecto del ruido en los modelos obtenidos con iModfit, NMFF y YUPSCX**

En primer lugar, se han utilizando los casos de ajuste flexible empleados en la comparativa con el método Zheng para estudiar el comportamiento de *iModfit* con distintos niveles de ruido. En esta ocasión se han incluido los dos sentidos de ajuste, lo que hace un total de 20 casos diferentes. En ausencia de ruido (SNR=∞) el número de colisiones, el porcentaje de ángulos

fuera de las regiones permitidas de Ramachandran y la resolución cristalográfica equivalente apenas aumentan de 29 a 44-45, del 0.8 al 1.1-1.2% y de 2.7 a 3.1 Å, respectivamente (Tabla 30). Con una SNR de 5, los resultados son muy similares, *i.e.* se alcanzan 44-48 colisiones, un 1.1-1.3% de ángulos diedros en regiones no permitidas y la resolución equivalente se mantiene en 3.1 Å. Al alcanzar niveles de ruido similares a los de los mapas experimentales (SNR~2), y aunque a 10 Å la calidad de las estructuras es igual que en los casos sin ruido, con los mapas de 15 y 20 Å las colisiones aumentan hasta 52 y 62, respectivamente, y el porcentaje de ángulos fuera de las regiones permitidas alcanza el 1.3% a 15 Å y 1.5% a 20 Å. Incluso al emplear los mapas de 10 Å de resolución y una SNR=1, la calidad de los modelos finales también resulta prácticamente idéntica a la obtenida en ausencia de ruido, *i.e.* tan solo las colisiones aumentan ligeramente (48). En cambio, a 15 y 20 Å aparece un cierto deterioro de las estructuras, *i.e.* las colisiones, porcentajes de ángulos en regiones no permitidas y resoluciones equivalentes aumentan hasta 78, 1.8% y 3.4 Å (a 15 Å) y hasta 106, 3.4% y 3.6 Å (a 20 Å), respectivamente. Nótese que los porcentajes de longitudes y ángulos de enlace no se ven alterados en ningún caso. Estos resultados demuestran la robustez de *iModfit* al emplear mapas simulados con ruido en el rango de resoluciones de 10 a 20 Å. Además, cabe destacar que a 10 Å de resolución tanto la calidad estructural de los modelos finales como sus desviaciones con las estructuras empleadas para generar los mapas resultan prácticamente idénticas a las obtenidas en ausencia

**Tabla 30.** Evaluación de la calidad de las estructuras obtenidas con *iModfit* utilizando mapas simulados y distintos niveles de ruido.

SNR <sup>a</sup>	Resolución [Å]	RMSD <sup>b</sup> [Å]	Colisiones [%]	Rama. [%]	Enlaces [%]	Ángulos [%]	Molprobability [Å]
∞	10	1.03 ± 1.39	44 ± 17	1.1 ± 0.8	0.3 ± 0.8	7.1 ± 11.9	3.1 ± 0.4
	15	1.13 ± 1.41	44 ± 17	1.2 ± 0.8	0.3 ± 0.8	7.1 ± 11.9	3.1 ± 0.4
	20	1.26 ± 1.48	45 ± 17	1.2 ± 0.7	0.3 ± 0.8	7.1 ± 11.9	3.1 ± 0.4
5	10	1.04 ± 1.40	44 ± 17	1.1 ± 0.8	0.3 ± 0.8	7.1 ± 11.9	3.1 ± 0.4
	15	1.20 ± 1.48	46 ± 17	1.2 ± 0.8	0.3 ± 0.8	7.1 ± 11.9	3.1 ± 0.4
	20	1.36 ± 1.53	48 ± 17	1.3 ± 0.7	0.3 ± 0.8	7.1 ± 11.9	3.1 ± 0.4
2	10	1.08 ± 1.45	45 ± 17	1.1 ± 0.8	0.3 ± 0.8	7.1 ± 11.9	3.1 ± 0.4
	15	1.47 ± 1.70	52 ± 17	1.3 ± 0.8	0.3 ± 0.8	7.1 ± 11.9	3.2 ± 0.4
	20	1.83 ± 2.50	62 ± 17	1.5 ± 0.7	0.3 ± 0.8	7.1 ± 11.9	3.3 ± 0.4
1	10	1.24 ± 1.53	48 ± 17	1.1 ± 0.8	0.3 ± 0.8	7.1 ± 11.9	3.1 ± 0.4
	15	2.20 ± 2.17	78 ± 17	1.8 ± 0.8	0.3 ± 0.8	7.1 ± 11.9	3.4 ± 0.5
	20	3.55 ± 2.39	106 ± 17	3.4 ± 0.7	0.3 ± 0.8	7.1 ± 11.9	3.6 ± 0.5
-	Inicial	6.10 ± 4.28	29 ± 25	0.8 ± 1.0	0.3 ± 0.8	7.1 ± 11.9	2.7 ± 0.6

<sup>a</sup> Relación señal/ruido (SNR= $\mu_{\text{señal}}/\sigma_{\text{ruido}}$ ).

<sup>b</sup> La descripción de estas métricas se detalla en el pie de la Tabla 11. También se han incluido los promedios correspondientes a las estructuras iniciales. Los promedios no solo corresponden a los resultados individuales mostrados en la Tabla 29, sino que también incluyen los resultados en ambos sentidos.

de ruido. Por todo ello, podemos concluir que *iModfit* es un método muy robusto frente a la presencia de ruido simulado.

En segundo lugar, para estudiar cómo afecta el ruido a los ajustes obtenidos con otras herramientas de ajuste flexible basadas en NMA (*NMFF*) y MD (*YUPSCX*), y comparar sus resultados con los de *iModfit*, también se han empleado los casos ajuste flexible con distintos niveles de ruido utilizados en la comparativa con el método Zheng. En la Tabla 31 se recogen los promedios de las desviaciones finales obtenidas con los programas *iModfit*, *NMFF* y *YUPSCX* en función de la resolución y la SNR de los mapas. Aunque las desviaciones obtenidas con *YUPSCX* y los mapas de 10Å son bastante bajas (de 1.78 a 1.92Å), con los de 15 y 20Å las desviaciones son más altas, *i.e.* de 3.25 a 3.30Å con los de 15Å, y de 3.88 a 4.20Å con los de 20Å, lo que apenas supone una reducción en la desviación inicial del 50%. No obstante, los valores obtenidos en estos casos resultan prácticamente independientes del nivel de ruido de los mapas. Cuando se utilizan los métodos basados en NMA (*NMFF* e *iModfit*) las desviaciones se reducen en comparación con las de MD (*YUPSCX*), *i.e.* las reducciones son de ~0.5Å al utilizar mapas de 10Å, y de hasta ~2Å con los de 15 y 20Å. Sin embargo, a pesar de que las desviaciones obtenidas con los métodos de NMA y los mapas de 10Å apenas dependen de la SNR, los ajustes se vuelven más sensibles al ruido conforme disminuye la resolución. Por ejemplo, mientras que en los casos a 15Å, las desviaciones aumentan gradualmente con el ruido de 1.48 a 2.22Å con *NMFF* y de 1.13 a 2.20Å con *iModfit*, al utilizar los mapas de 20Å el incremento es mayor, *i.e.* los valores aumentan de 1.59 a 3.05Å y de 1.26 a 3.55Å con *NMFF* e *iModfit*, respectivamente. No obstante, cabe destacar que con los programas basados en NMA las desviaciones únicamente superan los 2Å cuando se emplean los máximos niveles de ruido del ensayo (SNR=1). Al comparar entre sí las desviaciones promedio obtenidas por los distintos

**Tabla 31.** Desviaciones finales de los modelos obtenidos con *iModfit*, *NMFF* y *YUPSCX* utilizando mapas simulados y distintos niveles de ruido.

Resolución [Å]	RMSD <sup>a</sup> [Å]	SNR <sup>b</sup>											
		<i>iModfit</i>				<i>NMFF</i>				<i>YUPSCX</i>			
		∞	5	2	1	∞	5	2	1	∞	5	2	1
10	6.10	1.03	1.04	1.08	1.24	1.25	1.22	1.54	1.37	1.78	1.79	2.04	1.92
15	6.10	1.13	1.20	1.47	2.20	1.48	1.33	1.58	2.22	3.25	3.30	3.29	3.26
20	6.10	1.26	1.36	1.83	3.55	1.59	1.56	1.95	3.05	3.88	3.92	3.97	4.20

<sup>a</sup> Las desviaciones finales se han obtenido utilizando mapas con distintas SNR ( $\mu_{\text{señal}}/\sigma_{\text{ruido}}$ ). También se muestran las desviaciones iniciales (I). Los resultados de *iModfit* y *YUPSCX* se han efectuado con sus parámetros por defecto, y los de *NMFF* utilizando 30 modos y 0.001 de umbral de convergencia. La orientación inicial corresponde a la superposición de las estructuras atómicas iniciales y objetivo utilizando el método de Kabsch (Kabsch, 1978).

<sup>b</sup> Los promedios corresponden al ajuste en ambos sentidos.

métodos puede apreciarse claramente que las de *iModfit* son mejores que las de los demás en todo el rango de resoluciones y niveles de ruido ensayados, y que *NMFF* obtiene generalmente mejores promedios que *YUPSCX*.

En las Tablas 32 y 33 se recogen los valores promedio de las métricas de validación estructural correspondientes a los modelos finales obtenidos con los programas *YUPSCX* y *NMFF*, respectivamente. Con *YUPSCX* (Tabla 32) el número de colisiones importantes se mantiene en niveles aceptables en todas las condiciones, *i.e.* entre 42 y 54, lo que supone un incremento mínimo respecto del valor inicial (29). El porcentaje de ángulos fuera de las regiones permitidas de Ramachandran permanece prácticamente constante pues tan solo aumenta como máximo hasta el 1.1% desde el 0.8% inicial. En cuanto al porcentaje de distancias y ángulos de enlace inadecuados, los incrementos también son mínimos en comparación con los valores iniciales, *i.e.* se producen un aumento desde el 0.3% inicial hasta el 0.7% para las longitudes de enlace y del 7.1 hasta el 12.4% para los ángulos. La resolución cristalográfica equivalente apenas varía de los 2.7Å iniciales hasta los 3.0-3.2Å de los modelos finales. Todos estos valores van en consonancia con los resultados obtenidos por *YUPSCX* en ausencia de ruido (sección 5.8.8). Sorprendentemente, se observa una ligera mejoría en todas las métricas de calidad estructural conforme aumenta el nivel de ruido de los mapas, aunque su magnitud está muy por debajo de las correspondientes desviaciones estándar. Al utilizar *NMFF* (Tabla 33), el número de colisiones es menor que con *YUPSCX* en todos los casos, excepto en los de SNR=1 y mapas de 15 y 20Å de resolución. Sin embargo, el resto de métricas empeoran con *NMFF*.

**Tabla 32.** Evaluación de la calidad de las estructuras obtenidas con *YUPSCX* utilizando mapas simulados y distintos niveles de ruido.

SNR	Resolución [Å]	RMSD <sup>a</sup> [Å]	Colisiones [%]	Rama. [%]	Enlaces [%]	Ángulos [%]	Molprobability [Å]
$\infty$	10	1.78 ± 0.49	46 ± 10	0.9 ± 1.0	0.7 ± 0.8	12.3 ± 13.4	3.1 ± 0.4
	15	3.25 ± 1.72	52 ± 13	1.0 ± 1.1	0.7 ± 0.8	12.4 ± 13.2	3.2 ± 0.4
	20	3.88 ± 1.80	54 ± 14	1.0 ± 1.1	0.7 ± 0.8	12.4 ± 13.2	3.2 ± 0.4
5	10	1.79 ± 0.52	45 ± 10	1.0 ± 1.0	0.6 ± 0.8	11.8 ± 13.3	3.1 ± 0.4
	15	3.30 ± 1.87	50 ± 13	1.0 ± 1.0	0.7 ± 0.9	11.8 ± 13.3	3.2 ± 0.4
	20	3.92 ± 1.92	52 ± 13	1.1 ± 1.2	0.7 ± 0.9	12.1 ± 13.4	3.2 ± 0.4
2	10	2.04 ± 1.16	42 ± 11	1.0 ± 0.9	0.5 ± 0.8	10.7 ± 13.3	3.1 ± 0.4
	15	3.29 ± 1.72	47 ± 14	1.0 ± 1.0	0.5 ± 0.7	10.7 ± 13.0	3.1 ± 0.4
	20	3.97 ± 1.81	50 ± 13	1.1 ± 1.2	0.7 ± 0.9	11.2 ± 13.4	3.2 ± 0.4
1	10	1.92 ± 0.53	38 ± 11	0.9 ± 0.9	0.4 ± 0.7	9.9 ± 12.9	3.0 ± 0.4
	15	3.26 ± 1.04	43 ± 11	0.9 ± 1.0	0.5 ± 0.9	10.2 ± 13.2	3.1 ± 0.4
	20	4.20 ± 1.58	45 ± 13	1.0 ± 1.1	0.5 ± 0.8	10.4 ± 12.8	3.1 ± 0.4
-	Inicial	6.10 ± 4.28	29 ± 25	0.8 ± 1.0	0.3 ± 0.8	7.1 ± 11.9	2.7 ± 0.6

<sup>a</sup> Valores promedio obtenidos según Tabla 30 utilizando *YUPSCX* con sus parámetros por defecto.

Los porcentajes de ángulos diedros en regiones no permitidas por norma general se mantienen entre el 1.4 y el 1.6%, aunque en los casos de menor resolución y mayor nivel de ruido alcanzan valores del 2.0 al 2.7%. Asimismo, la resolución cristalográfica equivalente se mantiene en el rango de 3.0 a 3.3 Å en todos los casos. A pesar de que todos estos valores son relativamente aceptables, los porcentajes de longitudes y ángulos de enlace inadecuados con *NMFF* son muy altos, como ya se ha observado en la sección 5.8.7, alcanzando en los casos sin ruido valores del 16.6 al 19.0% para las longitudes de enlace y del 21.9 al 23.9% para los ángulos. Estos valores aumentan aún más conforme sube el nivel de ruido y disminuye la resolución de los mapas hasta alcanzar valores máximos del 33.6 y del 37.5% para las longitudes y los ángulos de enlace, respectivamente.

Según estos resultados, *iModfit* no solo obtiene las mejores desviaciones con las estructuras utilizadas para generar los mapas en prácticamente todas las condiciones de ruido ensayadas, sino que también conserva la estructura covalente de los modelos iniciales gracias al empleo de los modos normales en CI. Además, mantiene dentro de límites razonables tanto el número de colisiones importantes como el porcentaje de ángulos diedros en regiones no permitidas de Ramachandran. Por todo ello, aunque los programas examinados en esta sección también obtienen buenos ajustes, *iModfit* es la herramienta con el mejor balance entre calidad estructural y similitud con las estructuras objetivo, incluso en presencia de elevados niveles de ruido gaussiano.

**Tabla 33.** Evaluación de la calidad de las estructuras obtenidas con *NMFF* utilizando mapas simulados y distintos niveles de ruido.

SNR	Resolución [Å]	RMSD <sup>a</sup> [Å]	Colisiones [%]	Rama. [%]	Enlaces [%]	Ángulos [%]	Molprobability [Å]
$\infty$	10	1.25 ± 0.62	35 ± 24	1.5 ± 1.9	16.6 ± 10.8	21.9 ± 13.8	3.0 ± 0.4
	15	1.48 ± 0.80	36 ± 25	1.4 ± 1.5	17.1 ± 10.5	22.2 ± 15.4	3.0 ± 0.4
	20	1.59 ± 0.75	37 ± 24	1.5 ± 1.6	19.0 ± 14.9	23.9 ± 15.7	3.0 ± 0.4
5	10	1.22 ± 0.54	36 ± 23	1.4 ± 1.7	16.9 ± 10.3	22.4 ± 15.1	3.0 ± 0.4
	15	1.33 ± 0.48	37 ± 23	1.4 ± 1.6	19.6 ± 11.1	24.5 ± 15.7	3.0 ± 0.4
	20	1.56 ± 0.59	37 ± 24	1.5 ± 1.7	19.1 ± 12.8	24.0 ± 14.8	3.1 ± 0.4
2	10	1.54 ± 1.51	36 ± 24	1.5 ± 1.7	17.7 ± 11.2	22.4 ± 14.2	3.0 ± 0.4
	15	1.58 ± 0.62	39 ± 24	1.6 ± 1.5	20.9 ± 13.6	25.4 ± 17.4	3.1 ± 0.4
	20	1.95 ± 0.58	46 ± 25	2.0 ± 1.9	26.3 ± 17.5	31.1 ± 20.8	3.2 ± 0.4
1	10	1.37 ± 0.48	36 ± 24	1.5 ± 1.6	18.9 ± 12.3	23.5 ± 15.6	3.0 ± 0.4
	15	2.22 ± 0.67	53 ± 46	2.5 ± 2.7	27.4 ± 21.4	31.5 ± 22.2	3.2 ± 0.5
	20	3.05 ± 0.86	61 ± 41	2.7 ± 2.2	33.6 ± 18.0	37.5 ± 22.7	3.3 ± 0.5
-	Inicial	6.10 ± 4.28	29 ± 25	0.8 ± 1.0	0.3 ± 0.8	7.1 ± 11.9	2.7 ± 0.6

<sup>a</sup> Obtenido utilizando *NMFF* con 30 modos y 0.001 de convergencia (véase Tabla 30).



### 5.8.12. Comparación entre los tiempos de cálculo de iModfit, NMFF y YUPSCX

Para comparar la velocidad relativa entre los ajustes efectuados con *iModfit*, *NMFF* y *YUPSCX*, en la Tabla 34 se muestran los tiempos que tardan los distintos métodos en ajustar todas las estructuras atómicas del conjunto de pruebas de proteínas de *MolMovDB* utilizando los mapas a 10Å de resolución. Además, para comprobar si existe alguna diferencia entre realizar

**Tabla 34.** Comparación de los tiempos de ajuste flexible con *iModfit*, *NMFF* y *YUPSCX*.

Id <sup>a</sup>	N <sup>b</sup>	RMSD <sup>c</sup> [Å]	Tiempo <sup>d</sup> [min]					
			Cierre			Apertura		
			<i>iModfit</i>	<i>NMFF</i>	<i>YUPSCX</i>	<i>iModfit</i>	<i>NMFF</i>	<i>YUPSCX</i>
1ex6	186	4.58	1	9	2	1	5	2
4ake	214	8.26	1	9	3	5	15	3
1ggg	220	10.26	3	9	3	3	12	3
2lao	238	8.67	1	7	3	1	7	3
1urp	271	7.69	1	5	3	2	11	3
1ram	273	4.96	2	7	4	2	10	4
5at1	310	2.41	2	10	4	1	13	4
1ckm	317	4.35	1	11	4	3	18	4
3dap	317	5.83	2	11	4	3	9	4
1bp5	329	12.16	2	11	4	5	12	4
1jql	366	2.55	3	7	5	3	12	5
1omp	370	7.29	2	24	5	2	8	5
8adh	374	2.19	1	6	5	1	6	5
9aat	387	2.19	1	6	5	1	3	5
1bnc	452	5.72	5	19	6	4	13	6
1dpe	507	12.43	4	26	7	8	47	7
1rkm	517	5.76	3	15	8	4	26	8
1sx4	524	15.83	17	126	7	20	61	7
1i7d	609	4.20	9	50	9	12	65	9
1lfg	691	8.08	13	66	10	12	33	10
1oao	728	13.60	12	48	11	16	59	11
1ih7	890	7.20	23	92	14	27	111	14
1su4	994	17.99	59	257	15	85	325	15
Total	-	-	167	831	142	222	881	142
Promedio	-	7.57	7	36	6	10	38	6
Desv.Est.	-	4.39	13	57	4	18	68	4

<sup>a-c</sup> Código PDB correspondiente a la estructura abierta del conjunto de pruebas de *MolMovDB* (<sup>a</sup>), número de aminoácidos (<sup>b</sup>) y desviación entre las estructuras abierta y cerrada (<sup>c</sup>).

<sup>d</sup> Tiempos de ejecución para llevar a cabo los ajustes. Con *NMFF* se han utilizado 30 modos y 0.001 de umbral de convergencia, y en el resto, los parámetros por defecto. Se muestran los valores correspondientes al ajuste de las estructuras abiertas en los mapas (10Å) cerrados (cierre) y *viceversa* (apertura) .

\* Se han empleado procesadores Intel® Xeon™ E5410 a 2.33GHz en el caso de *NMFF*, e Intel® Core™2 Quad Q6600 a 2.44GHz en el resto.

los ajustes a partir de estructuras abiertas o cerradas, se han separado los tiempos en función de la conformación de la estructura inicial. La herramienta *YUPSCX* es capaz de terminar los 46 ajustes en 4.7 horas, seguido de cerca por *iModfit* con 6.5 horas. El método más lento es *NMFF*, con una duración total de 28.5 horas. Observando los tiempos individuales puede comprobarse que las velocidades de *iModfit* y *YUPSCX* son muy similares en todos los casos excepto en el de mayor tamaño (*Isu4*), en el cual este último método es de 4 a 6 veces más rápido. Según esto, el coste computacional de *YUPSCX*, que emplea un potencial aproximado de MD, escala mejor con el tamaño del sistema que el de los métodos basados en NMA. Esto es debido principalmente a que la fracción de tiempo dedicada al NMA comienza a ser mayoritaria. Por otra parte, *iModfit* y *NMFF* tardan aproximadamente una hora menos en completar los ajustes a partir de estructuras abiertas. Este hecho concuerda con lo observado en las secciones 5.2 y 5.6.1 ya que los modos de baja frecuencia representan mejor las transiciones conformacionales cuando se calculan desde una conformación abierta. En cambio, los ajustes del método basado en MD (*YUPSCX*) tardan lo mismo independientemente del grado de apertura de la conformación inicial. Con todo esto puede concluirse que el método de ajuste basado en NMA y CI (*iModfit*) es 4-5 veces más rápido que su equivalente en CC (*NMFF*) y, además, que su velocidad es similar a la del método más rápido (*YUPSCX*) en casi todos los casos. No obstante, cabe destacar que *iModfit* es más rápido que *YUPSCX* si se fija el 50% de los ángulos diedros. De este modo, es capaz de efectuar los ajustes en 4.4 horas manteniendo la calidad de los resultados (véase la sección 5.8.3).

#### **5.8.13. Validación de iModfit en casos experimentales de ajuste flexible**

Hasta ahora, todos los mapas objetivo empleados para dirigir los ajustes fueron simulados a partir de estructuras a resolución atómica en ausencia o presencia de niveles controlados de ruido gaussiano. Sin embargo, este tipo de ruido simulado tan solo es una primera aproximación al ruido real presente en las reconstrucciones 3D de microscopía electrónica y SP. Estas reconstrucciones están sujetas a multitud de factores que pueden distorsionar tanto la estructura de las macromoléculas como la información estructural que proviene de ellos, *i.e.* las imágenes. Algunos de estos factores pueden ser las fuerzas mecánicas que actúan durante la preparación de la muestra (vitricación, tinción negativa o crio-tinción), el daño por radiación o las imperfecciones en las lentes del microscopio o el detector. Además, la asignación de orientaciones correctas a las partículas individuales está fuertemente condicionada por la naturaleza ruidosa de las imágenes, la presencia de orientaciones preferentes, la variabilidad estructural en la muestra, el algoritmo de reconstrucción utilizado, etc. (Frank, 2006). Dado que resulta muy difícil (o imposible) simular adecuadamente y en condiciones controladas la influencia que tienen todos estos factores en las reconstrucciones 3D, resulta esencial la utilización de un conjunto de

pruebas experimental para comprobar el funcionamiento de *iModfit* en situaciones reales. Con este fin, se han tomado de la base de datos de microscopía electrónica (EMDB, del inglés Electron Microscopy Data Bank) (Lawson, *et al.*, 2011) una serie de mapas experimentales de proteínas para las cuales existen estructuras atómicas en el PDB. Los tamaños de las proteínas y las resoluciones de los mapas abarcan el rango accesible a microscopía electrónica y SP, *i.e.* masas moleculares de ~50 a ~500 kDa y resoluciones de 8 a 24 Å (véase la Tabla 35). Tras ajustar rígidamente estas estructuras en los mapas correspondientes con *ADP\_EM*, se ha obtenido un conjunto representativo de 12 casos experimentales de ajuste flexible. A diferencia de los casos que utilizan mapas simulados de las secciones anteriores, en los ajustes experimentales no existe una estructura atómica con la cual evaluar directamente la desviación de los modelos finales. Por este motivo, para valorar los modelos finales adecuadamente, además de la correlación, también hay que considerar tanto la calidad de las estructuras como la apariencia visual de los ajustes. Para ello, en la Tabla 36 se muestran las correlaciones entre los mapas experimentales y los mapas simulados de los modelos finales junto con las métricas de validación estructural de *Molprobability*. Para medir la amplitud del cambio conformacional también se han incluido las desviaciones entre los modelos iniciales y finales. Además, de la Figura 50 a la 66 se han

**Tabla 35.** Casos de ajuste flexible con mapas experimentales.

Ajuste <sup>a</sup>	N <sup>b</sup>	Mm <sup>c</sup> [kDa]	Res. <sup>d</sup> [Å]	Umbral <sup>e</sup>	PDB <sup>f</sup>	EMDB <sup>g</sup>
Ndc80/Nuf2	316	35	8.0	0.008	2ve7 <sup>*</sup>	5223
Tubulina	838	92	8.0	0.008	1jff	5223
GroEL <sub>14</sub> /ES <sub>7</sub> /ADP <sub>7</sub> <sup>*</sup>	524	58	8.7	0.010	1sx4	1181
GroEL <sub>14</sub> <sup>*</sup>	524	58	10.3	0.010	1sx4	1042
EFG <sup>†</sup>	655	72	10.8	7.000	1fnm	(Valle, <i>et al.</i> , 2003)
Prefoldina	698	77	20.0	0.010	1fxk	(Martin-Benito, <i>et al.</i> , 2002)
Clamp	810	89	17.0	0.000	3a1j	(E.Nogales, com.personal)
ATPasa-Ca	994	109	8.0	1.000	1su4	(Zhang, <i>et al.</i> , 1998)
CopAdC	1386	152	17.0	1.000	2voy <sup>§</sup>	5004
Termosoma	3836	422	10.0	0.201	1a6d <sup>‡</sup>	1396
ARNP (cerrada)	4101	451	20.0	0.100	1y1v	1283
ARNP (abierta)	4101	451	20.0	0.100	1y1v	1284

<sup>a-g</sup> Caso de ajuste flexible (<sup>a</sup>). Número de aminoácidos (<sup>b</sup>). Masa molecular aproximada (<sup>c</sup>). Resolución empleada en los ajustes (<sup>d</sup>). Umbral a partir del cual son tenidas en cuenta las densidades de los mapas (<sup>e</sup>). Código PDB de la estructura atómica (<sup>f</sup>). Código del mapa experimental de la base de datos EMDB, o en su defecto referencia dónde ha sido publicado (<sup>g</sup>).

<sup>\*</sup> Estos mapas han sido segmentados de los complejos correspondientes utilizando como máscara un mapa filtrado a 20Å de la estructura atómica de un monómero. Dicha estructura fue orientada previamente en el mapa experimental después de ser ajustada con *ADP\_EM* utilizando todo el complejo.

<sup>†</sup> Este mapa se ha segmentado a partir del mapa del ribosoma completo.

<sup>‡</sup> Se han eliminado las regiones para las cuales no existe claramente densidad en los mapas.

<sup>§</sup> Estructura atómica utilizada para construir manualmente un modelo dimérico compatible con el mapa experimental.

**Tabla 36.** Resultados del ajuste flexible con *iModfit* utilizando mapas experimentales.

Ajuste <sup>a</sup>	Correlación <sup>b</sup>		RMSD <sup>c</sup> [Å]		Colisiones <sup>d</sup>		Rama.		Enlaces		Ángulos		MolProb	
	I	F	F	F <sub>min</sub>	I	F	I	F	I	F	I	F	I	F
GroEL <sub>14</sub> /ES <sub>7</sub> /ADP <sub>7</sub>	0.64	0.87	13.62	10.46	34	63	0.8	2.3	0.8	0.8	0.2	0.2	3.3	3.7
GroEL <sub>14</sub>	0.59	0.94	15.22	12.28	33	56	0.6	1.9	0.4	0.4	1.2	1.2	3.3	3.7
EFG	0.68	0.89	9.93	8.35	41	64	0.5	1.1	1.8	1.8	0.6	0.6	3.2	3.5
Prefoldina	0.86	0.92	5.17	4.75	15	37	0.3	0.6	0.0	0.0	0.0	0.0	1.9	2.4
Clamp*	0.86	0.91	8.14	7.69	30	59	0.4	1.1	0.3	0.3	0.9	0.9	2.9	3.4
ATPasa-Ca	0.67	0.90	14.15	13.08	35	88	0.8	3.1	0.6	0.6	0.6	0.6	3.0	3.5
CopAdC†	0.78	0.84	5.11	3.62	106	142	1.1	1.7	0.2	0.2	1.2	1.2	3.6	3.7
Termosoma	0.29	0.60	18.39	16.87	28	36	1.0	1.3	0.0	0.0	1.1	1.1	3.0	3.2
ARNP abierta	0.88	0.91	7.82	7.37	84	147	9.0	11.2	0.5	0.5	1.8	1.8	4.0	4.2
ARNP cerrada	0.88	0.92	10.26	9.52	84	166	9.0	12.6	0.4	0.4	1.8	1.8	4.0	4.3
Ndc80/Nuf2	0.23	0.24	2.84	1.73	13	45	0.0	0.0	0.0	0.0	0.3	0.3	2.6	2.9
Tubulina	0.37	0.40	2.67	1.62	110	103	10.7	11.2	0.0	0.0	0.6	0.6	4.3	4.3
Promedio	0.64	0.78	9.44	8.11	51	84	2.9	4.0	0.4	0.4	0.9	0.9	3.3	3.6

<sup>a</sup> Los ajustes se han llevado a cabo con *iModfit* utilizando los parámetros por defecto, excepto en los siguientes casos: en (\*) se utilizan los 20 primeros modos y en (†) se emplea el 2% de ellos considerando rígidas las hélices- $\alpha$  y las láminas- $\beta$ .

<sup>b</sup> Correlación normalizada entre los mapas experimentales y los mapas simulados de las estructuras iniciales (I) y finales (F) correspondientes.

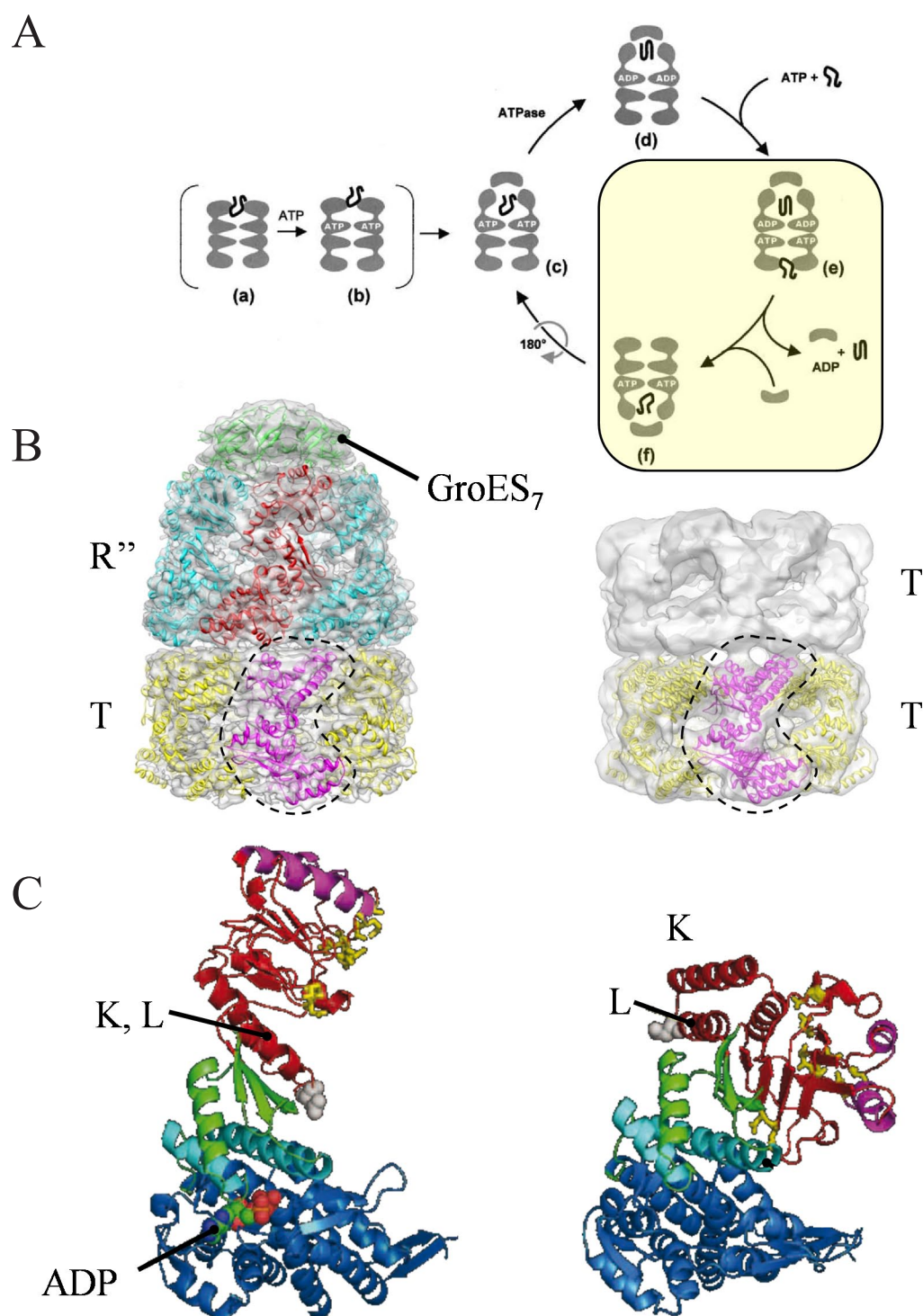
<sup>c</sup> Desviación ( $C_\alpha$ ) entre la estructura inicial y el modelo final (F). También se muestra la desviación mínima obtenida minimizando el RMSD ( $C_\alpha$ ) entre ambas estructuras ( $F_{\min}$ ) (Kabsch, 1978).

<sup>d</sup> Métricas de validación estructural de *Molprobit*.

ilustrado individualmente cada uno de los ajustes. Nótese que, gracias a la utilización de CI, las longitudes y ángulos de enlace inadecuados se mantienen en sus niveles basales en todos los casos. A continuación se comentan en detalle cada uno de los ajustes experimentales.

### Sistema GroEL/GroES

La chaperona GroEL interviene en el plegamiento de multitud de proteínas bacterianas. Actúa capturando las cadenas polipeptídicas en proceso de plegamiento y les proporciona el entorno adecuado para alcanzar su estructura nativa. Su ciclo de funcionamiento consta de múltiples cambios conformacionales cooperativos de gran amplitud dirigidos por la hidrólisis de ATP y las interacciones con el sustrato y la co-chaperona GroES<sub>7</sub> (Figura 50A). En los ajustes se han utilizado dos mapas de densidad obtenidos por crio-microscopía de 8 y 10Å de resolución. Estos mapas corresponden, respectivamente, al complejo GroEL<sub>14</sub>-ADP<sub>7</sub>-GroES<sub>7</sub> y a GroEL<sub>14</sub> en ausencia de ligandos (Figura 50B). El primero consta de dos anillos de GroEL<sub>7</sub>, en conformaciones abierta (R'') (Figura 50C, izquierda) y cerrada (T) (Figura 50C, derecha), junto con un anillo de GroES<sub>7</sub> en el extremo del anillo R''. El segundo está formado únicamente por dos anillos de GroEL<sub>7</sub> en la misma conformación que la del anillo T del primer complejo.



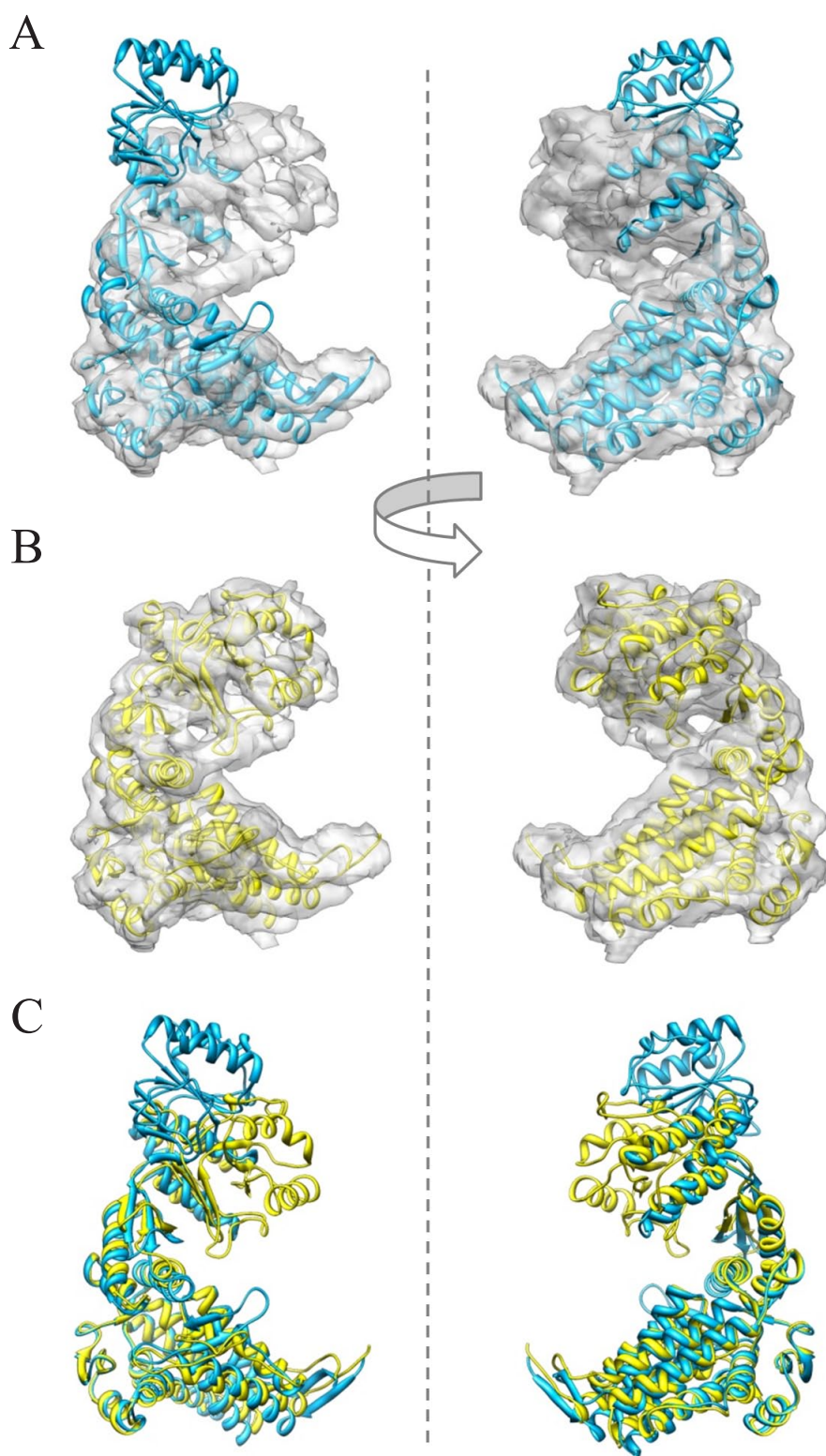
**Figura 50.** Sistema GroEL/GroES. En el panel A se ilustra esquemáticamente su ciclo de funcionamiento. En B se muestra el resultado del ajuste rígido entre la estructura *Isx4* completa y el mapa de GroEL<sub>14</sub>-ADP<sub>7</sub>-GroES<sub>7</sub> (izquierda), y entre el anillo en conformación T de *Isx4* y el mapa de GroEL<sub>14</sub> en ausencia de ligandos (derecha). También se ilustra la región segmentada de los mapas (línea punteada). En C se han representado la conformación R'' (izquierda) y T (derecha) de los monómeros indicados en rojo y magenta, respectivamente, en el panel B. Los dominios apical, intermedio y ecuatorial aparecen en rojo, verde y azul, respectivamente. La representación del panel A ha sido tomada de Ranson *et al.* (Ranson *et al.* 2001) y la del panel C de Hyeon *et al.* (Hyeon, *et al.*, 2006).



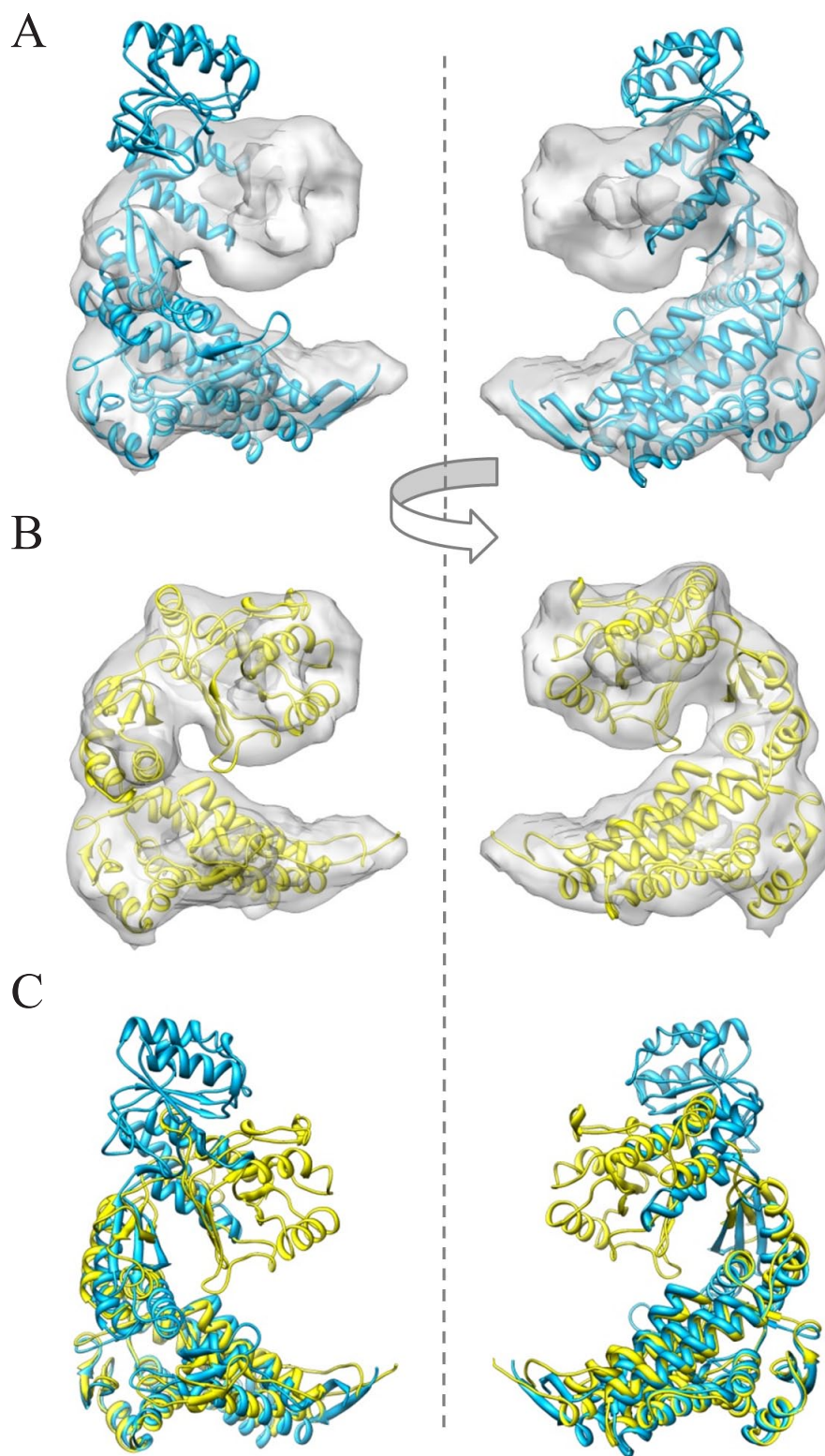
Para llevar a cabo los ajustes se ha segmentado un único monómero de los anillos T de cada mapa. Como se ilustra en la Figura 50B, después de ajustar rigidamente la estructura atómica del complejo GroEL<sub>14</sub>-ADP<sub>7</sub>-GroES<sub>7</sub> (*Ixx4*) en cada uno de los mapas completos, se han segmentado las regiones situadas dentro de una máscara obtenida por filtración a 20Å de la estructura atómica de uno de los monómeros en conformación T (líneas punteadas). En estos casos, la orientación inicial se ha determinado superponiendo la estructura atómica del monómero abierto de GroEL sobre la estructura cerrada utilizada para segmentar los mapas con el programa *Chimera* (Pettersen *et al.*, 2004), que superpone las regiones más similares de ambas estructuras. Así, se obtienen dos casos de ajuste flexible experimental del monómero de GroEL que representan la transición de la conformación abierta (R'') a la cerrada (T). Este movimiento corresponde a la liberación del sustrato, GroES<sub>7</sub> y ADP de uno de los anillos, y a la unión de un nuevo sustrato y GroES<sub>7</sub> en el anillo opuesto (véase el sombreado de la Figura 50A).

En las Figuras 51 y 52 se muestran dos vistas laterales de las estructuras iniciales (A) y los modelos finales (B) de los ajustes flexibles correspondientes a los mapas (en transparencia) de los complejos GroEL<sub>14</sub>-ADP<sub>7</sub>-GroES<sub>7</sub> y GroEL<sub>14</sub> sin ligandos, respectivamente. En concordancia con el cambio conformacional mostrado en la Figura 50C, en ambos casos el dominio apical (en colores rojo y magenta) gira hacia abajo en el sentido de las agujas del reloj para introducirse en la densidad de la estructura cerrada. Este cambio viene acompañado por el movimiento de dentro a fuera de las hélices K y L del dominio apical, tal y como se describe en Hyeon *et al.* (Hyeon, *et al.*, 2006). Los modelos finales se adaptan muy bien la densidad de los mapas, lo que permite obtener correlaciones de 0.87 con GroEL<sub>14</sub>-ADP<sub>7</sub>-GroES<sub>7</sub> y 0.94 con GroEL<sub>14</sub> sin ligandos (Tabla 36). Como las conformaciones T de los dos mapas son muy parecidas a la conformación T de la estructura cristalográfica, es posible estimar la calidad del ajuste utilizando los RMSD entre esta última y los modelos ajustados. De este modo se puede comprobar que los ajustes del complejo GroEL<sub>14</sub>-ADP<sub>7</sub>-GroES<sub>7</sub> y de GroEL<sub>14</sub> sin ligandos tan solo se desvían de la estructura cristalográfica 3.89 y 3.76Å, respectivamente. Más aún, si en lugar de calcular las desviaciones directamente a partir de la estructura cristalográfica T, se ajusta ésta previamente con *iModfit* a cada uno de los mapas, las desviaciones se reducen hasta 1.34 y 2.91Å, respectivamente. Nótese que las estructuras no han sido alineadas previamente para calcular estas desviaciones ( $C_\alpha$ ). A pesar de la gran amplitud del cambio conformacional (>13Å) que tiene lugar en los ajustes de las Figuras 51 y 52 (C), la calidad de las estructuras se mantiene. Tan solo se producen incrementos moderados en el número de colisiones, *i.e.* de las 33~34 iniciales hasta unas 60, y en el porcentaje de ángulos diedros fuera de las regiones permitidas de Ramachandran que alcanza el ~2% (Tabla 36).





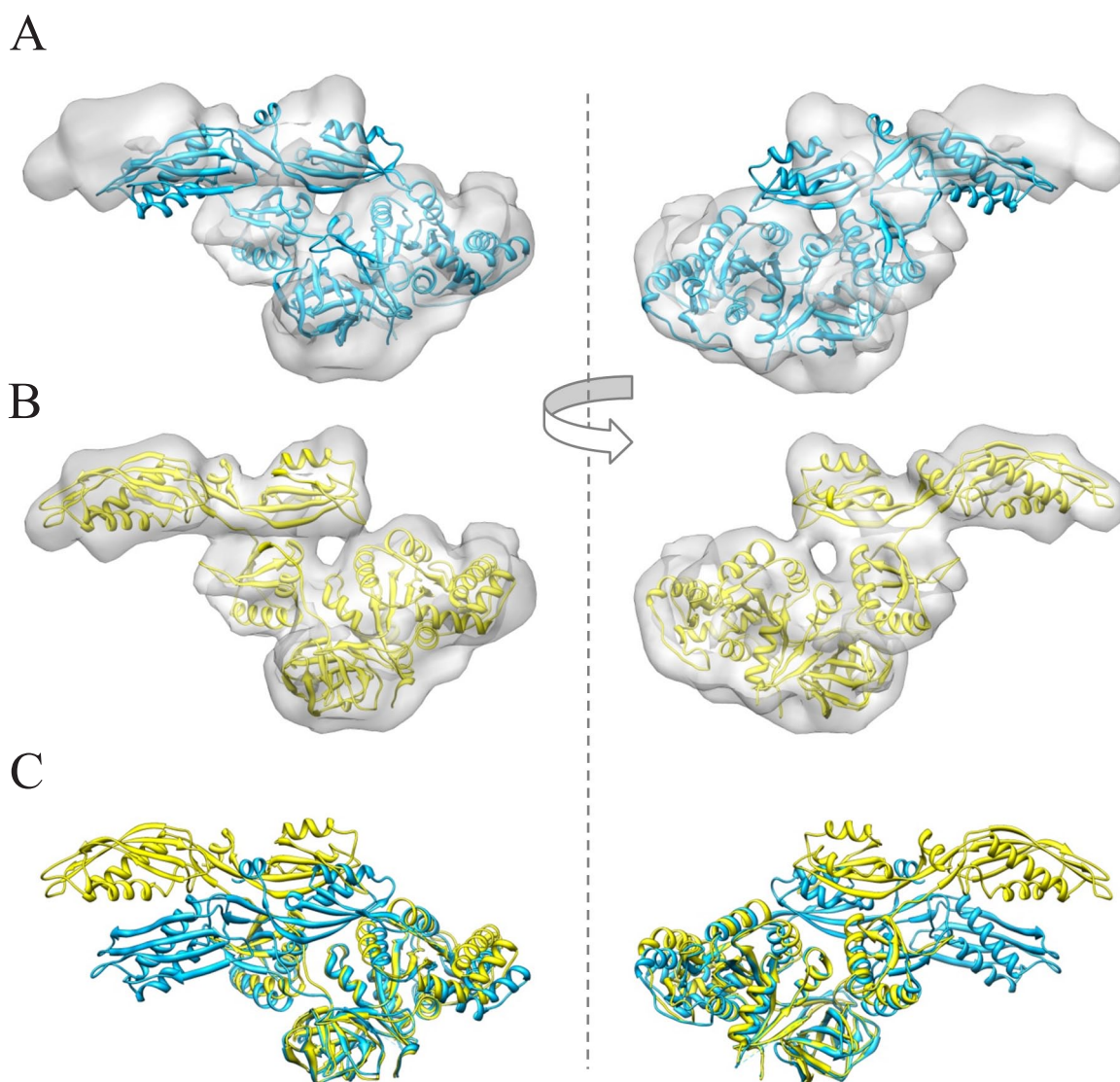
**Figura 51.** Ajuste flexible de la estructura atómica de GroEL en un mapa experimental de microscopía electrónica a 8Å. En los paneles A y B se han representado dos vistas laterales de las estructuras inicial (azul) y ajustada (amarillo), respectivamente. En C se muestra la estructura inicial alineada sobre la ajustada con *Chimera*.



**Figura 52.** Ajuste flexible de la estructura atómica de GroEL en un mapa experimental de microscopía electrónica a 10Å. En los paneles A y B se han representado dos vistas laterales de las estructuras inicial (azul) y ajustada (amarillo), respectivamente. En C se muestra la estructura inicial alineada sobre la estructura final con *Chimera*.

### Factor de Elongación G

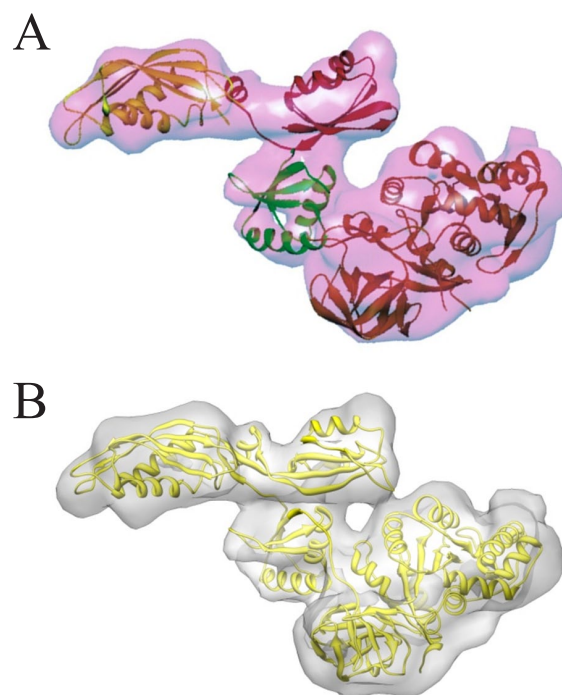
El factor de elongación G (EFG, del inglés *Elongation Factor G*) interacciona con el ribosoma durante la traducción de la información genética contenida en el ARNm para producir proteínas. Este factor cataliza la translocación del ARNt y del ARNm en cada ciclo de elongación de la cadena polipeptídica naciente utilizando GTP. Cuando el EFG (con GTP) se une al sitio A (*Aminoacil*) del ribosoma, desplaza el ARNt que previamente ocupaba esa posición hasta una posición intermedia A/P (sitios *Aminoacil*-30S y *Peptidil*-50S ocupados) y el ARNt del sitio P hasta la posición intermedia P/E (*Peptidil/Exit*). Entonces, la hidrólisis del GTP provoca un cambio conformacional que empuja a los ARNt hasta las posiciones P/P y E/E,



**Figura 53.** Ajuste flexible de la estructura atómica del EFG en un mapa experimental de microscopía electrónica a 11Å. En los paneles A y B se han representado dos vistas laterales de las estructuras inicial (azul) y ajustada (amarillo), respectivamente. En C se muestra la estructura inicial alineada con Chimera sobre el modelo ajustado.



y al ARNm tres nucleótidos más adelante. Después, el EFG (con GDP) es liberado dejando libre el sitio A para un nuevo ARNt. Para ajustar la estructura atómica del EFG se ha utilizado un mapa segmentado de dicho factor a partir de la estructura completa del ribosoma (70S·MFTI-ARNt<sup>lle</sup>·EFG·GDPNP·puromicina) a unos 11Å de resolución. Como orientación inicial se ha utilizado el ajuste de mayor correlación obtenido con *ADP\_EM*. En la Figura 53 se muestran dos vistas laterales del mapa experimental (en transparencia) sobre las estructuras iniciales (A) y los modelos finales (B). A pesar de que inicialmente una buena parte de la estructura atómica se encuentra fuera del mapa, al efectuar el ajuste flexible el modelo atómico se adapta completamente al mapa de densidad y la correlación aumenta de 0.68 a 0.89 (Tabla 36). Después de este cambio conformacional de gran amplitud (~10Å, panel C), los elementos de estructura secundaria permanecen inalterados y tanto el número de colisiones como el porcentaje de ángulos diedros en regiones no permitidas tan solo aumentan de 41 a 61 y de 0.5 a 1.1%, respectivamente. Este caso de ajuste flexible experimental ya ha sido caracterizado con distintos métodos de ajuste flexible (Kovacs, *et al.*, 2008; Tama, *et al.*, 2004b; Valle, *et al.*, 2003; Velazquez-Muriel, *et al.*, 2006). Los modelos producidos por estos métodos son muy similares entre sí y resultan compatibles con el modelo obtenido automáticamente con *iModfit*. Por ejemplo, en la Figura 54 puede comprobarse el gran parecido entre este último y el modelo obtenido por Valle *et al.* (Valle, *et al.*, 2003) mediante el ajuste rígido manual de cada dominio.

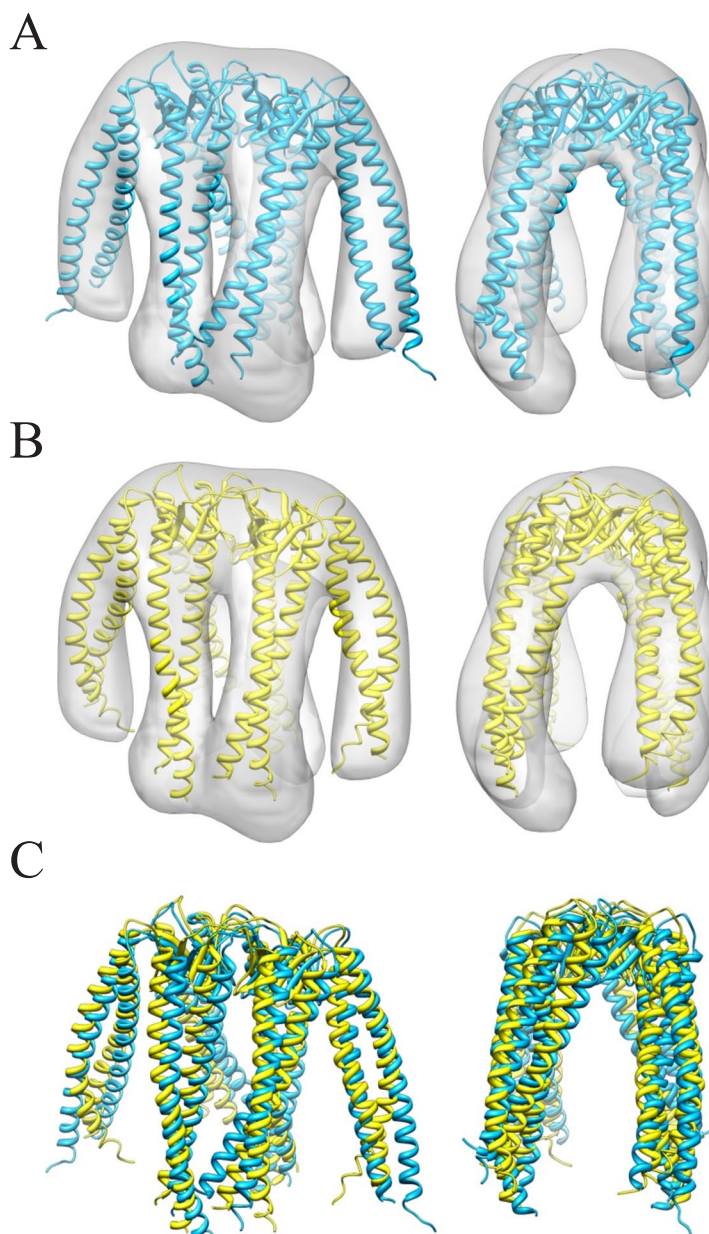


**Figura 54.** Comparación de los ajustes flexibles de la estructura atómica del EFG en un mapa experimental. En el panel A se ha reproducido el modelo ajustado publicado en Valle *et al.* (Valle, *et al.*, 2003) y en el panel B el obtenido con *iModfit*.

### Prefoldina

La prefoldina colabora con otra chaperona, CCT, para facilitar el plegamiento de algunas proteínas de arqueobacterias y células eucariotas, como por ejemplo, de actinas y tubulinas. Su estructura se asemeja a una medusa de seis tentáculos cuyas puntas cuentan con residuos hidrofóbicos que permiten estabilizar proteínas en conformaciones no-nativas de diferentes tamaños y formas (Martin-Benito, *et al.*, 2002; Martin-Benito, *et al.*, 2007). El ajuste flexible de la estructura atómica de la prefoldina en el mapa experimental a 20Å de resolución se ilustra

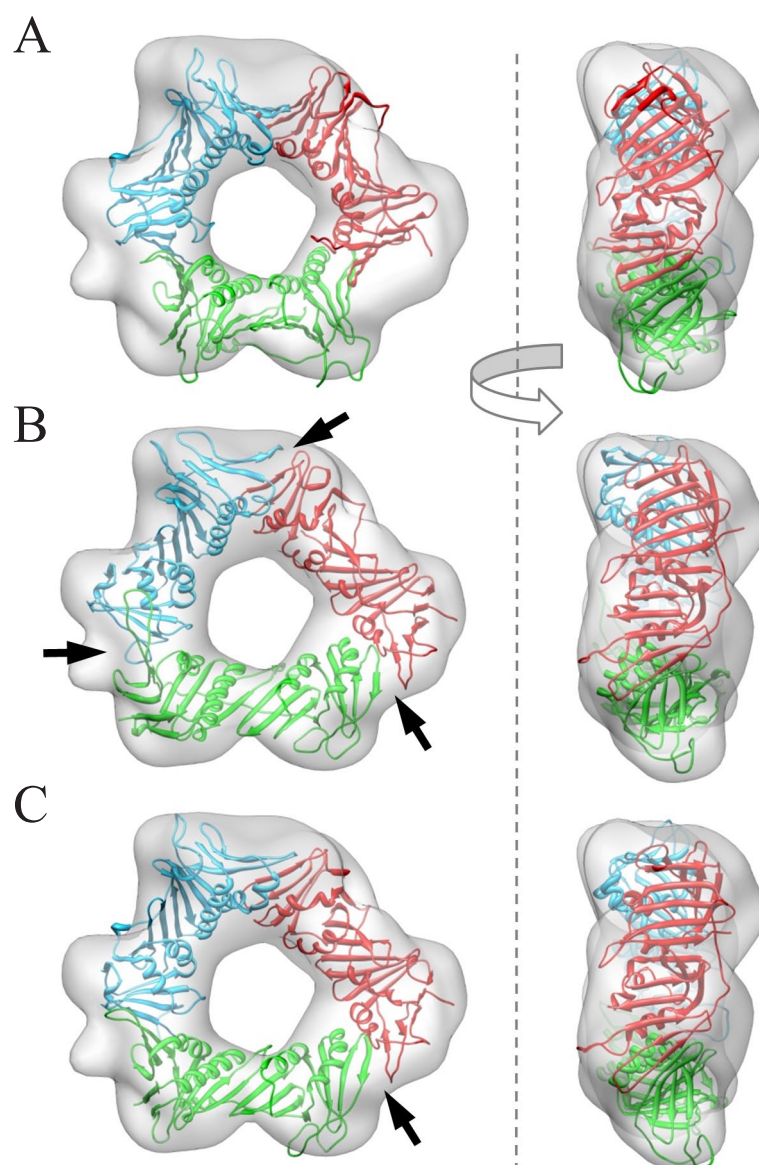
en la Figura 55. Se han representado dos vistas de las estructuras iniciales (A) y finales (B) superpuestas sobre el mapa (en transparencia). Después de efectuar el ajuste flexible, los tentáculos que inicialmente sobresalen del mapa (A) se introducen completamente en la densidad (B). Este movimiento de unos 5Å de RMSD (C) permite incrementar la correlación de 0.86 a 0.92 sin alterar significativamente la calidad del modelo final, *i.e.* las colisiones aumentan de 15 a 37 y el porcentaje de ángulos diedros no permitidos pasa de 0.3 al 0.6%. Dicho movimiento está en consonancia con la flexibilidad conformacional necesaria para que esta proteína pueda unirse a distintos sustratos de diferentes tamaños.



**Figura 55.** Ajuste flexible de la estructura atómica de prefoldina en un mapa experimental de microscopía electrónica a 20Å. En los paneles A y B se muestran dos vistas del mapa en transparencia con las estructuras inicial (azul) y ajustada (amarillo), respectivamente. En C se han representado únicamente las estructuras ajustadas.

### Clamp (Rad9-Hus1-Rad1)

El complejo heterotrimérico clamp (Rad9-Hus1-Rad1) tiene un papel central en las respuestas celulares desencadenadas al producirse daños en el ADN. Su función consiste en activar puntos de control del ciclo celular y reclutar enzimas que reparan las lesiones cuando las ADN polimerasas replicativas se paran dejando sin aparear grandes segmentos ADN de hebra única (Sohn y Cho, 2009). Para determinar la orientación inicial entre la estructura atómica y el mapa de clamp se ha llevado a cabo el ajuste rígido con *ADP\_EM*. En la Figura 56A se muestran dos vistas ortogonales de este ajuste. Dada la forma anular de esta estructura y la baja



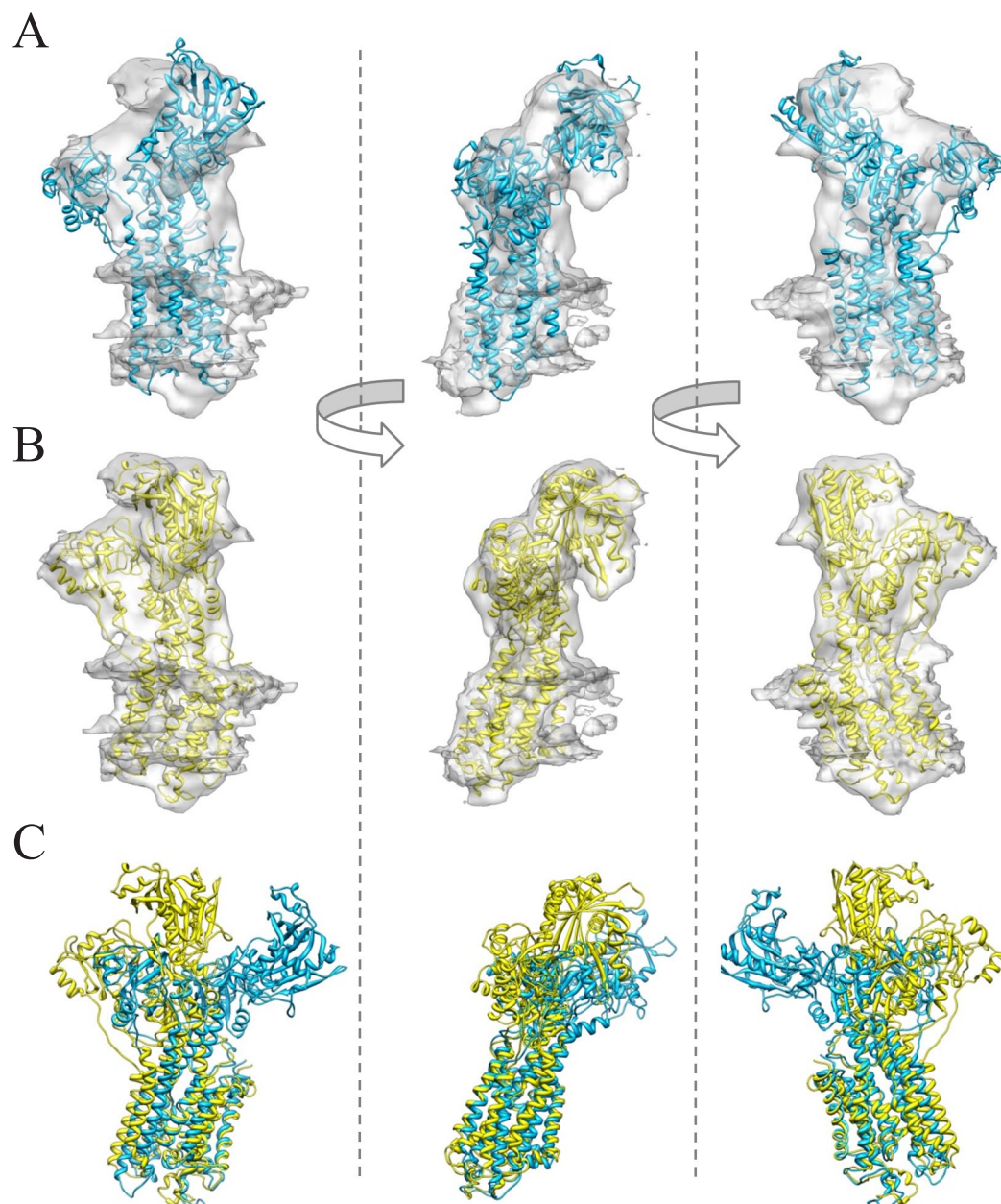
**Figura 56.** Ajuste flexible de la estructura atómica de la proteína clamp en un mapa experimental de microscopía electrónica a 17Å. En el panel A se muestra el ajuste rígido inicial entre la estructura atómica y el mapa de densidad (en transparencia). Se han representado los resultados del ajuste flexible utilizando: los parámetros por defecto (B) y los 20 primeros modos con los valores por defecto de los demás parámetros (C). Los monómeros Rad9, Hus1 y Rad1 se muestran en colores azul, verde y rojo, respectivamente.



resolución del mapa ( $\sim 17\text{\AA}$ ), se obtuvieron múltiples soluciones prácticamente equivalentes (datos no mostrados). Al efectuar los ajustes flexibles con los parámetros por defecto (B) se aprecian ligeras distorsiones en las estructuras de lámina- $\beta$  que contribuyen al mantenimiento del anillo (flechas). La calidad estructural de este modelo es relativamente baja dado que aumentan mucho tanto el número de colisiones, de 30 a  $\sim 100$ , como el porcentaje de ángulos diedros fuera de las regiones permitidas, del 0.4 a 5.1% (datos no mostrados). Estas distorsiones seguramente sean debidas a la presencia en el mapa de varias protuberancias y diferencias que rompen la pseudo-simetría de la estructura atómica del hetero-trímero y provocan la distorsión del anillo. Para minimizar este problema se ha repetido el ajuste flexible, pero utilizando únicamente los 20 primeros modos. En el panel C puede comprobarse que así las distorsiones disminuyen significativamente, por lo que la reducción del número de modos en los ajustes permite limitar fácilmente el sobreajuste cuando la resolución y la calidad de los mapas experimentales son bajas. En la Tabla 36 se muestra que incluso con 20 modos es posible alcanzar correlaciones de 0.91 y que después de un movimiento de unos  $8\text{\AA}$  (RMSD) la calidad del modelo final sigue siendo buena, *i.e.* 59 colisiones y un 1.1% de ángulos fuera de las regiones permitidas de Ramachandran. El movimiento observado podría estar relacionado con su deslizamiento a lo largo del ADN o con la formación de complejos con las ADN polimerasas.

### ATPasa de calcio

La ATPasa de calcio del retículo sarco/endoplasmático del músculo esquelético es una proteína de membrana que mantiene y restaura los niveles de calcio citoplasmáticos después de cada contracción muscular bombeando calcio hacia el retículo sarco/endoplasmático gracias a la energía química del ATP (Moller, *et al.*, 2010). Su estructura consta de cuatro dominios: transmembrana (T), de unión a nucleótido (N), de fosforilación (P) y actuador (A). Todos ellos sufren grandes cambios conformacionales durante el bombeo de calcio hacia el otro lado de la membrana. En la Figura 57 se ilustran los resultados del ajuste flexible de la estructura atómica de la ATPasa de calcio en un mapa experimental de crio-microscopía a  $8\text{\AA}$  de resolución obtenido a partir de cristales tubulares (Zhang, *et al.*, 1998). En ella se muestran tres vistas ortogonales entre sí de la estructura inicial (A) y del modelo final (B) con el mapa en transparencia, así como también de ambas estructuras atómicas después de ser alineadas entre sí con *Chimera* (C). La orientación de la estructura inicial (A) corresponde al ajuste de mayor correlación obtenido con *ADP\_EM*. Tras efectuar el ajuste flexible con los parámetros por defecto, todos los dominios de la estructura atómica que inicialmente están fuera del mapa quedan ahora totalmente cubiertos por la densidad (B). La mayor parte del cambio conformacional sucede en los dominios N y A, aunque las hélices del dominio T también experimentan cambios significativos. El ajuste

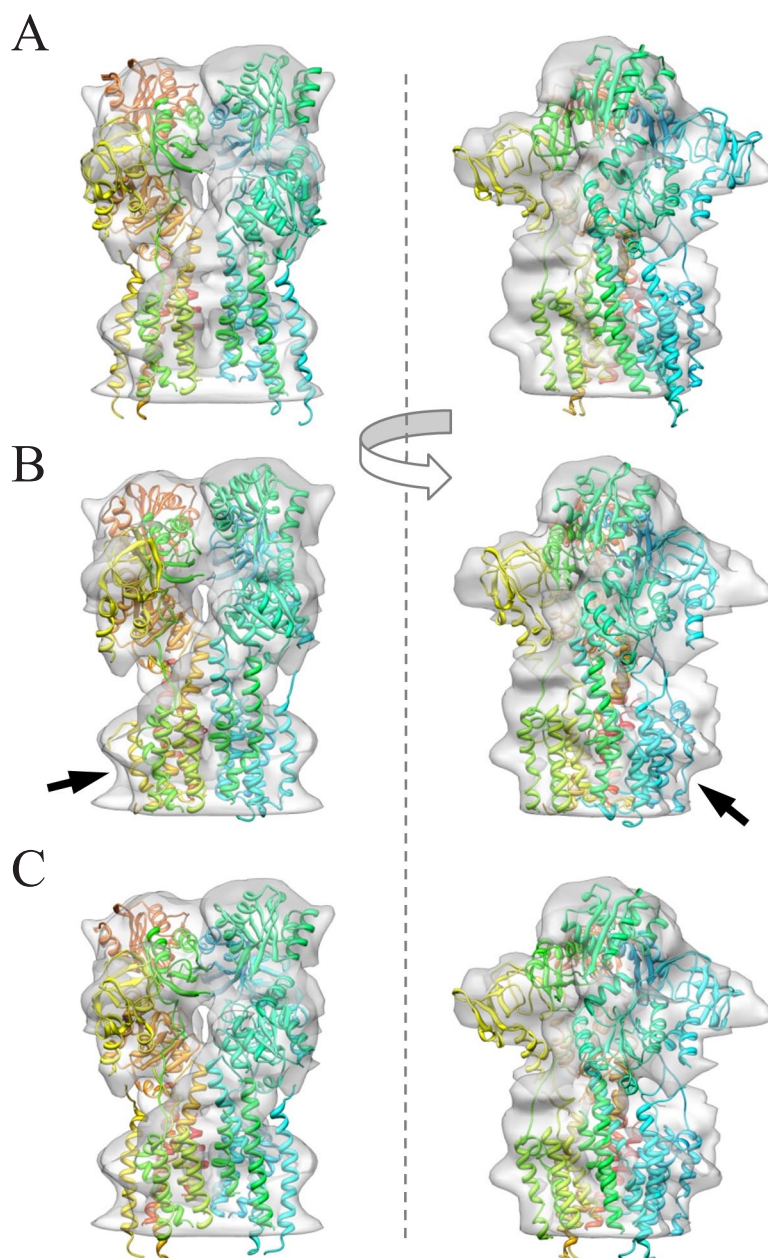


**Figura 57.** Ajuste flexible de la estructura atómica de la ATPasa de calcio en un mapa experimental de microscopía electrónica de 8Å. En los paneles A y B se han representado tres vistas ortogonales entre sí de las estructuras inicial (azul) y ajustada (amarillo), respectivamente. En C se muestra la estructura inicial alineada con *Chimera* sobre el modelo final.

implica un movimiento de unos 14Å de RMSD que eleva la correlación de 0.67 a 0.90. Estos resultados son totalmente compatibles con los obtenidos utilizando el método de Hinsén (Hinsén, *et al.*, 2005). En cuanto a la calidad del modelo generado, se produce un incremento tanto en el número de colisiones (de 35 a 88) como en el porcentaje de ángulos diedros en regiones no permitidas de Ramachandran (de 0.8 a 3.1%). No obstante, estas distorsiones resultan razonables teniendo en cuenta la gran amplitud del cambio conformacional y que los elementos de estructura secundaria permanecen prácticamente inalterados (C).

### Bomba de cobre

Al igual que el anterior complejo, la bomba de cobre es una proteína de membrana con un papel central en el mantenimiento de las concentraciones adecuadas de cobre intracelular, *i.e.* permite la presencia de suficiente cobre para sintetizar las metalo-enzimas necesarias, mientras que previene el daño oxidativo que una concentración elevada podría ocasionar a la célula. En



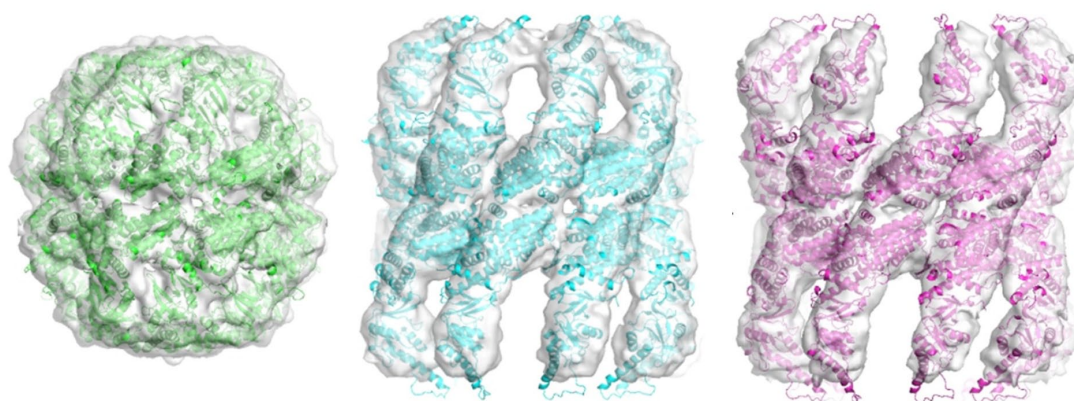
**Figura 58.** Ajuste flexible de la estructura atómica de la bomba de cobre en un mapa experimental de microscopía electrónica a 17Å. En el panel A se muestra el ajuste rígido inicial entre la estructura atómica y el mapa de densidad (en transparencia). Se han representado los resultados del ajuste flexible utilizando tanto los parámetros por defecto (B) como el 2% de los modos fijando las hélices- $\alpha$  y láminas- $\beta$  (C). Cada color representa una cadena diferente.

este caso de ajuste se ha utilizado como estructura inicial el modelo pseudo-atómico obtenido por Wu *et al.* (Wu, *et al.*, 2008) a partir un mapa experimental a 17Å de resolución. Este modelo fue generado utilizando dos estructuras distintas de la ATPasa de calcio como moldes sobre los que se superpusieron manualmente las estructuras atómicas de cada dominio de forma compatible con el mapa de densidad. Dicho mapa experimental, que fue obtenido por crio-microscopía a partir de cristales tubulares, es el mapa empleado en estos ajustes. En el panel A de la Figura 58 se muestran dos vistas ortogonales de la estructura atómica inicial superpuesta con *ADP\_EM* sobre el mapa experimental a 17Å de resolución (en transparencia). Como puede apreciarse claramente, el ajuste de la estructura inicial todavía puede mejorarse dado que existen regiones fuera de la densidad. Para mejorar el ajuste se ha refinado su conformación mediante el ajuste flexible con *iModfit*. Al efectuar el ajuste flexible con los parámetros por defecto (panel B) se obtiene una correlación de 0.86. Sin embargo, algunas hélices- $\alpha$  del modelo final (flechas) se deforman excesivamente para introducirse en la densidad. Además, las colisiones aumentan de 106 hasta unas 190 colisiones (dato no mostrado). En ocasiones, la falta de densidad para justificar toda la estructura atómica o la baja resolución del modelo inicial pueden provocar sobreajuste. Para minimizar las distorsiones se ha repetido el ajuste utilizando únicamente el 2% de los modos (~60) y fijando los ángulos diedros de todas las hélices- $\alpha$  y láminas- $\beta$  (panel C). Así, aunque se obtiene una correlación ligeramente menor (0.84), únicamente se alcanzan 142 colisiones (partiendo de 106) y el porcentaje de ángulos diedros no permitidos tan solo aumenta del 1.1 al 1.7%. Como puede apreciarse en el panel C, las restricciones impuestas mantienen muy bien la integridad de las hélices del modelo final, a pesar de que no existe densidad suficiente para acomodarlas completamente y de que el modelo inicial ha sido generado a partir de dominios de estructuras atómicas diferentes.

### **Termosoma**

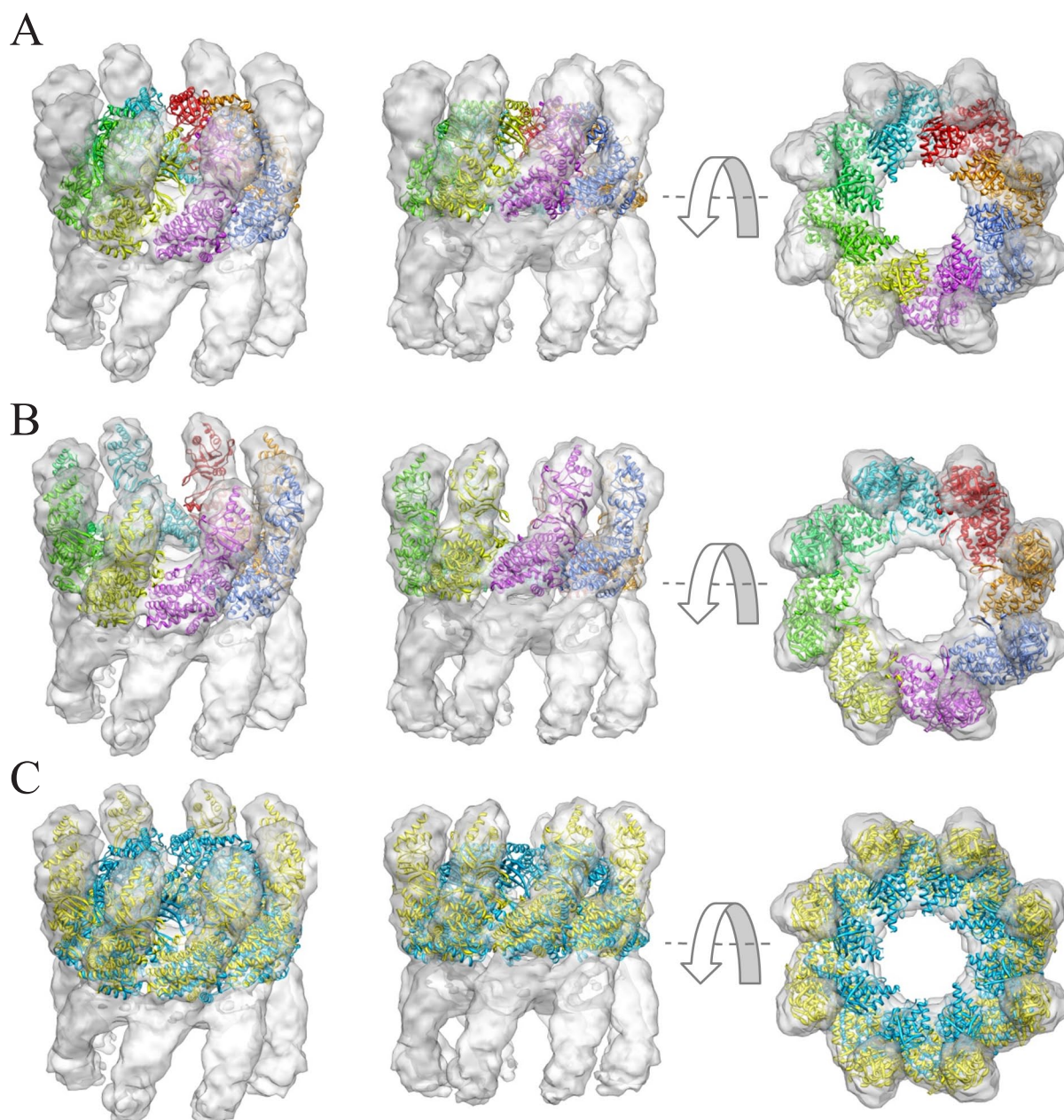
El termosoma es una chaperona molecular de arqueobacterias, homóloga de la chaperona eucariota CCT, que asiste a las proteínas durante su proceso de plegamiento para que alcancen correctamente su estructura nativa. Al igual que GroEL en bacterias, el termosoma posee sitios de unión hidrofóbicos en sus dominios apicales para unirse a los sustratos y un entorno hidrofílico en su interior; sin embargo, no requiere una co-chaperonina (GroES) para llevar a cabo su función gracias a una inserción en el dominio apical que actúa a modo de tapa. Aunque el termosoma experimenta grandes cambios conformacionales y existen varias estructuras a media/baja resolución de distintos intermediarios, únicamente se ha resuelto su estructura atómica en conformación cerrada (véase la Figura 59). Esta estructura consta de dos anillos idénticos que se unen para formar una estructura esférica altamente simétrica de unos 900 kDa





**Figura 59.** Mapas experimentales del termosoma. Los mapas se muestran superpuestos sobre la estructura cristalográfica (verde) y los modelos pseudo-atómicos generados por Clare *et al.* (Clare *et al.* 2008) ajustando rígidamente cada dominio por separado, en azul y magenta, respectivamente.

de masa molecular (Figura 59 izquierda). Los dominios apicales terminan en unas hélices- $\alpha$  que se entrelazan entre sí en los polos cerrando la esfera. En el ajuste se ha utilizado el mapa experimental en conformación abierta a 10Å de resolución mostrado en la Figura 59 (derecha) (Clare, *et al.*, 2008). Dado que en este mapa no se aprecia densidad para dichas hélices, éstas han sido eliminadas de la estructura. En la Figura 60A se han representado tres vistas de uno de los anillos de la estructura atómica después de ser ajustado rígidamente en el mapa completo con *ADP\_EM*. Al efectuar el ajuste flexible utilizando los parámetros por defecto se obtiene un modelo final que incrementa la correlación del 0.29 inicial hasta 0.60. Este valor es bastante elevado teniendo en cuenta que se ha considerado todo el mapa en el cálculo de la correlación y tan solo la mitad de la estructura atómica correspondiente. Como puede observarse en el panel B, el ajuste entre el modelo atómico y el mapa es excelente en todos los monómeros, la zona ecuatorial del anillo se mantiene en su posición, y no se aprecian distorsiones en ningún elemento de estructura secundaria. A pesar de que la amplitud del cambio conformacional es de unos 18Å de RMSD (panel C), la calidad estructural del modelo final también resulta excelente, con un ligero aumento de 28 a 36 en el número de colisiones y del 1.0 al 1.3% en el porcentaje de ángulos diedros fuera de las regiones permitidas de Ramachandran. Nótese que el modelo final es compatible con el modelo pseudo-atómico obtenido por Clare *et al.* (Clare, *et al.*, 2008) a partir del ajuste rígido de cada dominio por separado (Figura 59 derecha).



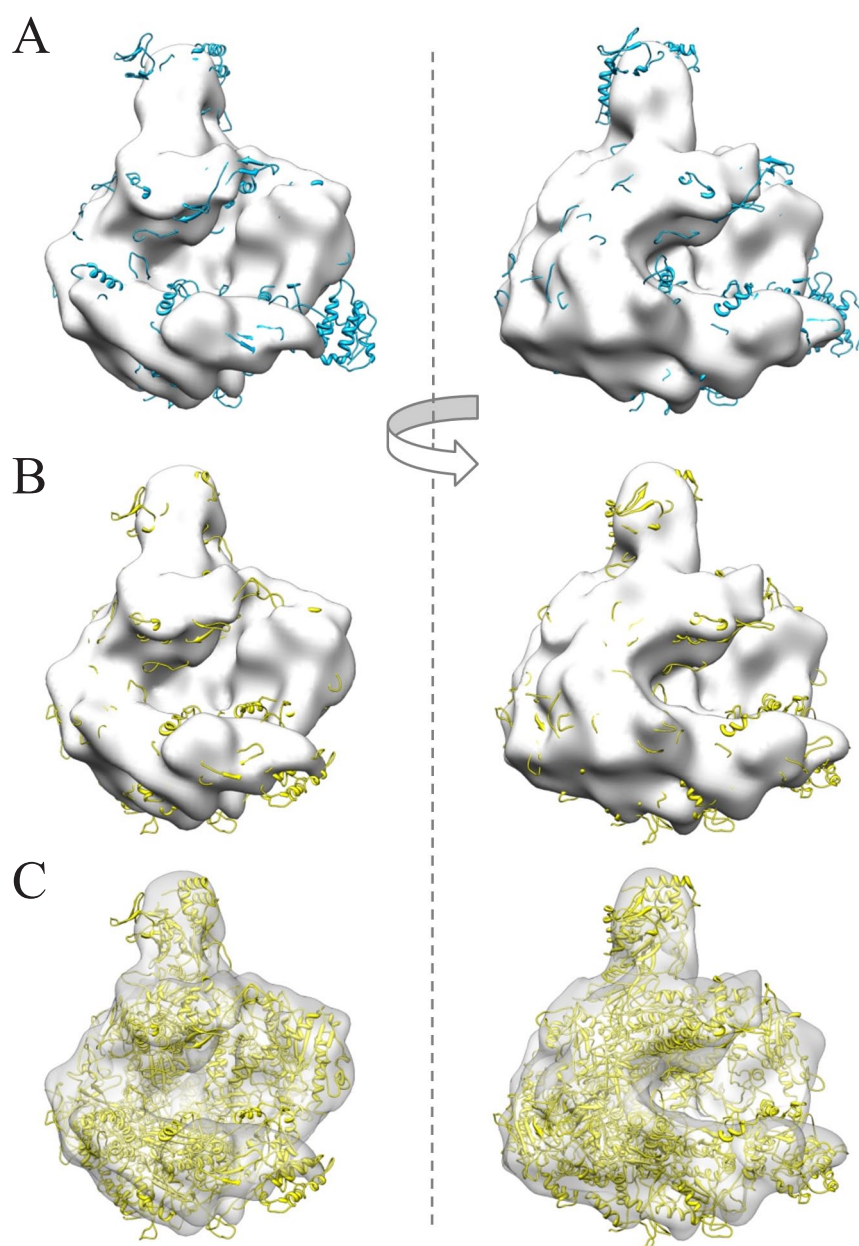
**Figura 60.** Ajuste flexible de la estructura atómica del termosoma en un mapa experimental de microscopía electrónica a 10Å. En el panel A se muestra la orientación inicial obtenida con *ADP\_EM*, en B, el modelo ajustado con *iModfit*, y en C, la superposición de la estructura inicial (azul) sobre dicho modelo (amarillo). El mapa se muestra en en transparencia.

### ARN polimerasa II

La ARN polimerasa II (ARNP) es la enzima encargada de sintetizar ARN mensajero a partir del ADN del núcleo de las células eucariotas para que pueda ser traducido en proteínas por los ribosomas. Su funcionamiento está altamente regulado por activadores y represores que interaccionan, o bien con ella directamente, o bien con las regiones del ADN donde ésta se une. Además, también juega un papel central en el proceso de reparación del ADN acoplado

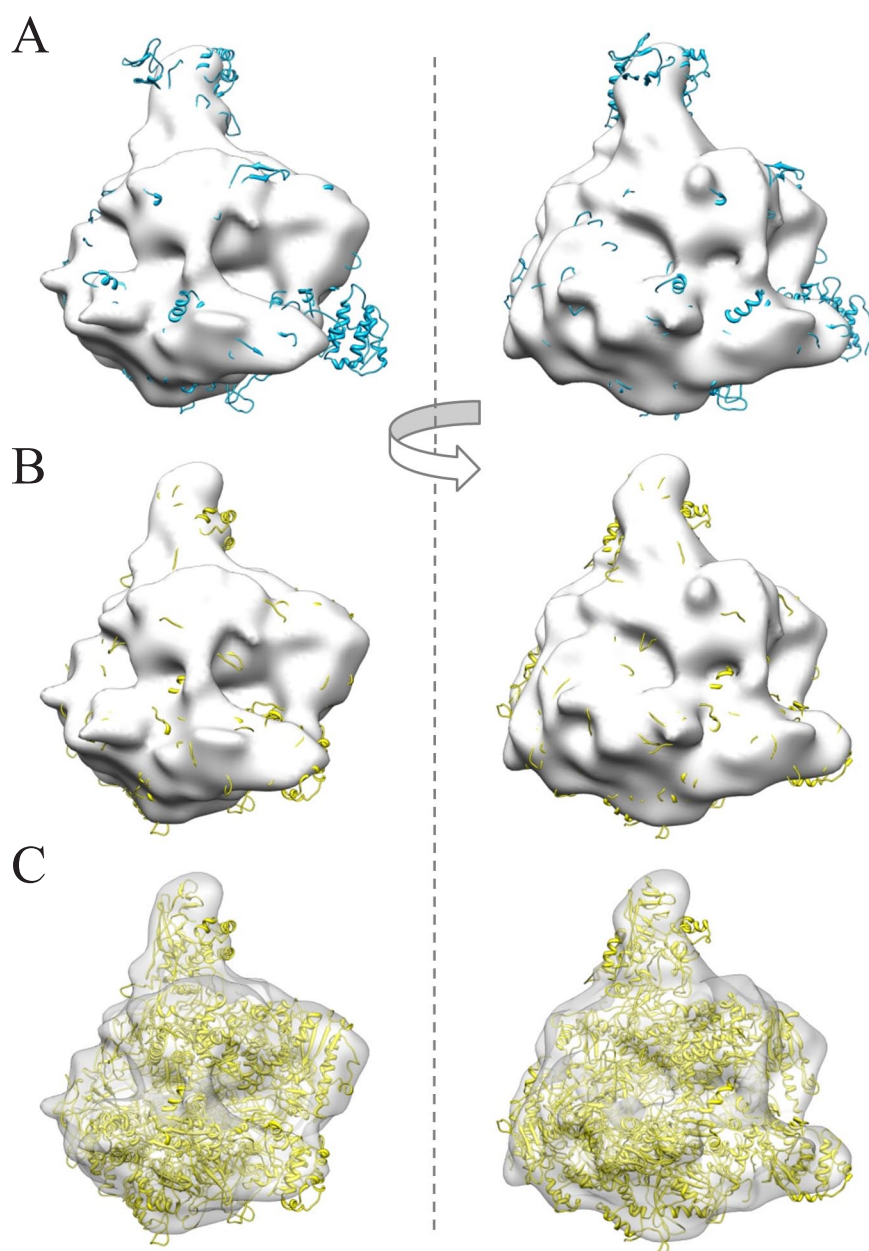


a la transcripción. La ARN polimerasa II de eucariotas es un complejo macromolecular de 12 subunidades y unos 500kDa que carece de simetría. Utilizando técnicas de crio-microscopía y SP ha sido posible reconstruir dos estructuras de la ARNP humana, una abierta y otra cerrada, a 22Å de resolución que coexisten en disolución (Kostek, *et al.*, 2006). Se han llevado a cabo sendos ajustes flexibles entre estos mapas y la estructura atómica de la ARNP de levadura. Aunque la forma humana se diferencia de la de levadura en varias inserciones y deleciones, sus estructuras son bastante parecidas. En ambos ajustes se ha utilizado como orientación inicial la del mejor ajuste obtenido con *ADP\_EM*. En el panel A de las Figuras 61 y 62 se muestran dos

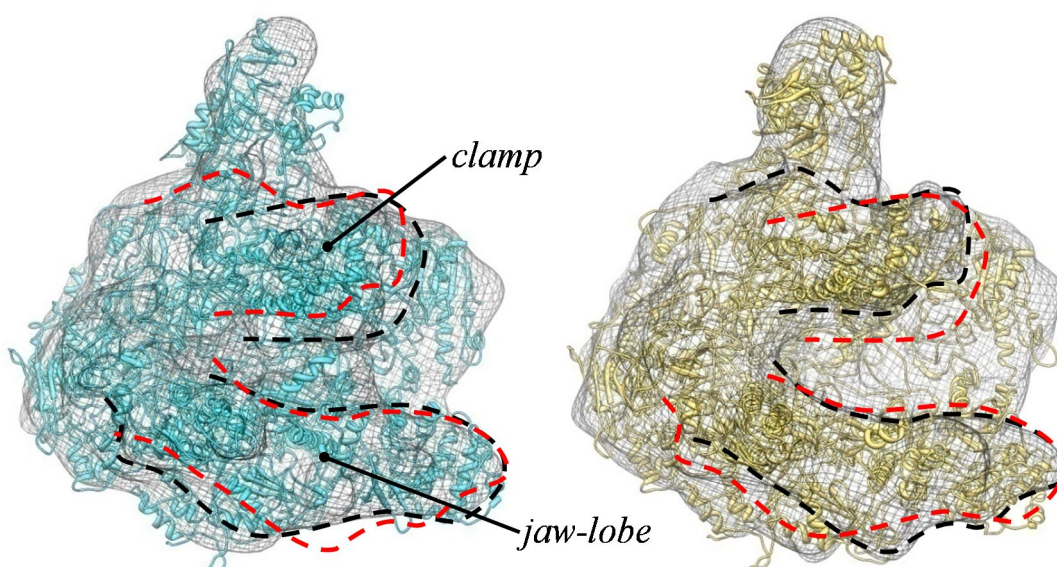


**Figura 61.** Ajuste flexible de la estructura atómica de la ARN polimerasa II humana en un mapa experimental abierto de microscopía electrónica a 22Å.

vistas ortogonales de las estructuras iniciales (azul) sobre sus respectivos mapas (blanco). Tras realizar los ajustes flexibles con los parámetros por defecto se observa que la estructura atómica cubre mejor las densidades (paneles B) alcanzando correlaciones de 0.91 y 0.92 con los mapas abierto y cerrado, respectivamente. Como puede observarse en los paneles C de ambas figuras, la integridad de los elementos de estructura secundaria se mantiene. No obstante, se alcanzan valores bastante elevados de colisiones (147 y 166) y de porcentajes de ángulos diedros no permitidos (11.2 y 12.6%). Esto puede deberse tanto a la baja resolución del mapa (22Å) como a la baja calidad de la estructura atómica inicial, que tiene 84 colisiones y un 9.0% de ángulos



**Figura 62.** Ajuste flexible de la estructura atómica de la ARN polimerasa II humana en un mapa experimental cerrado de microscopía electrónica a 22Å.



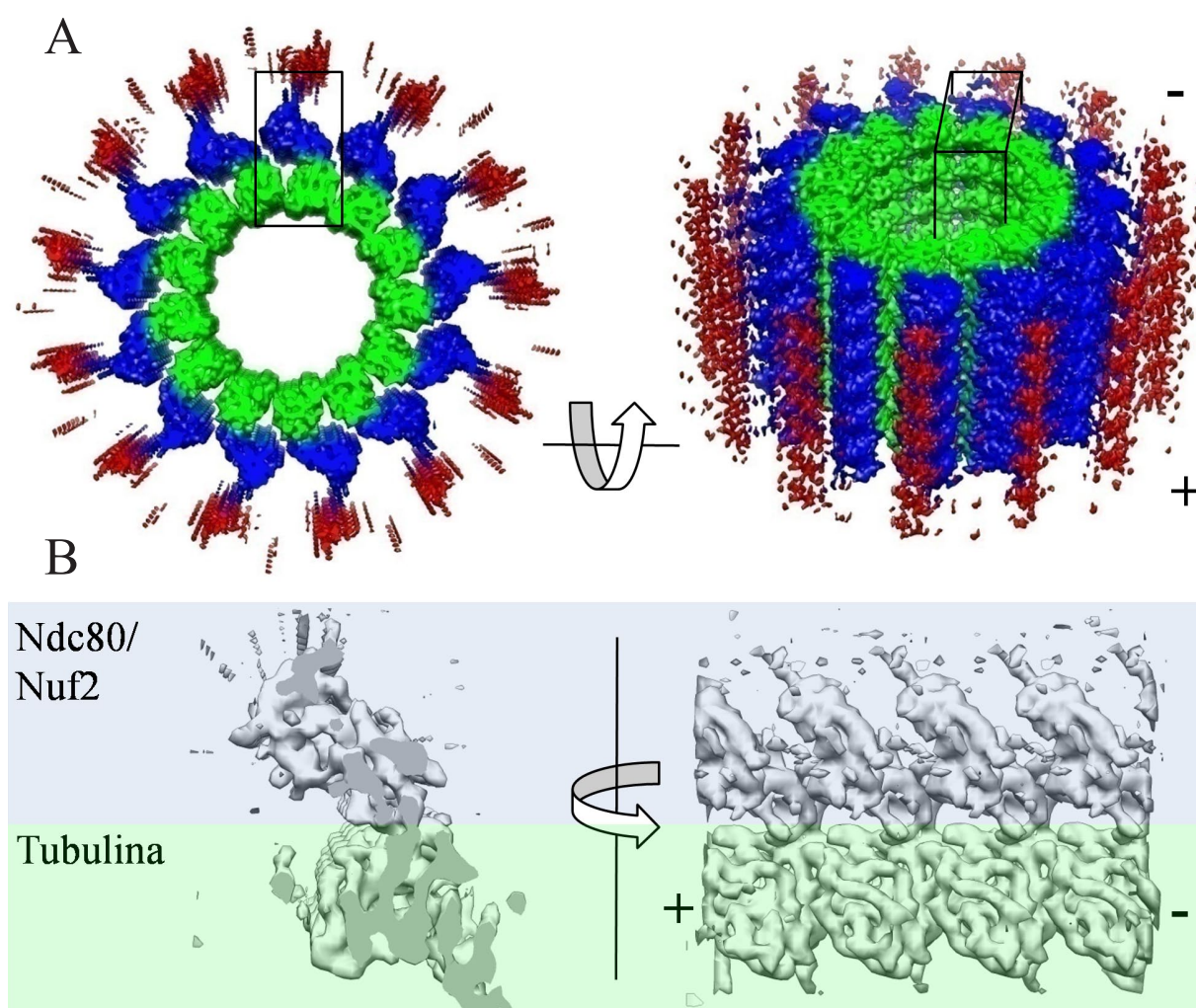
**Figura 63.** Cambios conformacionales de la ARN polimerasa II humana. Se han ilustrado los modelos finales obtenidos a partir de los mapas cerrado (izquierda) y abierto (derecha). Cada mapa se muestra en representación de malla. Por claridad, en cada mapa se han resaltado con líneas punteadas de color negro los contornos de las regiones *clamp* y *jaw-lobe*. También se muestran los contornos correspondientes a la otra conformación en color rojo.

diedros no permitidos. Además, las inserciones y deleciones que diferencian las secuencias de las formas humana y de levadura también pueden inducir deformaciones adicionales que reduzcan la calidad de estos modelos. En la Figura 63 se han resaltado las diferencias entre los modelos finales producto del ajuste flexible. La forma ajustada al mapa cerrado se diferencia principalmente de la ajustada al mapa abierto en el cierre de las regiones *clamp* y *jaw-lobe* para ocluir parcialmente la hendidura donde se une el ADN. En cualquier caso, el ajuste flexible de la estructura atómica de la ARNP permite modelar a nivel atómico estos cambios conformacionales y está en consonancia con los movimientos descritos en Kostek *et al.* (Kostek, *et al.*, 2006).

### Microtúbulos decorados con Ndc80/Nuf2

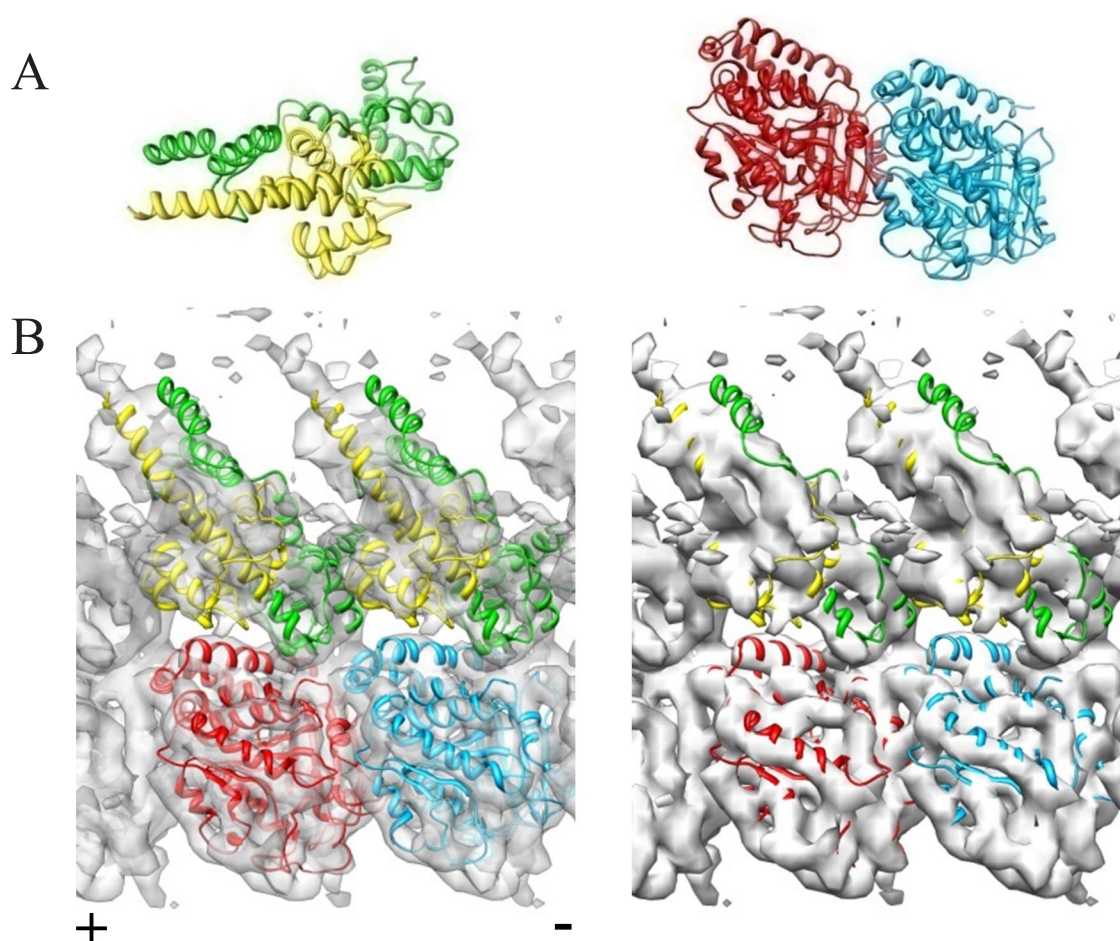
En el panel A de la Figura 64 se muestra la arquitectura general del mapa experimental de un microtúbulo (verde) decorado con una construcción de la proteína Ndc80/Nuf2 (azul y rojo) a 8Å de resolución (Alushin, *et al.*, 2010). La resolución del mapa es lo suficientemente alta como para identificar la mayor parte de los elementos de estructura secundaria en las regiones próximas al centro del microtúbulo (en verde y azul); sin embargo, en la periferia (en rojo) no es posible distinguir ninguna estructura con claridad. Por este motivo, en el ajuste se han eliminado las regiones de la estructura atómica de Ndc80/Nuf2 para las cuales no se observa densidad con nitidez. Dado el gran tamaño del complejo y su simetría tubular, se ha utilizado el mapa





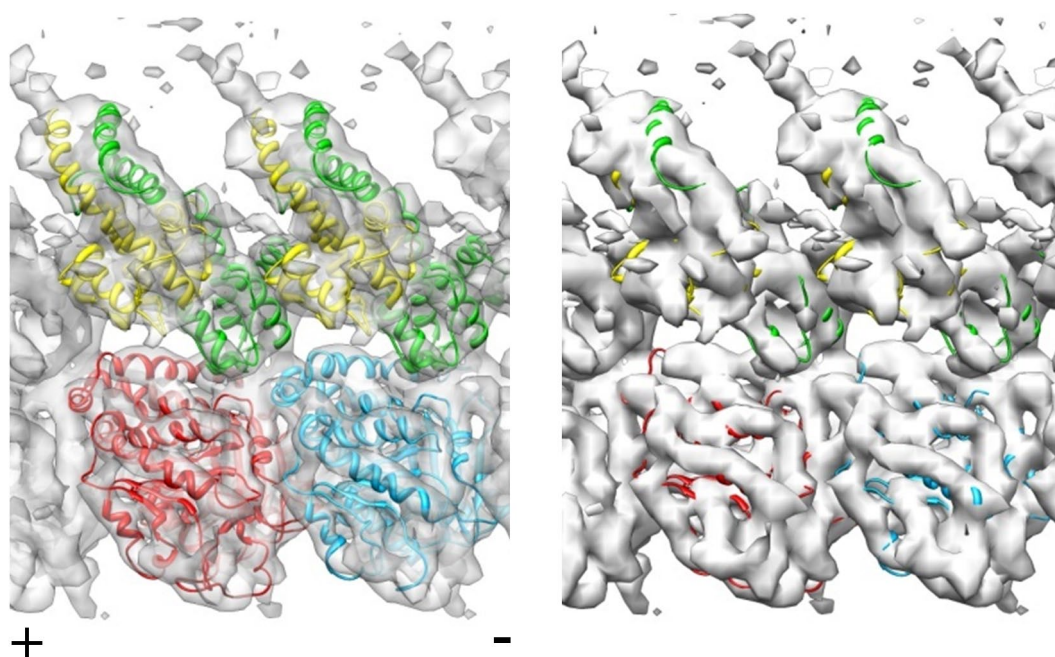
**Figura 64.** Estructura de los microtúbulos decorados con Ndc80/Nuf2 ilustrada sobre un mapa de microscopía electrónica a 8Å. El panel A muestra el complejo desde el extremo (-) del microtúbulo (izquierda) junto con una vista en perspectiva de (derecha). En el panel B se muestran dos vistas de un protofilamento segmentado del mapa completo (véase el recuadro del panel A): una desde el extremo (-) (izquierda) y otra lateral (derecha).

segmentado de un único proto-filamento (panel B) (comunicación personal de E. Nogales). Los ajustes tanto de cada una de las dos estructuras atómicas de Ndc80/Nuf2 como del dímero de tubulina se han llevado a cabo por separado. En la Figura 65 se muestran las distintas estructuras atómicas iniciales (A) y el resultado de los ajustes rígidos (B). Como se muestra en Alushin *et al.* (Alushin, *et al.*, 2010) el ajuste rígido de las estructuras atómicas permite la adecuada interpretación del mapa dada la similitud de ambas conformaciones. Sin embargo, todavía es posible acomodar mejor las estructuras atómicas en el mapa. En el panel B puede verse claramente como al utilizar un umbral algo más alto que el necesario para justificar la masa total del complejo, algunos elementos de estructura secundaria no se encuentran totalmente



**Figura 65.** Ajuste rígido de las estructuras de tubulina y Ndc80/Nuf2 en un mapa experimental del microtúbulo decorado con Ndc80/Nuf2 a 8Å. En el panel A se ha ilustrado la estructura atómica de Ndc80 en verde y Nuf2 en amarillo (2ve7 ΔCt), junto con las cadenas α (azul) y β (rojo) del dímero de tubulina (*Ijff*). En B se muestra el resultado del ajuste rígido con *ADP\_EM* de estas estructuras. Los ajustes se han llevado a cabo utilizando el mapa mostrado en el panel B de la Figura 64 y cada estructura atómica por separado.

centrados en la densidad. En cambio, tras el ajuste flexible con *iModfit* (Figura 66) es evidente que todos los elementos de estructura secundaria se ajustan mucho mejor a la densidad del mapa. Nótese que aún efectuando los ajustes con cada estructura por separado, ni los modelos invaden las densidades adyacentes ni se detectan distorsiones estructurales. Los valores de las métricas de *Molprobability* corroboran la ausencia de distorsiones (Tabla 36) e indican un excelente mantenimiento de la calidad de las estructuras. Los cambios observados son pequeños, pero muestran la capacidad de captar detalles atómicos cuando la calidad y resolución del mapa son elevadas.



**Figura 66.** Ajuste flexible de las estructuras de tubulina y Ndc80/Nuf2 un el mapa experimental del microtúbulo decorado con Ndc80/Nuf2 a 8Å. Modelos atómicos resultantes del ajuste flexible empleando *iModfit* con sus parámetros por defecto. El mapa se ha representado en transparencia (izquierda) y en color sólido (derecha) utilizando un umbral relativamente alto para resaltar los elementos de estructura secundaria. Estos ajustes se han llevado a cabo con el mapa y las estructuras atómicas procedentes del ajuste rígido mostrados en la Figura 65. El código de colores es el mismo.

#### 5.8.14. Comparación de iModfit con NMFF y YUPSCX en casos experimentales

Se han llevado a cabo todos los ajustes del conjunto de pruebas experimental utilizando las herramientas *NMFF* y *YUPSCX* para comparar *iModfit* con otros métodos actuales de ajuste flexible basados tanto en NMA como en MD. Con *NMFF* se han utilizado los parámetros optimizados, *i.e.* 30 modos y 0.001 de umbral de convergencia, y con *YUPSCX* los valores por defecto. En la Tabla 37 se muestran las correlaciones finales obtenidas con cada uno de los métodos junto con las desviaciones entre las estructuras iniciales y los modelos finales correspondientes. En términos generales, las correlaciones son similares y los modelos tienden a ocupar toda la densidad de los mapas. Al comparar *NMFF* con *iModfit*, puede apreciarse que con esta última herramienta todas las correlaciones son iguales o ligeramente más altas, aunque con *YUPSCX* éstas son generalmente algo mayores que con los otros dos métodos. Las desviaciones entre estructuras iniciales y finales, que indican la amplitud del cambio conformacional asociado al ajuste, son muy similares ( $\pm 1\text{Å}$ ) en casi todos los casos y con los tres métodos. Únicamente en los casos de clamp y ARNP (abierto y cerrado) las diferencias son algo mayores (2~3Å). Para comprobar el parecido entre los modelos finales de los distintos métodos, en la Tabla 37 también se han calculado las desviaciones entre ellos. Estas desviaciones son similares (1~2Å)



**Tabla 37.** Comparativa entre los métodos *YUPSCX*, *NMFF* e *iModfit* utilizando mapas experimentales.

Ajuste <sup>a</sup>	Corr.I <sup>b</sup>	<i>iModfit</i>		<i>NMFF</i>		<i>YUPSCX</i>		RMSD <sup>c</sup> [Å]		
		Corr. <sup>c</sup>	F <sup>d</sup> [Å]	Corr.	F [Å]	Corr.	F [Å]	<i>iM-N</i>	<i>iM-Y</i>	<i>N-Y</i>
GroEL <sub>14</sub> /ES <sub>7</sub> /ADP <sub>7</sub>	0.64	0.87	13.62	0.84	14.03	<u>0.89</u>	13.34	2.79	<u>2.40</u>	3.24
GroEL <sub>14</sub>	0.59	0.94	15.22	0.94	15.31	<u>0.95</u>	15.02	2.17	2.12	<u>1.89</u>
EFG	0.68	0.89	9.93	0.89	10.26	<u>0.90</u>	9.36	<u>2.29</u>	2.58	3.75
Prefoldina	0.86	0.92	5.17	0.91	5.41	<u>0.95</u>	5.74	<u>1.96</u>	3.56	3.87
Clamp*	0.86	0.91	8.14	0.88	6.89	<u>0.93</u>	5.45	4.50	4.83	<u>4.20</u>
ATPasa-Ca	0.68	0.90	14.15	0.88	14.46	<u>0.91</u>	13.44	3.17	<u>3.10</u>	4.36
CopAdC†	0.78	0.84	5.11	0.84	5.94	<u>0.87</u>	6.13	<u>3.15</u>	4.13	4.38
Termosoma	0.29	0.60	18.39	0.52	16.74	<u>0.63</u>	17.78	6.92	<u>3.91</u>	6.58
ARNP abierta	0.88	<u>0.91</u>	7.82	0.89	6.11	<u>0.91</u>	5.97	6.23	<u>6.21</u>	6.36
ARNP cerrada	0.88	<u>0.92</u>	10.26	0.89	8.11	0.90	8.17	9.46	<u>6.34</u>	7.63
Ndc80/Nuf2	0.23	0.24	2.84	0.24	2.86	<u>0.25</u>	2.82	1.88	<u>1.55</u>	1.99
Tubulina	0.37	0.40	2.67	0.40	2.50	<u>0.42</u>	2.87	<u>1.19</u>	1.75	1.77
Promedio	0.65	0.78	9.44	0.76	9.05	0.79	8.84	3.81	3.54	4.17

<sup>a</sup> Todos los ajustes se han efectuado utilizando los parámetros por defecto, excepto en algunos casos llevados a cabo con *iModfit* (\*,†) (véase el pie de la Tabla 36).

<sup>b-c</sup> Correlación normalizada entre los mapas experimentales y los mapas simulados de las estructuras iniciales (Corr.I) y finales (Corr.) correspondientes. Se han subrayado los valores del método que obtiene mayor correlación.

<sup>d</sup> Desviación entre la estructura inicial y el modelo final. Todas las desviaciones están basadas los átomos C<sub>α</sub>.

<sup>e</sup> Desviación entre las estructuras finales obtenidas con los métodos: *iModfit* y *NMFF* (*iM-N*), *iModfit* y *YUPSCX* (*iM-Y*), y *NMFF* y *YUPSCX* (*N-Y*). Se han subrayado los valores de menor desviación.

para los tres métodos en 6 de los 12 casos: Ndc80/Nuf2, Tubulina, GroEL<sub>14</sub>-ADP<sub>7</sub>-GroES<sub>7</sub>, GroEL<sub>14</sub>, EFG y ATPasa-Ca. La desviación entre los modelos finales de la prefoldina obtenidos con *iModfit* y *NMFF* también es baja (1.96Å). En el caso del termosoma, *iModfit* y *YUPSCX* también obtienen modelos finales parecidos (3.91Å) teniendo en cuenta la gran amplitud del cambio conformacional (~18Å). En los restantes casos, *i.e.* clamp, CopAdC y ARNP (abierta y cerrada), las diferencias son superiores a ~4Å con todos los métodos. Esto probablemente sea debido a la baja resolución de los mapas (17-22Å).

Para evaluar comparativamente la calidad de los modelos finales, en la Tabla 38 se muestran algunas de las métricas de validación de *Molprobity*. Aunque en general, el número de colisiones y de ángulos diedros no permitidos es bastante similar entre los métodos basados en NMA, con *NMFF* se obtienen valores mejores o iguales que con *iModfit* en 10 y en 6 de los 12 casos, respectivamente. En cambio, *YUPSCX* mejora las colisiones obtenidas con los métodos de NMA en 10 ocasiones y el porcentaje de ángulos diedros fuera de las regiones permitidas en todas ellas. Por lo general, el método basado en MD tan solo obtiene entre 40 y 60 colisiones y menos del 2% de ángulos diedros en regiones no permitidas. A pesar de que en los casos de ajuste simulado los métodos basados en NMA consiguen menos colisiones que *YUPSCX*, según estos resultados, en los casos experimentales esta tendencia se invierte. Esto se debe

**Tabla 38.** Comparativa entre la calidad estructural de los modelos finales obtenidos con *YUPSCX*, *NMFF*, e *iModfit* utilizando mapas experimentales.

Ajuste <sup>a</sup>	<i>iModfit</i>				<i>NMFF</i>				<i>YUPSCX</i>				Inicial	
	C <sup>b</sup>	R	E	A	C	R	E	A	C	R	E	A	C	R
GroEL <sub>14</sub> /ES <sub>7</sub> /ADP <sub>7</sub>	63	2.3	<u>0.8</u>	<u>0.2</u>	52	1.9	28.4	23.1	<u>45</u>	<u>1.2</u>	1.2	4.8	34	0.8
GroEL <sub>14</sub>	56	1.9	<u>0.4</u>	<u>1.2</u>	63	2.9	38.4	35.5	<u>48</u>	<u>1.2</u>	2.5	6.3	33	0.6
EFG	64	1.1	1.8	<u>0.6</u>	57	1.6	34.7	31.8	<u>38</u>	<u>0.6</u>	<u>0.0</u>	5.3	41	0.5
Prefoldina	37	0.6	<u>0.0</u>	<u>0.0</u>	<u>18</u>	0.6	8.3	7.9	33	<u>0.4</u>	0.1	1.9	15	0.3
Clamp*	59	1.1	<u>0.3</u>	<u>0.9</u>	43	1.9	28.2	26.4	<u>42</u>	<u>0.5</u>	<u>0.3</u>	5.1	30	0.4
ATPasa-Ca	88	3.1	0.6	<u>0.6</u>	61	3.5	31.8	34.5	<u>43</u>	<u>1.3</u>	<u>0.5</u>	5.1	35	0.8
CopAdC <sup>†</sup>	142	1.7	<u>0.2</u>	<u>1.2</u>	129	2.0	13.9	9.7	<u>64</u>	<u>1.6</u>	3.4	16.5	106	1.1
Termosoma	36	1.3	<u>0.0</u>	<u>1.1</u>	30	1.3	8.6	4.1	<u>29</u>	<u>0.9</u>	0.2	3.1	28	1.0
ARNP (abierta)	147	11.2	<u>0.5</u>	<u>1.8</u>	89	9.1	5.1	4.5	<u>69</u>	<u>7.9</u>	2.3	11.0	84	9.0
ARNP (cerrada)	166	12.6	<u>0.4</u>	<u>1.8</u>	93	9.2	6.7	6.2	<u>72</u>	<u>7.9</u>	2.2	10.9	84	9.0
Ndc80/Nuf2	45	<u>0.0</u>	<u>0.0</u>	<u>0.3</u>	<u>13</u>	<u>0.0</u>	2.5	2.5	32	<u>0.0</u>	<u>0.0</u>	1.3	13	0.0
Tubulina	103	11.2	<u>0.0</u>	<u>0.6</u>	107	10.7	4.1	3.2	<u>57</u>	<u>9.6</u>	2.0	11.1	110	10.7
Promedio	84	4.0	0.4	0.9	63	3.7	17.6	15.8	48	2.8	1.2	6.9	51	2.9

<sup>a</sup> Todos los ajustes se han llevado a cabo utilizando los parámetros por defecto, excepto en algunos casos efectuados con *iModfit* (\*,†) (véase el pie de la Tabla 36).

<sup>b</sup> Métricas de *Molprobability*: Número de colisiones importantes cada 1000 átomos (C), porcentajes de ángulos diedros fuera de las regiones permitidas de Ramachandran (R), de longitudes (E) y ángulos (A) de enlace inadecuados. Los valores iniciales de E y A pueden encontrarse en la Tabla 36.

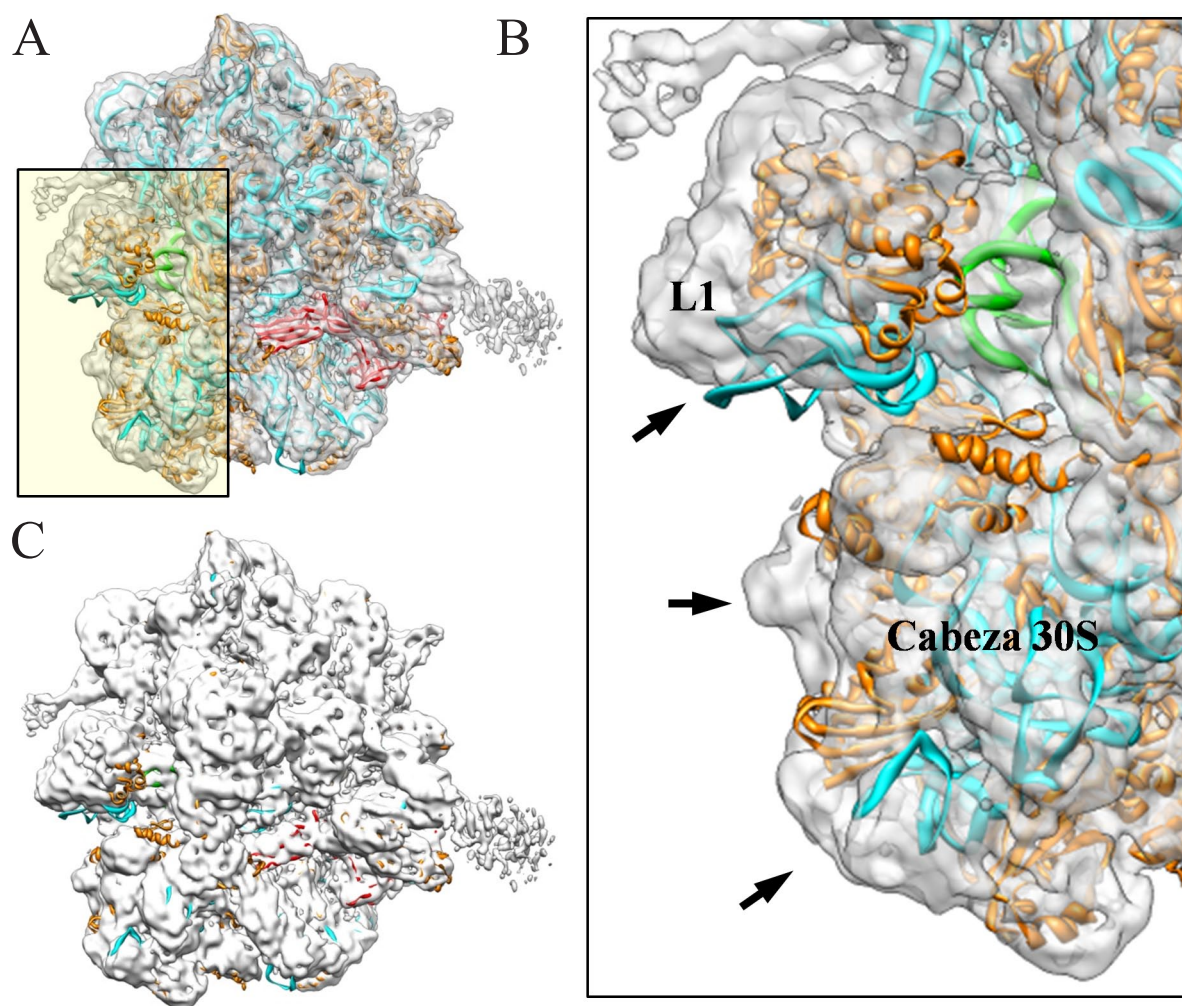
<sup>§</sup> Se han subrayado los valores del método que obtiene mejor puntuación.

probablemente a las restricciones impuestas por el término energético adicional que incorpora este último método para evitar explícitamente la aproximación excesiva entre los átomos (véase la sección 3.5.4 de la Introducción). Como se ha visto a lo largo de las pruebas realizadas, dicho método no solo permite mantener un nivel aproximadamente constante de colisiones (40~60) independientemente del tipo de ruido presente en los mapas, sino que en algunas ocasiones incluso es capaz de reducirlo. Esto puede comprobarse al comparar el número de colisiones de los modelos finales de *YUPSCX* con los valores iniciales. Por otra parte, cuando se consideran los porcentajes de longitudes y ángulos de enlace inadecuados, con *NMFF* y *YUPSCX* los valores empeoran mucho, alcanzándose valores superiores al 10% en la mayoría de los casos con el primero y entre el 5 y el 10% con el segundo. Por el contrario, *iModfit* siempre obtiene los mejores resultados con estas métricas, excepto en dos de los porcentajes de longitudes de enlace (EFG y ATPasa-Ca) que son ligeramente mejorados por *YUPSCX*. En general puede concluirse que los distintos métodos consiguen estructuras bastante similares. No obstante, a pesar de que *YUPSCX* es capaz de obtener las mejores correlaciones y menos colisiones que los demás, *iModfit* es el único que consigue, a la vez, buenas correlaciones manteniendo intacta la estructura covalente.

### 5.8.15. Validación de iModfit con mapas experimentales del ribosoma

Con el fin de demostrar que *iModfit* es capaz de llevar a cabo el ajuste flexible con grandes complejos macromoleculares se han utilizado como ejemplo dos mapas del ribosoma procariota ( $\sim 2.5$ MDa) a  $7\sim 8\text{\AA}$  de resolución. Estos mapas corresponden a dos conformaciones distintas que adquiere el complejo del ribosoma de *Thermus thermophilus* en presencia de EFG, GDP y del antibiótico ácido fusídico (Ratje, *et al.*, 2010). En cada ciclo de elongación de la síntesis de proteínas se une un aminoacil-ARNt en el sitio A (*Aminoacil*) del ribosoma, después se forma el enlace peptídico y, por último, se produce la translocación de los ARNt a través del ribosoma para que se abra el sitio A, se libere el ARNt usado y se pueda reperir el ciclo de nuevo. Uno de los mapas corresponde al ribosoma en un estado intermedio previo a la translocación, denominado *Tlpre* (del inglés, *pre-Translocational Intermediate*), donde ya se ha formado el enlace peptídico y el ARNt ocupa el sitio P (*Peptidil*) de la subunidad 30S y el E (*Exit*) de la 50S (P/E ARNt). En cambio, en el otro mapa, la región del ARNt que interacciona con el anticodón se ha movido  $8\text{-}10\text{\AA}$  respecto de la posición P/E. Aunque esta nueva conformación es muy parecida a la E/E observada en el estado post-translocacional, corresponde a un estado intermedio diferente denominado *pe/E* en el cual cambian los contactos con la subunidad 30S y se mantienen con la 50S. Por ello, a esta conformación se la conoce como *Tlpost* (del inglés, *post-Translocational Intermediate*). Nótese que estas estructuras solo poseen un ARNt puesto que de otro modo no es posible obtener un complejo pre-translocacional estable con EFG. No obstante, es posible ajustar un segundo ARNt en posición ap/P sin impedimentos estéricos significativos (Ratje, *et al.*, 2010). Ambas conformaciones se diferencian principalmente en la rotación relativa de las subunidades, en la posición de la protuberancia L1 y en el giro de la cabeza de la subunidad 30S.

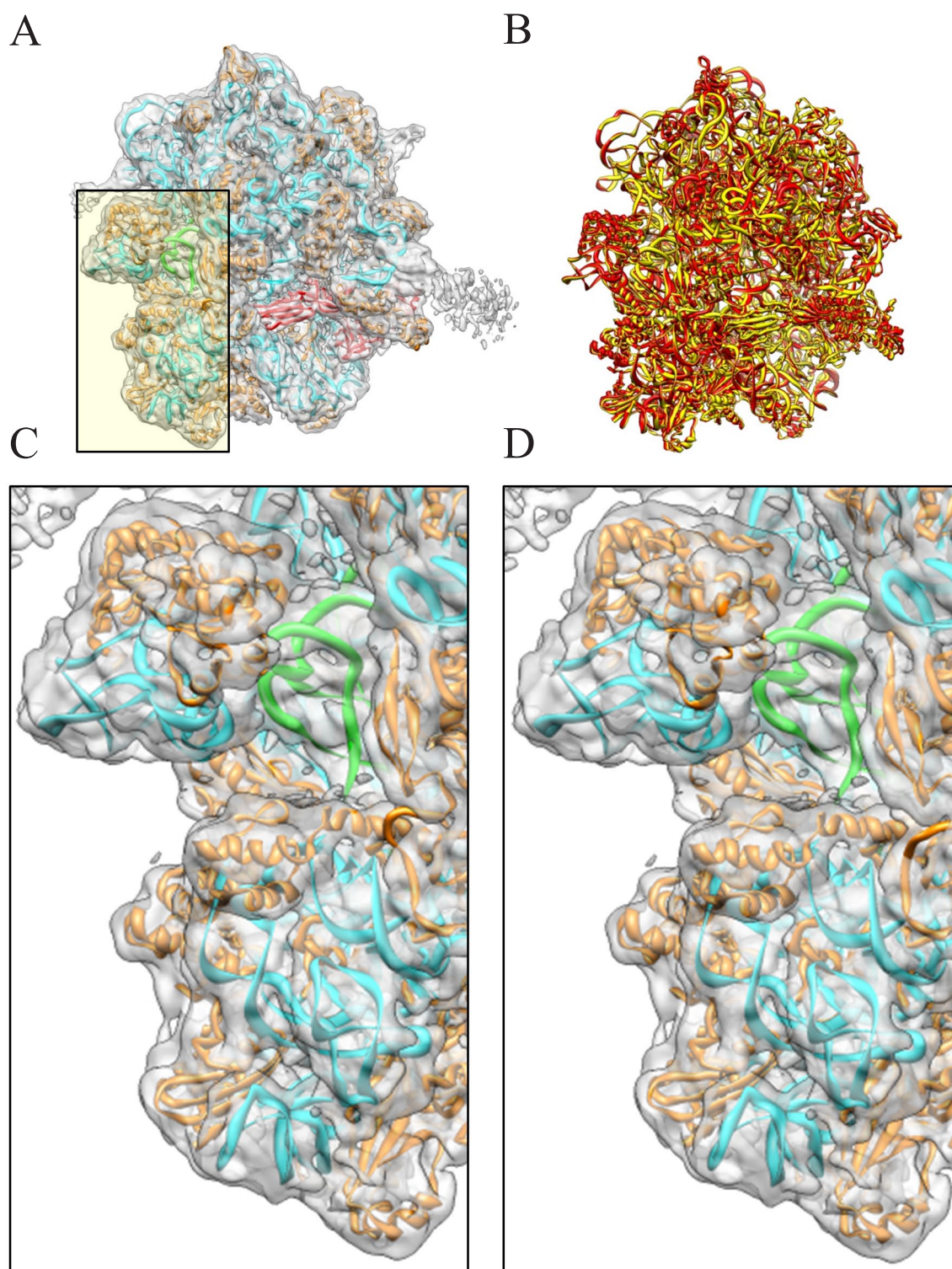
Para interpretar estos mapas, Ratje *et al.* (Ratje, *et al.*, 2010) llevaron a cabo sendos ajustes con una estructura atómica del ribosoma (*2wri/2wrj*) utilizando el método de ajuste flexible basado en MD denominado *MDFIT* (véase la sección 3.5.4 de la Introducción). En nuestro caso, los ajustes se han efectuado utilizando *iModfit* con estas estructuras ajustadas (*Tlpre*: *2xsy+2xtg*, y *Tlpost*: *2xuy+2xux*). En la Figura 67 se ha representado el ajuste rígido inicial entre la estructura atómica *Tlpre* y el mapa experimental en conformación *Tlpost* del ribosoma. Aunque en general, ambas conformaciones son bastante similares (A), pueden apreciarse claramente las diferencias en la protuberancia L1 y en la cabeza de la subunidad 30S (B y C). Las estructuras atómicas utilizadas constan de  $\sim 4500$  nucleótidos,  $\sim 7000$  aminoácidos y dos ligandos pequeños (ácido fusídico y GDP), *i.e.*  $\sim 36500$  variables. El NMA en CI de un sistema de este tamaño requiere cerca de 15GB de memoria RAM y unas 9 horas de cálculo. Para



**Figura 67.** Ajuste rígido inicial del ribosoma. Se ha representado el modelo atómico inicial en conformación T<sub>l</sub>pre ajustado con ADP\_EM sobre el mapa experimental (en transparencia) a 7 Å de resolución del ribosoma 70S en conformación T<sub>l</sub>post (A). La región donde se aprecian las mayores diferencias entre ambas conformaciones se ha resaltado en el recuadro (B). En C también se muestran las mismas estructuras pero con el mapa en color blanco. El ARNr se ha representado en azul, las proteínas en naranja, el EFG en rojo, el ARNt en verde, la molécula de ácido fusídico en magenta y la de GDP en amarillo.

llevar a cabo los ajustes flexibles con *iModfit* y el modelo atómico HA en tiempos razonables utilizando un ordenador corriente con unos 4GB de RAM, se ha fijado aleatoriamente el 50% de los ángulos diedros. Como ya se ha demostrado anteriormente, esta estrategia no afecta significativamente ni a los modos de baja frecuencia (sección 5.3.2), ni a las simulaciones de transiciones conformacionales (sección 5.6.4), ni a los ajustes flexibles (sección 5.8.3). El resto de parámetros se ha dejado en sus valores por defecto. Los resultados del ajuste flexible se muestran en la Figura 68. Tras el ajuste, la correlación aumenta de 0.82 a 0.88 y toda la estructura atómica se encuentra en el interior del mapa manteniendo simultáneamente la integridad de los elementos de estructura secundaria y la estructura covalente (panel A). En el panel C puede





**Figura 68.** Ajuste flexible del ribosoma 70S. En el panel A se ha representado el modelo final obtenido con *iModfit* utilizando el 50% de las variables (el resto de parámetros se dejan en los valores por defecto). En B se muestra el modelo final obtenido con *MDFIT* (rojo) superpuesto sobre el modelo final de *iModfit* (amarillo). En C se muestra en detalle la región recuadrada en A, y en D esta misma región, pero del modelo final obtenido con *MDFIT*. En estos casos, el modelo de *MDFIT* ha sido superpuesto sobre el de *iModfit* minimizando el RMSD (basado en los átomos Cα y P).

verse en detalle cómo la región L1 y la cabeza de la subunidad 30S se han desplazado para cubrir totalmente la densidad. En el panel B se ha comparado este modelo (amarillo) con el obtenido con *MDFIT* (rojo) utilizando el mismo mapa. Para ello, ambos modelos han sido alineados entre sí minimizando el RMSD y empleando el de *iModfit* como referencia. El parecido entre ambas estructuras es excelente, con un RMSD (basado en los átomos C<sub>α</sub> y P) de tan solo 1.00Å. Nótese que aunque la amplitud del cambio conformacional asociado al ajuste es de unos 4Å (RMSD entre el modelo inicial y final), los movimientos de la region L1, la cabeza de la subunidad 30S y el ARNt son de más de ~10Å. Estos movimientos concuerdan perfectamente con los observados en Ratje *et al.* (Ratje, *et al.*, 2010). En el panel D también se ha representado el modelo obtenido con *MDFIT* y el mapa correspondiente en transparencia. Al comparar los paneles B y D puede apreciarse claramente que los modelos generados con ambos programas son prácticamente iguales. También se obtienen resultados equivalentes efectuando los ajustes en sentido contrario, *i.e.* entre la estructura atómica *TIpost* y el mapa *TIpre* (datos no mostrados). En cuanto al coste computacional, cada ajuste con *iModfit* ha tardado unos 3 días en un ordenador de sobremesa. No obstante, los tiempos pueden ser menores disminuyendo aún más el número de variables o incrementando el umbral para recalcular los modos. Téngase en cuenta que el 86% del tiempo de cálculo ha sido dedicado al NMA. Por otro lado, también se ha llevado a cabo el ajuste de la estructura atómica *TIpre* en el mapa *TIpost* utilizando *NMFF*. A pesar de que este programa considera automáticamente bloques rígidos de 10 residuos en el cálculo de los modos del ribosoma, el ajuste es más lento. *NMFF* tarda unos 5 días en obtener un modelo ajustado que alcanza una desviación de 1.76Å de la estructura atómica ajustada con *MDFIT*. Sin embargo, *iModfit* no solo es capaz de obtener una desviación menor (1.00Å), sino que además lo hace en menos tiempo y utilizando un nivel de representación mucho menos reduccionista que *NMFF*. Estas observaciones consolidan la superioridad del NMA en CI frente al NMA en CC para tratar situaciones reales de ajuste con grandes complejos macromoleculares. Asimismo, *iModfit* es capaz de obtener modelos atómicos similares a los de *MDFIT* (1.00Å RMSD) con un ordenador corriente en lugar de un superordenador (Ratje, *et al.*, 2010).

### 5.9. Reconstrucciones tridimensionales del sistema RepB

RepB es una proteína hexamérica de unos 145 kDa involucrada en la iniciación de la replicación por el mecanismo del círculo rodador (RC, del inglés *Rolling Circle*) en los plásmidos de la familia pMV158. En la Figura 69A se muestra esquemáticamente este mecanismo de replicación. Dentro de los plásmidos que replican por RC se distinguen cuatro familias de acuerdo con su secuencia de control e iniciación de la replicación, cuyos modelos son: pT181, pC194,

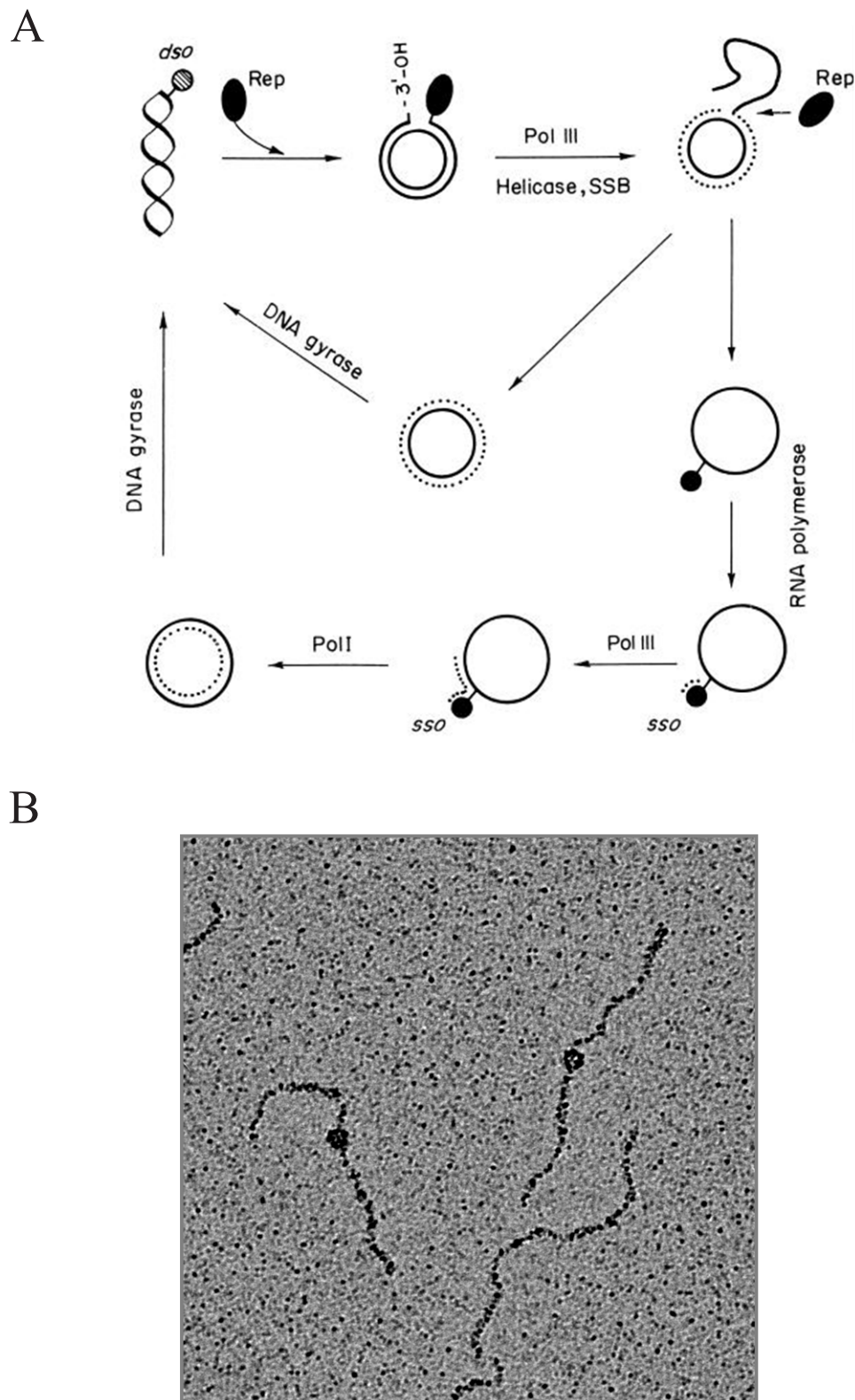


pMV158 y pSN2. En estos plásmidos, la proteína Rep se une al ADN en dos lugares próximos: las regiones *bind* y *nic* (del Solar, *et al.*, 1993). En la Figura 69B se puede ver una imagen del hexámero de RepB unido a un ADN que contiene la secuencia *bind* (tres repeticiones directas de 11 pb) y que ha sido obtenida mediante la técnica de sombreado con platino (comunicación personal de J.A. Ruiz-Masó). RepB se encarga de dos tareas: *i*) el corte específico de secuencia en la región *nic* (repeticiones inversas capaces de formar estructuras tipo tallo-lazo en ADN superenrollado), y *ii*) el posterior empalme (reacción de transferencia de cadena) tras la síntesis de la nueva hebra (Ruiz-Maso, *et al.*, 2004). La proteína RepB del plásmido pMV158 posee varias características que la distinguen de las de pT181 y pC194. La primera diferencia es la capacidad de cortar *in vitro* sobre un ADNse que consta únicamente de la secuencia *nic* (aunque *in vivo* son necesarias ambas para la replicación). La distancia que separa las regiones *bind* y *nic* en el plásmido pMV158 es muy grande, *i.e.* 84 pb ( $\sim 300\text{\AA}$ ) (Ruiz-Maso, *et al.*, 2007), mientras que en las otras familias estas regiones son prácticamente contiguas. Asimismo, aunque la rotura del enlace fosfodiéster parece llevada a cabo por una tirosina, en la familia pMV158 todavía no se ha observado la presencia de un intermediario covalente estable (del Solar, *et al.*, 1998). En cuanto a la estructura cuaternaria, RepB de pMV158 es un hexámero (Ruiz-Maso, *et al.*, 2004), mientras que en las otras familias, las proteínas Rep suelen ser dímeros. Además, la estequiometría de unión de RepB a ADN de doble cadena es 6:1, mientras que en el resto suele ser de 2:1 (G. del Solar, comunicación personal).

Uno de los objetivos de esta tesis se enmarca en el estudio y caracterización estructural a media/baja resolución del la proteína RepB del plásmido pMV158 en ausencia y presencia de ADN mediante microscopía electrónica de transmisión y SP. Estas tareas se llevaron a cabo en colaboración con los grupos de los profesores Gloria del Solar, Oscar Llorca (ambos del CIB-CSIC) y Mikel Coll (del IRB). El primer grupo proporcionó las muestras y llevó a cabo los ensayos de unión con ADN, el segundo asistió y contribuyó en las tareas de MET, SP y reconstrucción 3D, y el tercero obtuvo dos estructuras cristalográficas del hexámero de RepB, una de ellas con simetría C2 (*3dkx*) y la otra C3 (*3dky*) (Boer, *et al.*, 2009). Mis esfuerzos, como se verá más adelante, serán complementarios a los estudios cristalográficos y permitirán resolver la ambigüedad sobre cuál es la simetría de la estructura existente en disolución y caracterizar la estructura de RepB en medio acuoso en ausencia y presencia de ADN.

### 5.9.1. Resolviendo la ambigüedad cristalográfica

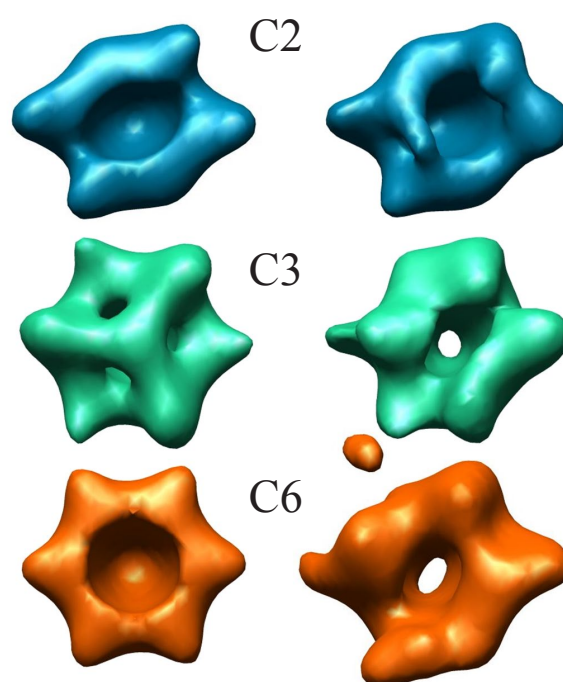
En una primera aproximación, se ha llevado a cabo un estudio preliminar utilizando las imágenes tomadas perpendicularmente a la rejilla. El análisis de simetría rotacional de las partículas individuales con *XMIPP* (Sorzano, *et al.*, 2004) revela la presencia de componentes



**Figura 69.** Función de RepB. Representación esquemática del mecanismo de replicación por círculo rodador (A). Imagen de RepB unida a un ADN con la secuencia *bind* obtenida con la técnica de sombreado con platino (B) (comunicación personal de J.A. Ruiz-Masó).

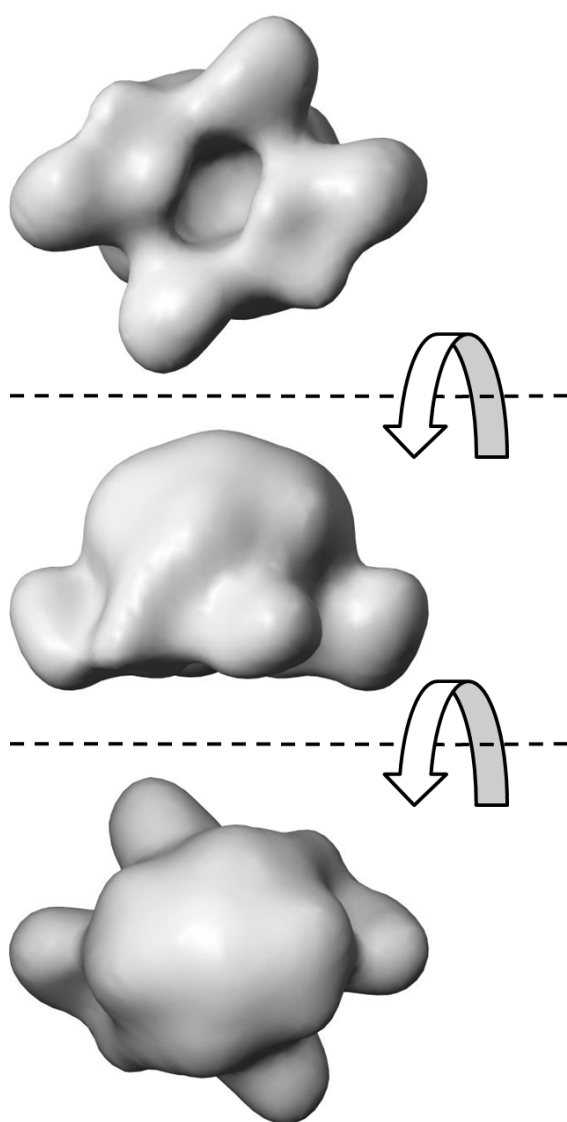
rotacionales de simetría 2, 3 y 6, lo que concuerda con las simetrías observadas en las estructuras cristalinas. No obstante, el análisis de las medias procedentes del alineamiento y clasificación libre de referencia sugiere que RepB se adhiere al soporte según un número limitado de orientaciones. A pesar de este inconveniente, se han utilizado distintos modelos iniciales, tales como *blobs* gaussianos o modelos generados por líneas comunes a partir de las medias libres de referencia, para obtener reconstrucciones 3D utilizando la técnica de encaje de proyecciones (véase la sección 3.4.3 de la Introducción). Aunque estos refinamientos se han efectuado tanto en presencia como en ausencia de restricciones de simetría, los modelos refinados no son compatibles con las estructuras atómicas y pierden rápidamente la simetría al quitar las restricciones.

En una segunda aproximación, se han recogido datos en dos orientaciones distintas de las mismas regiones de las micrografías siguiendo el protocolo RCT (Radermacher, 1988) para obtener un buen modelo inicial de referencia. También se han tomado nuevos datos en una sola orientación con distintas inclinaciones de la rejilla (0, 20 y 40°) para prevenir la presencia de orientaciones preferentes en la muestra. El modelo de RCT filtrado a 45Å (Figura 26) se ha utilizado para refinar los nuevos datos con la técnica del encaje de proyecciones aplicando restricciones de simetría C2, C3 y C6. Para estudiar la convergencia relativa de estas reconstrucciones se han comparado los modelos obtenidos primero imponiendo y luego relajando cada simetría (Figura 70). De esta forma se ha determinado que la única simetría consistente con los datos es la C2. Además, las reconstrucciones obtenidas al relajar las simetrías C3 y C6 muestran cierta desviación hacia C2. Todos estos resultados sugieren que la mayor parte de las partículas de RepB<sub>6</sub> se caracterizan por una simetría rotacional 2, al menos en las condiciones de la muestra. Después de varios ciclos de refinamiento, la reconstrucción 3D final de RepB<sub>6</sub> (Figura 71) alcanza una resolución de 22Å (criterio FSC=0.5). La estructura 3D tiene forma de copa con simetría rotacional 2 y una

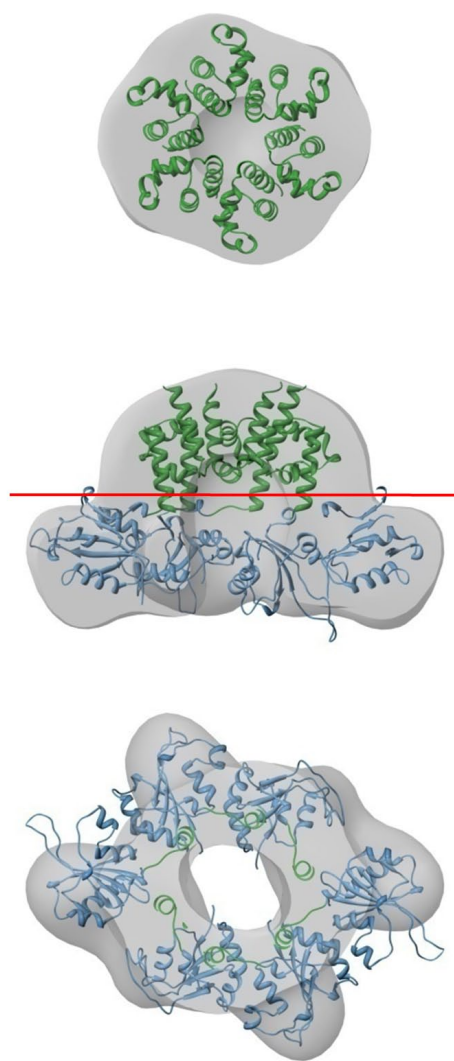


**Figura 70.** Pruebas de relajación de simetría durante la reconstrucción 3D de RepB. En la columna de la izquierda se muestran las reconstrucciones obtenidas con simetría C2 (azul), C3 (verde) y C6 (naranja), y a su derecha, las mismas estructuras reconstruidas sin restricciones.

región estrecha y otra ancha que encierran una cavidad. Además, la hendidura característica de la estructura cristalográfica con simetría C2 también se observa en el mapa. En la Figura 72 se ha superpuesto esta estructura en el mapa de RepB<sub>6</sub> utilizando el programa de ajuste rígido *ADP\_EM*. En ella se comprueba que tanto la forma como la simetría de las regiones estrecha (C6) y ancha (C2) del mapa coinciden con los dominios C- (verde) y N-terminales (azul) de la estructura cristalográfica (Ct y Nt, respectivamente).



**Figura 71.** Reconstrucción 3D por microscopía electrónica y SP de RepB libre. Las tres vistas son ortogonales entre sí.



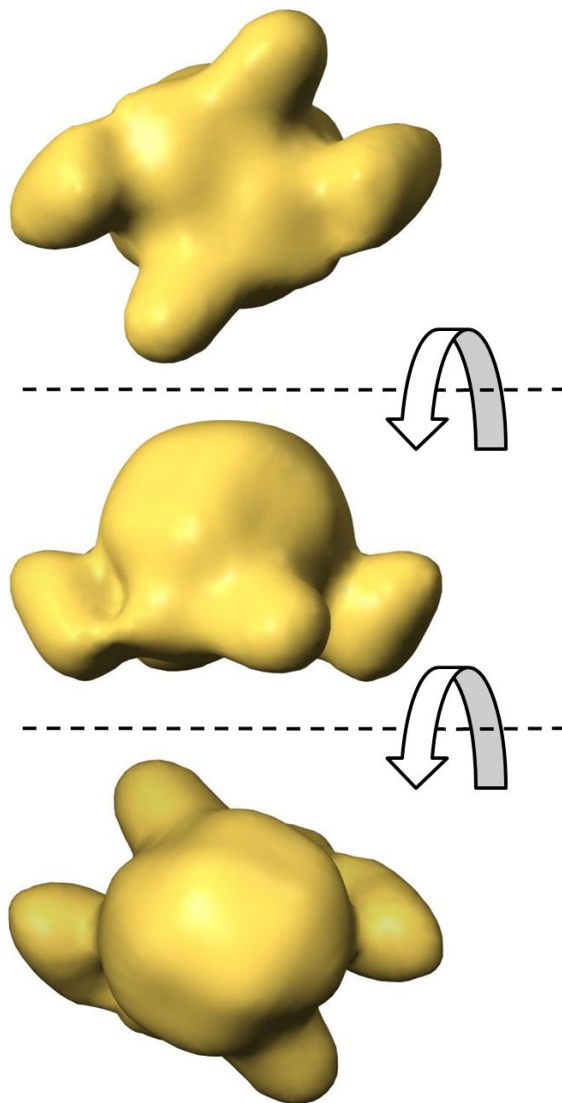
**Figura 72.** Ajuste rígido de la estructura atómica de RepB en el mapa de microscopía. Las regiones estrecha (C6) y ancha (C2) se han ilustrado coloreando la estructura atómica en verde y azul, respectivamente. El mapa (en transparencia) se ha cortado en dos por claridad. La línea roja indica la altura del corte.



### 5.9.2. Reconstrucción en presencia de ADN

La obtención de una reconstrucción 3D del complejo de RepB<sub>6</sub> con un ADN de 42 pares de bases de la región *bind* se ha llevado a cabo siguiendo estrategias equivalentes, *i.e.* a partir del modelo inicial de RepB<sub>6</sub> libre filtrado a 45Å, pero utilizando los datos con ADN en el refinamiento por encaje de proyecciones. Tras varias iteraciones de refinamiento se ha obtenido una reconstrucción final de RepB<sub>6</sub> en presencia de ADN (Figura 73) a 24Å de resolución (FSC=0.5). Aunque globalmente su forma es similar a la de la reconstrucción de RepB<sub>6</sub> libre y de la estructura C2 cristalográfica, presenta una densidad adicional que ocluye la hendidura y cierra la cavidad.

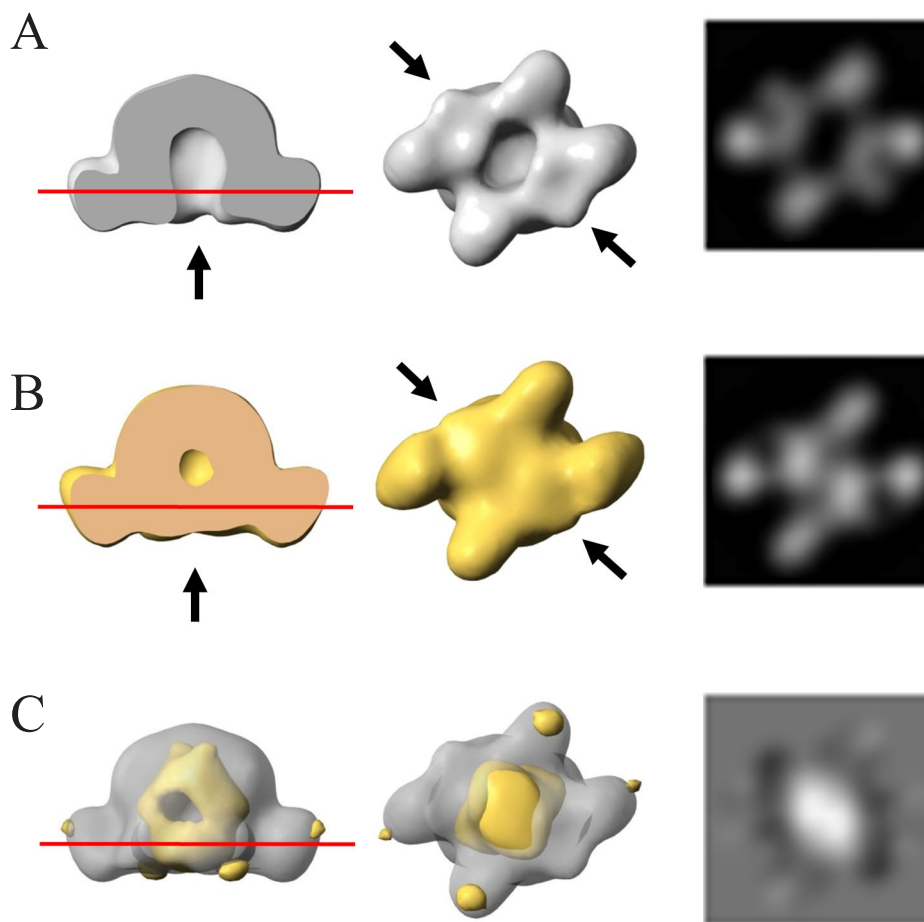
En la Figura 74 se ilustran las diferencias entre las reconstrucciones de RepB<sub>6</sub> obtenidas en ausencia (A) y presencia de ADN (B). La presencia de esta densidad extra se hace evidente al observar tanto los cortes de sección (derecha) como el mapa de diferencia entre ambas reconstrucciones (C). El hecho de que la densidad extra solo aparezca en presencia de ADN y en la región de la hendidura sugiere claramente que esta es la zona de unión entre la proteína y la región *bind* del ADN. Su presencia se ha confirmado mediante un control en el cual se ha refinado la reconstrucción final sin ADN utilizando datos con ADN y *viceversa*. Sin embargo, no puede descartarse que también contribuyan a esta densidad los dominios Nt al moverse hacia el centro de la hendidura.



**Figura 73.** Reconstrucción 3D de microscopía electrónica y SP de RepB en presencia de ADN. Se muestran tres vistas ortogonales entre sí.

### 5.9.3. Ajuste flexible de las reconstrucciones de RepB con iModfit

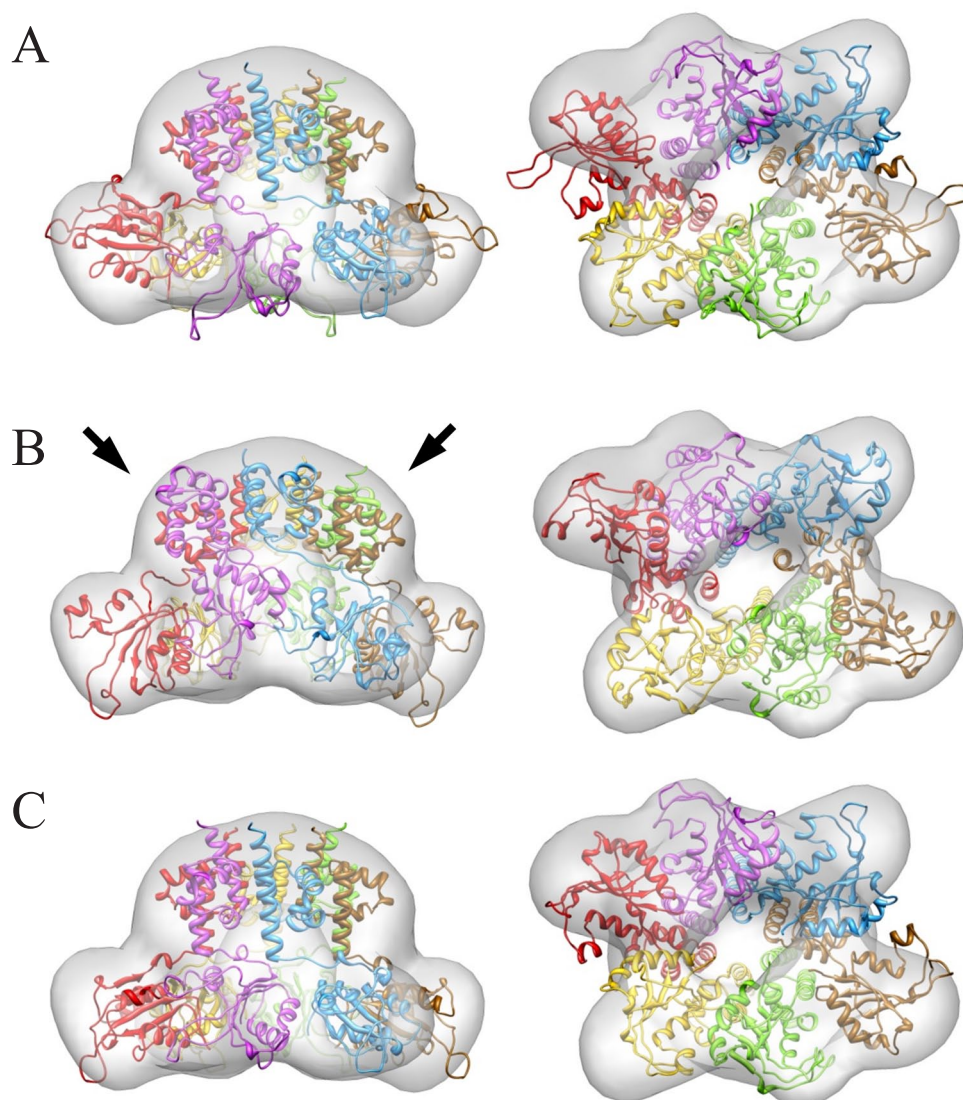
Para obtener modelos atómicos a partir de los dos mapas experimentales de RepB<sub>6</sub> se ha utilizado la estructura cristalográfica con simetría C2, dado que es la única compatible. En primer lugar, para determinar las orientaciones iniciales entre la estructura atómica y los mapas



**Figura 74.** Comparación entre los mapas de la proteína RepB libre y en presencia de ADN. Se ilustran distintas representaciones de RepB<sub>6</sub> libre (A), con ADN (B) y del mapa de diferencia entre ambas en amarillo (C). Se han elegido dos vistas ortogonales: lateral (izquierda) y frontal (derecha), junto con un corte de sección (derecha) tomado a nivel de las líneas rojas. Nótese que en (C) también se ha representado el mapa sin ADN en transparencia.

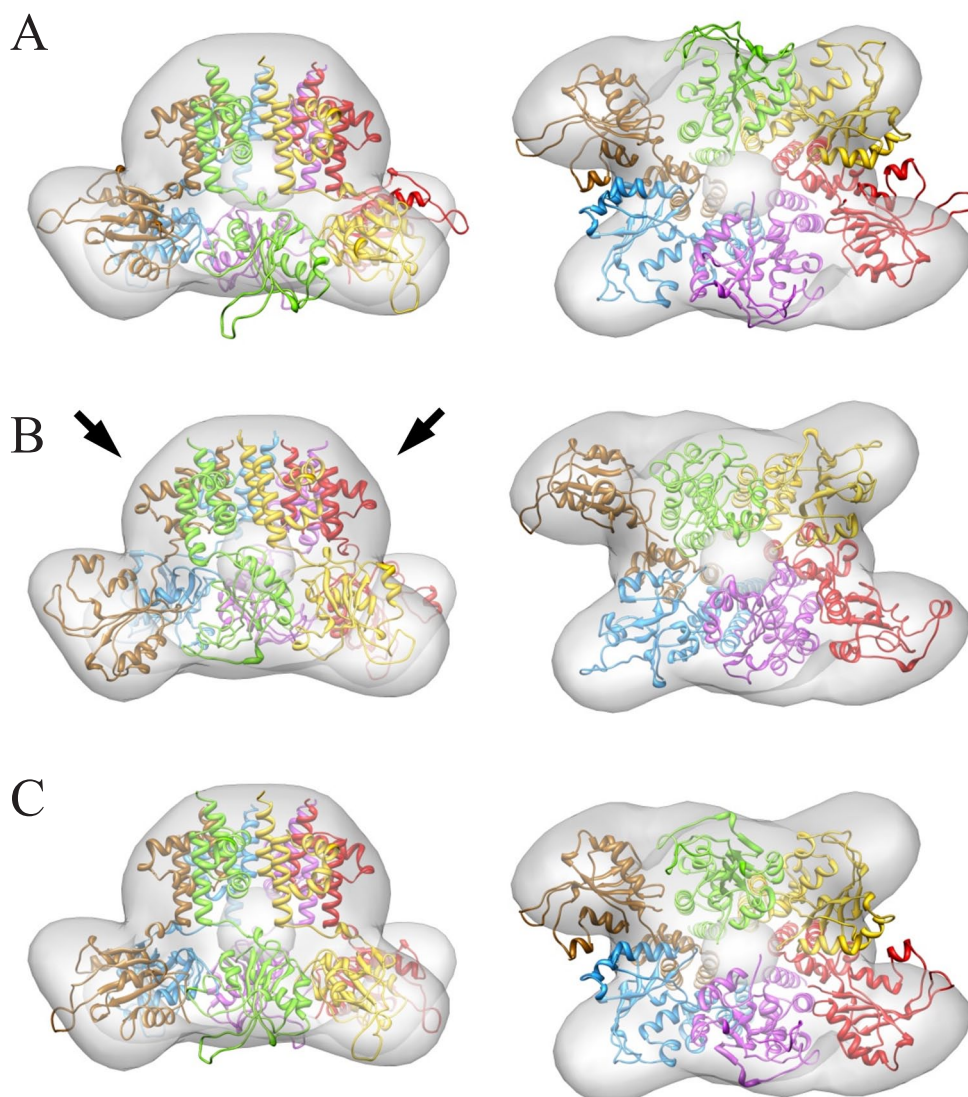
se han llevado a cabo los ajustes rígidos con *ADP\_EM*. En los paneles A de las Figuras 75 y 76 se muestran las orientaciones iniciales de mayor correlación obtenidas con los mapas en ausencia y presencia de ADN, respectivamente. En ambos casos la región Ct (de simetría C6) de la estructura atómica coincide con la región estrecha del mapa y la Nt (de simetría C2) con la región ancha. Después de efectuar el ajuste flexible utilizando *iModfit* con los parámetros por defecto (paneles B) se observan distorsiones en el dominio Ct en los dos casos (flechas). Como ya se ha venido señalando en algunos de los casos de validación experimentales, este tipo de deformaciones indican la presencia de sobreajuste y seguramente sean debidas a la baja resolución de los mapas. Para solucionar este inconveniente simplemente se han repetido los ajustes utilizando los 10 modos de menor energía (paneles C). De este modo, las correlaciones aumentan de 0.81 a 0.86, se preserva adecuadamente la simetría de la región Ct, y se consigue un buen ajuste de los dominios Nt. No obstante, no es posible ajustar algunos elementos pequeños de los dominios Nt que sobresalen de la densidad.





**Figura 75.** Ajuste flexible del hexámero de la proteína RepB libre utilizando *iModfit*. La orientación de la estructura inicial (panel A) se ha determinado con *ADP\_EM*. En el panel B se muestra el modelo final del ajuste flexible llevado a cabo con los parámetros por defecto, y en C el modelo obtenido empleando únicamente los 10 primeros modos y los valores por defecto de los demás parámetros.

Gracias al ajuste flexible se puede comprobar que la estructura de RepB<sub>6</sub> en ausencia de ADN obtenida por microscopía corresponde a una forma con la hendidura algo más abierta que en la estructura cristalográfica (Figura 75C). Por otra parte, en presencia de ADN la estructura se encuentra algo más cerrada (Figura 76C) de forma que se aproximan entre sí dos de los dominios Nt situados a ambos lados de la hendidura (en color verde y magenta). Sin descartar que este movimiento sea debido únicamente a la presencia del ADN en la hendidura, es probable que la capacidad de abrir y cerrar estos dominios tenga significado biológico.

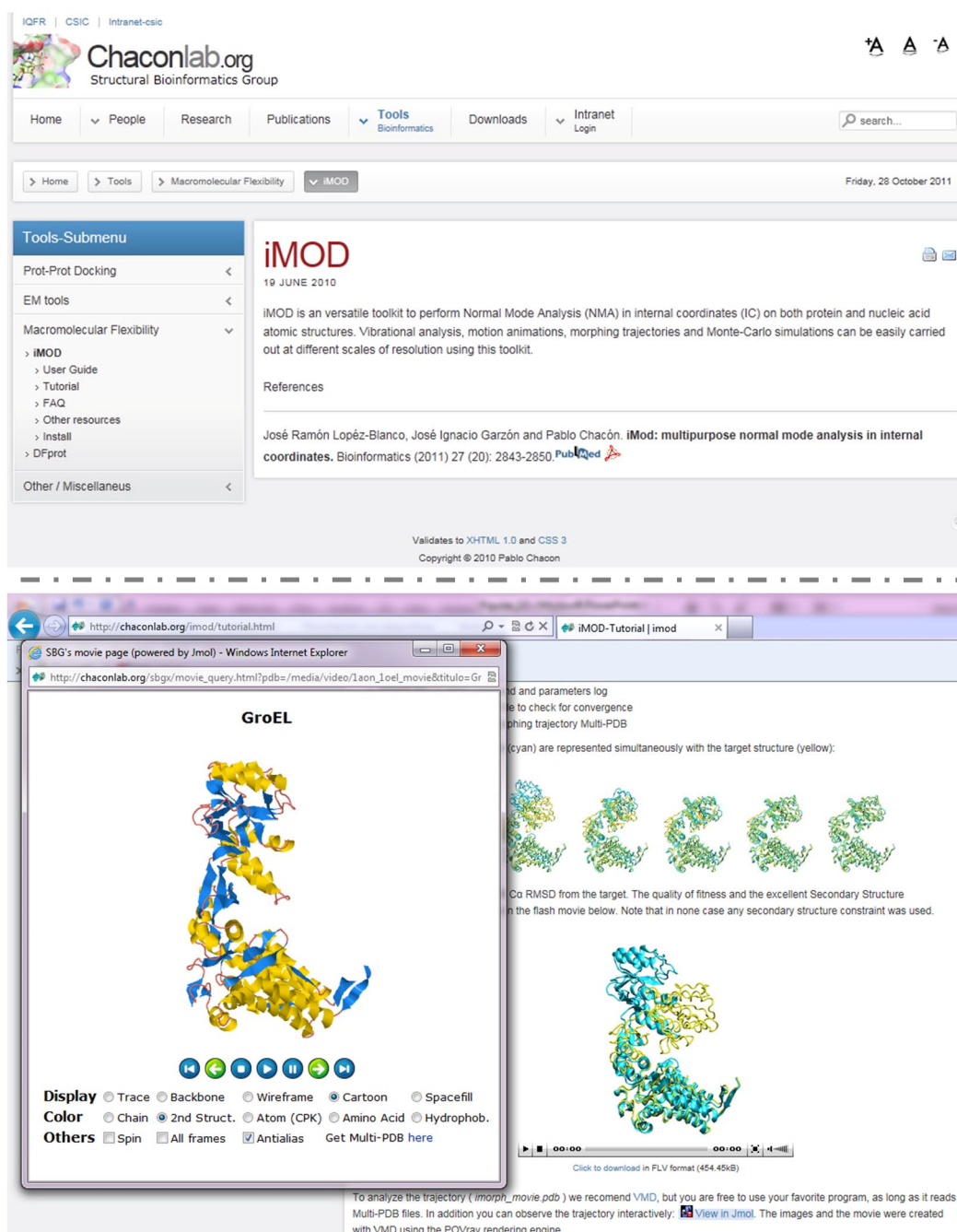


**Figura 76.** Ajuste flexible del complejo de la proteína RepB en presencia de ADN utilizando *iModfit*. La orientación de la estructura inicial (panel A) se ha determinado con *ADP\_EM*. En el panel B se muestra el modelo final del ajuste flexible llevado a cabo con los parámetros por defecto, y en C el modelo obtenido empleando únicamente los 10 primeros modos y los valores por defecto de los demás parámetros.

### 5.10. Disseminación de la metodología desarrollada

Para dar a conocer y poner a disposición de la comunidad las herramientas bioinformáticas desarrolladas se ha creado un sitio web dedicado <http://chaconlab.org/> donde el usuario puede descargar libremente las herramientas de *iMod* (<http://chaconlab.org/imod/index.html>) (Lopez-Blanco, *et al.*, 2011) y el programa *iModfit* (<http://chaconlab.org/imodfit/index.html>). En estas páginas web el usuario puede encontrar toda la documentación necesaria organizada en las siguientes secciones:

- *Guía de usuario*, donde se explican las opciones básicas y avanzadas de las distintas herramientas.
- *Tutorial*. Son ejemplos prácticos comentados con consejos de utilización e instrucciones para llevar a cabo de forma rápida y sencilla las tareas más comunes, como por ejemplo: el cálculo de los modos normales de vibración, la animación y visualización de modos, la simulación de trayectorias de cambios conformacionales (Figura 77), la exploración



**Figura 77.** Diseminación de *iMod*. En la parte superior se ilustra la página principal de *iMod*, donde puede accederse a las distintas secciones. En la parte inferior se muestra en primer plano una animación interactiva con *Jmol* de la transición conformacional entre la forma abierta y cerrada de GroEL sobre un fragmento del tutorial para calcular trayectorias de transiciones conformacionales con *iMorph*.



conformacional mediante simulaciones de MC o el ajuste flexible de estructuras atómicas en mapas de densidad.

- *Preguntas frecuentes* (FAQ), en esta sección se encuentran respuestas a las preguntas más habituales sobre cómo utilizar y personalizar las distintas herramientas, como por ejemplo, cómo definir las distintas representaciones reduccionistas. Además, se han incluido los procedimientos más comunes para tratar sistemas macromoleculares de gran tamaño tales como cápsides virales o ribosomas (Figura 78).
- *Otros recursos*, aquí se incluyen enlaces a programas y servidores externos que pueden

Tools-Submenu  
Prot-Prot Docking <  
EM tools ▾  
  > iMOD-Fit  
  > **User Guide**  
  > Tutorial  
  > FAQ  
  > Other resources  
  > Install  
  > ADP\_EM  
Macromolecular Flexibility <  
Other / Miscellaneous <

## iMODFIT-User Guide

19 JUNE 2010

Here we give a brief overview of the necessary steps to run the program but we strongly encourage to follow the tutorials. The user guide is partitioned in three levels of usage:

- [Minimal options](#)
- [Basic options](#)
- [Advanced options](#)

### Minimal options

For flexible fitting a single atomic structure into a EM density map enter the following command at the prompt.

```
imodfit <pdb> <map> <resolution> <cutoff>
```


where:

pdb	PDB initial model (required).
map	Target EM map file (required). Situs and CCP4 formats are accepted; but extensions ".sit" and ".cpp4" are mandatory.
resolution	Resolution (required). Our resolution criterion follows those used by the standard processing package EMAN (e.g. PDB2MRC tool)
Cutoff	EM map threshold (required). All density levels below this value will not be considered.

---

### Fitting the closed CCMV capsid into the swollen 10Å map.

To illustrate iMODFIT's performance on a huge system we are going to fit the [closed](#) capsid structure into a [simulated 10Å map](#) obtained from the swollen structure [pdb](#). This "swelling", represents a huge concerted motion about 24.0 Å Cα RMSD where 2nm sized openings appear in the protein shell. Atomic structures were obtained from [VIPERdb](#). Input structures are shown below.



As a consequence of both motion and system size we had to carry it out in two steps using some extra options:

```
imodfit 1cwp_prot.pdb ccmv_swln_10A3.cpp4 10 0 -e 0.05 --nowrmsd -6Dref_delay 20000 -i 200000 -r 1 -t -o ccmvfit
```

**Figura 78.** Diseminación de *iModfit*. En la parte superior se ilustra la guía de usuario de *iModfit*. En la parte inferior se muestra uno de los artículos de la sección de Preguntas Frecuentes que trata sobre ajuste de estructuras de gran tamaño.

resultar de utilidad para, por ejemplo, reparar modelos atómicos, obtener estructuras atómicas y mapas de densidad o evaluar la calidad de las estructuras.

- *Instalación*, donde se ponen a disposición de la comunidad científica, y de manera gratuita, las distintas versiones compiladas de las herramientas desarrolladas y los conjuntos de prueba utilizados en esta tesis.

En estas secciones también se incluyen ejemplos cuyos resultados pueden ser descargados o visualizados directamente en forma de animaciones o de representaciones interactivas en tres dimensiones (Figura 68) gracias al *applet* de visualización molecular *Jmol* (<http://www.jmol.org/>).

## 6. Conclusiones

Se ha diseñado e implementado un conjunto de herramientas, denominado *iMod*, para el cálculo, análisis, comparación y visualización de los modos normales de vibración en coordenadas internas (CI), que son:

- Aplicables a grandes macromoléculas de proteínas y ácidos nucleicos con múltiples cadenas y pequeños ligandos rígidos.
- Capaces de emplear distintas aproximaciones reduccionistas.
- Eficientes y fáciles de configurar.
- Accesibles y documentadas vía web.

Se ha validado sistemáticamente el **NMA en CI** con estructuras de proteínas y ácidos nucleicos comparándolo con el NMA en coordenadas cartesianas (CC).

- ✓ Los modos normales de menor energía obtenidos en CI y CC son equivalentes (valores de solapamiento altos, ej.  $\gamma_{50} \approx 0.9$ )
- ✓ El espacio conformacional descrito en CI es más compacto que el descrito en CC, requiriéndose 30% menos modos para expresar el 90% de la varianza total.
- ✓ Gracias a la reducción del número de variables, el coste computacional para obtener los modos en CI es mucho menor. Con el conjunto de prueba de *MolMovDB* se han observado aceleraciones de 2 (C $\alpha$ ) y 600 (HA) veces respecto del NMA en CC. El aumento de eficiencia permite extender el rango de aplicación del NMA a estructuras de mayor tamaño (ej. 28600 residuos en un PC).
- ✓ Se ha comprobado la eficacia de los modos normales obtenidos en CI para describir las **transiciones conformacionales** de gran amplitud de proteínas (de la base de datos *MolMovDB*) y ácidos nucleicos (del PDB), siendo ésta la primera validación sistemática del NMA de ARN que incluye un conjunto representativo de transiciones conformacionales:



- Empleando el 5-10% de los modos se han podido generar trayectorias para ir de una conformación a otra sin distorsiones estructurales significativas. En promedio se han obtenido desviaciones finales de  $\sim 1\text{\AA}$  partiendo de  $\sim 7\text{\AA}$  de desviación inicial.
- Los modos normales de menor energía describen satisfactoriamente la mayor parte de los cambios conformacionales colectivos observados en los conjuntos de prueba. Por ejemplo, los 5 primeros modos de menor energía codifican el 82% del movimiento.
- ✓ La herramienta desarrollada para la **exploración conformacional** basada en modos vibracionales de baja frecuencia y muestreo de Monte Carlo ha sido validada satisfactoriamente con un conjunto de proteínas que experimentan el cierre de sus dominios tras la unión de sus ligandos.
  - Se han obtenidos modelos similares a las estructuras holo ( $\text{RMSD} < 1.5\text{\AA}$ ) manteniendo la calidad estructural utilizando tan solo los 5 primeros modos de las estructuras apo.
  - Ha sido posible enriquecer el muestreo  $\sim 10$  veces en conformaciones más próximas a la estructura holo restringiendo el radio de giro de las conformaciones generadas.
- ✓ La utilización de **modelos reduccionistas** de proteínas no afecta significativamente a los modos normales de baja energía e incrementa la velocidad de los métodos:
  - Los primeros modos con los modelos C $\alpha$  y C5 son prácticamente iguales a los de HA (los solapamientos entre ellos son elevados, p. ej.  $\gamma_{10} \approx 0.9$ ).
  - El empleo de los modelos atómicos C $\alpha$  y C5 en simulaciones de transiciones conformacionales de proteínas permite generar estructuras con desviaciones idénticas a las obtenidas con el modelo HA ( $\sim 0.8\text{\AA}$  RMSD) y disminuir los tiempos de cálculo un 26 y un 12%, respectivamente.
  - La nueva estrategia de fijar aleatoriamente ángulos diedros permite obtener modos equivalentes a los originales con un coste computacional mucho menor. La fijación aleatoria del 50% de los ángulos diedros permite alcanzar solapamientos elevados ( $\gamma > 0.9$ ) acelerando el NMA  $\sim 10$  veces y las simulaciones de transiciones conformacionales entre 2 y 4 veces. También se obtienen resultados iguales con ARN.

Se ha desarrollado y validado *iModfit*, una herramienta de **ajuste flexible** de estructuras atómicas en mapas de microscopía electrónica basada en el NMA en CI.

- ✓ En nuestros bancos de pruebas con mapas simulados se han obtenidos ajustes que solo se desvían  $\sim 1\text{\AA}$  de las estructuras utilizadas para simular los mapas manteniéndose además la estereoquímica inicial.
- ✓ El ajuste flexible resulta muy robusto y se obtienen buenos ajustes:
  - Con proteínas y ácidos nucleicos.
  - Con mapas de 8 a  $25\text{\AA}$  de resolución.
  - En presencia elevados niveles de ruido (SNR de 1 a 5).
  - Utilizando las distintas representaciones reduccionistas.
  - Partiendo de distintas orientaciones iniciales.
  - Cuando la calidad estructural de los modelos iniciales es baja, como en el caso de los ajustes efectuados a partir de modelos de homología.
- ✓ Es una herramienta muy eficiente. Por ejemplo, es posible efectuar el ajuste flexible de una gran macromolécula de 1000 residuos en  $\sim 1$  hora utilizando un PC corriente con un procesador a 2.44GHz. La utilización de los modelos reduccionistas permite una mayor eficiencia sin afectar significativamente a la calidad de las estructuras generadas.
  - En proteínas, los modelos C5 y C $\alpha$  incrementan la velocidad de los ajustes entre un 10 y un 30%, respectivamente, y permiten reconstruir modelos HA directamente.
  - La fijación aleatoria del 50% permite aceleraciones de un 30-40% en proteínas, y de 2 a 4 veces en ARN, manteniendo la misma calidad estructural.
- ✓ De forma general y empleando datos simulados, *iModfit*:
  - Obtiene ajustes con menores desviaciones y estructuras de mayor calidad que las generadas con las herramientas basadas en NMA en CC (*NORMA* y *NMFF*).

- Consigue ajustes comparables o mejores que los obtenidos con las herramientas de MD (*YUPSCX*, *MDFE*, *Flex-EM*, *MDFIT* y Tama 1 y 2).
  - Conserva mejor la estructura covalente y es más robusto frente a ruido simulado que las demás aproximaciones.
  - Es más rápido que los métodos basados en NMA y MD, a excepción de la herramienta *YUPSCX* que alcanza velocidades similares.
- ✓ *iModfit* permite obtener buenos modelos ajustados con mapas experimentales de media/baja resolución y estructuras de 35 kDa a 2.5 MDa. Los resultados obtenidos son compatibles con los ajustes efectuados con otras herramientas.
- ✓ La restricción del número de modos utilizados en los ajustes permite reducir el sobreajuste, especialmente cuando la resolución de los mapas experimentales es baja.

Mediante microscopía electrónica en tinción negativa y análisis de partículas individuales se han obtenido dos reconstrucciones 3D de la proteína **RepB** libre y su complejo con ADN a 22 y 24 Å de resolución, respectivamente. Estas reconstrucciones han permitido:

- ✓ Resolver la ambigüedad de estructuras atómicas disponibles, siendo la estructura cristalográfica con simetría C2 la única compatible en condiciones cercanas a las fisiológicas.
- ✓ Localizar aproximadamente las regiones de interacción con el ADN y postular un modelo según el cual la región N-terminal podría estrecharse para facilitar la unión del ADN cerca de la hendidura.
- ✓ Modelar con *iModfit* el cambio conformacional que acompañaría a la unión del ADN.

## 7. Apéndice

**Tabla A1.** Solapamientos individuales de los vectores de transición con los modos normales de proteínas obtenidos en CI.

Test	Abierta						Cerrada									
	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\delta_3$	$\delta_5$	$\delta_{10}$	$N\alpha_1$	$N\sigma_{90\%}$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\delta_3$	$\delta_5$	$\delta_{10}$	$N\alpha_1$	$N\sigma_{90\%}$
lex6	0.87	0.19	0.19	0.91	0.92	0.95	1	48	0.88	0.19	0.15	0.91	0.92	0.95	1	42
4ake	0.82	0.32	0.26	0.92	0.95	0.96	1	32	0.65	0.44	0.18	0.80	0.82	0.84	1	81
lggg	0.87	0.25	0.20	0.92	0.93	0.94	1	48	0.44	0.37	0.35	0.58	0.69	0.85	2	68
2lao	0.92	0.25	0.10	0.95	0.96	0.96	1	57	0.59	0.31	0.26	0.65	0.65	0.80	3	94
lurp	0.88	0.35	0.16	0.96	0.97	0.97	2	70	0.81	0.38	0.24	0.93	0.93	0.95	1	105
lram	0.58	0.48	0.37	0.84	0.92	0.96	2	7	0.68	0.49	0.24	0.86	0.91	0.96	3	28
5at1	0.55	0.30	0.26	0.66	0.73	0.81	2	97	0.36	0.34	0.32	0.50	0.61	0.70	3	111
lckm	0.88	0.33	0.16	0.94	0.95	0.96	1	38	0.54	0.43	0.41	0.63	0.83	0.89	5	92
3dap	0.86	0.22	0.21	0.92	0.94	0.97	1	38	0.58	0.44	0.42	0.20	0.64	0.91	7	75
1bp5	0.93	0.11	0.09	0.94	0.95	0.95	1	73	0.45	0.41	0.31	0.51	0.71	0.81	1	125
1jql	0.59	0.47	0.25	0.77	0.77	0.82	1	59	0.57	0.48	0.24	0.75	0.75	0.80	1	63
1omp	0.91	0.17	0.08	0.93	0.94	0.94	2	135	0.81	0.37	0.26	0.92	0.94	0.95	1	145
8adh	0.52	0.49	0.47	0.85	0.85	0.86	1	148	0.64	0.49	0.25	0.84	0.85	0.86	3	144
9aat	0.79	0.23	0.22	0.85	0.88	0.90	1	137	0.80	0.25	0.22	0.82	0.89	0.90	1	142
1bnc	0.85	0.18	0.17	0.85	0.88	0.90	1	88	0.69	0.35	0.28	0.78	0.84	0.86	1	150
1dpe	0.88	0.32	0.15	0.94	0.95	0.96	1	173	0.66	0.46	0.28	0.77	0.89	0.90	1	210
1rkm	0.95	0.10	0.09	0.96	0.96	0.98	1	145	0.72	0.24	0.24	0.77	0.82	0.89	1	177
1sx4	0.80	0.29	0.23	0.84	0.89	0.91	1	37	0.76	0.28	0.24	0.80	0.86	0.90	3	53
1i7d	0.60	0.33	0.29	0.61	0.73	0.87	2	114	0.50	0.45	0.25	0.51	0.72	0.77	2	130
1lfg	0.52	0.51	0.37	0.64	0.81	0.91	1	135	0.53	0.40	0.38	0.68	0.82	0.92	3	150
1oao	0.59	0.49	0.37	0.85	0.93	0.96	2	154	0.72	0.51	0.20	0.75	0.91	0.94	3	209
1ih7	0.81	0.28	0.23	0.85	0.87	0.94	1	128	0.55	0.36	0.34	0.66	0.70	0.79	3	220
1su4	0.71	0.27	0.26	0.79	0.79	0.87	1	99	0.62	0.36	0.29	0.62	0.65	0.76	1	253
Promedio	0.77	0.30	0.23	0.86	0.89	0.92	1.3	89.6	0.63	0.38	0.28	0.71	0.80	0.86	2.2	124.7
Desv.Est.	0.15	0.12	0.10	0.10	0.08	0.05	0.4	47.9	0.13	0.09	0.07	0.17	0.11	0.07	1.5	60.4

Véase el pie de la Tabla 6.

**Tabla A2.** Solapamientos individuales de los vectores de transición con los modos normales de ARN obtenidos en CI.

Test	Abierta								Cerrada							
	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\delta_3$	$\delta_5$	$\delta_{10}$	$N\alpha_1$	$N\sigma_{90\%}$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\delta_3$	$\delta_5$	$\delta_{10}$	$N\alpha_1$	$N\sigma_{90\%}$
1a51	0.74	0.49	0.19	0.89	0.92	0.93	1	15	0.67	0.39	0.35	0.79	0.87	0.93	1	23
2i7z	0.61	0.52	0.24	0.81	0.85	0.88	2	13	0.77	0.32	0.24	0.83	0.87	0.90	1	27
1z2j	0.68	0.48	0.34	0.90	0.93	0.94	2	10	0.79	0.35	0.28	0.91	0.91	0.93	2	14
1u63	0.46	0.40	0.21	0.61	0.62	0.63	1	25	0.53	0.24	0.19	0.56	0.60	0.68	1	21
3knj	0.53	0.33	0.24	0.64	0.66	0.76	2	20	0.71	0.23	0.13	0.73	0.74	0.78	1	20
3fih	0.58	0.48	0.22	0.78	0.82	0.84	3	20	0.64	0.43	0.18	0.47	0.81	0.85	4	17
1p5o	0.59	0.49	0.38	0.86	0.87	0.90	2	9	0.64	0.42	0.39	0.78	0.88	0.91	2	12
3cul	0.54	0.52	0.29	0.79	0.81	0.88	1	19	0.69	0.28	0.23	0.74	0.77	0.88	1	29
1mfq	0.83	0.36	0.18	0.83	0.91	0.94	1	23	0.90	0.12	0.11	0.92	0.92	0.94	1	31
2wdj	0.64	0.55	0.39	0.85	0.95	0.96	1	5	0.87	0.20	0.18	0.90	0.92	0.94	1	5
3e1a	0.53	0.32	0.22	0.63	0.70	0.73	1	364	0.52	0.50	0.20	0.72	0.74	0.78	3	264
Promedio	0.61	0.45	0.26	0.78	0.82	0.85	1.5	48	0.70	0.32	0.23	0.76	0.82	0.87	1.6	42
Desv.Est.	0.10	0.08	0.07	0.11	0.11	0.10	0.7	105	0.12	0.11	0.09	0.14	0.10	0.08	1.0	74

Véase el pie de la Tabla 7.

**Tabla A3.** Resultados individuales de las simulaciones de transiciones conformacionales de proteínas.

Caso	Cierre												t [s]	Apertura												t [s]
	RMSD		Col.			%R			Molprobability			RMSD		Col.			%R			Molprobability						
	I	F	I	A	F	I	A	F	I	A	F			F	I	A	F	I	A	F						
1ex6	4.58	0.58	32	33	42	0.0	0.0	0.0	3.3	3.3	3.4	22	0.57	13	16	20	0.5	0.5	0.5	2.4	2.5	2.5	21			
4ake	8.26	0.83	24	32	55	1.4	1.4	1.4	2.4	2.6	3.0	43	0.92	6	16	29	0.0	0.4	0.5	1.6	2.0	2.0	84			
1ggg	10.26	0.77	24	31	53	0.9	0.7	0.0	3.4	3.5	3.8	40	0.84	23	41	41	0.9	1.1	1.8	3.3	3.5	3.5	45			
2lao	8.67	0.46	8	12	22	0.0	0.3	0.4	2.4	2.5	3.0	33	0.54	4	15	13	0.0	0.0	0.0	1.7	2.3	2.3	46			
1urp	7.69	0.38	7	11	14	0.4	0.4	0.4	2.1	2.2	2.3	36	0.43	7	13	11	0.4	0.4	0.4	1.7	2.1	2.1	49			
1ram	4.96	0.61	51	52	56	1.5	1.6	1.9	3.6	3.6	3.7	43	0.61	35	35	38	1.5	1.5	1.5	3.3	3.3	3.3	42			
5at1	2.41	1.01	19	29	50	2.6	2.8	2.9	3.1	3.3	3.6	48	1.02	18	33	52	0.7	0.9	1.3	2.5	2.8	2.8	70			
1ckm	4.35	0.66	9	12	20	0.3	0.3	0.3	2.2	2.3	2.6	48	0.69	10	17	15	0.0	0.0	0.0	2.4	2.5	2.5	66			
3dap	5.83	0.40	13	16	20	0.3	0.3	0.3	2.4	2.5	2.6	53	0.49	31	35	36	1.0	1.0	1.0	3.2	3.3	3.3	94			
1bp5	12.16	0.64	24	31	48	0.0	0.1	0.6	2.8	2.9	3.1	101	0.72	7	25	33	0.6	0.6	0.6	1.8	2.3	2.3	122			
1jql	2.55	0.85	34	37	43	1.1	1.0	0.8	3.0	3.1	3.2	43	0.88	19	19	20	0.6	0.5	0.3	3.0	3.0	3.0	49			
1omp	7.29	0.50	17	20	27	0.8	0.8	0.8	2.7	2.8	3.0	92	0.61	29	35	36	0.3	0.3	0.3	3.1	3.2	3.2	74			
8adh	2.19	0.55	29	30	31	0.3	0.3	0.3	3.3	3.3	3.3	30	0.55	4	5	9	0.0	0.0	0.0	1.5	1.5	1.5	31			
9aat	2.19	0.44	10	12	16	0.3	0.3	0.3	2.2	2.3	2.4	34	0.42	25	26	27	0.0	0.0	0.0	3.1	3.1	3.1	33			
1bnc	5.72	1.20	38	42	55	0.4	0.7	2.0	3.4	3.4	3.6	147	1.54	39	42	45	0.2	0.6	1.1	3.4	3.5	3.5	165			
1dpe	12.43	0.54	16	23	35	0.0	0.0	0.0	2.4	2.6	2.9	233	0.58	19	32	29	0.0	0.1	0.0	2.7	3.0	3.0	301			
1rkm	5.76	0.42	14	16	24	0.2	0.3	0.4	2.7	2.8	3.0	134	0.48	5	11	16	0.2	0.2	0.2	1.6	1.9	1.9	171			
1sx4	15.83	1.10	33	39	51	0.6	0.9	1.7	3.3	3.4	3.6	637	1.12	25	31	41	0.2	0.2	0.4	2.6	2.7	2.7	690			
1i7d	4.20	0.97	25	28	41	1.0	1.0	0.8	2.6	2.7	2.9	298	0.95	19	28	54	1.3	1.4	1.5	3.2	3.3	3.3	364			
1lfg	8.08	0.68	32	37	44	1.3	1.4	1.6	3.3	3.3	3.4	723	0.70	36	39	44	1.7	1.7	1.6	3.5	3.5	3.5	623			
1oao	13.60	0.59	18	20	28	0.4	0.3	0.4	2.5	2.6	2.8	750	0.65	13	20	21	0.4	0.3	0.3	2.4	2.6	2.6	866			
1ih7	7.20	0.73	20	23	33	0.5	0.5	0.7	2.4	2.5	2.7	1121	0.79	26	33	40	1.0	1.1	1.1	2.9	3.0	3.0	1254			
1su4	17.99	2.08	35	43	80	0.8	0.9	2.6	3.0	3.1	3.5	3820	2.21	22	39	69	2.0	2.2	2.8	3.0	3.3	3.3	5150			
Promedio	7.57	0.74	23	27	38	0.7	0.7	0.9	2.8	2.9	3.1	371	0.80	19	26	32	0.6	0.6	0.7	2.6	2.8	2.8	453			
Desv.Est.	4.39	0.37	11	12	16	0.6	0.6	0.8	0.5	0.4	0.4	809	0.40	11	11	15	0.6	0.6	0.7	0.7	0.6	0.6	1073			

Véase el pie de la Tabla 9.

**Tabla A4.** Resultados individuales de las simulaciones de transiciones conformacionales de ARN.

Caso	Cierre						t [s]	Apertura						t [s]
	RMSD		Col.			RMSD		Col.			t [s]			
	I	F	I	A	F			F	I	A		F		
1a51	8.72	0.63	28	33	36	16	1.18	30	30	38	38			
2i7z	10.79	1.18	87	87	93	23	1.37	70	66	70	26			
1z2j	10.49	0.61	33	34	36	15	0.53	33	34	35	12			
1u63	4.65	1.41	45	49	53	24	1.49	3	8	21	24			
3knj	6.34	2.01	39	41	48	40	1.81	7	7	8	49			
3fih	7.95	1.59	8	10	14	37	1.54	24	26	27	41			
1p5o	8.31	1.53	33	33	33	31	1.56	32	33	36	40			
3cul	3.64	0.69	21	20	20	38	0.70	25	25	25	41			
1mfq	5.23	0.99	34	33	32	57	1.06	55	56	58	53			
2wdj	5.27	0.64	28	28	28	1051	0.64	32	33	33	581			
3e1a	8.12	3.02	4	8	25	7586	3.22	31	33	45	7298			
Promedio	7.23	1.30	33	34	38	811	1.37	31	32	36	746			
Desv.Est.	2.37	0.74	22	21	22	2268	0.74	19	18	17	2179			

Véase el pie de la Tabla 9.

## 8. Bibliografía

- Adcock, S.A. y McCammon, J.A. (2006) *Molecular dynamics: survey of methods for simulating the activity of proteins*, Chem Rev, 106, 1589-1615.
- Adrian, M., Dubochet, J., Fuller, S.D. y Harris, J.R. (1998) *Cryo-negative staining*, Micron, 29, 145-160.
- Alamo, L., Wriggers, W., Pinto, A., Bartoli, F., Salazar, L., Zhao, F.Q., Craig, R. y Padron, R. (2008) *Three-dimensional reconstruction of tarantula myosin filaments suggests how phosphorylation may regulate myosin activity*, J Mol Biol, 384, 780-797.
- Alushin, G.M., Ramey, V.H., Pasqualato, S., Ball, D.A., Grigorieff, N., Musacchio, A. y Nogales, E. (2010) *The Ndc80 kinetochore complex forms oligomeric arrays along microtubules*, Nature, 467, 805-810.
- Amadei, A., Linssen, A.B. y Berendsen, H.J. (1993) *Essential dynamics of proteins*, Proteins, 17, 412-425.
- Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A. y Sorensen, D. (1999) *LAPACK Users' Guide*. Society for Industrial and Applied Mathematics, Philadelphia, PA.
- Andrusier, N., Mashinch, E., Nussinov, R. y Wolfson, H.J. (2008) *Principles of flexible protein-protein docking*, Proteins, 73, 271-289.
- Atilgan, A.R., Durell, S.R., Jernigan, R.L., Demirel, M.C., Keskin, O. y Bahar, I. (2001) *Anisotropy of fluctuation dynamics of proteins with an elastic network model*, Biophys J, 80, 505-515.
- Bahar, I., Atilgan, A.R. y Erman, B. (1997) *Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential*, Fold Des, 2, 173-181.
- Bahar, I., Lezon, T.R., Bakan, A. y Shrivastava, I.H. (2010a) *Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins*, Chem Rev, 110, 1463-1497.
- Bahar, I., Lezon, T.R., Yang, L.W. y Eyal, E. (2010b) *Global dynamics of proteins: bridging between structure and function*, Annu Rev Biophys, 39, 23-42.
- Bahar, I. y Rader, A.J. (2005) *Coarse-grained normal mode analysis in structural biology*, Curr Opin Struct Biol, 15, 586-592.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. y Bourne, P.E. (2000) *The Protein Data Bank*, Nucleic Acids Res, 28, 235-242.
- Best, R.B., Chen, Y.G. y Hummer, G. (2005) *Slow protein conformational dynamics from multiple experimental structures: the helix/sheet transition of arc repressor*, Structure, 13, 1755-1763.
- Boer, D.R., Ruiz-Maso, J.A., Lopez-Blanco, J.R., Blanco, A.G., Vives-Llaser, M., Chacon, P., Uson, I., Gomis-Ruth, F.X., Espinosa, M., Llorca, O., del Solar, G. y Coll, M. (2009) *Plasmid replication initiator RepB forms a hexamer reminiscent of ring helicases and has mobile nuclease domains*, EMBO J, 28, 1666-1678.
- Bonvin, A.M. (2006) *Flexible protein-protein docking*, Curr Opin Struct Biol, 16, 194-200.
- Bottegoni, G., Rocchia, W., Rueda, M., Abagyan, R. and Cavalli, A. (2011) *Systematic exploitation of multiple receptor conformations for virtual ligand screening*, PLoS One, 6, e18845.



- Braun, W., Yoshioki, S. y Go, N. (1984) *Formulation of Static and Dynamic Conformational Energy Analysis of Biopolymer Systems Consisting of Two or More Molecules*, Journal of the Physical Society of Japan, 53, 3269-3275.
- Brooks, B. y Karplus, M. (1983) *Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor*, Proc Natl Acad Sci U S A, 80, 6571-6575.
- Brooks, B.R., Janežič, D. y Karplus, M. (1995) *Harmonic analysis of large systems. I. Methodology*, Journal of Computational Chemistry, 16, 1522-1542.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. y Madden, T. (2009) *BLAST+: architecture and applications*, BMC Bioinformatics, 10, 421.
- Camacho, C.J. y Vajda, S. (2002) *Protein-protein association kinetics and protein docking*, Curr Opin Struct Biol, 12, 36-40.
- Cantor, C.R. y Schimmel, P.R. (1980) *Biophysical Chemistry: Part II: Techniques for the Study of Biological Structure and Function*. W. H. Freeman
- Cao, Z.W., Xue, Y., Han, L.Y., Xie, B., Zhou, H., Zheng, C.J., Lin, H.H. y Chen, Y.Z. (2004) *MoViES: molecular vibrations evaluation server for analysis of fluctuational dynamics of proteins and nucleic acids*, Nucleic Acids Res, 32, W679-685.
- Carlson, H.A. y McCammon, J.A. (2000) *Accommodating protein flexibility in computational drug design*, Mol Pharmacol, 57, 213-218.
- Carugo, O. y Argos, P. (1999) *Reliability of atomic displacement parameters in protein crystal structures*, Acta Crystallogr D Biol Crystallogr, 55, 473-478.
- Cavasotto, C.N., Kovacs, J.A. y Abagyan, R.A. (2005) *Representing receptor flexibility in ligand docking through relevant normal modes*, J Am Chem Soc, 127, 9632-9640.
- Ceulemans, H. y Russell, R.B. (2004) *Fast fitting of atomic structures to low-resolution electron density maps by surface overlap maximization*, J Mol Biol, 338, 783-793.
- Clare, D.K., Stagg, S., Quispe, J., Farr, G.W., Horwich, A.L. y Saibil, H.R. (2008) *Multiple states of a nucleotide-bound group 2 chaperonin*, Structure, 16, 528-534.
- Cozzini, P., Kellogg, G.E., Spyralis, F., Abraham, D.J., Costantino, G., Emerson, A., Fanelli, F., Gohlke, H., Kuhn, L.A., Morris, G.M., Orozco, M., Pertinhez, T.A., Rizzi, M. y Sotriffer, C.A. (2008) *Target flexibility: an emerging consideration in drug discovery and design*, J Med Chem, 51, 6237-6255.
- Crowther, R.A. (1971) *Procedures for three-dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs*, Philos Trans R Soc Lond B Biol Sci, 261, 221-230.
- Cui, Q. y Bahar, I. (2006) *Normal mode analysis: theory and applications to biological and chemical systems*. Chapman & Hall/CRC.
- Chacon, P., Tama, F. y Wriggers, W. (2003) *Mega-Dalton biomolecular motion captured from electron microscopy reconstructions*, J Mol Biol, 326, 485-492.
- Chacon, P. y Wriggers, W. (2002) *Multi-resolution contour-based fitting of macromolecular structures*, J Mol Biol, 317, 375-384.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. y Richardson, D.C. (2010) *MolProbity: all-atom structure validation for macromolecular crystallography*, Acta Crystallogr D Biol Crystallogr, 66, 12-21.
- Chen, X., Poon, B.K., Dousis, A., Wang, Q. y Ma, J. (2007) *Normal-mode refinement of anisotropic thermal parameters for potassium channel KcsA at 3.2 Å crystallographic*

- resolution*, Structure, 15, 955-962.
- Cheng, L.S., Amaro, R.E., Xu, D., Li, W.W., Arzberger, P.W. y McCammon, J.A. (2008) *Ensemble-based virtual screening reveals potential novel antiviral compounds for avian influenza neuraminidase*, J Med Chem, 51, 3878-3894.
- Cheng, Y. y Walz, T. (2009) *The advent of near-atomic resolution in single-particle electron microscopy*, Annu Rev Biochem, 78, 723-742.
- Cheng, Y., Wolf, E., Larvie, M., Zak, O., Aisen, P., Grigorieff, N., Harrison, S.C. y Walz, T. (2006) *Single particle reconstructions of the transferrin-transferrin receptor complex obtained with different specimen preparation techniques*, J Mol Biol, 355, 1048-1065.
- Choi, V. (2006) *On updating torsion angles of molecular conformations*, J Chem Inf Model, 46, 438-444.
- Damm, K.L. y Carlson, H.A. (2006) *Gaussian-weighted RMSD superposition of proteins: a structural comparison for flexible proteins and predicted protein structures*, Biophys J, 90, 4558-4573.
- Darst, S.A., Opalka, N., Chacon, P., Polyakov, A., Richter, C., Zhang, G. y Wriggers, W. (2002) *Conformational flexibility of bacterial RNA polymerase*, Proc Natl Acad Sci U S A, 99, 4296-4301.
- del Solar, G., Giraldo, R., Ruiz-Echevarria, M.J., Espinosa, M. y Diaz-Orejas, R. (1998) *Replication and control of circular bacterial plasmids*, Microbiol Mol Biol Rev, 62, 434-464.
- del Solar, G., Moscoso, M. y Espinosa, M. (1993) *In vivo definition of the functional origin of replication (ori(+)) of the promiscuous plasmid pLS1*, Mol Gen Genet, 237, 65-72.
- Delarue, M. y Dumas, P. (2004) *On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models*, Proc Natl Acad Sci U S A, 101, 6957-6962.
- Diamond, R. (1990) *On the use of normal modes in thermal parameter refinement: theory and application to the bovine pancreatic trypsin inhibitor*, Acta Crystallogr A, 46 ( Pt 6), 425-435.
- Dominguez, C., Boelens, R. y Bonvin, A.M. (2003) *HADDOCK: a protein-protein docking approach based on biochemical or biophysical information*, J Am Chem Soc, 125, 1731-1737.
- Dominguez, C., Bonvin, A.M., Winkler, G.S., van Schaik, F.M., Timmers, H.T. y Boelens, R. (2004) *Structural model of the UbcH5B/CNOT4 complex revealed by combining NMR, mutagenesis, and docking approaches*, Structure, 12, 633-644.
- Dubochet, J., Adrian, M., Chang, J.J., Homo, J.C., Lepault, J., McDowell, A.W. y Schultz, P. (1988) *Cryo-electron microscopy of vitrified specimens*, Q Rev Biophys, 21, 129-228.
- Durand, P., Trinquier, G. y Sanejouand, Y.-H. (1994) *A new approach for determining low-frequency normal modes in macromolecules*, Biopolymers, 34, 759-771.
- Dykeman, E.C. y Sankey, O.F. (2010) *Normal mode analysis and applications in biological physics*, J Phys Condens Matter, 22, 423202.
- Eckart, C. (1935) *Some Studies Concerning Rotating Axes and Polyatomic Molecules*, Phys Rev, 47, 552.
- Echols, N., Milburn, D. y Gerstein, M. (2003) *MolMovDB: analysis and visualization of conformational change and structural flexibility*, Nucleic Acids Res, 31, 478-482.
- Emekli, U., Schneidman-Duhovny, D., Wolfson, H.J., Nussinov, R. y Haliloglu, T. (2008) *HingeProt: automated prediction of hinges in protein structures*, Proteins, 70, 1219-1227.
- Erman, B. (2006) *The gaussian network model: precise prediction of residue fluctuations and application to binding problems*, Biophys J, 91, 3589-3599.

- Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.Y., Pieper, U. y Sali, A. (2007) *Comparative protein structure modeling using MODELLER*, Curr Protoc Protein Sci, Chapter 2, Unit 2 9.
- Eyal, E. y Bahar, I. (2008) *Toward a molecular understanding of the anisotropic response of proteins to external forces: insights from elastic network models*, Biophys J, 94, 3424-3435.
- Eyal, E., Chennubhotla, C., Yang, L.W. y Bahar, I. (2007) *Anisotropic fluctuations of amino acids in protein structures: insights from X-ray crystallography and elastic network models*, Bioinformatics, 23, i175-184.
- Eyal, E., Dutta, A. y Bahar, I. (2011) *Cooperative dynamics of proteins unraveled by network models*, Wiley Interdisciplinary Reviews: Computational Molecular Science, 1, 426-439.
- Fabiola, F. y Chapman, M.S. (2005) *Fitting of high-resolution structures into electron microscopy reconstruction images*, Structure, 13, 389-400.
- Falke, S., Tama, F., Brooks, C.L., 3rd, Gogol, E.P. y Fisher, M.T. (2005) *The 13 angstroms structure of a chaperonin GroEL-protein substrate complex by cryo-electron microscopy*, J Mol Biol, 348, 219-230.
- Fanconi, B., Small, E.W. y Peticolas, W.L. (1971) *Phonon dispersion curves and normal coordinate analysis of -poly-L-alanine*, Biopolymers, 10, 1277-1298.
- Feig, M. y Burton, Z.F. (2010) *RNA polymerase II flexibility during translocation from normal mode analysis*, Proteins: Struct Funct Bioinformatics, 78, 434-446.
- Fernandez-Recio, J., Totrov, M. y Abagyan, R. (2002) *Soft protein-protein docking in internal coordinates*, Protein Sci, 11, 280-291.
- Floquet, N., Marechal, J.D., Badet-Denisot, M.A., Robert, C.H., Dauchez, M. y Perahia, D. (2006) *Normal mode analysis as a prerequisite for drug design: application to matrix metalloproteinases inhibitors*, FEBS Lett, 580, 5130-5136.
- Flores, S., Echols, N., Milburn, D., Hespenheide, B., Keating, K., Lu, J., Wells, S., Yu, E.Z., Thorpe, M. y Gerstein, M. (2006) *The Database of Macromolecular Motions: new features added at the decade mark*, Nucleic Acids Res, 34, D296-301.
- Flory, P.J., Gordon, M. y McCrum, N.G. (1976) *Statistical Thermodynamics of Random Networks [and Discussion]*, Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences, 351, 351-380.
- Frank, J. (2006) *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state*. Oxford University Press.
- Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M. y Leith, A. (1996) *SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields*, J Struct Biol, 116, 190-199.
- Franklin, J., Koehl, P., Doniach, S. y Delarue, M. (2007) *MinActionPath: maximum likelihood trajectory for large-scale structural transitions in a coarse-grained locally harmonic energy landscape*, Nucleic Acids Res, 35, W477-482.
- Fulle, S. y Gohlke, H. (2010) *Molecular recognition of RNA: Challenges for modelling interactions and plasticity*, J Mol Recogn, 23, 220-231.
- Gabel, F., Bicout, D., Lehnert, U., Tehei, M., Weik, M. y Zaccai, G. (2002) *Protein dynamics studied by neutron scattering*, Q Rev Biophys, 35, 327-367.
- Galkin, V.E., Orlova, A., Schroder, G.F. y Egelman, E.H. (2010) *Structural polymorphism in F-actin*, Nat Struct Mol Biol, 17, 1318-1323.
- Garzon, J.I., Kovacs, J., Abagyan, R. y Chacon, P. (2007a) *ADP\_EM: fast exhaustive multi-*

- resolution docking for high-throughput coverage*, Bioinformatics, 23, 427-433.
- Garzon, J.I., Kovacs, J., Abagyan, R. y Chacon, P. (2007b) *DFprot: a webtool for predicting local chain deformability*, Bioinformatics, 23, 901-902.
- Gerstein, M. y Krebs, W. (1998) *A database of macromolecular motions*, Nucleic Acids Res, 26, 4280-4290.
- Go, N., Noguti, T. y Nishikawa, T. (1983) *Dynamics of a small globular protein in terms of low-frequency vibrational modes*, Proc Natl Acad Sci U S A, 80, 3696-3700.
- Goldstein, H., Poole, C.P. y Safko, J. (2002) *Classical Mechanics*. Addison-Wesley.
- Grater, F., Shen, J., Jiang, H., Gautel, M. y Grubmuller, H. (2005) *Mechanically induced titin kinase activation studied by force-probe molecular dynamics simulations*, Biophys J, 88, 790-804.
- Gray, J.J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C.A. y Baker, D. (2003) *Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations*, J Mol Biol, 331, 281-299.
- Grubisic, I., Shokhirev, M.N., Orzechowski, M., Miyashita, O. y Tama, F. (2010) *Biased coarse-grained molecular dynamics simulation approach for flexible fitting of X-ray structure into cryo electron microscopy maps*, J Struct Biol, 169, 95-105.
- Halperin, I., Ma, B., Wolfson, H. y Nussinov, R. (2002) *Principles of docking: An overview of search algorithms and a guide to scoring functions*, Proteins, 47, 409-443.
- Halle, B. (2002) *Flexibility and packing in proteins*, Proc Natl Acad Sci U S A, 99, 1274-1279.
- Hinsen, K. (1998) *Analysis of domain motions by approximate normal mode calculations*, Proteins, 33, 417-429.
- Hinsen, K. (2008) *Structural flexibility in proteins: impact of the crystal environment*, Bioinformatics, 24, 521-528.
- Hinsen, K., Petrescu, A.-J., Dellerue, S., Bellissent-Funel, M.-C. y Kneller, G.R. (2000) *Harmonicity in slow protein dynamics*, Chemical Physics, 261, 25-37.
- Hinsen, K., Reuter, N., Navaza, J., Stokes, D.L. y Lacapere, J.J. (2005) *Normal mode-based fitting of atomic structure into electron density maps: application to sarcoplasmic reticulum Ca-ATPase*, Biophys J, 88, 818-827.
- Hinsen, K., Thomas, A. y Field, M.J. (1999) *Analysis of domain motions in large proteins*, Proteins, 34, 369-382.
- Hollup, S.M., Salensminde, G. y Reuter, N. (2005) *WEBnm@: a web application for normal mode analyses of proteins*, BMC Bioinformatics, 6, 52.
- Humphrey, W., Dalke, A. y Schulten, K. (1996) *VMD: visual molecular dynamics*, J Mol Graph, 14, 33-38, 27-38.
- Hyeon, C., Lorimer, G.H. y Thirumalai, D. (2006) *Dynamics of allosteric transitions in GroEL*, Proc Natl Acad Sci U S A, 103, 18939-18944.
- Ilari, A. y Savino, C. (2008) *Protein structure determination by x-ray crystallography*, Methods Mol Biol, 452, 63-87.
- Itoh, K. y Shimanouchi, T. (1970) *Vibrational frequencies and modes of alpha-helix*, Biopolymers, 9, 383-399.
- Jiang, W., Baker, M.L., Ludtke, S.J. y Chiu, W. (2001) *Bridging the information gap: computational tools for intermediate resolution structure interpretation*, J Mol Biol, 308, 1033-1044.
- Jolley, C.C., Wells, S.A., Fromme, P. y Thorpe, M.F. (2008) *Fitting low-resolution cryo-EM*



- maps of proteins using constrained geometric simulations*, Biophys J, 94, 1613-1621.
- Jonic, S., Sorzano, C.O. y Boisset, N. (2008) *Comparison of single-particle analysis and electron tomography approaches: an overview*, J Microsc, 232, 562-579.
- Kabsch, W. (1976) *A solution for the best rotation to relate two sets of vectors*, Acta Crystallographica Section A, 32, 922-923.
- Kabsch, W. (1978) *A discussion of the solution for the best rotation to relate two sets of vectors*, Acta Crystallographica Section A, 34, 827-828.
- Kamiya, K., Sugawara, Y. y Umeyama, H. (2003) *Algorithm for normal mode analysis with general internal coordinates*, J Comput Chem, 24, 826-841.
- Karplus, M. y McCammon, J.A. (2002) *Molecular dynamics simulations of biomolecules*, Nat Struct Biol, 9, 646-652.
- Keskin, O., Ma, B. y Nussinov, R. (2005) *Hot regions in protein--protein interactions: the organization and contribution of structurally conserved hot spot residues*, J Mol Biol, 345, 1281-1294.
- Kidera, A. y Go, N. (1990) *Refinement of protein dynamic structure: normal mode refinement*, Proc Natl Acad Sci U S A, 87, 3718-3722.
- Kidera, A., Inaka, K., Matsushima, M. y Go, N. (1992) *Normal mode refinement: crystallographic refinement of protein dynamic structure applied to human lysozyme*, Biopolymers, 32, 315-319.
- Kim, M.K., Jernigan, R.L. y Chirikjian, G.S. (2002) *Efficient generation of feasible pathways for protein conformational transitions*, Biophys J, 83, 1620-1630.
- Kitao, A., Hayward, S. y Go, N. (1994) *Comparison of normal mode analyses on a small globular protein in dihedral angle space and Cartesian coordinate space*, Biophys Chem, 52, 107-114.
- Kostek, S.A., Grob, P., De Carlo, S., Lipscomb, J.S., Garczarek, F. y Nogales, E. (2006) *Molecular architecture and conformational flexibility of human RNA polymerase II*, Structure, 14, 1691-1700.
- Kovacs, J.A., Cavasotto, C.N. y Abagyan, R. (2005) *Conformational sampling of protein flexibility in generalized coordinates: Application to ligand docking*, J Comput Theor Nanosci, 2, 354-361.
- Kovacs, J.A., Chacon, P. y Abagyan, R. (2004) *Predictions of protein flexibility: first-order measures*, Proteins, 56, 661-668.
- Kovacs, J.A., Yeager, M. and Abagyan, R. (2008) *Damped-dynamics flexible fitting*, Biophys J, 95, 3192-3207.
- Krebs, W.G., Alexandrov, V., Wilson, C.A., Echols, N., Yu, H. y Gerstein, M. (2002) *Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic*, Proteins, 48, 682-695.
- Krebs, W.G. y Gerstein, M. (2000) *The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework*, Nucleic Acids Res, 28, 1665-1675.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. y Higgins, D.G. (2007) *Clustal W and Clustal X version 2.0*, Bioinformatics, 23, 2947-2948.
- Lawson, C.L., Baker, M.L., Best, C., Bi, C., Dougherty, M., Feng, P., van Ginkel, G., Devkota, B., Lagerstedt, I., Ludtke, S.J., Newman, R.H., Oldfield, T.J., Rees, I., Sahni, G., Sala, R.,

- Velankar, S., Warren, J., Westbrook, J.D., Henrick, K., Kleywegt, G.J., Berman, H.M. y Chiu, W. (2011) *EMDataBank.org: unified data resource for CryoEM*, *Nucleic Acids Research*, 39, D456-D464.
- Lensink, M.F. y Mendez, R. (2008) *Recognition-induced conformational changes in protein-protein docking*, *Curr Pharm Biotechnol*, 9, 77-86.
- Lensink, M.F., Mendez, R. y Wodak, S.J. (2007) *Docking and scoring protein complexes: CAPRI 3rd Edition*, *Proteins*, 69, 704-718.
- Leo-Macias, A., Lopez-Romero, P., Lupyan, D., Zerbino, D. y Ortiz, A.R. (2005a) *An analysis of core deformations in protein superfamilies*, *Biophys J*, 88, 1291-1299.
- Leo-Macias, A., Lopez-Romero, P., Lupyan, D., Zerbino, D. y Ortiz, A.R. (2005b) *Core deformations in protein families: a physical perspective*, *Biophys Chem*, 115, 125-128.
- Levitt, M., Sander, C. y Stern, P.S. (1983) *The normal modes of a protein: Native bovine pancreatic trypsin inhibitor*, *International Journal of Quantum Chemistry*, 24, 181-199.
- Levitt, M., Sander, C. y Stern, P.S. (1985) *Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme*, *J Mol Biol*, 181, 423-447.
- Levy, R.M. y Karplus, M. (1979) *Vibrational approach to the dynamics of an  $\alpha$ -helix*, *Biopolymers*, 18, 2465-2495.
- Lindahl, E., Azuara, C., Koehl, P. y Delarue, M. (2006) *NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis*, *Nucleic Acids Res*, 34, W52-56.
- Lindahl, E. y Delarue, M. (2005) *Refinement of docked protein-ligand and protein-DNA structures using low frequency normal mode amplitude optimization*, *Nucleic Acids Res*, 33, 4496-4506.
- Lopez-Blanco, J.R., Garzon, J.I. y Chacon, P. (2011) *iMod: multipurpose normal mode analysis in internal coordinates*, *Bioinformatics*.
- Lu, M. y Ma, J. (2005) *The role of shape in determining molecular motions*, *Biophys J*, 89, 2395-2401.
- Lu, M., Poon, B. y Ma, J. (2006) *A New Method for Coarse-Grained Elastic Normal-Mode Analysis*, *J Chem Theory Comput*, 2, 464-471.
- Lucic, V., Leis, A. y Baumeister, W. (2008) *Cryo-electron tomography of cells: connecting structure and function*, *Histochem Cell Biol*, 130, 185-196.
- Ludtke, S.J. (2010) *3-D structures of macromolecules using single-particle analysis in EMAN*, *Methods Mol Biol*, 673, 157-173.
- Ludtke, S.J., Baldwin, P.R. y Chiu, W. (1999) *EMAN: semiautomated software for high-resolution single-particle reconstructions*, *J Struct Biol*, 128, 82-97.
- Llorca, O. (2005) *Introduction to 3D reconstruction of macromolecules using single particle electron microscopy*, *Acta Pharmacol Sin*, 26, 1153-1164.
- Ma, J. (2004) *New advances in normal mode analysis of supermolecular complexes and applications to structural refinement*, *Curr Protein Pept Sci*, 5, 119-123.
- Ma, J. (2005) *Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes*, *Structure*, 13, 373-380.
- Marabini, R., Masegosa, I.M., San Martin, M.C., Marco, S., Fernandez, J.J., de la Fraga, L.G., Vaquerizo, C. y Carazo, J.M. (1996) *Xmipp: An Image Processing Package for Electron Microscopy*, *J Struct Biol*, 116, 237-240.
- Maragakis, P. y Karplus, M. (2005) *Large amplitude conformational change in proteins explored*



- with a plastic network model: adenylate kinase*, J Mol Biol, 352, 807-822.
- Marques, O. y Sanejouand, Y.H. (1995) *Hinge-bending motion in citrate synthase arising from normal mode calculations*, Proteins, 23, 557-560.
- Marti-Renom, M.A., Pieper, U., Madhusudhan, M.S., Rossi, A., Eswar, N., Davis, F.P., Al-Shahrour, F., Dopazo, J. y Sali, A. (2007) *DBAli tools: mining the protein structure space*, Nucleic Acids Res, 35, W393-397.
- Martin-Benito, J., Boskovic, J., Gomez-Puertas, P., Carrascosa, J.L., Simons, C.T., Lewis, S.A., Bartolini, F., Cowan, N.J. y Valpuesta, J.M. (2002) *Structure of eukaryotic prefoldin and of its complexes with unfolded actin and the cytosolic chaperonin CCT*, EMBO J, 21, 6377-6386.
- Martin-Benito, J., Gomez-Reino, J., Stirling, P.C., Lundin, V.F., Gomez-Puertas, P., Boskovic, J., Chacon, P., Fernandez, J.J., Berenguer, J., Leroux, M.R. y Valpuesta, J.M. (2007) *Divergent substrate-binding mechanisms reveal an evolutionary specialization of eukaryotic prefoldin compared to its archaeal counterpart*, Structure, 15, 101-110.
- May, A. y Zacharias, M. (2005) *Accounting for global protein deformability during protein-protein and protein-ligand docking*, Biochim Biophys Acta, 1754, 225-231.
- May, A. y Zacharias, M. (2008a) *Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein-protein docking*, Proteins, 70, 794-809.
- May, A. y Zacharias, M. (2008b) *Protein-ligand docking accounting for receptor side chain and global flexibility in normal modes: evaluation on kinase inhibitor cross docking*, J Med Chem, 51, 3499-3506.
- Mayr, L.M. y Bojanic, D. (2009) *Novel trends in high-throughput screening*, Curr Opin Pharmacol, 9, 580-588.
- McCammon, J.A., Gelin, B.R. y Karplus, M. (1977) *Dynamics of folded proteins*, Nature, 267, 585-590.
- Meagher, K.L. y Carlson, H.A. (2004) *Incorporating protein flexibility in structure-based drug discovery: using HIV-1 protease as a test case*, J Am Chem Soc, 126, 13276-13281.
- Mendez, R. y Bastolla, U. (2010) *Torsional network model: normal modes in torsion angle space better correlate with conformation changes in proteins*, Phys Rev Lett, 104, 228103.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. y Teller, E. (1953) *Equation of state calculations by fast computing machines*, J Chem Phys, 21, 1087-1092.
- Michel, F., Crucifix, C., Granger, F., Eiler, S., Mouscadet, J.-F., Korolev, S., Agapkina, J., Ziganshin, R., Gottikh, M., Nazabal, A., Emiliani, S., Benarous, R., Moras, D., Schultz, P. y Ruff, M. (2009) *Structural basis for HIV-1 DNA integration in the human genome, role of the LEDGF/P75 cofactor*, EMBO J, 28, 980-991.
- Ming, D., Kong, Y., Lambert, M.A., Huang, Z. y Ma, J. (2002) *How to describe protein motion without amino acid sequence and atomic coordinates*, Proc Natl Acad Sci U S A, 99, 8620-8625.
- Mitra, K., Schaffitzel, C., Shaikh, T., Tama, F., Jenni, S., Brooks, C.L., 3rd, Ban, N. y Frank, J. (2005) *Structure of the E. coli protein-conducting channel bound to a translating ribosome*, Nature, 438, 318-324.
- Miyashita, O., Onuchic, J.N. y Wolynes, P.G. (2003) *Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins*, Proc Natl Acad Sci U S A, 100, 12570-12575.
- Moller, J.V., Olesen, C., Winther, A.M. y Nissen, P. (2010) *The sarcoplasmic Ca<sup>2+</sup>-ATPase:*

- design of a perfect chemi-osmotic pump*, Q Rev Biophys, 43, 501-566.
- Navaza, J., Lepault, J., Rey, F.A., Alvarez-Rua, C. y Borge, J. (2002) *On the fitting of model electron densities into EM reconstructions: a reciprocal-space formulation*, Acta Crystallogr D Biol Crystallogr, 58, 1820-1825.
- Neylon, C. (2008) *Small angle neutron and X-ray scattering in structural biology: recent examples from the literature*, Eur Biophys J, 37, 531-541.
- Noguti, T. y Go, N. (1982) *Collective variable description of small-amplitude conformational fluctuations in a globular protein*, Nature, 296, 776-778.
- Noguti, T. y Go, N. (1983a) *Dynamics of Native Globular Proteins in Terms of Dihedral Angles*, J. Phys. Soc. Jpn. , 52, 3283-3288.
- Noguti, T. y Go, N. (1983b) *A Method of Rapid Calculation of a Second Derivative Matrix of Conformational Energy for Large Molecules*, J. Phys. Soc. Jpn. , 52, 3685-3690.
- Noguti, T. y Go, N. (1985) *Efficient Monte Carlo method for simulation of fluctuating conformations of native proteins*, Biopolymers, 24, 527-546.
- Noy, A., Meyer, T., Rueda, M., Ferrer, C., Valencia, A., Perez, A., de la Cruz, X., Lopez-Bes, J.M., Pouplana, R., Fernandez-Recio, J., Luque, F.J. y Orozco, M. (2006) *Data mining of molecular dynamics trajectories of nucleic acids*, J Biomol Struct Dyn, 23, 447-456.
- Oberhauser, A.F. y Carrion-Vazquez, M. (2008) *Mechanical biochemistry of proteins one molecule at a time*, J Biol Chem, 283, 6617-6621.
- Orans, J., McSweeney, E.A., Iyer, R.R., Hast, M.A., Hellinga, H.W., Modrich, P. y Beese, L.S. (2011) *Structures of human exonuclease I DNA complexes suggest a unified mechanism for nuclease family*, Cell, 145, 212-223.
- Orellana, L., Rueda, M., Ferrer-Costa, C., López-Blanco, J.R., Chacón, P. y Orozco, M. (2010) *Approaching elastic network models to molecular dynamics flexibility*, J Chem Theory Comput, 6, 2910-2923.
- Orozco, M., Noy, A. y Pérez, A. (2008) *Recent advances in the study of nucleic acid flexibility by molecular dynamics*, Curr Opin Struct Biol, 18, 185-193.
- Orzechowski, M. y Tama, F. (2008) *Flexible fitting of high-resolution x-ray structures into cryoelectron microscopy maps using biased molecular dynamics simulations*, Biophys J, 95, 5692-5705.
- Palma, P.N., Krippahl, L., Wampler, J.E. y Moura, J.J. (2000) *BiGGER: a new (soft) docking algorithm for predicting protein interactions*, Proteins, 39, 372-384.
- Petrone, P. y Pande, V.S. (2006) *Can conformational change be described by only a few normal modes?*, Biophys J, 90, 1583-1593.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. y Ferrin, T.E. (2004) *UCSF Chimera--a visualization system for exploratory research and analysis*, J Comput Chem, 25, 1605-1612.
- Pons, C., Grosdidier, S., Solernou, A., Perez-Cano, L. and Fernandez-Recio, J. (2010) *Present and future challenges and limitations in protein-protein docking*, Proteins, 78, 95-108.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T. y Flannery, B.P. (1992) *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press.
- Radermacher, M. (1988) *Three-dimensional reconstruction of single particles from random and nonrandom tilt series*, J Electron Microscop Tech, 9, 359-394.
- Radon, J. (1986) *On the Determination of Functions from Their Integral Values along Certain Manifolds*, IEEE Trans Med Imaging, 5, 170-176.

- Rajamani, R. y Reynolds, C.H. (2004) *Modeling the binding affinities of beta-secretase inhibitors: application to subsite specificity*, Bioorg Med Chem Lett, 14, 4843-4846.
- Ratje, A.H., Loerke, J., Mikolajka, A., Brunner, M., Hildebrand, P.W., Starosta, A.L., Donhofer, A., Connell, S.R., Fucini, P., Mielke, T., Whitford, P.C., Onuchic, J.N., Yu, Y., Sanbonmatsu, K.Y., Hartmann, R.K., Penczek, P.A., Wilson, D.N. y Spahn, C.M. (2010) *Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites*, Nature, 468, 713-716.
- Renault, L., Chou, H.T., Chiu, P.L., Hill, R.M., Zeng, X., Gipson, B., Zhang, Z.Y., Cheng, A., Unger, V. y Stahlberg, H. (2006) *Milestones in electron crystallography*, J Comput Aided Mol Des, 20, 519-527.
- Ritchie, D.W. (2008) *Recent progress and future directions in protein-protein docking*, Curr Protein Pept Sci, 9, 1-15.
- Roseman, A.M. (2000) *Docking structures of domains into maps from cryo-electron microscopy using local correlation*, Acta Crystallogr D Biol Crystallogr, 56, 1332-1340.
- Rosenthal, P.B. y Henderson, R. (2003) *Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy*, J Mol Biol, 333, 721-745.
- Rossmann, M.G. (2000) *Fitting atomic models into electron-microscopy maps*, Acta Crystallogr D Biol Crystallogr, 56, 1341-1349.
- Rueda, M., Bottegoni, G. y Abagyan, R. (2009) *Consistent improvement of cross-docking results using binding site ensembles generated with elastic network normal modes*, J Chem Inf Model, 49, 716-725.
- Rueda, M., Chacon, P. y Orozco, M. (2007a) *Thorough validation of protein normal mode analysis: a comparative study with essential dynamics*, Structure, 15, 565-575.
- Rueda, M., Ferrer-Costa, C., Meyer, T., Perez, A., Camps, J., Hospital, A., Gelpi, J.L. y Orozco, M. (2007b) *A consensus view of protein dynamics*, Proc Natl Acad Sci U S A, 104, 796-801.
- Ruiz-Maso, J.A., Lopez-Zumel, C., Menendez, M., Espinosa, M. y del Solar, G. (2004) *Structural features of the initiator of replication protein RepB encoded by the promiscuous plasmid pMV158*, Biochim Biophys Acta, 1696, 113-119.
- Ruiz-Maso, J.A., Lurz, R., Espinosa, M. y del Solar, G. (2007) *Interactions between the RepB initiator protein of plasmid pMV158 and two distant DNA regions within the origin of replication*, Nucleic Acids Res, 35, 1230-1244.
- Russel, D., Lasker, K., Phillips, J., Schneidman-Duhovny, D., Velazquez-Muriel, J.A. y Sali, A. (2009) *The structural dynamics of macromolecular processes*, Curr Opin Cell Biol, 21, 97-108.
- Rusu, M., Birmanns, S. y Wriggers, W. (2008) *Biomolecular pleiomorphism probed by spatial interpolation of coarse models*, Bioinformatics, 24, 2460-2466.
- Sander, T., Liljefors, T. y Balle, T. (2008) *Prediction of the receptor conformation for iGluR2 agonist binding: QM/MM docking to an extensive conformational ensemble generated using normal mode analysis*, Journal of Molecular Graphics and Modelling, 26, 1259-1268.
- Scheres, S.H., Valle, M., Nunez, R., Sorzano, C.O., Marabini, R., Herman, G.T. y Carazo, J.M. (2005) *Maximum-likelihood multi-reference refinement for electron microscopy images*, J Mol Biol, 348, 139-149.
- Schroder, G.F., Brunger, A.T. y Levitt, M. (2007) *Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution*, Structure, 15, 1630-1641.

- Seeliger, D. y de Groot, B.L. (2010) *Conformational transitions upon ligand binding: holo-structure prediction from apo conformations*, PLoS Comput Biol, 6, e1000634.
- Shaw, D.E., Dror, R.O., Salmon, J.K., Grossman, J.P., Mackenzie, K.M., Bank, J.A., Young, C., Deneroff, M.M., Batson, B., Bowers, K.J., Chow, E., Eastwood, M.P., Ierardi, D.J., Klepeis, J.L., Kuskin, J.S., Larson, R.H., Lindorff-Larsen, K., Maragakis, P., Moraes, M.A., Piana, S., Shan, Y. y Towles, B. (2009) *Millisecond-scale molecular dynamics simulations on Anton. Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*. ACM, Portland, Oregon, pp. 1-11.
- Shaw, D.E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R.O., Eastwood, M.P., Bank, J.A., Jumper, J.M., Salmon, J.K., Shan, Y. y Wriggers, W. (2010) *Atomic-level characterization of the structural dynamics of proteins*, Science, 330, 341-346.
- Shoichet, B.K. (2004) *Virtual screening of chemical libraries*, Nature, 432, 862-865.
- Skjaerven, L., Hollup, S.M. y Reuter, N. (2009) *Normal mode analysis for proteins*, Journal of Molecular Structure: THEOCHEM, 898, 42-48.
- Skjaerven, L., Jonassen, I. y Reuter, N. (2007) *TMM@: a web application for the analysis of transmembrane helix mobility*, BMC Bioinformatics, 8, 232.
- Smith, G.R. y Sternberg, M.J. (2002) *Prediction of protein-protein interactions by docking methods*, Curr Opin Struct Biol, 12, 28-35.
- Smith, G.R., Sternberg, M.J. y Bates, P.A. (2005) *The relationship between the flexibility of proteins and their conformational states on forming protein-protein complexes with an application to protein-protein docking*, J Mol Biol, 347, 1077-1101.
- Soheilifard, R., Makarov, D.E. y Rodin, G.J. (2008) *Critical evaluation of simple network models of protein dynamics and their comparison with crystallographic B-factors*, Phys Biol, 5, 026008.
- Sohn, S.Y. y Cho, Y. (2009) *Crystal structure of the human rad9-hus1-rad1 clamp*, J Mol Biol, 390, 490-502.
- Sorzano, C.O., Marabini, R., Velazquez-Muriel, J., Bilbao-Castro, J.R., Scheres, S.H., Carazo, J.M. y Pascual-Montano, A. (2004) *XMIPP: a new generation of an open-source image processing package for electron microscopy*, J Struct Biol, 148, 194-204.
- Sternberg, M.J., Gabb, H.A. y Jackson, R.M. (1998) *Predictive docking of protein-protein and protein-DNA complexes*, Curr Opin Struct Biol, 8, 250-256.
- Stumpff-Kane, A.W., Maksimiak, K., Lee, M.S. y Feig, M. (2008) *Sampling of near-native protein conformations during protein structure refinement using a coarse-grained model, normal modes, and molecular dynamics simulations*, Proteins, 70, 1345-1356.
- Suhre, K., Navaza, J. y Sanejouand, Y.H. (2006) *NORMA: a tool for flexible fitting of high-resolution protein structures into low-resolution electron-microscopy-derived density maps*, Acta Crystallogr D Biol Crystallogr, 62, 1098-1100.
- Suhre, K. y Sanejouand, Y.H. (2004a) *ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement*, Nucleic Acids Res, 32, W610-614.
- Suhre, K. y Sanejouand, Y.H. (2004b) *On the potential of normal-mode analysis for solving difficult molecular-replacement problems*, Acta Crystallogr D Biol Crystallogr, 60, 796-799.
- Tama, F. (2003) *Normal mode analysis with simplified models to investigate the global dynamics of biological systems*, Protein Pept Lett, 10, 119-132.
- Tama, F. y Brooks, C.L., 3rd (2002) *The mechanism and pathway of pH induced swelling in*



- cowpea chlorotic mottle virus*, J Mol Biol, 318, 733-747.
- Tama, F., Gadea, F.X., Marques, O. y Sanejouand, Y.H. (2000) *Building-block approach for determining low-frequency normal modes of macromolecules*, Proteins, 41, 1-7.
- Tama, F., Miyashita, O. y Brooks, C.L., 3rd (2004a) *Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis*, J Mol Biol, 337, 985-999.
- Tama, F., Miyashita, O. y Brooks, C.L., 3rd (2004b) *Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM*, J Struct Biol, 147, 315-326.
- Tama, F. y Sanejouand, Y.H. (2001) *Conformational change of proteins arising from normal mode calculations*, Protein Eng, 14, 1-6.
- Tama, F., Wriggers, W. y Brooks, C.L., 3rd (2002) *Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory*, J Mol Biol, 321, 297-305.
- Tan, R.K., Devkota, B. y Harvey, S.C. (2008) *YUP.SCX: coaxing atomic models into medium resolution electron density maps*, J Struct Biol, 163, 163-174.
- Tasumi, M., Takeuchi, H., Ataka, S., Dwivedi, A.M. y Krimm, S. (1982) *Normal vibrations of proteins: glucagon*, Biopolymers, 21, 711-714.
- Tatsumi, R., Fukunishi, Y. y Nakamura, H. (2004) *A hybrid method of molecular dynamics and harmonic dynamics for docking of flexible ligand to flexible receptor*, J Comput Chem, 25, 1995-2005.
- Teague, S.J. (2003) *Implications of protein flexibility for drug discovery*, Nat Rev Drug Discov, 2, 527-541.
- Teodoro, M.L. y Kavraki, L.E. (2003) *Conformational flexibility models for the receptor in structure based drug design*, Curr Pharm Des, 9, 1635-1648.
- Teodoro, M.L., Phillips, G.N., Jr. y Kavraki, L.E. (2003) *Understanding protein flexibility through dimensionality reduction*, J Comput Biol, 10, 617-634.
- Tirion, M.M. (1996) *Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis*, Phys Rev Lett, 77, 1905-1908.
- Topf, M., Lasker, K., Webb, B., Wolfson, H., Chiu, W. y Sali, A. (2008) *Protein structure fitting and refinement guided by cryo-EM density*, Structure, 16, 295-307.
- Totrov, M. y Abagyan, R. (2008) *Flexible ligand docking to multiple receptor conformations: a practical alternative*, Curr Opin Struct Biol, 18, 178-184.
- Trabuco, L.G., Villa, E., Mitra, K., Frank, J. y Schulten, K. (2008) *Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics*, Structure, 16, 673-683.
- Vakser, I.A. y Kundrotas, P. (2008) *Predicting 3D structures of protein-protein complexes*, Curr Pharm Biotechnol, 9, 57-66.
- Valle, M., Zavialov, A., Sengupta, J., Rawat, U., Ehrenberg, M. y Frank, J. (2003) *Locking and unlocking of ribosomal motions*, Cell, 114, 123-134.
- van Heel, M., Gowen, B., Matadeen, R., Orlova, E.V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M. y Patwardhan, A. (2000) *Single-particle electron cryo-microscopy: towards atomic resolution*, Q Rev Biophys, 33, 307-369.
- van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R. y Schatz, M. (1996) *A new generation of the IMAGIC image processing system*, J Struct Biol, 116, 17-24.
- van Heel, M. y Schatz, M. (2005) *Fourier shell correlation threshold criteria*, J Struct Biol, 150, 151-154.

- 151, 250-262.
- van Vlijmen, H.W. y Karplus, M. (2005) *Normal mode calculations of icosahedral viruses with full dihedral flexibility by use of molecular symmetry*, J Mol Biol, 350, 528-542.
- van Wynsberghe, A.W. y Cui, Q. (2005) *Comparison of mode analyses at different resolutions applied to nucleic acid systems*, Biophys J, 89, 2939-2949.
- Velazquez-Muriel, J.A. y Carazo, J.M. (2007) *Flexible fitting in 3D-EM with incomplete data on superfamily variability*, J Struct Biol, 158, 165-181.
- Velazquez-Muriel, J.A., Valle, M., Santamaria-Pang, A., Kakadiaris, I.A. y Carazo, J.M. (2006) *Flexible fitting in 3D-EM guided by the structural variability of protein superfamilies*, Structure, 14, 1115-1126.
- Volkman, N. y Hanein, D. (2003) *Docking of atomic models into reconstructions from electron microscopy*, Methods Enzymol, 374, 204-225.
- Vonrhein, C., Schlauderer, G.J. y Schulz, G.E. (1995) *Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinases*, Structure, 3, 483-490.
- Wagner, G. (1993) *NMR relaxation and protein mobility*, Current Opinion in Structural Biology, 3, 748-754.
- Wako, H., Kato, M. y Endo, S. (2004) *ProMode: a database of normal mode analyses on protein molecules with a full-atom model*, Bioinformatics, 20, 2035-2043.
- Weiss, D.R. y Levitt, M. (2009) *Can morphing methods predict intermediate structures?*, J Mol Biol, 385, 665-674.
- Wells, S., Menor, S., Hespenheide, B. y Thorpe, M.F. (2005) *Constrained geometric simulation of diffusive motion in proteins*, Phys Biol, 2, S127-136.
- Wilson, E.B. (1939) *A Method of Obtaining the Expanded Secular Equation for the Vibration Frequencies of a Molecule*, The Journal of Chemical Physics, 7, 1047-1052.
- Wilson, E.B., Decius, J.C. y Cross, P.C. (1955) *Molecular vibrations: the theory of infrared and Raman vibrational spectra*. Dover Publications.
- Wriggers, W. (2010) *Using Situs for the integration of multi-resolution structures*, Biophys Rev, 2, 21-27.
- Wriggers, W. y Birmanns, S. (2001) *Using situs for flexible and rigid-body fitting of multiresolution single-molecule data*, J Struct Biol, 133, 193-202.
- Wriggers, W. y Chacon, P. (2001) *Modeling tricks and fitting techniques for multiresolution structures*, Structure, 9, 779-788.
- Wriggers, W., Chacón, P., Kovacs, J., Tama, F. y Birmanns, S. (2004) *Topology representing neural networks reconcile biomolecular shape, structure, and dynamics*, Neurocomputing, 56, 365-379.
- Wriggers, W., Milligan, R.A. y McCammon, J.A. (1999) *Situs: A package for docking crystal structures into low-resolution maps from electron microscopy*, J Struct Biol, 125, 185-195.
- Wriggers, W., Milligan, R.A., Schulten, K. y McCammon, J.A. (1998) *Self-organizing neural networks bridge the biomolecular resolution gap*, J Mol Biol, 284, 1247-1254.
- Wu, C.C., Rice, W.J. y Stokes, D.L. (2008) *Structure of a copper pump suggests a regulatory role for its metal-binding domain*, Structure, 16, 976-985.
- Yang, L.W. y Chng, C.P. (2008) *Coarse-grained models reveal functional dynamics--I. Elastic network models--theories, comparisons and perspectives*, Bioinform Biol Insights, 2, 25-45.
- Yang, L.W., Eyal, E., Chennubhotla, C., Jee, J., Gronenborn, A.M. y Bahar, I. (2007) *Insights into equilibrium dynamics of proteins from comparison of NMR and X-ray data with*



- computational predictions*, Structure, 15, 741-749.
- Yang, L.W., Liu, X., Jursa, C.J., Holliman, M., Rader, A.J., Karimi, H.A. y Bahar, I. (2005) *iGNM: a database of protein functional motions based on Gaussian Network Model*, Bioinformatics, 21, 2978-2987.
- Yang, L.W., Rader, A.J., Liu, X., Jursa, C.J., Chen, S.C., Karimi, H.A. y Bahar, I. (2006) *oGNM: online computation of structural dynamics using the Gaussian Network Model*, Nucleic Acids Res, 34, 24-31.
- Zaccai, G. (2003) *Proteins as nano-machines: dynamics–function relations studied by neutron scattering*, Journal of Physics: Condensed Matter, 15, S1673.
- Zacharias, M. (2004) *Rapid protein-ligand docking using soft modes from molecular dynamics simulations to account for protein deformability: binding of FK506 to FKBP*, Proteins, 54, 759-767.
- Zacharias, M. (2010) *Accounting for conformational changes during protein-protein docking*, Curr Opin Struct Biol, 20, 180-186.
- Zacharias, M. y Sklenar, H. (1999) *Harmonic modes as variables to approximately account for receptor flexibility in ligand–receptor docking simulations: Application to DNA minor groove ligand complex*, Journal of Computational Chemistry, 20, 287-300.
- Zhang, J., Baker, M.L., Schroder, G.F., Douglas, N.R., Reissmann, S., Jakana, J., Dougherty, M., Fu, C.J., Levitt, M., Ludtke, S.J., Frydman, J. y Chiu, W. (2010) *Mechanism of folding chamber closure in a group II chaperonin*, Nature, 463, 379-383.
- Zhang, P., Toyoshima, C., Yonekura, K., Green, N.M. y Stokes, D.L. (1998) *Structure of the calcium pump from sarcoplasmic reticulum at 8-Å resolution*, Nature, 392, 835-839.
- Zheng, W. (2011) *Accurate flexible fitting of high-resolution protein structures into cryo-electron microscopy maps using coarse-grained pseudo-energy minimization*, Biophys J, 100, 478-488.
- Zheng, W. y Brooks, B.R. (2005) *Normal-modes-based prediction of protein conformational changes guided by distance constraints*, Biophys J, 88, 3109-3117.
- Zheng, W., Brooks, B.R. y Hummer, G. (2007) *Protein conformational transitions explored by mixed elastic network models*, Proteins, 69, 43-57.
- Zheng, W. y Doniach, S. (2003) *A comparative study of motor-protein motions by using a simple elastic-network model*, Proc Natl Acad Sci U S A, 100, 13253-13258.
- Zhou, Z.H. (2011) *Atomic resolution cryo electron microscopy of macromolecular complexes*, Adv Protein Chem Struct Biol, 82, 1-35.