

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS QUÍMICAS
Departamento de Bioquímica y Biología Molecular I



**CARACTERIZACIÓN ESTRUCTURAL DE LA
PROTEÍNA MODULAR CbpF DE
“STREPTOCOCCUS PNEUMONIAE”:
APLICACIÓN DE COMPLEJOS DE Gd PARA LA
OBTENCIÓN DE FASES EN CRISTALOGRAFÍA
MACROMOLECULAR.**

**MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR**

Rafael Molina Monterrubio

Bajo la dirección del doctor

Juan Antonio Hermoso Domínguez

Madrid, 2010

• ISBN: 978-84-693-4091-2

© Rafael Molina Monterrubio, 2009

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS QUÍMICAS
DEPARTAMENTO DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR I

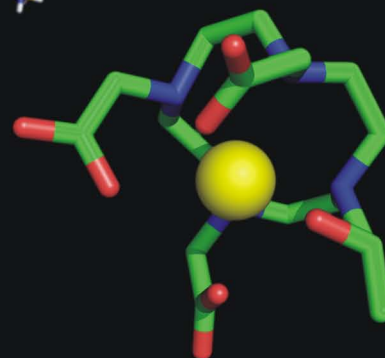
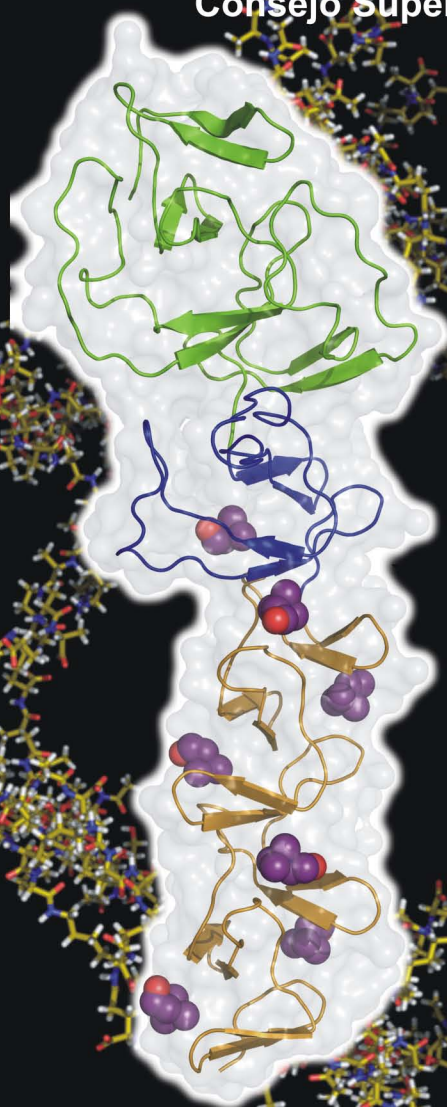
CARACTERIZACIÓN ESTRUCTURAL DE LA PROTEÍNA
MODULAR CbpF DE *Streptococcus pneumoniae*.
APLICACIÓN DE COMPLEJOS DE Gd PARA LA OBTENCIÓN
DE FASES EN CRISTALOGRAFÍA MACROMOLECULAR.

TESIS DOCTORAL

RAFAEL MOLINA MONTEERRUBIO

Instituto de Química-Física Rocasolano
Grupo de Cristalografía Macromolecular y Biología Estructural
Consejo Superior de Investigaciones Científicas

2009



UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS QUÍMICAS
Departamento de BIOQUÍMICA Y BIOLOGÍA MOLECULAR I



CARACTERIZACIÓN ESTRUCTURAL DE LA PROTEÍNA
MODULAR CbpF DE *Streptococcus pneumoniae*.
APLICACIÓN DE COMPLEJOS DE Gd PARA LA
OBTENCIÓN DE FASES EN CRISTALOGRAFÍA
MACROMOLECULAR.

TESIS DOCTORAL

Rafael Molina Monterrubio

Madrid, 2009



Instituto de Química-Física Rocasolano
Grupo de Cristalografía Macromolecular y Biología Estructural
Consejo Superior de Investigaciones Científicas

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS QUÍMICAS
Departamento de BIOQUÍMICA Y BIOLOGÍA MOLECULAR I



CARACTERIZACIÓN ESTRUCTURAL DE LA PROTEÍNA
MODULAR CbpF DE *Streptococcus pneumoniae*.
APLICACIÓN DE COMPLEJOS DE Gd PARA LA
OBTENCIÓN DE FASES EN CRISTALOGRAFÍA
MACROMOLECULAR.



Memoria presentada por Rafael Molina
Monterrubio para optar al grado de Doctor por
la Universidad Complutense de Madrid.

Director: Dr. Juan Antonio Hermoso Domínguez

A mis padres. A mi familia.

A los que siempre están.

AGRADECIMIENTOS

Quisiera dedicar mi más sincero agradecimiento a todas aquellas personas que de una forma u otra han hecho posible que este trabajo pudiera realizarse.

En primer lugar quisiera dar las gracias al director de esta tesis, el Dr. Juan Antonio Hermoso Domínguez, no sólo por todo lo que me ha enseñado del mundo de la *Cristalografía de Proteínas*, por ser más un compañero que un jefe, por sus dosis de optimismo siempre necesarias en este trabajo, sino fundamentalmente y por encima de todo, por ser capaz de contagiar la pasión que pone en su trabajo haciéndote partícipe de ese entusiasmo desde el primer día. De verdad Juan, personal y profesionalmente ha sido un lujo trabajar contigo. ¡Muchas gracias por darme la oportunidad de vivir todo esto!

Uno de los pilares en los que se ha cimentado esta tesis ha sido sin duda la “conexión franco-alemana”. De ella sólo puedo recordar buenos momentos a todos los niveles y desde luego para mí ha sido una de las mejores experiencias vividas durante todo el doctorado. A Richard, mi maestro en las artes del faseado con Gd, compañero de conciertos, más un padre que un jefe, camarada de mojitos (aún sigo buscando ron blanco de 55° para copiar tu receta), siempre con temas interesantes que contar ya fuera de ciencia o de cualquier otra cosa (nunca se me olvidará tu historia con Georges Charpak) y desde luego uno de esos científicos que cumple el estereotipo al que siempre admiré. A Meike, compañera de faseado, suministradora de los “complejos mágicos”, siempre dispuesta a ayudar en todo lo posible con la sonrisa por bandera. ¡Muchas gracias por el esfuerzo final para llegar a tiempo! No quisiera olvidarme también de mi compañero de aventuras por tierras francesas, Eduard, con el que compartí excursiones por los Alpes (ya puedo decir que he subido al Alpe d’Huez...aunque fuera en coche), inquietudes e incluso mi primer toma de contacto con París (¡¡¡muchas gracias por todo Pauli!!). A Annie, por su hospitalidad, por su naturalidad, por hablarme siempre que podía en español....¡¡nunca te gané jugando a esa especie de Scrabble!!

A todos de miembros del Grupo de Cristalografía Macromolecular y Biología Estructural, los X-men, por ser una familia (y digo bien, familia) de excelentes profesionales y compañeros de fatiga. A Josemi, la “otra” cara conocida cuando llegué al departamento, por sus explicaciones magistrales y con el que acabé descubriendo que

somos pseudo-paisanos (¡viva Alcubillas!). A Félix, Julia y Armando, siempre dispuestos a ayudar y, sobre todo, a animar los ratos del café (Armando, no me puedo quitar de la cabeza la tertulia sobre la conciencia de la gamba...). A Martín, por conseguir que los doctorandos estén “como en casa” y fundamentalmente, por ser el germen de un grupo extraordinario del que he tenido la suerte de pertenecer. A mi amiga-compañera-enchufe Inma, juntos desde nuestros primeros pinitos con la Química (allá por el 95'...), responsable de la oportunidad de ponerme en contacto con mi idolatrado “Roca” y con la que he compartido tantos y tan buenos momentos. Te deseo mucha suerte en tu periplo post-doctoral porque sin duda que si hay alguien que “se lo ha currado” eres tú. A Celia, extremeña-sevillana integrante del comando “Venezuela”, por aceptar los buenos y malos resultados siempre con una sonrisa (maldita SKL...), por hacerme un informe exhaustivo de la tesis de 19 hojas, por esa carta de despedida que nos secó las lágrimas y fundamentalmente, por todos los buenos momentos que hemos compartido y que nunca olvidaré. A M^a José, por su vitalidad y buen rollo, a Rebeca nuestra zoóloga que terminó de perfeccionar su baile latino en Venezuela (supongo, aunque sólo fuera por el profesor que tuviste...), Pablo, Sandra, Juana-María (si no fuera por tus geles...), e Iván, por haberse involucrado tanto cuando lo he necesitado. A mi primera padawan, Reyes... ¡cuanto juego dabas y cuanto se te echa de menos! A Susana Lobos y Noella Silva, las primeras venezolanas en aterrizar en el “Roca”, por aportar al grupo su incansable capacidad de trabajo (Susana, si llegas a estar más tiempo en el labo, acabarías con el presupuesto del departamento para soluciones de screening!!!) y sobre todo su alegría. A las componentes del 121, Bea, Marta y Lourdes, por echar siempre una mano y a las nuevas generaciones incorporadas que no nombro porque seguro que me olvidaría de alguno/a...

A todas aquellas personas ajenas al GCMBE que también han formado parte del trabajo y de mi estancia en el instituto. A Marga y Laura, que me iniciaron en la purificación de las CBPs, a Pepe Saíz y Germán Rivas, por su ayuda con los experimentos de ultracentrifugación y a Rosa Lebrón, por su eficacia y rapidez con las masas. A Jaime, mi colega de la competencia, un gran tipo con el que compartía tantas y tantas cosas (de verdad, parece que llevábamos vidas paralelas) y al que le deseo lo mejor. Un día de tertulia depresiva me dijo que a veces el recorrido del camino merece más la pena que el fin de ese camino. Tenía razón. Nuestro viaje, aunque no fuera en motocicleta, desde luego que

mereció la pena. A Consuelito, por ser tan eficaz y sobre todo por esas fiestas de navidad que seguramente sin ella no serían lo mismo...

Del departamento de Microbiología Molecular del CIB, a Pedro García, Jose Luis García y Ana González, que trabajaron intensamente con nuestra “F” y que aunque se resistió hasta el final, nos proporcionó un final feliz.

A mis amigas/os, por tener que soportar mis paranoias tesineras, por estar ahí cuando las cosas se pusieron difíciles, por el apoyo necesario para compaginar el trabajo con la tesis y fundamentalmente, por todo lo que hemos vivido.

A Bea, por ser la auténtica sufridora de los “daños colaterales” de esta Tesis, por demostrarme en todo momento su apoyo incondicional, por asumir con abnegación mi devoción por este trabajo incluso cuando a veces lo hace incompatible con “hacer planes”, por todo lo que hemos compartido... ¡Ah!, ¡¡¡y por tus sesiones de Photoshop!!!

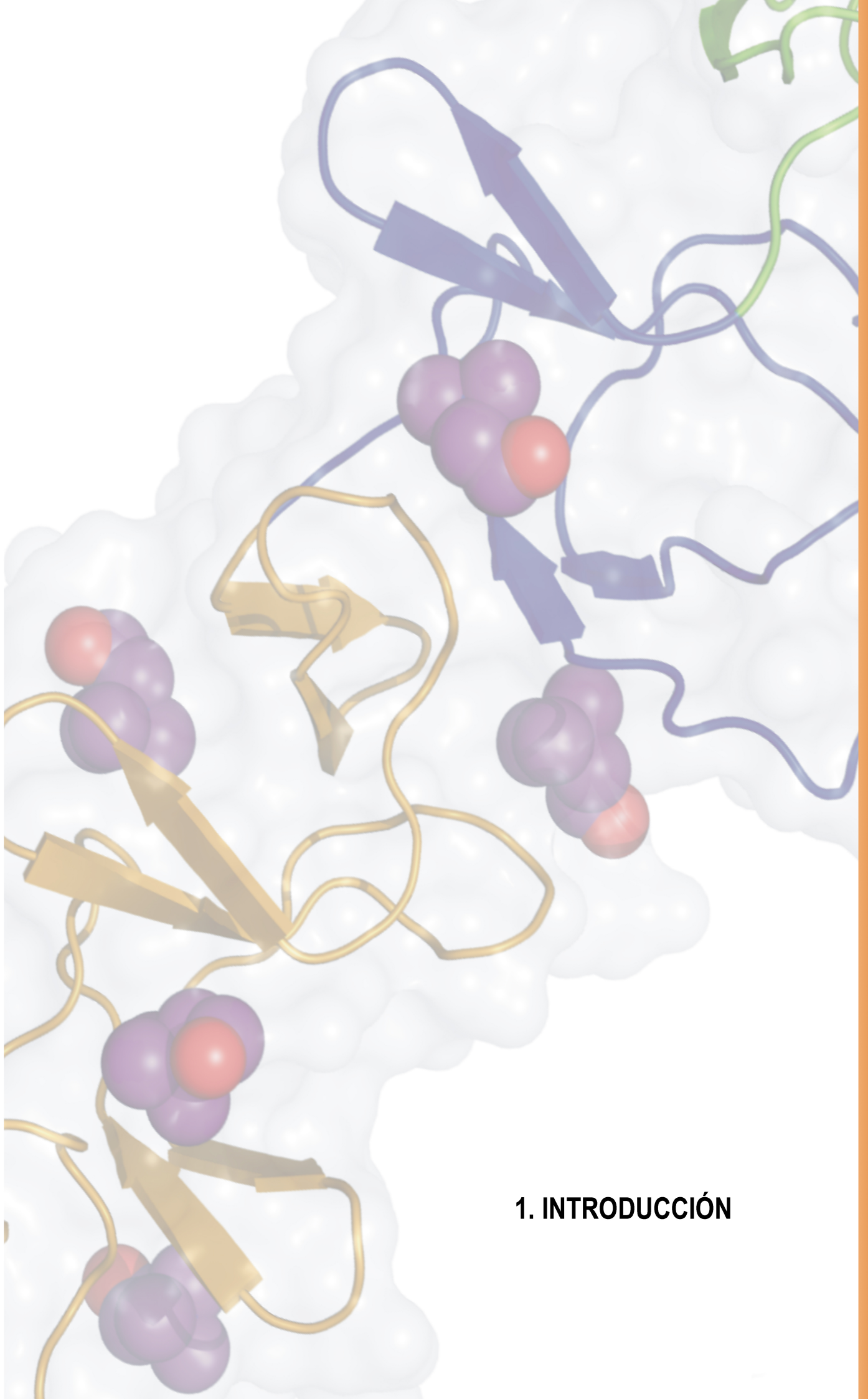
Por último, a mi familia, por estar desde siempre a mi lado ayudándome en todo lo necesario. A mi hermana Mayra que nunca ha entendido muy bien “para qué tanto esfuerzo” pero que siempre me ha apoyado, por su cariño, por todo lo que hemos vivido juntos y lo que no queda por vivir (sí, sí, incluso con tus sobris...). A mis padres, Arcángel y M^a Cruz, sin los cuales no hubiera llegado hasta aquí, por su sacrificio, por sus ánimos y consejos, por sufrir conmigo en los momentos más difíciles, y sobre todo, por recordarme en cada momento que es lo verdaderamente importante en la vida.

¡¡MUCHAS GRACIAS A TODOS!!

1. INTRODUCCIÓN	1
1.1. <i>STREPTOCOCCUS PNEUMONIAE</i>	1
1.1.1. CARACTERÍSTICAS GENERALES	1
1.1.2. LA CÁPSULA POLISACARÍDICA	3
1.1.3. LA PARED CELULAR DE NEUMOCOCO	4
1.1.3.1. El peptidoglicano	5
1.1.3.2. Otros componentes de la pared celular	8
1.1.4. PROTEÍNAS DE UNIÓN A COLINA	11
1.1.5. PAPEL DE LA ENVUELTA CELULAR Y DE LAS PROTEÍNAS DE SUPERFICIE EN LA VIRULENCIA DE NEUMOCOCO	15
1.1.6. ESTRUCTURAS TRIDIMENSIONALES DE CBPS DE NEUMOCOCO	17
1.1.6.1. C-LytA	17
1.1.6.2. Cpl-1	19
1.1.6.3. Pce	20
1.1.6.4. Dominio de adhesión R2 de CbpA	23
1.2. OBTENCIÓN DE FASES EN LA CRISTALOGRAFÍA MACROMOLECULAR	25
1.2.1. DETERMINACIÓN ESTRUCTURAL DE MACROMOLÉCULAS POR DIFRACCIÓN DE RAYOS X	25
1.2.2. DISPERSIÓN ANÓMALA	34
1.2.3. DETERMINACIÓN DE LA POSICIÓN DE LOS ÁTOMOS PESADOS Y OBTENCIÓN DE FASES	37
1.2.4. PRODUCCIÓN DE CRISTALES DERIVADOS PARA LA OBTENCIÓN DE FASES POR MÉTODOS DE REEMPLAZAMIENTO MOLECULAR ISOMORFO O POR MÉTODOS DE DISPERSIÓN ANÓMALA	41
1.2.5. UTILIZACIÓN DE COMPLEJOS DE LANTÁNIDOS PARA OBTENER DERIVADOS	50
2. OBJETIVOS	59
3. TÉCNICAS EXPERIMENTALES, RESULTADOS Y DISCUSIÓN	63
3.1. ESQUEMA GENERAL EN CRISTALOGRAFÍA DE RAYOS-X	63
3.2. EXPRESIÓN Y PURIFICACIÓN	63
3.2.1. INTRODUCCIÓN	63
3.2.2. EXPRESIÓN Y PURIFICACIÓN DE CbpF	66

3.3. ESPECTROMETRÍA DE MASAS.....	68
3.3.1. CONSIDERACIONES GENERALES	68
3.3.2. DETERMINACIÓN DE ESPECIES MOLECULARES EN MUESTRAS PURIFICADAS DE CBPF.....	69
3.4. ULTRACENTRIFUGACIÓN ANALÍTICA	70
3.4.1. EXPERIMENTOS DE EQUILIBRIO DE SEDIMENTACIÓN POR ABSORBANCIA	71
3.4.2. EXPERIMENTOS DE EQUILIBRIO DE SEDIMENTACIÓN POR REFRACTOMETRÍA.....	73
3.5. CRISTALIZACIÓN	75
3.5.1. INTRODUCCIÓN.....	75
3.5.1.1. Principios básicos de la cristalización.....	77
3.5.1.2. Condiciones ideales de la proteína a cristalizar	79
3.5.1.3. Proceso experimental de cristalización.....	81
3.5.1.4. Métodos de cristalización.....	82
3.5.2. CRISTALIZACIÓN DE CBPF Y SUS DERIVADOS LANTÁNIDOS.....	84
3.5.2.1. Cristalización de CbpF.....	84
3.5.2.2. Cristalización de derivados de CbpF. Derivados lantánidos.....	85
3.6. TÉCNICAS CRISTALOGRÁFICAS Y RESOLUCIÓN ESTRUCTURAL	87
3.6.1. CONCEPTOS GENERALES.....	87
3.6.1.1. Concepto de difracción	87
3.6.1.2. Toma de datos de difracción.....	88
3.6.1.3. Fuentes de radiación y equipos de difracción	89
3.6.1.4. Preparación de la muestra	91
3.6.1.5. Resolución de la estructura	93
3.6.1.6. Construcción del modelo y refinamiento.....	96
3.6.2. RESOLUCIÓN ESTRUCTURAL DE CBPF	97
3.6.2.1. Toma de datos de difracción.....	97
3.6.2.2. Obtención de fases con soluciones de átomo pesado convencionales. Obtención de fases con yoduro.....	99
3.6.2.3. Obtención de fases con complejos lantánidos.....	100
3.6.2.4. Construcción y refinamiento de la estructura del complejo CbpF-colina	104
3.6.3. ANÁLISIS DE LA ESTRUCTURA TRIDIMENSIONAL DEL COMPLEJO CBPF-COLINA	106
3.6.3.1. Organización y plegamiento general.....	106

3.6.3.2. Divergencia secuencial en las repeticiones de unión a colina	111
3.6.3.3. Modelo de unión de CbpF a la cadena de teicoico. Modelo de interacción de CbpF a la pared de neumococo.....	113
3.6.4. BÚSQUEDA DE LA FUNCIÓN DE CBPF.....	115
3.6.4.1. Ensayos de unión con diferentes análogos de la pared y análogos de teicoico	115
3.6.4.2. Estudios de acoplamiento	119
3.6.4.3. Ensayos in vitro e in vivo de actividad lítica de murein hidrolasas en presencia y ausencia de CbpF	121
3.6.5. ANÁLISIS DE LA OBTENCIÓN DE FASES MEDIANTE COMPLEJOS LANTÁNIDOS DE GD EN CRISTALOGRAFÍA MACROMOLECULAR.....	128
3.6.5.1. Consideraciones generales.....	129
3.6.5.2. Recogida de datos y procesamiento.....	130
3.6.5.3. Obtención de fases por SAD.....	131
3.6.5.4. Estudio comparativo del poder de obtención de fases de los distintos complejos de Gd.....	132
3.6.5.5. Estudio comparativo de la cocrystalización frente a la inmersión como método de obtención de derivados.....	136
3.6.5.6. Modo de unión de los diferentes complejos de Gd empleados.....	139
3.6.5.7. Estudio comparativo de los datos obtenidos por radiación sincrotrón frente al ánodo rotatorio.....	143
4. CONCLUSIONES	151
5. BIBLIOGRAFÍA	157
6. ARTÍCULOS	177



1. INTRODUCCIÓN

1. INTRODUCCIÓN.

1.1. *STREPTOCOCCUS PNEUMONIAE*.

1.1.1. CARACTERÍSTICAS GENERALES.

Streptococcus pneumoniae (neumococo) es una bacteria Gram positiva que posee una de las mayores tasas de mortalidad y morbilidad para el ser humano, siendo el causante de un gran número de procesos infecciosos (neumonía, otitis media, sinusitis, peritonitis, etc.) y de procesos invasivos severos (meningitis, septicemia, etc.), particularmente en ancianos, niños y personas inmunodeprimidas. Según la Organización Mundial de la Salud (OMS), la neumonía es la cuarta causa de mortalidad mundial, y el neumococo es el responsable del 10-25% de los casos en países desarrollados, y de hasta un 70% en países en vías de desarrollo, estimándose que anualmente esta bacteria provoca la muerte de casi un millón de niños en todo el mundo (OMS, 2007). En la actualidad, la relevancia adquirida por las infecciones causadas por neumococo unido a la inexistencia de una vacuna realmente eficaz o a la aparición de cepas resistentes a antibióticos, hacen necesario el desarrollo de nuevas estrategias terapéuticas. Dentro de estas nuevas estrategias cabe resaltar la búsqueda de determinantes antigénicos más selectivos que la cápsula polisacáridica de las actuales vacunas con el fin de desarrollar terapias más eficaces. Así, recientemente ha cobrado especial importancia el estudio de las proteínas de superficie de neumococo dada su implicación en la virulencia de la bacteria y su potencialidad como agentes antigénicos (Bergmann & Hammerschmidt, 2006). Otra de las estrategias cuya aplicación ha sido demostrada en estudios *in vivo*, consiste en el empleo de proteínas líticas codificadas por los bacteriófagos que infectan a neumococo como enzibióticos (Hermoso *et al.*, 2007).

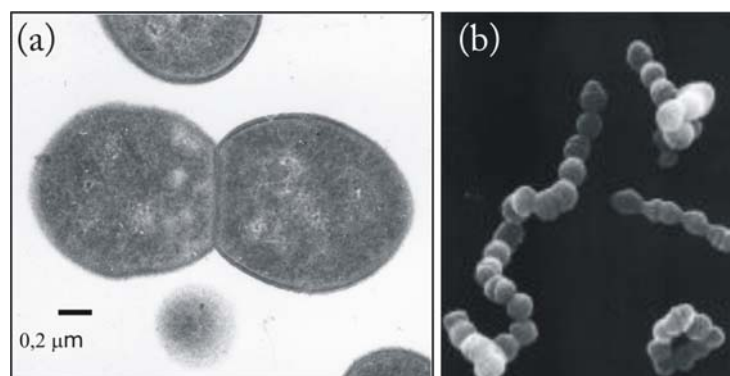


Figura 1.1.1: Variaciones morfológicas de *S. pneumoniae*. (a) Diplococo. (b) Largas cadenas.

El neumococo tiene como hábitat natural la nasofaringe humana y su colonización puede tener lugar incluso durante los primeros días de vida y permanecer en un estado latente. Sin embargo, bajo ciertas condiciones, el neumococo se vuelve virulento y es capaz de causar los procesos invasivos (Cartwright *et al.*, 2002). En 1928 Griffith puso de manifiesto que la cápsula es el principal factor de virulencia. Mediante la inoculación a ratones de neumococos encapsulados (estirpe lisa) y no encapsulados (estirpe rugosa), encontró que los animales tratados con la forma encapsulada morían, mientras que la inyección de la estirpe rugosa era inocua (Griffith, 1928). Posteriormente se aisló una nueva variedad que se manifestaba como no encapsulada, con superficies irregulares, y que mostraba alteraciones en el patrón de división, produciendo filamentos de gran longitud (Dawson, 1933). La variante morfológica más frecuentemente aislada en personas infectadas con neumococo es la forma “lisa” de Griffith.

Una de las características más importantes del neumococo es que presenta un requerimiento nutricional absoluto por colina, la cual es incorporada en forma de fosforilcolina (PC) presente en el ácido teicoico (TA) de la pared celular y en el ácido lipoteicoico (LTA) de la membrana (Tomasz, 1967). La ausencia de este aminoalcohol en el medio de cultivo produce alteraciones fisiológicas graves como la detención de la división celular y la formación de cadenas celulares, unidas mediante pequeñas expansiones de los polos (filamentos) de la pared celular (Tomasz *et al.*, 1975). Este efecto también se observa si se incluyen concentraciones elevadas de colina en el medio de cultivo (20 µg/ml) o si ésta se sustituye por análogos estructurales como la etanolamina (Tomasz, 1968) y otros compuestos de amonio cuaternario. Cuando se reemplaza la colina por etanolamina, además de no producirse la separación de las células hijas, aparecen otras alteraciones, como el bloqueo de la autólisis al final de la fase estacionaria, la resistencia a la lisis producida por desoxicolato sódico (DOC), la tolerancia a antibióticos β-lactámicos, la incapacidad de transformación y el aumento de la resistencia a la infección por fagos (Tomasz, 1970). Sin embargo, la adición de colina al medio revierte todos estos efectos. Estudios posteriores han permitido la obtención de mutantes de neumococo (cepas JY2190 y R6Col⁻) que no muestran este requerimiento por colina y que son capaces de multiplicarse en ausencia de este aminoalcohol (Yother *et al.*, 1998; Severin *et al.*, 1997).

Existen otros factores, tales como la presencia o la ausencia de iones metálicos (Austrian, 1953), que pueden alterar también la morfología de neumococo. En ausencia de cationes

divalentes las formas no filamentosas se comportan fenotípicamente como las filamentosas, lo que indica que los metales pueden influir en el proceso de división celular (Webb, 1951; Shankar & Bard, 1952). Basándose en este hecho se ha sugerido la presencia de una enzima, en las formas no filamentosas, que estaría implicada en la separación de las células hijas tras la división celular y cuya actividad sería dependiente de metales divalentes.

En un principio se describió que la virulencia de neumococo venía exclusivamente determinada por la presencia de la cápsula (Austrian, 1953) aunque, recientemente, se han descrito nuevos factores de virulencia, entre los que se encuentran diversas proteínas y componentes de la pared celular (McCullers & Tuomanen, 2001; Mitchell, 2003).

A continuación se describen detalladamente los distintos componentes estructurales de la envuelta celular de neumococo.

1.1.2. LA CÁPSULA POLISACARÍDICA.

La cápsula es la estructura más externa (Figura 1.1.2) y el principal factor de virulencia de neumococo. Está compuesta, principalmente, por polisacáridos cargados negativamente (Kamerling, 2000) que rodean a la célula y que se mantienen unidos a la superficie de la bacteria, posiblemente, mediante enlaces covalentes (Skov Sørensen *et al.*, 1988; Yeung & Mattingly, 1986). Existen aproximadamente 90 serotipos capsulares distintos con una composición química compleja y variable (Paton & Morona, 2000), en la que los polisacáridos confieren las propiedades inmunogénicas y los componentes no sacarídicos proporcionan el carácter antigénico.

A pesar de que la cápsula no parece tener ningún papel en los fenómenos de adherencia, invasión o inflamación, ésta es esencial para la virulencia de la bacteria debido a su capacidad para bloquear el reconocimiento del neumococo por parte del hospedador, impidiendo así fagocitosis (Jonsson *et al.*, 1985; Yother, 2004). Existen variantes morfológicas de cepas de neumococo con diferente capacidad de unión a las células de la nasofaringe, y que se clasifican en función del fenotipo que presentan (variación de fase) sobre placas transparentes de agar sólidas denominándose opacas, semi-transparentes y transparentes (Weiser *et al.*, 1994). Las primeras presentan mayor cantidad de cápsula polisacarídica, menor contenido de ácidos teicoicos en la pared celular (Kim & Weiser, 1998), y su mayor virulencia se ha asociado a una

mejora de su supervivencia en sangre. Las variantes transparentes poseen menor cantidad de cápsula y mayor contenido de ácidos teicoicos, lo que aumenta la capacidad para colonizar la nasofaringe (Weiser *et al.*, 1994).

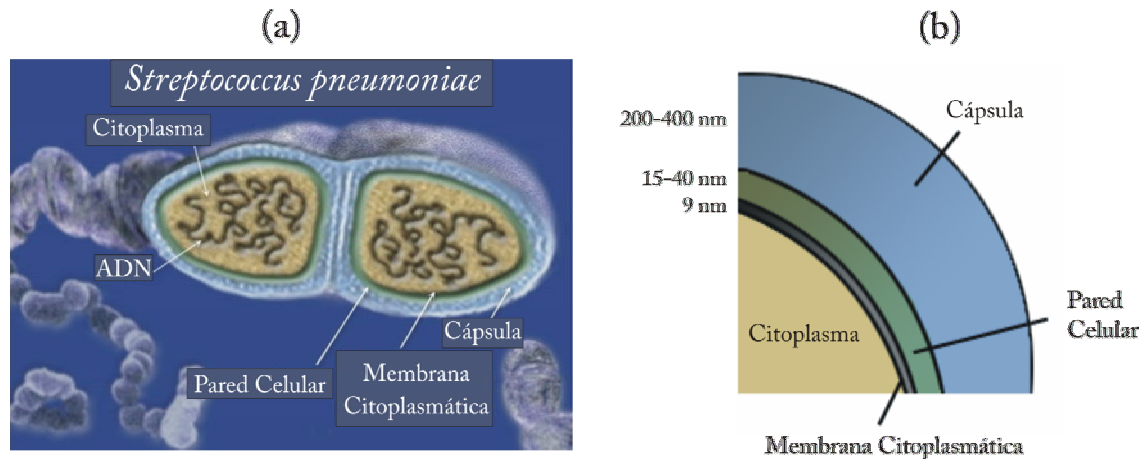


Figura 1.1.2: Organización celular en *Streptococcus pneumoniae*. (a) Corte celular. (b) Esquema de la estructura externa. Los espesores aproximados de cada capa se indican a la izquierda de la figura.

1.1.3. LA PARED CELULAR DE NEUMOCOCO.

La pared celular rodea a la membrana citoplasmática (Figuras 1.1.2 y 1.1.3) y confiere a la bacteria una morfología típica. La adquisición de este tipo de estructuras, a modo de exoesqueleto, ha servido a las bacterias como mecanismo de adaptación al medio externo, protegiéndola de su posible lisis, ya sea osmótica o mecánica (Salton, 1994). La pared bacteriana, además de ser el medio de intercambio de solutos entre el exterior y el interior celular, sirve de sistema de anclaje para toda una batería de proteínas implicadas en procesos de crecimiento y división celular, y en las interacciones de la célula con el medio exterior. De hecho, la pared desempeña un papel importante en los procesos de colonización, adherencia, inflamación e invasión bacteriana (McCullers & Tuomanen, 2001), ya que el neumococo modula, a través de la variación de fase, la distribución de subcomponentes de la pared relacionados con la internalización de la bacteria o la inducción de la respuesta inflamatoria durante el proceso de infección (Tuomanen *et al.*, 1985a; Tuomanen *et al.*, 1987).

La composición de la pared varía de unas especies a otras. En el caso de neumococo está constituida por un entramado de cadenas de peptidoglicano (mureína) y por los ácidos teicoicos asociados a ellas, formando este conjunto una estructura multilaminar con un espesor comprendido entre 15-40 nm (Figuras 1.1.2 y 1.1.3).

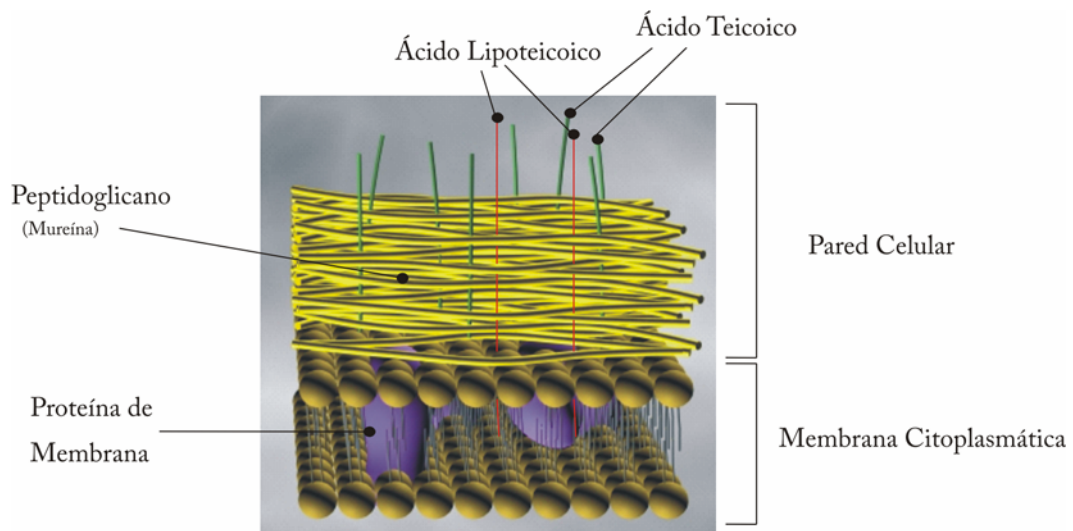


Figura 1.1.3: Representación esquemática de la pared celular de *S. pneumoniae*. Está constituida por el peptidoglicano (fibras amarillas), los ácidos teicoicos (trazos verdes) y lipoteicoicos (trazos rojos), así como la membrana citoplasmática (esferas marrones) y proteínas de membrana (morado).

1.1.3.1. El peptidoglicano.

El peptidoglicano de neumococo está formado por un entramado tridimensional de cadenas glicánicas constituidas por residuos alternados de ácido N-acetilmurámico y glucosamina unidos mediante enlaces glicosídicos $\beta 1 \rightarrow 4$ (Figuras 1.1.5a, 1.1.5c y 1.1.6), que se entrecruzan mediante cortos segmentos peptídicos. Cada cadena glicánica posee alrededor de unos 35 disacáridos (Beveridge, 1981), lo que supone, en su conformación más extendida, una longitud media de 35 nm (Formanek, 1983). El crecimiento de la pared se realiza intercalando nuevos anillos de peptidoglicano a los ya existentes (Pelzer, 1983). El nivel de acetilación de los residuos es del 90% para el murámico y del 16% para la glucosamina (Vollmer & Tomasz, 2000). La desacetilación de los residuos tiene lugar tras su incorporación a la pared celular y

constituye uno de los mecanismos de virulencia de la bacteria, al hacerla resistente a las lisozimas del hospedador (Vollmer & Tomasz, 2000).

La unión entre el polisacárido y las pequeñas cadenas peptídicas se realiza mediante el grupo D-láctico del ácido murámico (Figura 1.1.5a y 1.1.5c). La disposición helicoidal que adoptan las cadenas glicánicas hace que las unidades peptídicas queden proyectadas en las tres direcciones del espacio, lo que permite establecer enlaces cruzados con los péptidos de cadenas próximas (Barnickel *et al.*, 1983) (Figuras 1.1.5b y 1.1.6). Las capas de mureína se disponen formando ángulos de 68° con las adyacentes, favoreciendo así la formación del mayor número de conexiones interpeptídicas. Esto proporciona rigidez a la pared y aumenta la resistencia a la degradación por las enzimas líticas, manteniendo, por otra parte, cierta flexibilidad debido a la presencia de las cadenas peptídicas. Además, la estructura del peptidoglicano presenta poros que permiten la interacción de la pared con las proteínas de superficie de la bacteria (Beveridge, 1981). Recientemente se ha determinado la estructura tridimensional de un fragmento sintético del peptidoglicano (Meroueh *et al.*, 2006) a partir del cual se propone que las cadenas glicánicas se dispondrían perpendicularmente a la membrana, con las cadenas peptídicas giradas cada 120° , permitiendo así la interacción con los péptidos adyacentes, y formando en conjunto una estructura tipo panal de abeja (Figura 1.1.4).

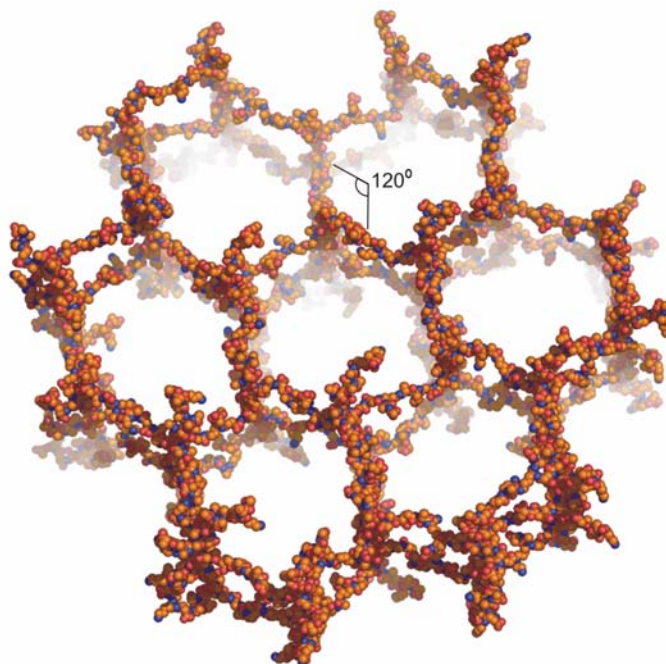


Figura 1.1.4: Estructura tridimensional de un fragmento sintético de peptidoglicano.

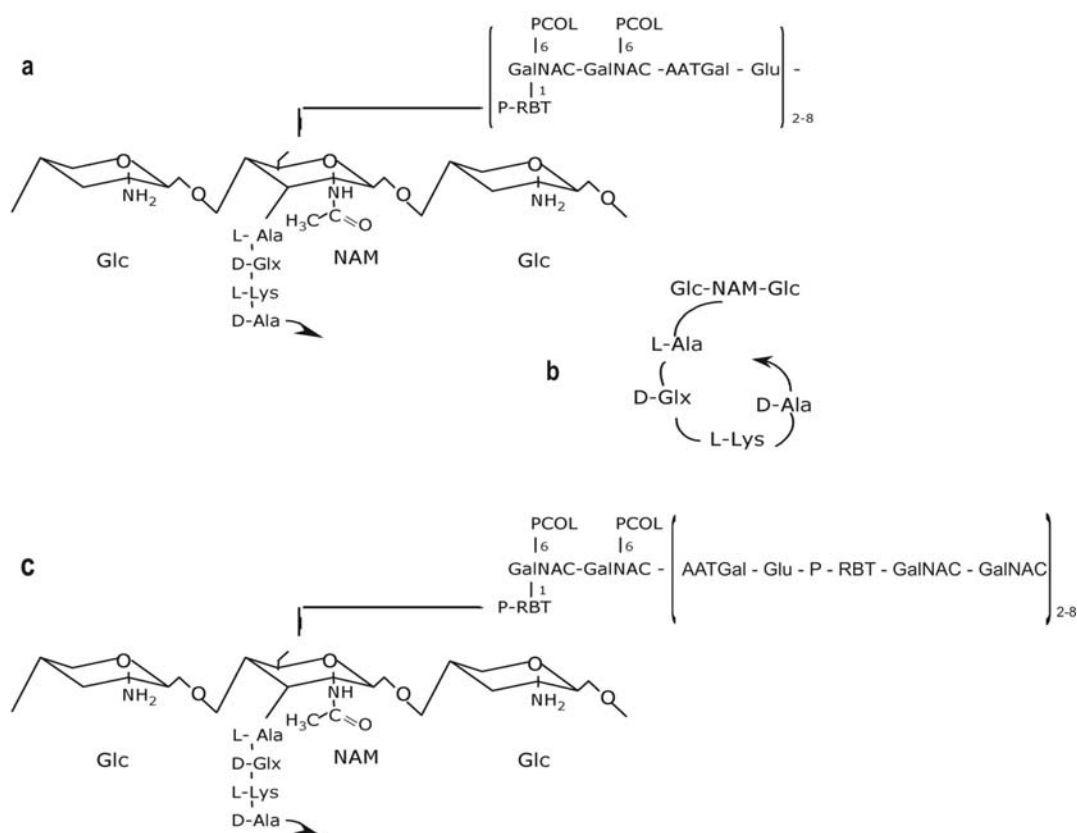


Figura 1.1.5: Estructura primaria del peptidoglicano (a). Modelo aceptado de cadena glicánica formada por la repetición de los residuos de ácido N-acilmurámico (NAM) y glucosamina (Glc). En la posición 6 del NAM se unen las cadenas de ácido teicoico (GalNAC, N-acetilgalactosamina; AATGal, diacetoamidotridesoxigalactosa; Glu, glucosa; PCOL, fosforilcolina; P-RBT, ribitol fosfato). Unida al grupo D-lactilo en posición 2 del NAM se posiciona la cadena peptídica que, en su conformación más estable, presenta un giro (b). Modelo recientemente propuesto de cadena glicánica (c) (Seong Seo *et al.*, 2008).

En neumococo las cadenas peptídicas contienen mayoritariamente tres (L-Ala-D-Glx-L-Lys) o cuatro (L-Ala-D-Glx-L-Lys-D-Ala) aminoácidos, apareciendo en mucha menor proporción el pentapéptido L-Ala-D-Glx-L-Lys-D-Ala-D-Ala (Schleifer & Kandler, 1972). La presencia del tripéptido parece indicar la existencia de D,D y D,L-carboxipeptidasas (García Bustos *et al.*, 1987; García Bustos & Tomasz, 1990). Existen dos tipos de conexiones interpeptídicas: la primera tiene lugar entre la D-Ala en posición terminal de la unidad donadora y la L-Lys en posición 3 de la unidad receptora mediante el puente L-Ala-L-Ser (o menos frecuentemente L-Ala-L-Ala); el segundo tipo es el enlace directo entre la D-Ala terminal y la L-Lys (Figura 1.1.6). El primer tipo es el más abundante y, aunque, estos son los aminoácidos que se han

encontrado en mayor proporción en las uniones interpeptídicas, se ha descrito también la presencia minoritaria de glicina y aspártico (Fischer y Tomasz, 1985). Por otro lado, se ha observado que las estirpes resistentes a penicilina presentan un contenido importante de péptidos ramificados, mientras que en cepas sensibles el 70% de los péptidos son lineales (García Bustos & Tomasz, 1990; Severin & Tomasz, 1996).

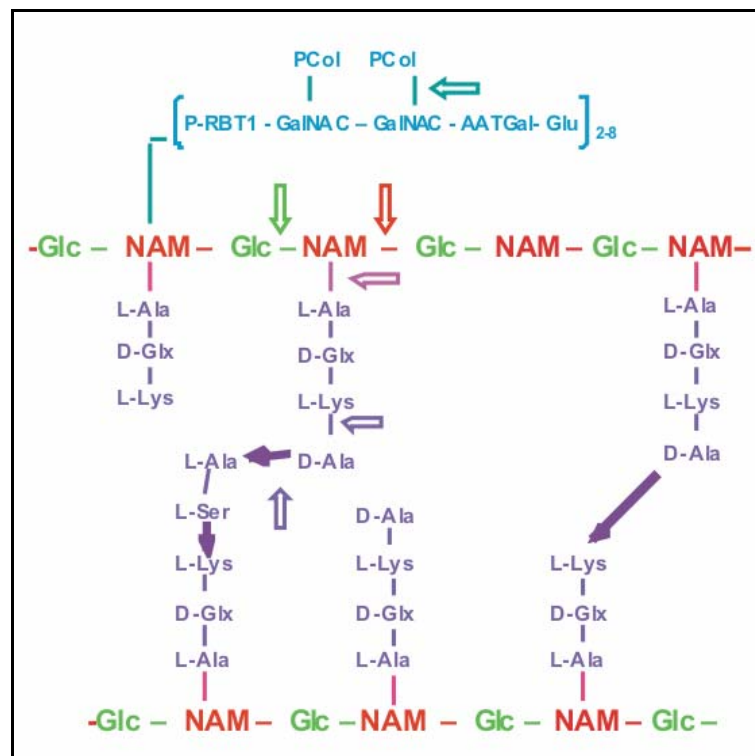


Figura 1.1.6: Esquema de la pared celular de neumococo. Cadenas glicánicas (rojo y verde); péptidos y uniones interpeptídicas (violeta); ácido teicoico (azul). Las flechas indican los enlaces de la pared que hidrolizan las amidasas (rosa), lisozimas (rojo), glucosaminidasas (verde), peptidasas (violeta), y fosforilcolín esterasa (azul).

1.1.3.2. Otros componentes de la pared celular.

La pared celular contiene otros polímeros sintetizados por la propia bacteria que le confieren características diferenciales. Entre estos compuestos se encuentran los ácidos teicoicos (TA), descritos por primera vez en 1930 como el polisacárido C de neumococo (Tilled *et al.*, 1930), y los ácidos lipoteicoicos (LTA) (Salton, 1994). En neumococo, los ácidos lipoteicoicos, denominados antígeno F, poseen la misma composición que los ácidos teicoicos, lo que genera una situación única dentro de las bacterias Gram-positivas (Fischer, 2000), donde los ácidos

teicoicos y lipoteicoicos son entidades estructural y biosintéticamente distintas. Mediante el empleo de técnicas inmunocitoquímicas se comprobó que los TA se localizan en las caras externas e internas de paredes celulares aisladas (Sørensen *et al.*, 1988), mientras que los LTA lo hacen sobre la membrana citoplasmática, uniéndose a ella a través de un glicolípido. Su síntesis se inicia en el citoplasma bacteriano por distintas rutas metabólicas (Bracha *et al.*, 1978; Glaser & Lindsay, 1974) y, una vez sintetizados, son translocados a través de la membrana citoplasmática (Lazarevic & Karamata, 1995; Fischer, 1993), aunque la incorporación al peptidoglicano insoluble tiene lugar durante la formación de la pared celular.

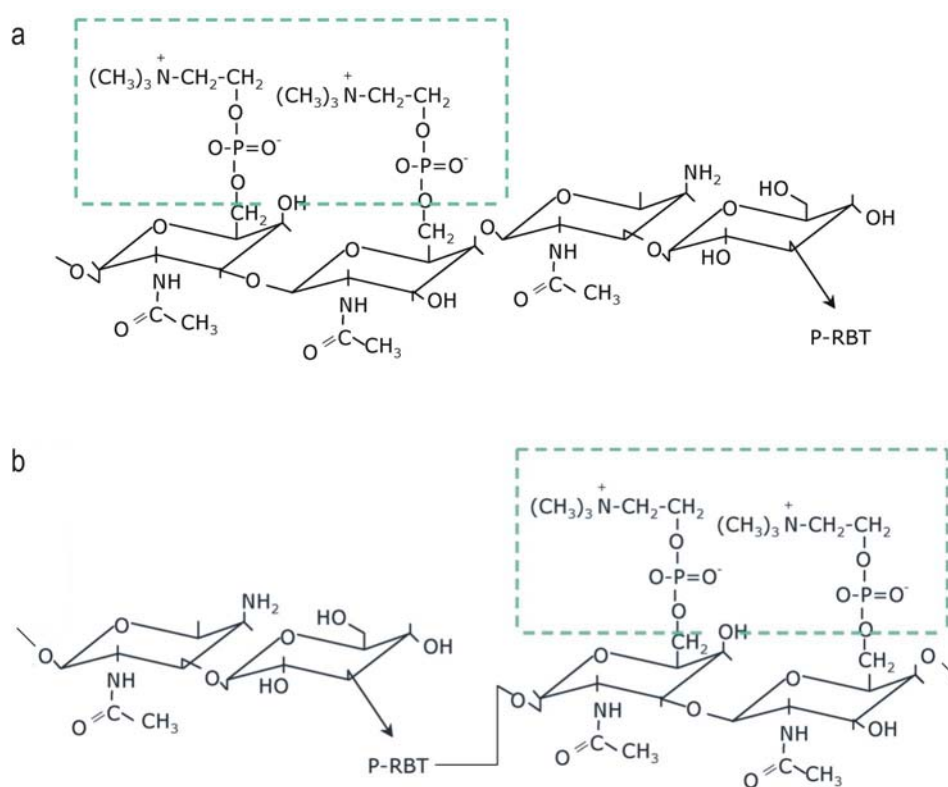


Figura 1.1.7: Estructura del pentasacárido que forma la unidad repetitiva básica del ácido teicoico. Se encuentra constituido por: P-RBT, ribitol fosfato; dos residuos de N-acetilgalactosamina (GalNAC) sustituidos en posición 6 por fosforilcolina (recuadro verde); diacetoamidotrioxogalactosa (AATGal), y glucosa (Glu). (a) Modelo aceptado. (b) Modelo propuesto recientemente.

El modelo actualmente aceptado de ácido teicoico es un polímero del pentasacárido [D-ribitol ($\beta 1 \rightarrow 1$) D-N-acetilgalactopiranososa ($\alpha 3 \rightarrow 1$) D-N-acetilgalactopiranososa ($\alpha 4 \rightarrow 1$) diacetamidotrioxogalactopiranososa ($\beta 3 \rightarrow 1$) α -D-glucopiranososa] (Figura 1.1.7a), en los que el

residuo de α -D-glucopiranososa de cada repetición se une, en posición 3, con el ribitolfosfato del siguiente pentasacárido a través de un enlace fosfodiéster (Fischer *et al.*, 1993). Sin embargo, este modelo (Figura 1.1.7a), no explica las propiedades del antígeno de Forssman del LTA y predice una masa molecular para LTA superior a la observada actualmente. Recientemente, se ha propuesto un nuevo modelo (Figura 1.1.7b) que solventa dichas inconsistencias (Seong Seo *et al.*, 2008). Los dos residuos de N-acetilgalactopiranososa, o sólo el unido al ribitol fosfato, pueden estar sustituidos en posición 6 por fosforilcolina (PC) (Karlsson *et al.*, 1999) (Figura 1.1.5). Las cadenas de teicoico contienen, como promedio, de 6 a 8 repeticiones del pentasacárido (Yother *et al.*, 1998) y constituyen entre un 40% y un 50% del peso seco de la pared. El contenido medio es de una cadena de TA por cada unidad de tetradisacárido-tetrapéptido (García Bustos *et al.*, 1987) lo cual quiere decir que el 25% de los murámicos están sustituidos. La distribución uniforme de los TA y LTA en la pared celular (Umeda *et al.*, 1992) le confiere carga negativa debido a sus numerosos grupos fosfato, capaces de formar quelatos con magnesio y otros cationes divalentes, lo que aumenta la rigidez del exoesqueleto.

La unión de los TA a la pared celular se establece a través de un conector formado por glicerol fosfato, ribitol fosfato o glucosil-fosfato que puede estar glicosilado o esterificado (su composición en neumococo no ha sido todavía determinada). Este conector se une, a su vez, a la cadena glicánica (Figuras 1.1.5 y 1.1.6) a través de un disacárido que establece un enlace fosfodiéster con la posición 6 del NAM (Behr *et al.*, 1992). Los ácidos lipoteicoicos se unen covalentemente a un glicolípido (D-Gal (α 1 \rightarrow 2)-D-Glu (α 1 \rightarrow 3)-*di*-O-acilglicerol) insertado en la cara externa de la membrana citoplasmática, y al intercalarse a través de la pared celular acceden hasta la superficie de la bacteria, donde pueden actuar como antígenos (Fischer, 2000).

Como ya se ha mencionado, la presencia de colina como componente estructural de la pared celular es una característica específica de neumococo (Tomasz, 1967). El 85% de la colina macromolecular presente en la célula se encuentra formando parte de los TA (Brundish & Baddiley, 1968), por lo que el 15% restante está unido a los LTA. El porcentaje de residuos de N-acetilgalactosamina monosustituidos por PC es de un 83%, correspondiendo el resto a la doble sustitución, aunque estos valores pueden variar de unas cepas a otras (Fischer *et al.*, 1993). Recientemente, se ha detectado la presencia de colina como componente de estructuras

periféricas en otros patógenos del tracto respiratorio como *Streptococcus mitis*, *Haemophilus influenzae*, *Neisseria meningitidis*, *Streptococcus oralis*, *Streptococcus constellatus*, *Clostridium beijerinckii* y *Clostridium NI-4* (Gosink *et al.*, 2000).

La colina en sí misma tiene un papel importante en la patogenicidad de neumococo actuando como adhesina. Participa directamente en la internalización de la bacteria, interaccionando con los receptores celulares del factor de activación plaquetaria (rPAF) expuestos en la superficie de las células epiteliales activadas (Cundell *et al.*, 1995). Además, durante el proceso de inflamación la colina es reconocida por la proteína C-reactiva humana (CRP) (Mold *et al.*, 1982), implicada en la respuesta inmune del hospedador (Szalai *et al.*, 1996; Hirschfield & Pepys, 2003). Por otra parte, los residuos de colina actúan como receptores para toda una familia de proteínas codificadas por la bacteria, denominadas CBPs (del inglés 'choline binding proteins'), entre las que se encuentran enzimas que hidrolizan la pared celular (mureín hidrolasas) y otros factores implicados en la infectividad de la bacteria (Höltje & Tomasz, 1975b; Rosenow *et al.*, 1997; Hammerschmidt *et al.*, 1999; Jedrzejewski, 2001).

Además de intervenir en el control de la actividad de las CBPs, tanto los TA como los LTA forman parte de los receptores de algunos bacteriófagos, y activan la ruta alternativa del complemento (Höltje & Tomasz, 1975a; Horne & Tomasz, 1985; Bierbaum & Shal, 1985; Rogers & Thurman, 1985; Tsuchido *et al.*, 1985; Fischer, 1988). Además, son altamente antigénicos, actúan como agentes quimiotácticos en la inflamación meníngea (Tuomanen *et al.*, 1985b) y estimulan la formación de interleucina 1 en los monocitos humanos (Riesenfeld-Orn *et al.*, 1989).

1.1.4. PROTEÍNAS DE UNIÓN A COLINA (CBPs).

La secuencia genómica de *S. pneumoniae* reveló la existencia de 3 grupos de proteínas de superficie: la familia de las proteínas de unión a colina (CBPs), las lipoproteínas y las proteínas con un motivo LPxTG (Leu-Pro-X-Thr-Gly) que se unen covalentemente a la pared celular tras la acción de una sortasa. Muchas de estas proteínas de superficie son factores de virulencia que contribuyen a la patogénesis de este organismo. Además, estas proteínas participan en interacciones específicas con los tejidos del hospedador humano, facilitando la supervivencia bacteriana, ayudando a la expansión de la bacteria dentro de los tejidos del hospedador, y

ocultando la superficie bacteriana a los mecanismos de defensa del hospedador. Debido a estas propiedades, numerosas proteínas expuestas en la superficie son dianas potenciales para el diseño de agentes profilácticos tales como vacunas y algunos de ellos podrían servir como dianas para el diseño de drogas terapéuticas.

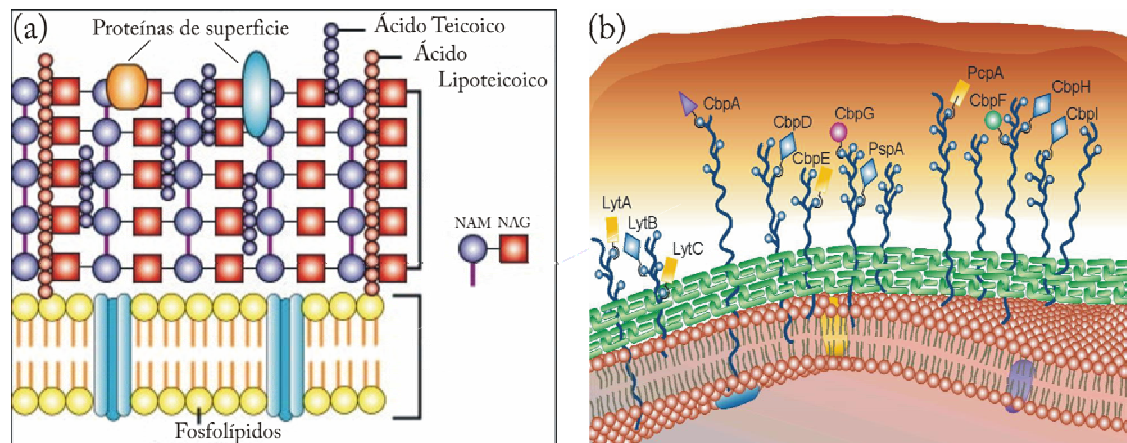


Figura 1.1.8: Proteínas de superficie. (a) Localización celular de las proteínas de superficie. (b) Proteínas de unión a colina como subgrupo de las proteínas de superficie.

S. pneumoniae sintetiza, dependiendo de la cepa, entre 12 y 15 CBPs (Hoskins *et al.*, 2001, Tettelin *et al.*, 2001), las cuales son cruciales en patogénesis y también como antígenos candidatos para desarrollar vacunas conjugadas. Las CBPs comparten una organización modular (Figura 1.1.9) tal que existen uno o varios módulos responsable/s de la función biológica y un módulo de unión a colina (CBM) (López & García, 2004), consistente éste en un tándem de entre 6 a 18 repeticiones de 20 aa cada una (García *et al.*, 2000; García *et al.*, 1999a & 1999b). Todas las CBPs, excepto LytB y LytC, tienen el CBM en la posición C-terminal. En la Tabla 1.1.1 se muestran las distintas CBPs identificadas en las dos cepas de neumococo cuyo genoma ha sido secuenciado completamente: TIGR4 (Tettelin *et al.*, 2001) y R6 (Hoskins *et al.*, 2001). Entre ellas se encuentran las mureína hidrolasas* codificadas por neumococo (LytA, LytB, LytC, Pce), diversos factores asociados con adhesión y virulencia (PspA y PspC), y otras proteínas con función desconocida (PcpA, PcpB, PcpC, CbpF). Además de las funciones específicas desarrolladas por cada una de ellas, las CBPs podrían ser importantes en el establecimiento de interacciones de contacto no específicas con las células del hospedador, y

* Proteínas que hidrolizan de forma específica los enlaces del peptidoglicano de la pared bacteriana

estabilizar electrostáticamente la cápsula por la carga neta positiva que aportan a la superficie celular (Swiatlo *et al.*, 2002).

La presencia de colina en la pared celular parece haber condicionado también la estructura de las enzimas líticas (endolisinas) codificadas por los bacteriófagos de neumococo que facilitan la salida de progenie fágica al final del ciclo vegetativo (López *et al.*, 1982). Todos los fagos de neumococo, con la excepción de Cpl-7, codifican para una endolisina constituida por un módulo catalítico en posición N-terminal y un módulo de unión a colina en el extremo C-terminal, altamente relacionado en secuencia con los de las mureín hidrolasas codificadas por la propia bacteria (García *et al.*, 1990). De hecho, las primeras CBPs identificadas fueron LytA, la autolisina mayoritaria de neumococo y Cpl-1, la lisozima codificada por el fago Cp-1 (García *et al.*, 1987a; García *et al.*, 1987b).

Organización Modular de las CBPs

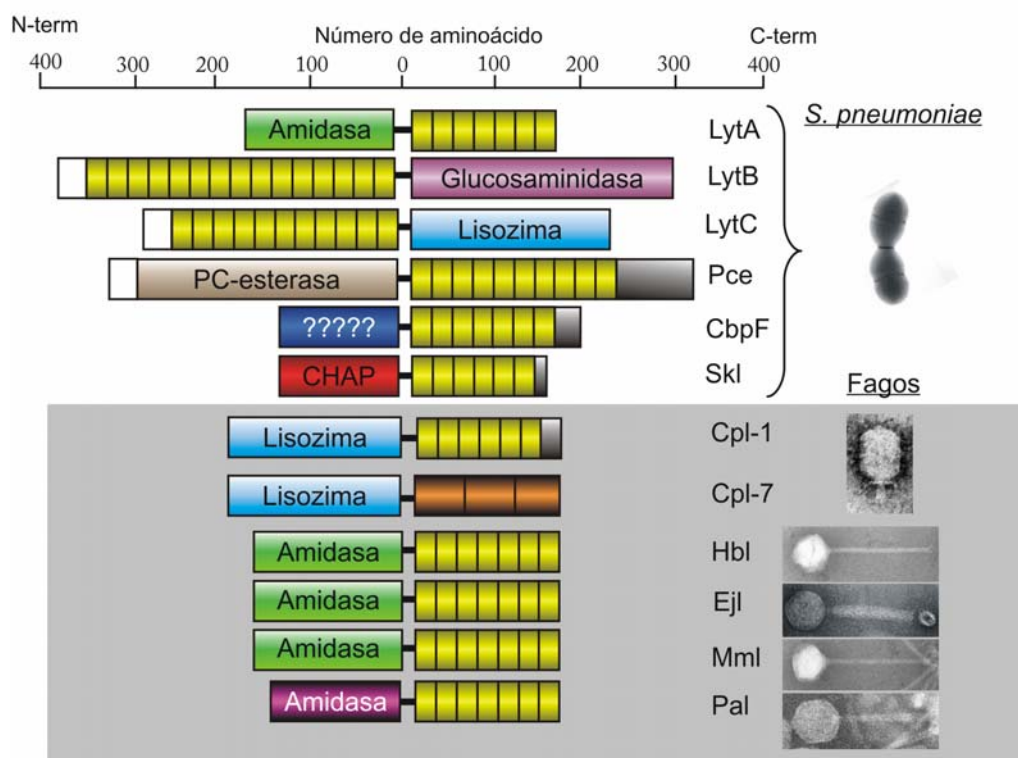


Figura 1.1.9: Esquema de la organización modular de las proteínas de unión a colina. Los dominios funcionales se identifican según su función; en verde (amidasa), rojo (amidasa de la familia CHAP), morado (glucosaminidasa), azul claro (lisozimas), azul oscuro (función desconocida), fucsia (amidasa de Pal) y marrón (fosforicolín esterasa). Las repeticiones de los CBMs se muestran en amarillo y Cpl-7 en naranja.

Introducción

TIGR4	R6	Función	Referencia
CbpI	--	n.d.	(Tettelin <i>et al.</i> , 2001)
PspA	PspA	Proteína de superficie. Inhibe la ruta del complemento y une la lactoferrina humana.	(Hammerschmidt <i>et al.</i> , 1999; Tu <i>et al.</i> , 1999)
CbpC	CbpF	n.d.	(Tettelin <i>et al.</i> , 2001; Hoskins <i>et al.</i> , 2001)
CbpJ	--	n.d.	(Tettelin <i>et al.</i> , 2001)
CbpG	CbpG truncada ¹	Serín-proteasa (CbpG-truncada); adhesina (CbpG)	(Tettelin <i>et al.</i> , 2001; Hoskins <i>et al.</i> , 2001 ; Gosink <i>et al.</i> , 2000 ; Mann <i>et al.</i> , 2006)
CbpF	PcpC	n.d.	--
PSP ²	Proteína hipotética	n.d.	(Tettelin <i>et al.</i> , 2001; Hoskins <i>et al.</i> , 2001)
CbpE	Pce ³	Hidrólisis de fosforilcolina de la pared celular (ver apartado 1.3.1.5)	(Höltje y Tomasz, 1974; de las Rivas <i>et al.</i> , 2001)
Glucosaminidasa	LytB	Separación de células hijas (ver apartado 1.3.1.2)	(de las Rivas <i>et al.</i> , 2002; García <i>et al.</i> , 1999a)
CbpA	PspC ⁴	Adhesina. Une la inmunoglobulina A y la proteína C3 del complemento	(Briles <i>et al.</i> , 1997; Hammerschmidt <i>et al.</i> , 1997; Rosenow <i>et al.</i> , 1997; Smith y Hostetter, 2000)
PspC, degenerada	SpsA truncada	n.d.	--
LytC	LytC	Lisozima. Autólisis a 30 °C (ver apartado 1.3.1.3)	(García <i>et al.</i> , 1999b)
Autolisina	LytA	Amidasa (ver apartado 1.3.1.1)	(García <i>et al.</i> , 1985; Höltje y Tomasz, 1976)
PcpA	PcpA	Adhesina	(Sánchez-Beato <i>et al.</i> , 1998)
CbpD	CbpD	Posible amidasa (ver apartado 1.3.1.4)	(Tettelin <i>et al.</i> , 2001; Hoskins <i>et al.</i> , 2001; Gosink <i>et al.</i> , 2000)

¹Nombre alternativo PcpB; ²*Pneumococcal surface protein*; ³ Nombre alternativo LytD; ⁴ Nombre alternativo SpsA y PbcA; n.d, no determinada;

Tabla 1.1.1: Proteínas de unión a colina (CBPs) identificadas en las cepas TIGR4 y R6 de *S. pneumoniae*

El motivo de unión a colina, inicialmente descrito en neumococo, ha sido identificado posteriormente en proteínas de otros microorganismos que contienen colina en su pared celular y en las de los bacteriófagos que las infectan (Sánchez-Beato, 1995a; Sánchez-Beato, 1995b; Katayama *et al.*, 1995). Por otro lado, se han encontrado secuencias relacionadas en glicosiltransferasas y proteínas de unión a carbohidratos de diversas especies de *Leuconostoc* y *Streptococcus* (Janecek *et al.*, 2000), y en toxinas de *Clostridium* (García *et al.*, 1988a; Wren,

B.W., 1991; Green *et al.*, 1995; Greco *et al.*, 2006) (<http://www.Sanger.ac.uk/Software/Pfam/número de acceso PF01473>).

1.1.5. PAPEL DE LA ENVUELTA CELULAR Y DE LAS PROTEÍNAS DE SUPERFICIE EN LA VIRULENCIA DE NEUMOCOCO (CBPs).

Como ya se ha mencionado, *S. pneumoniae* es un colonizador pasivo habitual de la nasofaringe humana y uno de los principales patógenos humanos (Musher, 1992). El mecanismo mediante el cual se produce la invasión hacia el tracto respiratorio inferior y el pulmón, o hacia el cerebro no se ha elucidado completamente (Novak & Tuomanen, 1999).

Entre los factores de virulencia más importantes se encuentran, además de la cápsula y de los componentes de la pared celular (secciones 1.1.2 y 1.1.3) diversas proteínas de la superficie celular, incluidas algunas CBPs, y la neumolisina (Mitchell, 2003; Figura 1.1.10).

Las proteínas de la familia de las CBPs consideradas como factores de virulencia son CbpA, PspA, CbpG y LytA (Tabla 1.1.1). CbpA fue la primera adhesina identificada en la superficie y está implicada en los procesos de adherencia y colonización de la nasofaringe a través de su interacción con citoquinas en células humanas activadas (Rosenow *et al.*, 1997). Además, se une a la IgA humana y a sus componentes de secreción (Hammerschmidt *et al.*, 1997), así como a la proteína del complemento C3. Por otro lado, PspA (proteína A de la superficie de neumococo) inhibe y reduce la efectividad del sistema del complemento del hospedador, además de prevenir la fagocitosis de neumococo (Neeleman *et al.*, 1999; Tu *et al.*, 1999). CbpG (Tabla 1.1.1) ha sido identificada como un determinante de la virulencia, operativo tanto en adhesión, durante la colonización, como en la sepsis (Gosink *et al.*, 2000), siendo junto con PspA las únicas CBPs conocidas que desempeñan un papel dominante en el desarrollo de infecciones sistémicas. Recientemente, se ha descrito una forma truncada de CbpG (Tabla 1.1.1) que carece del CBM (módulo de unión a colina) y presenta actividad proteásica '*in vitro*' cuando es secretada (Mann *et al.*, 2006). La autolisina LytA (sección 1.3.1.1), aunque de manera indirecta, coopera en la patogenicidad de neumococo al liberar componentes de la pared celular y otros factores de virulencia tan importantes como la neumolisina y la neuraminidasa (Berry *et al.*, 1989).

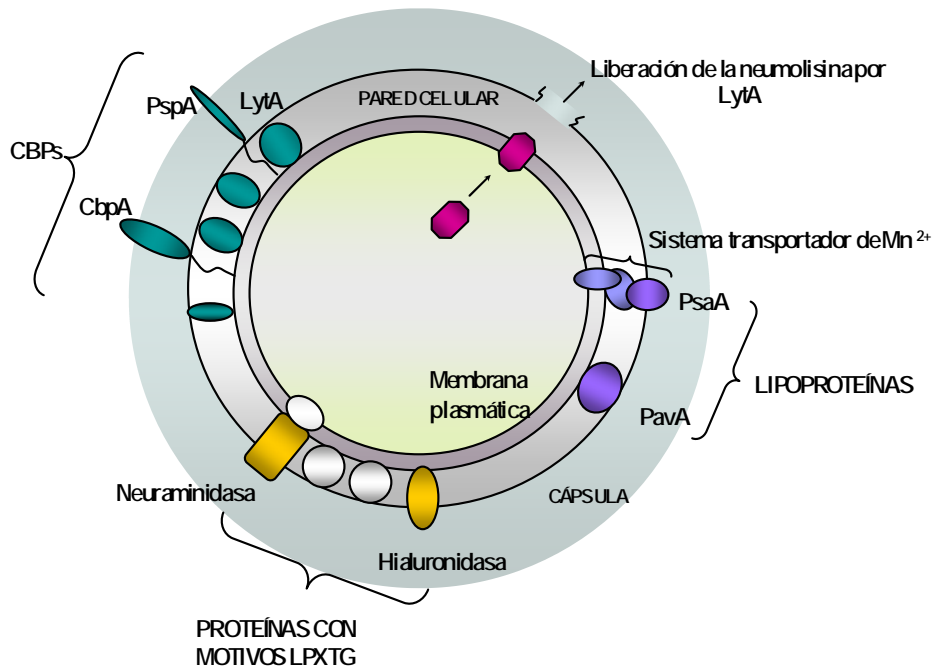


Figura 1.1.10: Esquema representativo de los principales factores de virulencia de *S. pneumoniae* (modificada de Mitchell, 2003).

Por otra parte, y aunque no se han descrito como tales factores de virulencia, CbpD, LytB, LytC, Pce y CbpG están implicadas en la colonización de la nasofaringe. A Pce y CbpG se les atribuye, también, un papel en el proceso de adherencia (Gosink *et al.*, 2000), mientras que CbpD asiste a LytA en la lisis celular (Kausmally *et al.*, 2005). LytC, al igual que LytA, tiene un efecto directo en la patogénesis derivado de su función lítica, que se sugiere localizada en la nasofaringe. En cuanto a LytB, la formación de cadenas de neumococo parece limitar la diseminación de la bacteria durante el proceso de infección, por lo que el interés hacia dicha enzima aumenta debido a su implicación en la separación de las células hijas tras la división celular (De las Rivas *et al.*, 2002).

A diferencia de las CBPs, la hialuronidasa y la neuraminidasa se encuentran unidas covalentemente a la pared celular mediante un motivo LPXTG localizado en posición C-terminal y su actividad parece relacionada con la degradación de la matriz extracelular y la exposición de posibles receptores celulares (Mitchell, 2003). Otro grupo de proteínas con el mismo tipo de

anclaje a la pared celular son las metaloproteasas IgA1, ZmpB, ZmpC, ZmpD y PrtA, cuya contribución a la virulencia se ha caracterizado recientemente mediante experimentos de infección en modelos animales (Oggioni *et al.*, 2003; Blue *et al.*, 2003; Chiavolini *et al.*, 2003; Bethe *et al.*, 2001).

Otras proteínas de superficie consideradas como factores de virulencia son las lipoproteínas PsaA y PavA unidas covalentemente a la membrana celular, cuya actividad está relacionada con los fenómenos de adhesión de neumococo a las células humanas (Tseng *et al.*, 2002; Holmes *et al.*, 2001).

Por último, la neumolisina, de origen citoplasmático, es la principal toxina codificada por neumococo y promueve la colonización reduciendo la capacidad de eliminación del 'mucus' (Jedrzejewski, 2001). Otras de sus acciones más relevantes son la formación de poros (Tilley *et al.*, 2005), la activación de la vía del complemento (Rubins *et al.*, 1996), la estimulación de la producción de mediadores inflamatorios (Houldsworth *et al.*, 1994), la inhibición fagocítica y leucocítica (Paton & Ferrante, 1983) y la inducción de apoptosis y daño en células cerebrales (Braun *et al.*, 2002).

1.1.6. ESTRUCTURAS TRIDIMENSIONALES DE CBPs DE NEUMOCOCO O FAGOS NEUMOCÓCICOS.

De todas las CBPs identificadas hasta el momento sólo se han caracterizado las estructuras tridimensionales de un fragmento del CBM de la autolisina LytA (Fernández-Tornero *et al.*, 2001), las estructuras completas de la lisozima del fago Cp1, Cpl-1, (Hermoso *et al.*, 2003) y de la fosforilcolinesterasa de neumococo, Pce, (Hermoso *et al.*, 2005), y el dominio R2 de CbpA (Luo *et al.*, 2005). Esta escasez de CBPs caracterizadas es debida principalmente, a la dificultad que conlleva la cristalización de proteínas modulares.

1.1.6.1. C-LytA

El CBM aislado de la amidasa LytA (C-LytA) fue la primera estructura resuelta de un módulo de unión a colina (Fernández-Tornero *et al.*, 2001 y 2002). La forma cristalizada corresponde al dímero obtenido en presencia de colina (Usobiaga *et al.*, 1996).

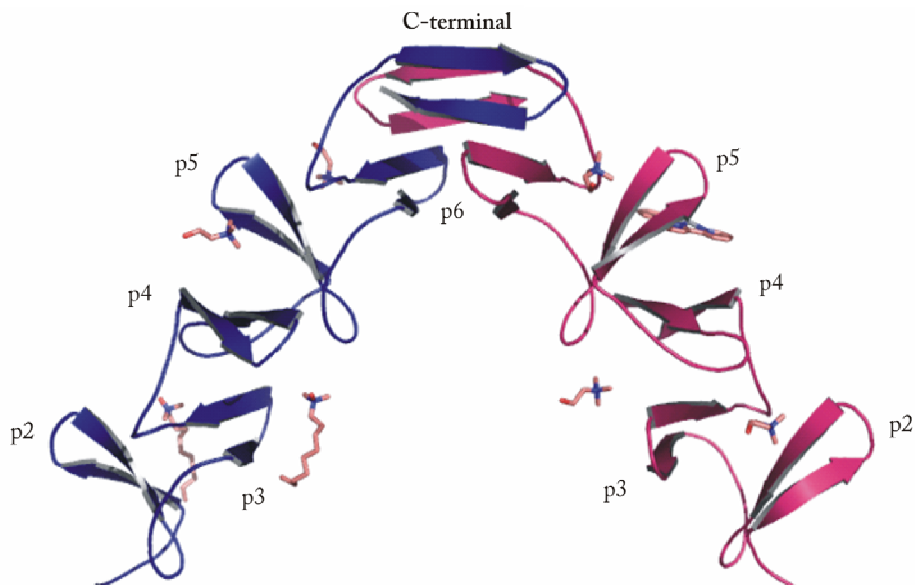


Figura 1.1.11: Estructura cristalográfica del dímero de C-LytA. Cada monómero aparece con distinto color y junto a cada horquilla se indica el motivo al que pertenece (datos tomados de Fernández-Tornero *et al.*, 2001). Los dos primeros sitios del monómero azul se encuentran ocupados por moléculas de DDAO (N-óxido de N,N-dimetil-decylamina), utilizadas en la cristalización. El sitio 3 del monómero rosa está ocupado por el derivado de platino utilizado para la resolución de la estructura (cloruro de 2,2':6',2"-terpiridina-platino II). En el resto de los sitios de ambos monómeros se encuentra situada la molécula de colina.

Las repeticiones presentes en la estructura (denominadas p2-p6) y el extremo C-terminal se pliegan formando una horquilla β seguida de un bucle que contiene un giro y una zona en conformación extendida. Las horquillas β están constituidas por unos 5 aminoácidos por hebra y se disponen formando una superhélice a izquierdas (solenoides β) de unos 25 Å de diámetro y 60 Å de altura, en la que cada horquilla está girada 120° con respecto a la anterior, de tal forma que las horquillas β_i y β_{i+3} adoptan una disposición paralela. A diferencia de otros β -solenoides (Kobe & Kajava, 2000), la estructura de C-lytA está formada por unidades individuales de estructura supersecundaria (horquillas β).

Cada monómero de la estructura cristalizada posee cuatro sitios de unión a colina, localizados en las interfases hidrofóbicas creadas entre dos horquillas β consecutivas (Fernández-Tornero *et al.*, 2001). Cada uno de los sitios de unión a colina está conformado por tres residuos aromáticos, situados en las horquillas β , y un aminoácido hidrofóbico localizado en el bucle que

las une. Las interacciones que se establecen en la cavidad son, por un lado, de tipo hidrofóbico con los grupos metilo de la colina, y por otro de tipo catión- π entre los anillos ricos en electrones de los aminoácidos aromáticos y la carga positiva de la colina.

Los monómeros se disponen formando un ángulo de unos 85° e interactúan a través de la región C-terminal del módulo de unión a colina (Figura 1.1.11), tal y como se había sugerido previamente (Varea *et al.*, 2000). La interfase está formada por las horquillas de p6 y la cola C-terminal de cada monómero, encontrándose ligeramente giradas para favorecer las interacciones con la otra subunidad. La molécula de colina no participa directamente en la unión entre monómeros, aunque potencia la dimerización (Varea *et al.*, 2000).

1.1.6.2. Cpl-1.

La estructura completa de la lisozima Cpl-1, la primera publicada de una CBP completa, ilustra claramente la naturaleza modular de las CBPs (Figura 1.1.12; Hermoso *et al.*, 2003). La cadena polipeptídica está formada por el módulo catalítico y el módulo de unión a colina, unidos mediante un conector ácido constituido por 11 residuos. El módulo catalítico está formado por un único dominio estructural que se pliega formando un barril α/β de estructura irregular (característico de la familia GH25 de las glicosilhidrolasas), al que le faltan las tres últimas hélices α y en el que la hebra $\beta 8$ adopta una disposición antiparalela. Las repeticiones del CBM se organizan en dos regiones estructurales bien definidas, denominadas CI y CII (Figura 1.1.12). CI está formada por las cuatro primeras repeticiones (p1-p4) que se disponen formando una superhélice a izquierdas con las horquillas extendidas perpendicularmente al eje de la superhélice, de modo análogo a lo que ocurre en C-LytA (Figura 1.1.11). Cada una de las repeticiones pertenecientes a este subdominio se encuentra localizada a 120° en sentido contrario a las agujas del reloj. Por otro lado, el dominio CII está constituido por una hoja β antiparalela de seis hebras que incluye las horquillas de las repeticiones p5 y p6 y dos hebras de la cola C-terminal. Esta región está directamente implicada en las interacciones existentes entre los dos módulos que forman Cpl-1.

La posición relativa de ambos módulos viene determinada por la estructura extendida del conector que los une, constituido mayoritariamente por residuos ácidos, y por las fuertes

interacciones hidrofóbicas que se establecen entre las hebras $\beta 6$ - $\beta 8$ del módulo catalítico y la región C-terminal del CBM (Hermoso *et al.*, 2003).

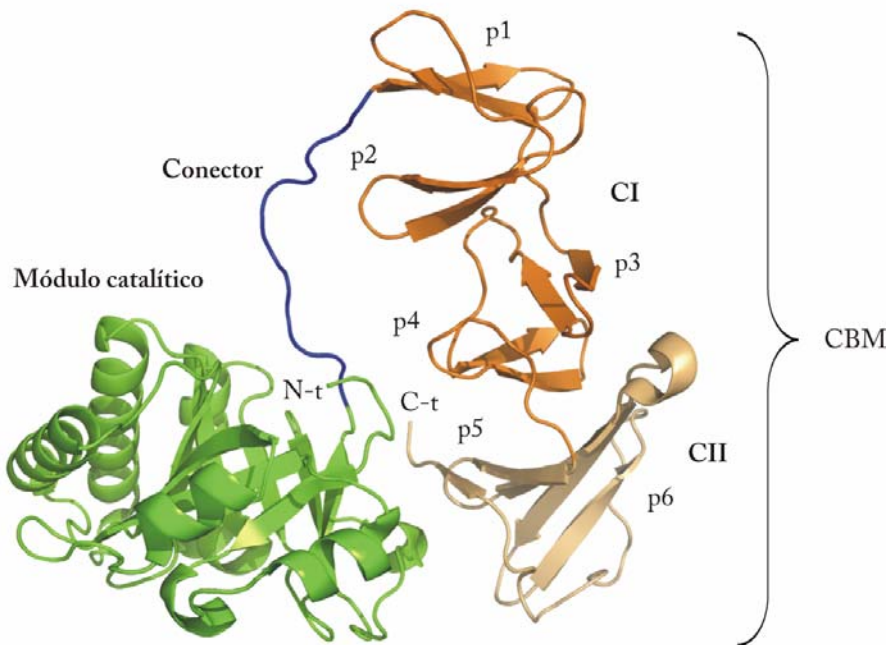


Figura 1.1.12: Estructura de Cpl-1. El código de colores es el siguiente: módulo catalítico en verde, conector en azul y CBM en naranja. Se indican, además, las repeticiones del CBM y los dos dominios (CI, naranja oscuro; y CII, naranja claro) que lo forman.

A pesar de que todas las repeticiones del CBM de Cpl-1 presentan el mismo tipo de estructura secundaria y de que los residuos implicados en la unión a colina están absolutamente conservados, la desviación del plegamiento global de CII con respecto a la superhélice a izquierdas hace que sólo dos de los cinco posibles sitios de unión a colina de Cpl-1 sean funcionales (Hermoso *et al.*, 2003). Este hecho pone de manifiesto la versatilidad de las repeticiones que forman los CBMs y su capacidad para actuar como receptores de colina o para mediar en las interacciones entre dominios y en la auto-asociación de las CBPs.

1.1.6.3. Pce.

La estructura de Pce (Hermoso *et al.*, 2005), la primera CBP resuelta de las codificadas por la bacteria, muestra una disposición modular (Figura 1.1.13), donde el módulo catalítico (azul) se encuentra unido al módulo de unión a colina (amarillo) por un conector formado por 12 residuos

(naranja). El extremo C-terminal está formado por 86 aminoácidos (residuos 517-602), de los que desconocemos el plegamiento y la estructura de los últimos 55 residuos.

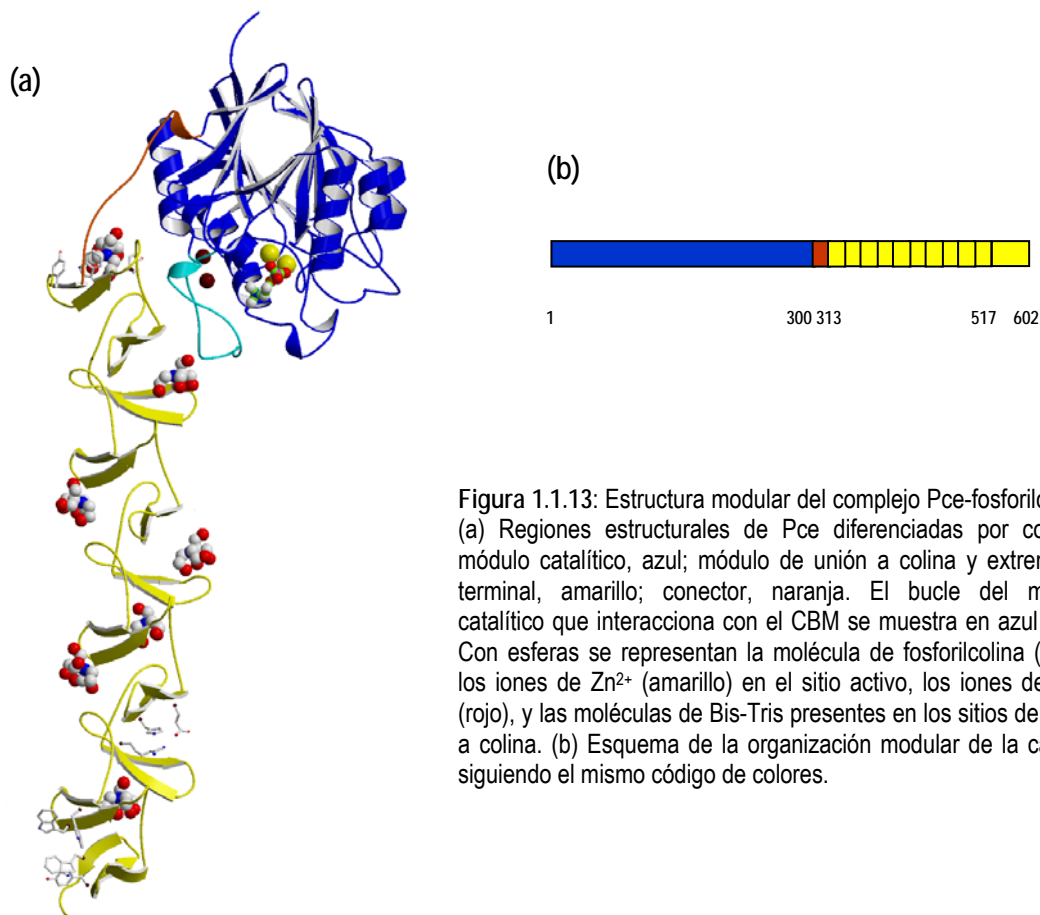


Figura 1.1.13: Estructura modular del complejo Pce-fosforilcolina. (a) Regiones estructurales de Pce diferenciadas por colores: módulo catalítico, azul; módulo de unión a colina y extremo C-terminal, amarillo; conector, naranja. El bucle del módulo catalítico que interacciona con el CBM se muestra en azul claro. Con esferas se representan la molécula de fosforilcolina (PC) y los iones de Zn^{2+} (amarillo) en el sitio activo, los iones de Ca^{2+} (rojo), y las moléculas de Bis-Tris presentes en los sitios de unión a colina. (b) Esquema de la organización modular de la cadena siguiendo el mismo código de colores.

El módulo catalítico de Pce está formado por un único dominio estructural constituido por un sándwich $\alpha\beta\text{-}\beta\alpha$ (Figura 1.1.14). Esta es la primera fosfodiesterasa resuelta cuya estructura presenta el plegamiento característico de la familia de las metalo- β -lactamasas, a pesar de tener una baja identidad de secuencia ($\leq 13\%$) con otras enzimas de esta familia y de poseer un tamaño considerablemente mayor. Al igual que ocurre con la secuencia, la similitud estructural entre Pce y otras α/β hidrolasas con este tipo de plegamiento es baja. Las dimensiones aproximadas del módulo catalítico son $55 \times 53 \times 45 \text{ \AA}^3$, dividiéndose en dos regiones equivalentes constituidas por láminas- β antiparalelas con 5 y 4 hebras β , respectivamente, seguidas de 3 motivos $\alpha\beta$ que enlazan con los conectores (Figura 1.1.14b). Las hélices α se exponen al solvente rodeando el núcleo compacto formado por las hebras β .

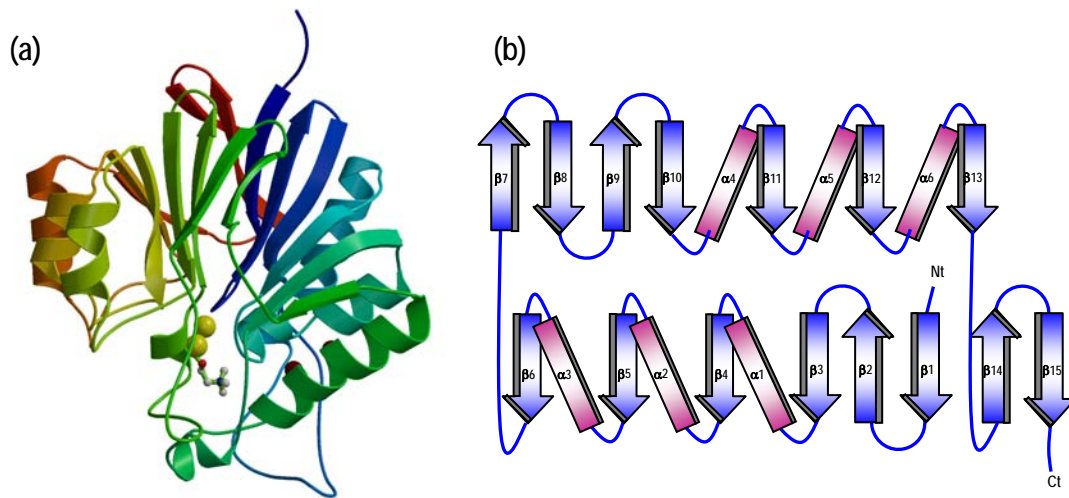


Figura 1.1.14: Estructura y topología del módulo catalítico de Pce. (a) Estructura del módulo catalítico. Los elementos de estructura secundaria se colorean desde el azul al rojo comenzando por el extremo N-terminal. Los iones de Zn^{2+} , Ca^{2+} , y la molécula de fosforilcolina se representan igual que en la figura 1.1.2. (b) Topología del módulo catalítico de Pce dividido según la disposición del sandwich, con las láminas β coloreadas en azul y las hélices α en rosa. Las hélices 3_{10} se han omitido por claridad.

El conector que une ambos módulos está formado por una región con varios residuos polares plegados en una hélice 3_{10} , seguida de una zona constituida por residuos aromáticos y prolinas que no presenta elementos de estructura secundaria.

El módulo de unión a colina está formado por 10 repeticiones (p1-p10) de 20 aminoácidos aproximadamente, con una secuencia consenso similar a la encontrada en otras CBPs (GWXK-X₄₋₅-WYY-hidrofóbico-X₃₋₅-GXMX₂₋₃). Cada repetición contiene una horquilla β seguida de un bucle que forma un giro (Figura 1.1.15a). Esquemáticamente, la forma del módulo puede asemejarse a un prisma triangular regular de 110 Å de altura y 26 Å de base, con los sitios de unión a colina situados a lo largo de las tres caras laterales del prisma (Figura 1.1.15b). El patrón de plegamiento global es el de una superhélice a izquierdas semejante a la descrita en C-LytA y el dominio CI del CBM de Cpl-1 (Fernández-Tornero *et al.*, 2001; Hermoso *et al.*, 2003). Las horquillas que la forman se disponen perpendicularmente al eje de la hélice, y cada repetición presenta un giro de 120° con respecto a la anterior, de tal forma que la horquilla β_i es paralela a la horquilla β_{i+3} (Figura 1.1.15b)

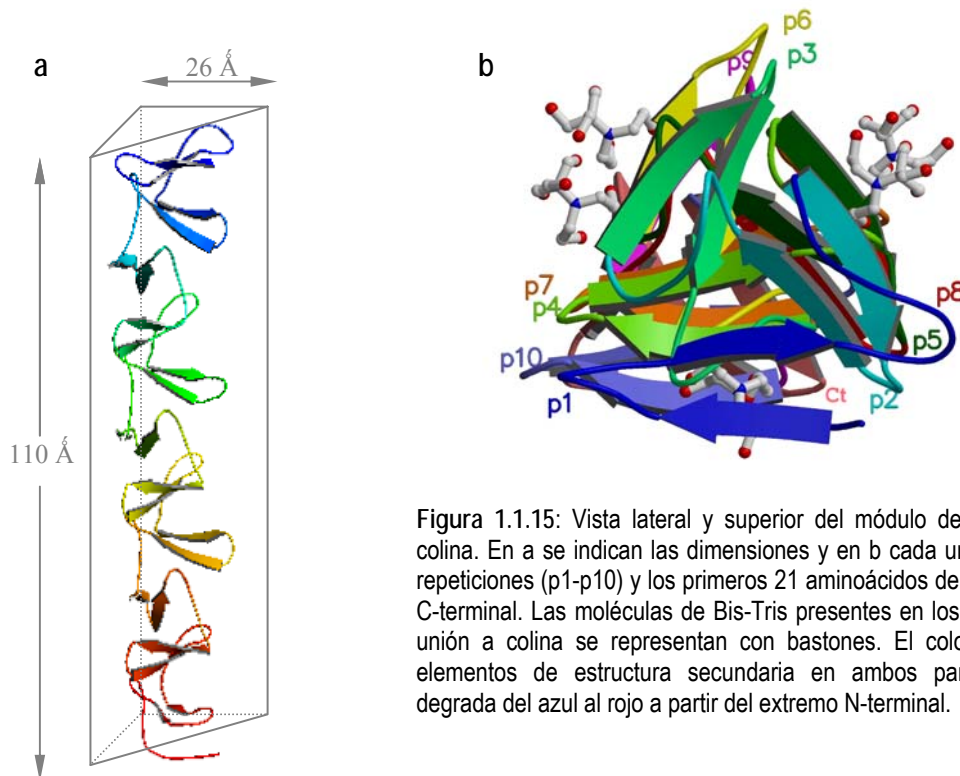


Figura 1.1.15: Vista lateral y superior del módulo de unión a colina. En a se indican las dimensiones y en b cada una de las repeticiones (p1-p10) y los primeros 21 aminoácidos del extremo C-terminal. Las moléculas de Bis-Tris presentes en los sitios de unión a colina se representan con bastones. El color de los elementos de estructura secundaria en ambos paneles se degrada del azul al rojo a partir del extremo N-terminal.

Curiosamente, los primeros 21 aminoácidos (residuos 517-534) del extremo C-terminal, a pesar de no tener identidad de secuencia con las repeticiones, adoptan el mismo tipo plegamiento, alargando la superhélice que forma el CBM (Figura 1.1.15a).

La elucidación tridimensional completa de la primera CBP de neumococo permitió, al igual que en el caso de la determinación estructural de Cpl-1 como primera endolisina completa, no sólo la confirmación de un plegamiento común del CBM, sino también la evaluación de las interacciones entre dominios de proteínas modulares. Además, a partir de la estructura 3D de Pce, se comprobó su implicación en la degradación del factor activador de plaquetas (PAF) y en la inhibición de la correcta respuesta del sistema inmune (Hermoso *et al.*, 2005).

1.1.6.4. Dominio de adhesión R2 de CbpA.

CbpA es una de las 15 proteínas CBPs identificadas en el genoma de la cepa TIGR4 de neumococo (Tettelin *et al.*, 2001). La resolución de la estructura completa de esta adhesina ha sido inviable hasta la fecha pero mediante técnicas de RMN se ha conseguido elucidar la estructura 3D de uno de sus dominios: el dominio R2 (Figura 1.1.16).

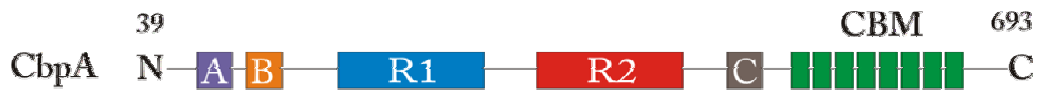


Figura 1.1.16: Organización modular de CbpA. Los dominios A, B, R1, R2 y C exhiben múltiples repeticiones del motivo cremallera de leucina mientras el dominio de unión a colina (CBM) representado en verde, contiene ocho repeticiones del motivo de unión a colina.

La estructura 3D del dominio R2 de CbpA (Luo *et al.*, 2005) está constituida por 3 α -hélices que adoptan un inusual triple α -hélice a través de interacciones antiparalelas α -hélice/ α -hélice (Figura 1.1.17a). Las α -hélices, aunque antiparalelas, no son exactamente coaxiales debido a un ligero desplazamiento (Figura 1.1.17b). La estructura del dominio R2 de CbpA reveló que una gran cantidad de residuos hidrofóbicos estaban escondidos entre las interfases α -hélice/ α -hélice y que en la superficie proteica aparecían abundantes residuos polares y cargados mostrando dos regiones con caracteres electrostáticos claramente diferenciados: una cara de elevado potencial electronegativo y otra cara electropositiva. Además entre las hélices 1 y 2 aparece el motivo YPT implicado en adhesión (Hammerschmidt *et al.*, 1997; Zhang *et al.*, 2000)

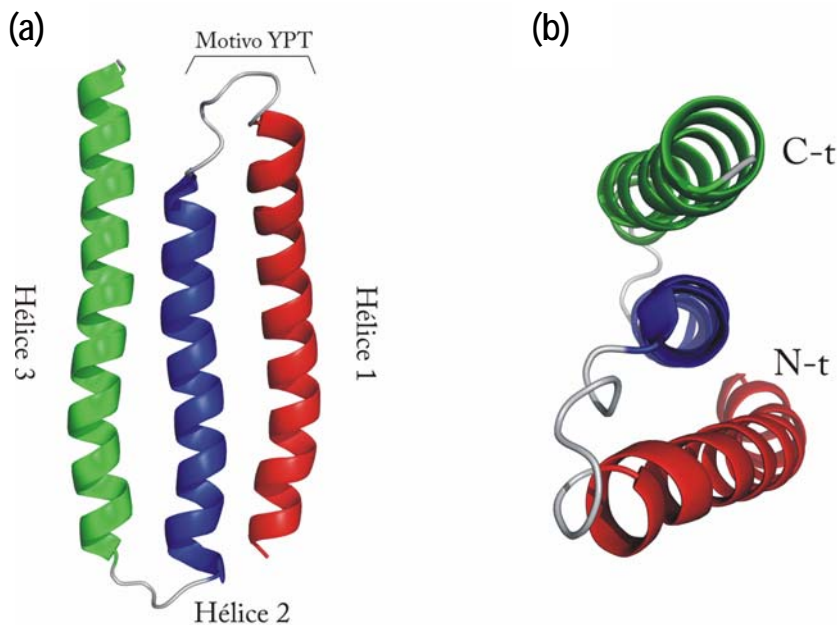


Figura 1.1.17: Estructura en solución del dominio R2 de CbpA. (a) Vista paralela al eje helicoidal de las α -hélice. (b) Vista perpendicular al eje helicoidal de las α -hélice.

1.2. OBTENCIÓN DE FASES EN LA CRISTALOGRAFÍA MACROMOLECULAR.

1.2.1. DETERMINACIÓN ESTRUCTURAL DE MACROMOLÉCULAS POR DIFRACCIÓN DE RAYOS X.

1.2.1.1. El problema de las fases en cristalografía de proteínas.

Resolver la estructura de una macromolécula por difracción de rayos X consiste, desde el punto de vista cristalográfico, en obtener una imagen de su densidad electrónica, con la cual construir un modelo a escala atómica. Para ello, el objetivo es determinar la función de densidad electrónica, $\rho(xyz)$, para cada punto de la celdilla unidad del cristal, partiendo de la información del espacio recíproco (espectro de difracción). La transformación del espacio recíproco al espacio cristalino (función de densidad electrónica) implica la resolución de una función, denominada transformada de Fourier (Ecuación 1.2.1). En ella intervienen los factores de estructura, $F(hkl)$, definidos en cada punto del espacio recíproco (hkl) , y que contienen la resultante de la dispersión de los rayos X, en esa dirección, que cooperativamente emiten los electrones de todos los átomos contenidos en la celdilla unidad (ver sección 3.5.1). La expresión analítica del factor de estructura se muestra en la ecuación 1.2.2, y en ella figura la magnitud f_j denominada factor de dispersión atómica, que da cuenta de la dispersión de los rayos X provocada por todos los electrones del átomo j .

$$\rho(xyz) = \frac{1}{V_c} \sum_h \sum_k \sum_l |F(hkl)| \cdot \cos 2\pi(hx + ky + lz - \alpha(hkl))$$

Ecuación 1.2.1: Función de densidad electrónica definida en cada punto de la celdilla unidad cristalina de coordenadas x, y, z . Esta ecuación representa la transformada de Fourier entre el espacio cristalino, definido por $\rho(xyz)$, y el espacio recíproco, definido por los factores de estructura. Es importante resaltar el hecho de que la densidad electrónica en un punto viene determinada por la información contenida en todos los factores de estructura, es decir, por la información de todo el espectro. Inversamente, el valor de un factor de estructura (módulo y fase) viene determinado por las posiciones de todos los átomos que forman la estructura, $\rho(xyz)$.

$$F(hkl) = \sum_j f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

Ecuación 1.2.2: El factor de estructura $F(hkl)$ es la resultante de la dispersión de todos los átomos de la celdilla. El factor f_j es el denominado factor de dispersión atómico y es la resultante de la dispersión de todos los electrones del átomo.

Los factores de estructura son ondas, y como tales se pueden describir por sus módulos, $|F(hkl)|$, y sus fases, $\alpha(hkl)$, respecto de un origen común de fases. Los módulos, $|F(hkl)|$, representan las amplitudes de los haces difractados y se obtienen directamente a partir de las intensidades que se recogen en el detector (la intensidad de una reflexión, $I(hkl)$, es proporcional al cuadrado de la amplitud, $|F(hkl)|$).

En resumen, a partir de las medidas del espectro de difracción se obtienen las intensidades $I(hkl)$. Estas intensidades nos proporcionan fácilmente las amplitudes $|F(hkl)|$ de los factores de estructura, pero, y aquí viene el problema, no sus fases $\alpha(hkl)$. Las fases son, pues, magnitudes que no pueden medirse experimentalmente, y que son necesarias para la determinación de la densidad electrónica en el cristal (ecuación 1.2.1). Por tanto es necesario recurrir a distintos métodos para poder resolver el denominado problema de la fase. Una vez obtenidas las fases asociadas a cada factor de estructura se podrá resolver de inmediato la ecuación 1.2.1, que nos proporcionará un mapa de la densidad electrónica y, por tanto, la estructura.

1.2.1.2. Métodos de resolución estructural.

Mientras el problema de la fase para cristales formados por moléculas de tamaños pequeño y medio ha sido resuelto mediante los llamados métodos directos; en el caso de los cristales de proteínas para resolver dicho problema hemos de recurrir a los siguientes métodos: (i) Reemplazo Molecular (MR, del inglés *Molecular Replacement*), haciendo uso de un modelo estructural de una proteína homóloga, previamente determinada, (ii) Reemplazo Isomorfo Múltiple (MIR, del inglés, *Multiple Isomorphous Replacement*), introduciendo átomos altamente dispersores o (iii) Dispersión Anómala Simple o Múltiple (SAD, del inglés *Single-wavelength Anomalous Diffraction* o MAD, del inglés *Multi-wavelength Anomalous Diffraction*), introduciendo átomos dispersores anómalos. Con cualquiera de estos tres métodos podemos calcular unas fases iniciales, $\alpha_c(hkl)$ que, junto con las amplitudes experimentales, $|F_{\text{exp}}(hkl)|$, nos permitirán calcular una función de densidad electrónica aproximada, sobre la cual podremos construir nuestro modelo estructural. El proceso general a seguir se resume en la figura 1.2.1.

$$\begin{array}{c}
 \text{xyz} \longrightarrow F_c(hkl) = \sum_j f_j e^{2\pi i(hx_j + ky_j + lz_j)} \longrightarrow |F_c(hkl)| + \alpha_c(hkl) \\
 \uparrow \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \downarrow \\
 \rho(xyz) = \frac{1}{V_c} \sum_h \sum_k \sum_l |F_{\text{exp}}(hkl)| \cdot \cos 2\pi(hx + ky + lz - \alpha_c(hkl)) \longleftarrow \\
 \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \uparrow \\
 \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad |F_{\text{exp}}(hkl)|
 \end{array}$$

Figura 1.2.1: Esquema del cálculo de un mapa de densidad electrónica, $\rho(xyz)$. A partir de las posiciones atómicas conocidas (xyz) se pueden obtener los módulos de factores de estructura y fases calculadas, $|F_c(hkl)|$ y $\alpha_c(hkl)$. La función de densidad electrónica se calcula con los módulos experimentales ($|F_{\text{exp}}(hkl)|$) y las fases calculadas ($\alpha_c(hkl)$).

1.2.1.3. Los métodos directos.

Bajo el nombre de métodos directos se engloba una serie de métodos para la resolución del problema de las fases empleando las relaciones entre fases derivadas a partir de los valores observados de las intensidades. El hecho de que la densidad electrónica deba ser cero o positiva, en cualquier punto de la celdilla cristalina, así como el hecho de la atomicidad de las moléculas, genera ciertas limitaciones en la distribución de fases asociada a los factores de estructura. Los métodos directos están basados en el establecimiento de sistemas de ecuaciones que usan las intensidades de los haces difractados y que describen dichas limitaciones. La resolución de dichos sistemas de ecuaciones proporciona información directa sobre la distribución de fases. Sin embargo, puesto que la validez de cada una de estas ecuaciones se establece en términos probabilísticos, es necesario disponer de un gran número de ecuaciones que sobredeterminen los valores de las incógnitas (fases). Esta metodología se denomina directa por el hecho de que, en contraste con otros métodos, proporciona la información estructural (densidad electrónica) directamente de los datos experimentales y se aplica con excelentes resultados para cristales y moléculas de tamaño pequeño y medio, es decir, hasta un centenar de átomos aproximadamente.

Estas relaciones introducen los factores de estructura normalizados, $E(hkl)$, números complejos obtenidos al dividir los factores de estructura $F(hkl)$ por la raíz cuadrada de la intensidad promedio obtenida para una resolución dada si los átomos estuvieran distribuidos al azar en la

celdilla. Las relaciones de fases más intensas se obtienen entre reflexiones con valores elevados del módulo del factor de estructura normalizado, $|E(hkl)|$ (Sayre, 1952; Karle & Hauptman, 1956; Bricogne, 1988).

El valor teórico máximo que puede poseer $|E(hkl)|$ es proporcional a la raíz cuadrada del número de átomos en la celdilla. Este valor será obtenido si las contribuciones al factor de estructura de los átomos de la celdilla están todas en fase para la reflexión (hkl) considerada. La probabilidad de este evento decrece muy rápidamente con el número de átomos de la celdilla. Del mismo modo, las relaciones entre fases se debilitan cuando el número de átomos en la celdilla aumenta. Este es el porqué mientras los métodos directos son cuasi-universalmente utilizados para resolver la estructura de pequeñas moléculas constituidas por centenares de átomos como máximo, su utilización para determinar la estructura de macromoléculas biológicas no comprende más que la localización de los átomos pesados, resolviendo estructuras constituidas por hasta miles de átomos y por aquellas que proporcionan datos de difracción a muy alta resolución ($d_{\text{resolución}} = \lambda/2\sin\theta < 1,2 \text{ \AA}$) (Usón & Sheldrick, 1999).

1.2.1.4. El método de reemplazo molecular.

Cuando se dispone del modelo estructural de una proteína con secuencia de aminoácidos homóloga, el problema de la fase se puede resolver mediante la técnica del reemplazo molecular (MR) (Rossmann, 1972; Machin, 1985; Navaza, 1994). La estructura de esta proteína homóloga se considera como si fuera la proteína que va a determinarse y sirve como un primer modelo que posteriormente será refinado. Este procedimiento está basado en la observación de que proteínas homólogas en su secuencia peptídica, muestran un plegamiento muy similar. El problema en este caso consiste en transferir la estructura molecular de la proteína conocida, desde su propio empaquetamiento cristalino hasta el cristal de la proteína con estructura desconocida. El posicionamiento de la molécula conocida en la celdilla unidad de la proteína desconocida, requiere determinar su correcta orientación y su posición precisa. Ambas operaciones se calculan mediante las funciones denominadas de rotación y de translación (Figura 1.2.2)

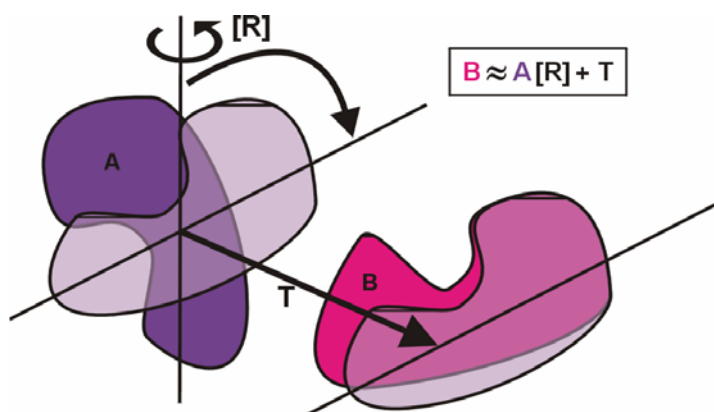


Figura 1.2.2: Esquema del proceso de reemplazo molecular (MR). La molécula de estructura conocida (A) se gira mediante la operación [R] y se traslada mediante [T], para llevarla a la posición que ocupa la molécula problema (A')

La bondad de las soluciones de estas funciones se expresa mediante los coeficientes de correlación entre las funciones de Patterson (funciones equivalentes a la ecuación 1.2.1., eliminando las fases y usando $|F(hkl)|^2$, que proporciona información sobre las distancias interatómicas) experimental y la calculada con la proteína conocida. Un coeficiente de correlación alto entre dichas funciones equivale a un buen acuerdo entre el espectro de difracción experimental y el calculado con la proteína conocida. Una vez orientada y trasladada convenientemente la molécula conocida, se calcula un mapa de densidad electrónica usando los factores de estructura experimentales y las fases obtenidas a partir de la estructura de la proteína homóloga.

Este método sólo necesita datos de difracción del cristal nativo, lo cual constituye una ventaja frente a los métodos de obtención de fases experimentales. Sin embargo, a pesar del número elevado de estructuras resueltas por reemplazo molecular y de que disponemos de un número razonable de estructuras conocidas, el método introduce errores. Uno de los problemas consiste en que la solución puede estar sesgada por el modelo y que una solución inicial aparentemente correcta no pueda ser posteriormente refinada. En el caso de ausencia de un modelo estructural conocido, tenemos que recurrir a métodos de obtención de fases *de novo* descritos a continuación.

1.2.1.5. Los métodos de reemplazo isomorfo.

El método de obtención de fases por reemplazo isomorfo consiste en comparar los datos de difracción $|F(hkl)|$ de un cristal de la proteína nativa con datos de cristales donde se han fijado átomos pesados a la molécula de proteína. Estos átomos pesados poseen un número de electrones más elevado que los que constituyen las moléculas biológicas, por lo que las intensidades de los picos de difracción del cristal derivado se diferencian significativamente de las intensidades de difracción del cristal nativo.

Un requisito esencial para la utilización de estos métodos es el isomorfismo entre las estructuras cristalinas de la proteína nativa y del derivado. La fijación de los átomos pesados sobre la molécula no debe modificar su estructura, su orientación o su posición en el cristal; los parámetros de celdilla deben ser muy próximos. La densidad electrónica ρ_{PH} del cristal derivado puede considerarse como la suma de la densidad electrónica ρ_P de la macromolécula sola, idéntica a la correspondiente en el cristal nativo, y de la densidad electrónica ρ_H de la subestructura de los átomos pesados. La transformada de Fourier de la densidad electrónica transfiere esta propiedad sobre los factores de estructura. El isomorfismo se traduce a partir de la igualdad:

$$\rho_{PH} = \rho_P + \rho_H \xrightarrow{TF} F_{PH} = F_P + F_H$$

donde F_{PH} , F_P y F_H representan los factores de estructura respectivos del cristal derivado, del cristal nativo y de la subestructura de átomos pesados.

Para determinar la fase de los factores de estructura, determinamos en primer lugar la subestructura de átomos pesados a partir de las diferencias entre las intensidades de las reflexiones del cristal derivado y las correspondientes al cristal nativo, empleando los denominados mapas de Patterson. Los mapas de Patterson representan el mapa de vectores de situación relativa entre los átomos pesados a partir del cual es relativamente sencillo determinar sus posiciones dentro de la celdilla unidad. Del conjunto de vectores interatómicos de cada átomo con sus simétricos, existe un grupo de especial relevancia para la correcta interpretación

de los mapas de Patterson que se denominan secciones de Harker, y que corresponden a determinadas concentraciones de máximos de la función de Patterson[†] (Figura 1.2.3).

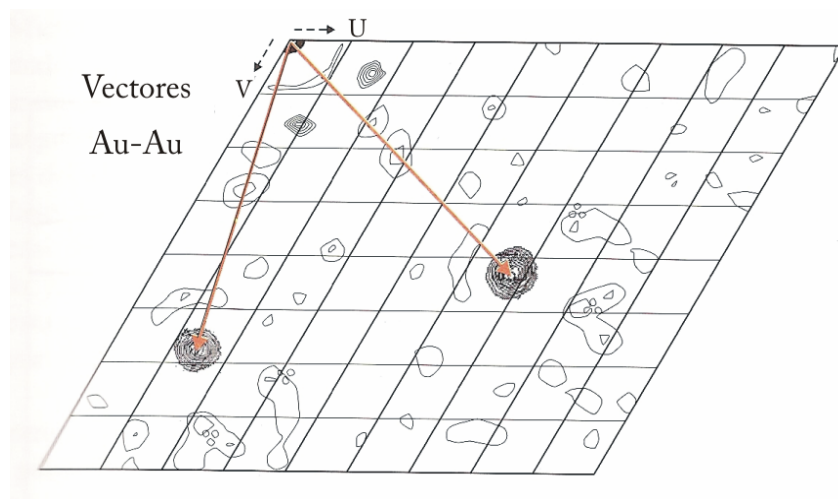


Figura 1.2.3: Sección Harker $w = 0$ del derivado AtHal3 con una sal de oro, mostrando los máximos prominentes para los vectores Au-Au (Albert *et al.*, 2000).

Una vez determinada la subestructura de los átomos pesados, podremos calcular F_H , en módulo y en fase. El factor de estructura buscado, F_P , vendrá dado por la relación $F_P = F_{PH} - F_H$ donde los módulos $|F_P|$ y $|F_{PH}|$ se conocen experimentalmente (Figura 1.2.4a). El miembro de la izquierda de la ecuación indica que la dirección donde apunta el vector F_P se encuentra en un círculo de radio $|F_P|$ centrado en el origen, mientras que los miembros de la derecha de la ecuación indican que se encuentra en un círculo de radio $|F_{PH}|$ y centrado sobre donde apunta $-F_H$. Mediante la intersección de ambos círculos obtendremos la dirección donde apunta F_P (Figura 1.2.4b) y por tanto, su fase. Sin embargo, como se observa en la figura 1.2.4b, el método denominado SIR (Single Isomorphous Replacement), que utiliza el cristal nativo y un solo cristal derivado, origina dos soluciones para cada fase resultante (de ahí que dicho método deba recurrir a fuentes adicionales para la obtención de fases). Para determinar la fase de los factores de estructura sin ambigüedad, es necesario al menos dos cristales derivados isomorfos diferentes lo cual nos conduce al método MIR (Multiple Isomorphous Replacement), que fue el

[†] Si en la función de densidad electrónica, $\rho(xyz)$, se omite la información de las fases y se reemplazan los módulos de los factores de estructura por sus cuadrados, se obtiene la denominada función de Patterson

primer método de obtención de fases utilizado para resolver la estructura de macromoléculas (Green *et al.*, 1954) (Figura 1.2.4c).

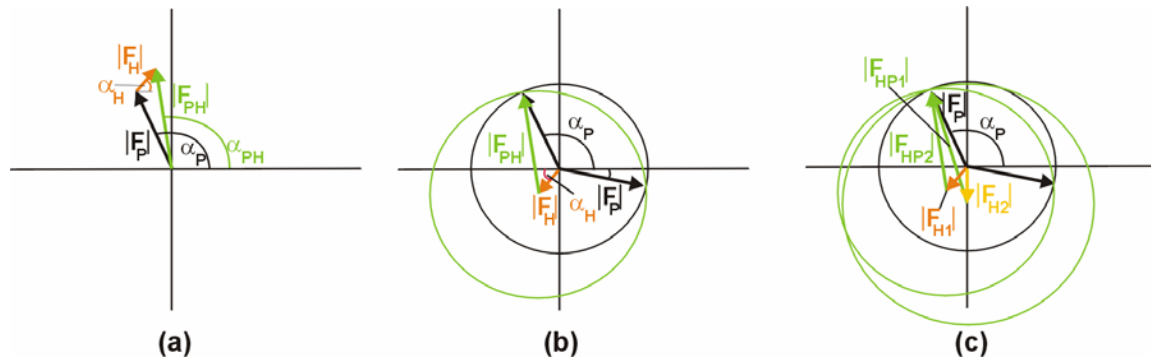


Figura 1.2.4: (a) Representación vectorial de los factores de estructura en el reemplazo isomorfo. (b) Construcción de Harker para un único derivado. (c) Construcción de Harker para un doble derivado.

Los métodos SIRAS/MIRAS (*Single/Multiple Isomorphous Replacement with Anomalous Scattering*) tienen en cuenta las diferencias anómalas de los datos de difracción como fuente de información de fase suplementaria.

1.2.1.6. Los métodos de dispersión anómala.

Los métodos que emplean la dispersión anómala se sirven del hecho de que el factor de dispersión atómica de un átomo dado (ecuación 1.2.2) varía fuertemente, en módulo y en fase, en la región de la línea de absorción de ese átomo: este es el fenómeno de la dispersión anómala. El factor de dispersión atómica de un difusor anómalo está representado por un número complejo: parte real y parte imaginaria (ver sección 1.2.2).

Estos métodos se liberan del problema del no-isomorfismo utilizando datos de difracción de un cristal derivado único. La estructura se determina a partir de este cristal derivado y no es necesario recoger los datos de un cristal de proteína nativo.

El método MAD (*Multiple-wavelength Anomalous Diffraction*) utiliza los datos de difracción recogidos a varias longitudes de onda, en torno a la región de la línea de absorción del dispersor anómalo. Su utilización es posible gracias al empleo de radiación sincrotrón en tanto que proporciona longitud onda sintonizable. Este es hoy en día el método de obtención de fases *de novo* más utilizado (Hendrickson, 1999).

Para el método SAD (*Single-wavelength Anomalous Diffraction*), los datos son recogidos a una sola longitud de onda elegida en la región donde el factor de dispersión atómica presenta una parte imaginaria elevada. Como en el caso del método SIR, la solución de fase en el método SAD es ambigua, pero, contrariamente al método SIR, las dos soluciones no son equiprobables y la ambigüedad de fase puede ser parcialmente salvada a través de la utilización de métodos de mejora posterior de fases como los métodos de modificación de densidad y la simetría no cristalográfica.

1.2.1.7. Comparación de los métodos de obtención de fases experimentales.

Los métodos de reemplazo isomorfo presentan la ventaja frente a los métodos de dispersión anómala que la diferencia isomorfa de las intensidades de difracción sobre las que depende la obtención de fases es generalmente mucho mayor que las diferencias anómalas en las que se basan los métodos MAD/SAD. Es más, el reemplazo isomorfo no requiere de una fuente de radiación sincrotrón que proporciona una longitud de onda sintonizable.

El mayor inconveniente del método es que la condición de isomorfismo entre los diferentes cristales utilizados, generalmente no se cumple, lo que provoca que las diferencias entre las intensidades de difracción no puedan ser interpretadas en términos de pura sustitución isomorfa, lo que finalmente puede impedir la obtención de fases. Por otro lado, el método MIR necesita los datos de difracción registrados con al menos dos cristales derivados, cristales que son frecuentemente difíciles de obtener.

Los métodos que utilizan la dispersión anómala permiten resolver el problema del no isomorfismo, ya que los datos de difracción pueden, en principio, ser registrados con un solo cristal. Por el contrario, éstos demandan medidas de las intensidades de difracción muy precisas como consecuencia de la relativa debilidad de las diferencias anómalas.

El método MAD puede ser utilizado a partir de datos recogidos con radiación de longitud de onda sintonizable. En relación al método SAD, presenta la ventaja de proporcionar datos de fase de muy buena calidad ya que resuelve experimentalmente la ambigüedad de la solución (Ver sección 1.2.2). Además, el método SAD presenta la ventaja de necesitar únicamente datos de difracción recogidos a una sola longitud de onda; y si el átomo pesado lo permite, los datos pueden, por ejemplo, ser recogidos con un generador de rayos X convencional. El método

permite la obtención de fases con los datos anómalos recogidos fuera de la línea de absorción de un dispersor anómalo determinado, en las condiciones donde la parte imaginaria del factor de dispersión atómica no sea desdeñable, sin ser máxima. Esto es particularmente interesante para los elementos cuyas líneas de absorción se encuentren en las regiones de longitud de onda difícilmente accesibles para los experimentos de difracción (S, Ca, Xe, U...). Por otra parte, como el método SAD sólo necesita un conjunto único de datos, el tiempo de exposición de un cristal a los rayos X es reducida, disminuyendo también los riesgos de daño por radiación. De hecho este problema es particularmente importante debido a la utilización de cristales cada vez más pequeños y fuentes de radiación sincrotrón cada vez más intensas. Además de la reducción del poder de difracción que se traduce en la pérdida de las reflexiones a alta resolución, los daños por radiación son una fuente de no isomorfismo intrínseca del registro de datos sobre un cristal, efecto que, si no es tenido en cuenta, puede impedir la obtención de fases. En lo que concierne a la precisión de los datos, el método SAD necesita una precisión elevada hasta el punto que las diferencias anómalas debidas a los elementos excitados lejos del pico de absorción puedan ser relativamente débiles.

1.2.2. DISPERSIÓN ANÓMALA.

1.2.2.1. Presentación del fenómeno.

Cuando la radiación X incidente tiene una frecuencia próxima a la frecuencia natural de oscilación de los electrones de un determinado átomo, se produce el fenómeno de la dispersión anómala. Este fenómeno modifica el factor de dispersión atómico (ver ecuación 1.2.2) de tal modo que su expresión se ve modificada con dos términos, f' y f'' , que dan cuenta de las componentes real e imaginaria, respectivamente, de la fracción anómala del factor de dispersión. De manera general, el factor de dispersión atómico f (número complejo que depende de la longitud de onda de la radiación incidente) vendrá dado por la expresión que se muestra en la ecuación 1.2.3:

$$f = f_0 + f'(\lambda) + if''(\lambda)$$

Ecuación 1.2.3: Componentes del factor de dispersión atómica. En presencia de dispersión anómala, el factor de dispersión atómico, f_0 , se ve modificado por dos términos adicionales que dan cuenta de las partes real e imaginaria de la dispersión anómala.

Las variaciones de f' y f'' frente a la energía de los rayos X se pueden calcular usando consideraciones teóricas, tal como se muestra en la figura 1.2.5 para el caso del cobre. Para los valores de la energía de los rayos X en los que existe resonancia el valor de f'' aumenta drásticamente, y al mismo tiempo el valor de f' decrece. Este hecho tiene gran importancia práctica si se tiene en cuenta que muchos de los átomos pesados usados en cristalografía tienen picos de absorción a energías (longitudes de onda) que se pueden obtener fácilmente mediante radiación sincrotrón. Así pues, los datos de difracción recogidos en estas condiciones tendrán una componente normal, debida fundamentalmente a los átomos ligeros (nitrógeno, carbono e hidrógeno), y una componente anómala, la cual será debida a los átomos pesados, que producirá una alteración en la fase global de cada reflexión. Todo esto se traducirá en un cambio de intensidad entre las llamadas reflexiones o pares de Friedel (parejas de reflexiones que en condiciones normales deberían tener la misma amplitud e idénticas fases, pero con signos opuestos). El cambio de intensidad detectable entre las parejas de Friedel recibe el nombre de difracción anómala, proporcionándonos información acerca de las fases generadas por los dispersores anómalos, como una primera aproximación de las fases globales.

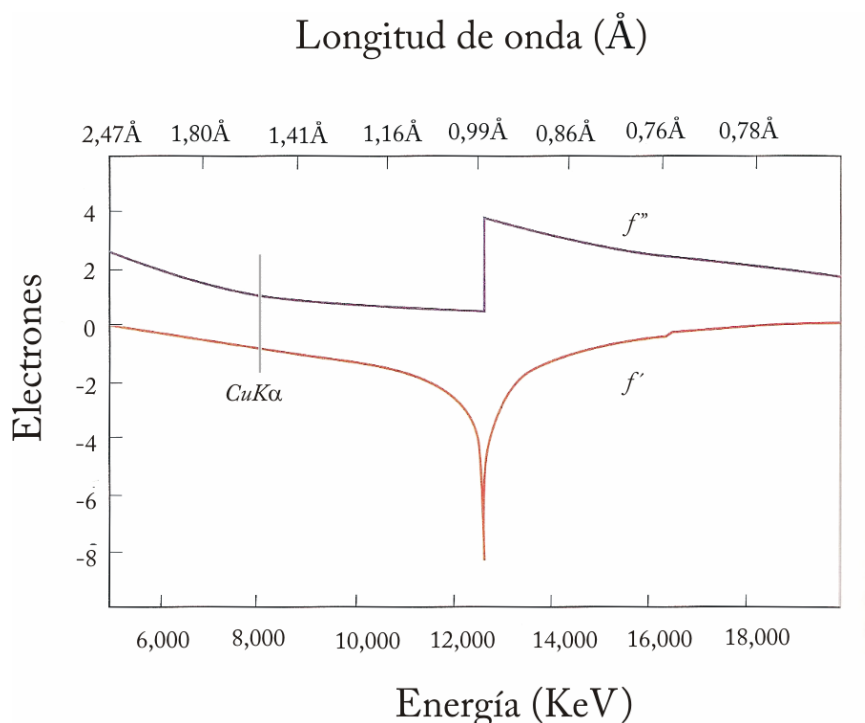


Figura 1.2.5: Variación de las componentes real e imaginaria del factor de dispersión atómico de selenio en función de la energía de los rayos X incidentes. La línea vertical señala la longitud de onda del CuK α .

Sin embargo, hemos de tener en cuenta que mientras los valores calculados teóricamente por la mecánica cuántica para f' y f'' son correctos lejos de las líneas de absorción, cerca de las líneas, los valores deben ser determinados de forma experimental.

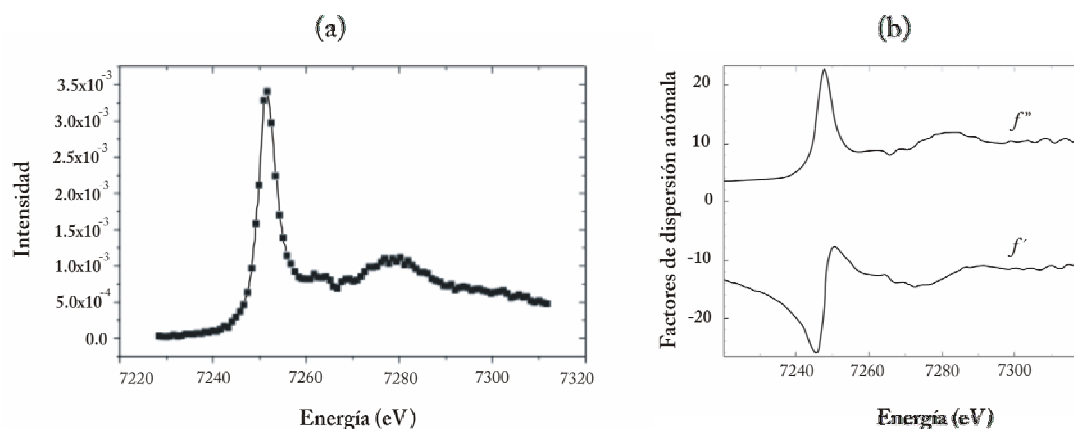


Figura 1.2.6: (a) Espectro de absorción, registrado en modo fluorescencia sobre una solución que contiene un complejo de Gd a la longitud de onda de la línea de absorción LIII del Gd. (b) Factores de dispersión anómala f' y f'' del Gd en función de la energía de la radiación incidente obtenidos a partir del espectro (a).

En el caso de la línea LIII de los lantánidos, los electrones de la capa L son excitados hacia un estado de la región de energía continua. Como se puede ver en la figura 1.2.6, el coeficiente de absorción de los lantánidos presenta un pico importante en la línea de absorción, tanto para f' como para f'' . Este pico es denominado línea blanca. Las líneas blancas son debidas a una densidad elevada de estados finales de la transición. En términos generales, la presencia de esta línea blanca está fuertemente influenciado por el entorno molecular del átomo considerado (Brown *et al.*, 1997). Esta influencia es, sin embargo, poco importante para la línea LIII de los lantánidos. Las variaciones de f'' con la energía de radiación X incidente son obtenidas a partir del espectro de absorción del elemento considerado recogiendo datos de fluorescencia alrededor de la línea de absorción (Figura 1.2.6a). Este espectro de excitación de fluorescencia es equivalente a un espectro de absorción, ya que la fluorescencia es directamente proporcional a la absorción: los electrones del elemento considerado son excitados por la absorción de la radiación incidente dejando vacantes en las capas inferiores que son rellenados por los electrones de capas de más alta energía y emiten a su vez radiación de fluorescencia. Los factores de dispersión anómalos f' y f'' son obtenidos a partir del espectro de absorción gracias a la ayuda del programa CHOOCH (Evans & Pettifer, 2001). El resultado está representado en la

figura 1.2.6b. Se puede recalcar que el máximo de f'' está desplazado aproximadamente 4eV en relación al mínimo de f' .

En el caso de los lantánidos, debido a que la posición y la presencia de la línea blanca son poco sensibles al entorno local del ion lantánido, basta determinar los valores de f' y f'' en función de la energía de radiación incidente para el elemento de interés independientemente del estado de coordinación del ion.

1.2.2.2. Las diferencias dispersivas.

La obtención de fases por el método MAD se basa en las diferencias entre los pares de Friedel medidas a una sola longitud de onda y las diferencias entre los pares de factores de estructura medidos a dos longitudes de onda diferentes debido a las variaciones de la parte real y de la parte imaginaria del factor de dispersión atómica con la longitud de onda. A fin de obtener valores dispersivos lo más elevados posible, para un experimento MAD se recogen datos al mínimo valor de f' y máximo de f'' (para obtener máximas diferencias anómalas). Generalmente, un tercer conjunto de datos se recoge a una longitud de onda lejos de la línea de absorción.

1.2.3. DETERMINACIÓN DE LA POSICIÓN DE LOS ÁTOMOS PESADOS Y OBTENCIÓN DE FASES.

1.2.3.1. Obtención de fases con el método SAD.

En el marco del método SAD el cálculo de la fase de los factores de estructura se realiza a partir de las posiciones de los átomos pesados en la celdilla y los factores de dispersión atómica $f = f_0 + f' + if''$, calculando así las contribuciones de los átomos pesados al factor de estructura. Representando vectorialmente los factores de estructura de modo análogo al caso del reemplazo isomorfo, observamos que la dispersión anómala provoca la ruptura de la ley de Friedel‡ (Figura 1.2.7a)

‡ La ley de Friedel expresa que, en condiciones normales, las reflexiones (hkl) y (-h-k-l) son idénticas en intensidad. Cuando hay dispersión anómala se produce la ruptura de la igualdad.

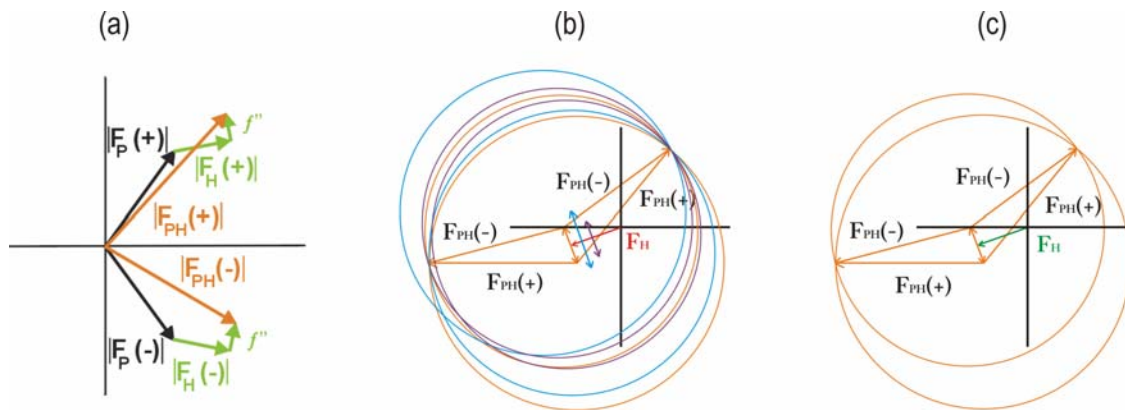


Figura 1.2.7: (a) Representación vectorial de los factores de estructura en la dispersión anómala. (b) Construcción de Harker para MAD. (c) Construcción de Harker para SAD.

Si bien en el caso del método MAD la ambigüedad de fase se solventa utilizando datos registrados a longitudes de onda diferentes (Figura 1.2.7b), en el método SAD (Figura 1.2.7c), dicha ambigüedad puede ser teóricamente resuelta gracias a la no equiprobabilidad de las fases obtenidas como solución (Dauter *et al.*, 2002). Sin embargo, desde el punto de vista práctico, la discriminación entre las soluciones suponía recoger datos de extraordinaria calidad y precisión haciendo este método inviable hasta hace muy poco tiempo. Hoy en día, gracias al avance en las instalaciones de recogida de datos cada vez más precisas y de mayor intensidad, potentes programas de procesamiento de datos así como en la mejora en algoritmos de obtención de fases y procedimientos de modificación de densidad, ha sido posible emplear este método para fasear.

1.2.3.2. Procedimientos de mejora de fase.

Antes de calcular el mapa de densidad electrónica experimental, las fases iniciales obtenidas son mejoradas por los procedimientos complementarios aplicando principalmente las modificaciones de densidad electrónica en el espacio directo (ver sección 3.6.1.5). Esto es particularmente importante para la obtención de fases por los métodos SAD o SIR donde la información de fase inicial es fuertemente bimodal. Las modificaciones de densidad en el espacio directo permiten solventar la ambigüedad resaltando la fase correspondiente a la densidad modificada.

1.2.3.3. La señal anómala.

Para el éxito de una obtención de fases basada en las diferencias anómalas, es primordial que la señal anómala sea superior a la incertidumbre de las medidas. La intensidad de los efectos anómalos a una longitud de onda dada puede ser estimada mediante la relación de Bijvoet a partir de las amplitudes de las reflexiones (hkl) y (-h-k-l) (Hendrickson & Ogata, 1997; Dauter *et al.*, 2002):

$$\frac{\langle \Delta F^{\pm} \rangle}{\langle F \rangle} = \left(\frac{2N_A}{N_P} \right)^{1/2} \frac{f_A''}{Z_{eff}},$$

en donde N_A es el número de dispersores anómalos con un factor anómalo f_A'' y N_P es el número de átomos de la proteína de factor de dispersión medio Z_{eff} . Esta relación asume que los dispersores anómalos son químicamente idénticos y que sus sitios están totalmente ocupados. Cuando diferentes dispersores anómalos con diferentes ocupaciones de los sitios están presentes en el cristal, la relación se transforma en:

$$\frac{\langle \Delta F^{\pm} \rangle}{\langle F \rangle} = \frac{1}{Z_{eff}} \left(\frac{2 \sum q_i^2 f_i''^2}{N_P} \right)^{1/2},$$

en donde q_i es la ocupación de los sitios y f_i'' es la parte imaginaria del factor de dispersión atómico del dispersor anómalo i .

Esta expresión muestra que, asumiendo una ocupación idéntica de los dispersores anómalos, se obtienen relaciones de Bijvoet idénticas para una proteína que tenga un tamaño cuatro veces mayor cuando f'' es simplemente el doble. De forma similar, asumiendo los valores dados para f_i'' , se obtienen relaciones de Bijvoet idénticas para una proteína cuatro veces mayor cuando las ocupaciones de los sitios se doblan. Conviene destacar también que si el número de dispersores anómalos se dobla, el tamaño de la proteína puede ser doblada alcanzando el mismo resultado.

La importancia de la tasa de ocupación de los sitios y del difusor anómalo elegido.

Con el fin de obtener una señal anómala máxima, es importante que los sitios de fijación de los dispersores anómalos estén fuertemente ocupados, y que el dispersor anómalo y la longitud de onda sean elegidas de forma que se obtenga el valor más elevado posible de f'' . Esto es muy importante para resolver estructuras de gran tamaño.

La importancia de la resolución de los datos.

La resolución de los datos de difracción viene dada por la calidad cristalina o, eventualmente, por la geometría del experimento. Dado que de las diferentes contribuciones al factor de dispersión atómica $f(\theta, \lambda) = f_0(\theta) + f'(\lambda) + i f''(\lambda)$ sólo $f_0(\theta)$ decrece fuertemente a medida que aumenta la resolución, la señal anómala debe ser relativamente más elevada a mayor resolución. Este efecto es sin embargo limitado por la agitación térmica del conjunto de los átomos y, en particular, de los dispersores anómalos, que hacen que a alta resolución las reflexiones sean mucho menos intensas y que su intensidad esté medida con una precisión menor que las reflexiones a baja resolución.

Esta es la razón por la que en la identificación de los sitios de fijación de los átomos pesados, los mejores resultados son a menudo obtenidos limitando la resolución de los datos utilizados a una resolución más baja que la resolución máxima de los datos (Dauter & Nagem, 2002). Este hecho se hace claramente patente cuando utilizamos los métodos directos que emplean los factores de estructura normalizados, más sensibles a los errores a alta resolución.

Por otro lado, una alta resolución de los datos es muy importante para interpretar el mapa de densidad electrónica experimental y para la construcción del modelo.

La importancia de la redundancia de los datos.

La señal anómala es débil en relación a las intensidades de difracción, siendo primordial, para obtener fases, que la señal anómala sea significativa debido a la inexactitud de las medidas. De este modo, para maximizar la señal anómala, hay que maximizar la exactitud de las medidas, es decir aumentar la relación señal ruido, I/σ , de las medidas. La relación señal sobre ruido depende, por supuesto, de la calidad cristalina, pero depende también de las condiciones de

recogida: tiempos de exposición elevados, y un número elevado de medidas por reflexión, mejoran la precisión estadística de los datos.

Una redundancia elevada de los datos (es decir, la medida independiente, repetida, de la intensidad de las reflexiones cuya equivalencia es impuesta por la simetría cristalina) permite mejorar la precisión estadística de las medidas, eliminando los errores sistemáticos debidos a artefactos del experimento (Dauter & Nagem, 2002). Esta es la razón por la que, para fasear utilizando la señal anómala, es esencial tomar datos de difracción con una redundancia elevada.

1.2.4. PRODUCCIÓN DE CRISTALES DERIVADOS PARA LA OBTENCIÓN DE FASES POR LOS MÉTODOS DE REEMPLAZO ISOMORFO Y POR MÉTODOS DE DISPERSIÓN ANÓMALA.

1.2.4.1. Elección del átomo pesado y de la longitud de onda de la radiación X incidente.

Criterios de dispersión.

Según el método de obtención de fases empleado, se emplean diferentes criterios para la elección del átomo pesado y de la longitud de onda de la radiación incidente.

Para el reemplazo isomorfo, el átomo con el número atómico Z más elevado maximiza las diferencias isomorfas, y éstas son independientes de la longitud de onda.

Para los métodos que utilizan la dispersión anómala, elegimos el átomo con un valor elevado de f'' a la longitud de onda disponible. Para el método MAD, necesitamos entonces que uno de los picos de absorción del dispersor anómalo sea accesible al experimento, ya que la variación de los factores anómalos f' y f'' son importantes. Para el método SAD, en el caso de que dispongamos de una fuente de rayos X de longitud de onda sintonizable que permita fijarla en el pico de absorción del dispersor anómalo, se situara en dicho pico, en el máximo de f'' . Si el pico de absorción del dispersor anómalo es inaccesible pero las medidas cerca del pico proporcionan un valor de f'' suficiente, es posible trabajar a dicha longitud de onda. Con el método SAD, utilizaremos una radiación de rayos X a longitud de onda fija. En el caso donde utilizamos dispersores anómalos intrínsecos de los cristales nativos (metaloproteínas, obtención de fases con azufre) elegiremos la longitud de onda que permita obtener el valor de f'' lo más elevado posible.

Criterios químicos.

Aparte de las propiedades físicas de la dispersión de rayos X, es importante que el átomo pesado elegido pueda fácilmente introducirse en el cristal de la proteína. Habrá que buscar un compuesto que incluya el átomo pesado y que origine una ocupación elevada en los sitios específicos de la macromolécula biológica, sin distorsionar el orden cristalino de la muestra. En el caso de los métodos de reemplazo isomorfo, además el cristal derivado ha de ser isomorfo al nativo. La ventaja de los dispersores anómalos intrínsecos es evidente: ningún esfuerzo complementario es necesario para la preparación del derivado y los sitios de los dispersores anómalos están, generalmente, fuertemente ocupados.

Desde el punto de vista práctico, para la elección de una solución de átomo pesado apropiada:

1. Si nuestra proteína tiene Cisteínas libres o Histidinas, debemos empezar por probar con compuestos de mercurio (Garman & Murray, 2003).
2. Si nuestra proteína presenta una distribución aminoacídica normal con Cisteínas, Metioninas o Histidinas, debemos probar compuestos de Platino (Garman & Murray, 2003)
3. Si existe abundancia de residuos cargados positivamente, debemos emplear $\text{Pt}(\text{CN})_2$.
4. Evaluar la solubilidad del compuesto elegido en la solución de cristalización de la proteína a resolver. La mayoría de estos compuestos son inestables en solución y únicamente en un rango determinado de pHs son solubles. Además, compuestos de la solución de cristalización pueden intervenir en la interacción del compuesto de átomo pesado con la proteína. El sulfato amónico tiene una elevada fuerza iónica disminuyendo las interacciones proteína – sal de átomo pesado (Drenth, 1999). De este modo, el compuesto de átomo pesado debe ser elegido con cuidado o reemplazar el sulfato amónico por otra sal (por ejemplo, sulfato de litio, sodio o potasio) o polietilenglicol. Otra de las incompatibilidades se produce con soluciones de cristalización que contengan ditiotreitil y las soluciones de átomo pesado mercuriales.
5. Preparar soluciones de átomo pesado inmediatamente antes de los ensayos de derivatización debido a su inestabilidad.
6. Verificar si la proteína posee Ca o Mg unido ya que en muchos casos puede intercambiarse por átomos más pesados, en particular lantánidos. Además debemos asegurarnos si la proteína

posee metales endógenos como el Fe o el Zn que puedan ser utilizados directamente en ensayos de dispersión anómala.

7. Las condiciones iniciales para la derivatización son concentraciones de átomo pesado entre 0.1 y 10 mM y tiempos de inmersión entre 10 minutos y varios días a la mayor concentración que tolere el cristal. Actualmente la tendencia en la derivatización por inmersión es emplear soluciones más concentradas (hasta 20 mM) y tiempos más cortos (segundos). En el caso de la cocrystalización debemos jugar con la concentración de solución de átomo pesado que a priori se preparará también entre 0.1 y 10 mM.

1.2.4.2. Accesibilidad del pico. Altas longitudes de onda.

En la mayoría de las líneas de sincrotrón dedicadas al método MAD, las longitudes de onda de los rayos X accesibles se sitúan entre 0.7 y 2.0 Å. Para ciertos elementos de interés, los picos de absorción se encuentran en la zona de los rayos X blandos (Tabla 1.2.1).

El mayor problema de utilizar altas longitudes de onda es la absorción. Así para las longitudes de onda superiores a 2.0 Å, la radiación es muy fuertemente absorbida por la muestra, y también por el aire: la intensidad difractada está fuertemente atenuada, lo que hace que degenere la calidad de las medidas y necesite correcciones de absorción. La radiación absorbida por la muestra induce además daño por radiación que hace variar las intensidades difractadas con el tiempo. Otro inconveniente de las altas longitudes de onda es que a la misma resolución, los ángulos de Bragg son más grandes: la resolución máxima proporcionada por los bordes del detector es con frecuencia inferior al límite de difracción del cristal. Así, en el pico de absorción LIII del Gd ($\lambda = 1.711$ Å) en la línea BM30 del ESRF, con un detector de 165 mm de diámetro situado a la distancia mínima de 110mm de la muestra, la resolución máxima que se puede recoger es de 2.7 Å.

Z	Elemento	Línea	$\lambda_{\text{línea}}$ [Å]	$f''_{\text{línea}}$ [e ⁻]	$f''_{\text{CuK}\alpha}$ [e ⁻] ($\lambda=1,5418$ Å)
15	P	K	5.78	4.1	0.43
16	S	K	5.02	4.1	0.56
17	Cl	K	4,39	4.1	0.70
20	Ca	K	3.07	4.1	1.3
25	Mn	K	1.90	9.7	2.8
26	Fe	K	1.74	4.5	3.2
27	Co	K	1.61	3.9	3.6
28	Ni	K	1.49	4.3	0.51
29	Cu	K	1.38	4.2	0.59
30	Zn	K	1.28	4.3	0.68
34	Se	K	0.98	6-10	1.1
35	Br	K	0.92	3.8	1.3
36	Kr	K	0.87	1.6-2.7	1.4
54	Xe	L _I	2.27	13.5	7.4
64	Gd	L _{III}	1.71	28	12.0
78	Pt	L _{III}	1.07	10.2	6.9
79	Au	L _{III}	1.04	10.2	7,3
80	Hg	L _{III}	1.01	10.1	7.7
82	Pb	L _{III}	0.95	10.1	8.5
92	U	L _{III}	0.72	10.3	16
92	U	M _V	3.49	120	

Tabla 1.2.1: Longitudes de onda de los picos de absorción y valores de f'' de ciertos elementos (Muller & Heinemann, 2005). Los valores de f'' para los elementos presentan una línea blanca en sus picos de absorción dependiente generalmente del entorno del dispersor anómalo, exceptuando los lantánidos. Los valores para el uranio provienen de Chesne (2002).

Para alcanzar los picos K de elementos pesados, los rayos X muy duros pueden ser utilizados (Schiltz *et al.*, 1997). Sin embargo, uno de los inconvenientes de los rayos X muy duros es la débil eficacia cuantitativa de la detección de los detectores usuales a esas energías. Otro inconveniente es la importancia que toma, a energías superiores a 30KeV, la dispersión Compton en relación a la dispersión elástica de los rayos X, lo que hace que aumente el ruido de fondo de manera significativa.

1.2.4.3. Picos de absorción y líneas blancas.

La figura 1.2.8 muestra la evolución de la longitud de onda de los picos de absorción K, L y M en función del número atómico Z de los elementos de la tabla periódica. Véase que los picos K de los elementos ligeros situados entre el Mn ($Z = 25$) y el Y ($Z = 39$) son fácilmente accesibles ($\lambda = 1.89\text{-}0.73 \text{ \AA}$). Los picos LIII de los elementos situados entre el Sm ($Z = 62$) y el uranio ($Z = 92$) son también fácilmente accesibles ($\lambda = 1.85 \text{ \AA} - 0.72 \text{ \AA}$).

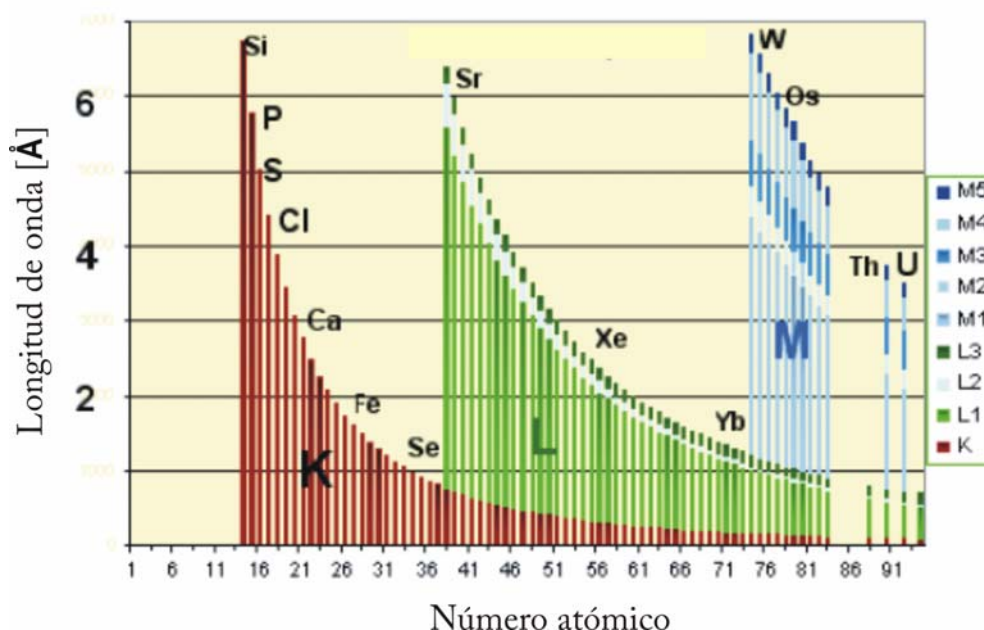


Figura 1.2.8: Picos de absorción K, L y M para los elementos estables de la tabla periódica.

La parte imaginaria f'' del factor de dispersión atómica es del orden de $4 e^-$ para la mayoría de los picos K y del orden de $10 e^-$ para los picos LIII (Muller & Heinemann, 2005), a excepción de aquellos elementos que presentan una línea blanca en su pico. Este es el caso para los picos K del Mn ($f'' \sim 10 e^-$) y del Se (f'' entre $6 e^-$ y $10 e^-$, según el estado de oxidación) y para los picos LIII de los lantánidos cuyos valores de f'' pueden alcanzar hasta $30 e^-$. Para los experimentos cerca de los picos, es más adecuado trabajar a una longitud de onda inferior a la del pico, ya que f'' no decrece tan lentamente para $\lambda < \lambda_{\text{umbral}}$. Así para la radiación $K\alpha$ del cobre, obtenemos valores de f'' de 2 a $4 e^-$ para los elementos situados entre el vanadio, V, y el cobalto, Co, valores de f'' que varían entre 9.6 y $13.3 e^-$ para los lantánidos que van del lantano, La, al disprosio, Dy, y un valor de $f'' = 16 e^-$ para el uranio.

1.2.4.4. Revisión de los métodos habituales para obtener los derivados para la obtención de fases por los métodos de reemplazamiento isomorfo o de dispersión anómala.

Utilización de proteínas con selenio o ácidos nucleicos bromados.

El método de obtención de fases *de novo* más utilizado con radiación sincrotrón es el método MAD aplicado a los cristales de proteínas con selenio (Hendrickson *et al.*, 1990; Dauter & Nagem, 2002). El método consiste en expresar la proteína de interés con selenometionina, estando el átomo de azufre de la metionina reemplazado por un átomo de selenio. Las ventajas del método son la ocupación completa de los sitios de selenio y la señal anómala relativamente elevada, que va desde 6 a 10 e⁻ (según el estado de oxidación del selenio, en el pico K del Se). Este pico, cuya longitud de onda es $\lambda = 0.98 \text{ \AA}$, es fácilmente accesible con las líneas de luz habituales, para las cuales, por otro lado, la intensidad ha sido, generalmente, optimizada a esta longitud de onda. El mayor inconveniente de este método es la laboriosidad de la puesta en práctica. Es necesario expresar la proteína en medio con selenio lo que conduce a bajas tasas de expresión y además, generalmente, la selenoproteína cristaliza en condiciones distintas a las de la proteína nativa.

Ciertas proteínas eucariotas no pueden ser expresadas en los sistemas bacterianos y en otros sistemas de expresión la incorporación de la selenometionina es mala. Finalmente, ciertas proteínas no contienen metionina.

Del mismo modo que la utilización de selenometioninas para resolver la estructura de proteínas, se pueden utilizar nucleótidos bromados o iodados para resolver la estructura de ácidos nucleicos o de sus complejos con las proteínas.

Difusión anómala de los átomos intrínsecos.

Uno de los métodos consiste en utilizar dispersores anómalos presentes de forma natural en las proteínas, como los metales de transición en las metaloproteínas, o el azufre presente en todas las proteínas que contienen metioninas o cisteínas. Esto presenta la ventaja de disponer de sitios totalmente ocupados y de no necesitar, por tanto, la preparación del derivado.

Metaloproteínas.

Más del 20% de las proteínas contienen intrínsecamente metales como Cu, Fe, Mn, Ni o Zn (Leslie *et al.*, 2004). Los picos de absorción K de un gran número de estos metales están situados entre 0.7 y 2.0 Å con los valores de f'' del orden de 4 e⁻. Así, estas proteínas se prestan a los experimentos de MAD o SAD al pico K, utilizando radiación sincrotrón. Ciertos elementos presentan una dispersión anómala no despreciable para la radiación K α del cobre y se prestan entonces a experimentos SAD con una fuente de radiación de laboratorio. Con una fuente de rayos X de longitud de onda variable se puede obtener un espectro de fluorescencia de emisión recogido a una energía de activación superior a la energía de los picos de absorción de los elementos presentes corrientemente. De este modo se puede detectar e identificar los metales presentes eventualmente en la proteína de interés.

Método SAD utilizado con la señal anómala del azufre.

Todas las proteínas que contienen cisteínas o metioninas contienen azufre. Estos dos aminoácidos representan de media el 4% de los aminoácidos. Para el método SAD, podemos utilizar la dispersión anómala cerca del pico (el pico K del azufre está a 5.02 Å). Fue en 1981 cuando se consiguió resolver la primera estructura de una macromolécula a partir de la señal anómala del azufre (Hendrickson & Teeter, 1981). Desde 1999, numerosas estructuras se han resuelto por el método SAD sirviéndose únicamente de la señal anómala del azufre (Ramagopal *et al.*, 2003b). Sin embargo, este método se aplica, por el momento, a los cristales que difractan muy bien. En efecto, la dispersión anómala del azufre a las longitudes de onda habituales es muy débil (para $\lambda_{Cu} = 1.5$ Å y $\lambda_{Cr} = 2.3$ Å los valores respectivos de f'' son 0.56 e⁻ y 1.14 e⁻). Una longitud de onda óptima con una línea de luz no optimizada para las grandes longitudes de onda sería de $\lambda \sim 1.8$ Å (Leslie *et al.*, 2004). Estas condiciones conducen a relaciones de Bijvoet muy débiles, alrededor del 1% (Leslie *et al.*, 2004). Utilizar esta señal anómala tan débil necesita datos de una gran exactitud, los datos deben entonces ser registrados con una redundancia muy elevada. Esto demanda cristales que difracten bien y que resistan los daños por radiación. El desarrollo del método va acompañado del desarrollo de fuentes de radiación sincrotrón de líneas de luz adaptadas a la utilización de altas longitudes de onda así como de la metodología adecuada para el tratamiento de los datos.

Método SAD utilizado con la señal anómala del fósforo.

Los ácidos nucleicos contienen una cantidad constante de fósforo (1 átomo de P por base). Del mismo modo que la obtención de fases con azufre se aplica a proteínas, la señal anómala del fósforo puede servir para resolver la estructura de ácidos nucleicos o de sus complejos con las proteínas. Esto conduce a una relación de Bijvoet del 2% para un polinucleótido con la radiación $K\alpha$ del cobre (Dauter, 2002).

Introducción de átomos pesados por inmersión o cocrystalización.

Para obtener derivados para los métodos MAD o MIR, utilizamos clásicamente las sales de los elementos pesados Pt, Au, Hg, U o Pb, las denominadas "las siete mágicas": K_2PtCl_4 , $KAu(CN)_2$, K_2HgI_4 , $UO_2(C_2H_3O_2)_2$, $HgCl_2$, $K_3UO_2F_5$ o el sulfato de paracloromercuribenzoico (PCMBs) (Garman & Murray, 2003). La línea LIII de cada uno de los elementos es muy accesible con un valor de f'' en torno a 10 e⁻, salvo para el platino que presenta una línea blanca en su pico con un valor de f'' superior a 20 e⁻. Si utilizamos una fuente de rayos X de laboratorio con un ánodo de cobre, los valores de f'' están comprendidos entre 7 e⁻ y 16 e⁻, este último valor obtenido con el uranio. Los dispersores pesados son introducidos de forma habitual en los cristales por inmersiones relativamente largas (de horas a varios días) de los cristales nativos en soluciones poco concentradas de sal (0.1-10 mM). El inconveniente principal del método es el modo de fijación, generalmente aleatorio, de los distintos centros metálicos que hace necesario repetir numerosos ensayos con diferentes compuestos para obtener cristales derivados. Sin embargo, el conocimiento de la secuencia de la proteína puede ayudar en la elección de la sal de átomo pesado a probar (ver sección 1.2.4.1). El problema radica en que la utilización de estas sales de átomos pesados no siempre es fácil puesto que estas sales son generalmente poco solubles e inestables en las soluciones acuosas, son incompatibles con las numerosas condiciones químicas de cristalización, son a menudo mal toleradas por los cristales de proteína y son altamente tóxicas. Desde hace poco, las inmersiones mucho más cortas (10 min-2h) son practicadas con esos mismos compuestos a concentraciones más elevadas (>10 mM), lo que parece dar derivados de mejor calidad cristalina, con un mejor isomorfismo que con inmersiones largas (Dauter, 2002). Las inmersiones cortas se practican igualmente con los iones que no forman parte de la unión específica con la proteína, como los iones metálicos no complejos, Ru o Cs, o los halogenuros. Así, las inmersiones rápidas (10-30s) en las soluciones de yoduro o de

yodo I₂/KI pueden producir derivados que incorporan los iones I⁻, I₃⁻ o I₅⁻ (Evans & Bricogne, 2002).

El empleo de yoduro como elemento pesado tiene muchas ventajas, que se pueden resumir en que no es tóxico, no produce alteraciones significativas en la estructura de la proteína ya que reemplaza moléculas de agua, es muy soluble, es barato, no altera las condiciones de cristalización y fundamentalmente, puede emplearse para resolver las estructuras sin necesidad de acudir a la radiación sincrotrón. A diferencia de lo que ocurre cuando empleamos las soluciones de átomo pesado convencionales, las sales de haluros originan una subestructura compuesta por un elevado número de átomos de yodo, con tasas de ocupación variables, como consecuencia del mencionado reemplazo de moléculas de agua de la esfera de solvatación y la localización de esta subestructura sólo se ha podido resolver gracias al desarrollo en los programas de búsqueda de átomos pesados como SHELXD. Las técnicas de obtención de fases que se aplican a los derivados de yoduro son diversas, MIR, MIRAS, SIR, SIRAS y SAD. La única técnica que no se puede aplicar es MAD, ya que el eje de absorción se encuentra a una longitud de onda fuera del rango disponible en el sincrotrón. Sin embargo, los derivados de bromuro pueden servir para los experimentos MAD (línea K a $\lambda = 0.92 \text{ \AA}$). En cualquier caso, la señal anómala del yodo a la longitud de onda de la radiación CuK α es alta, por lo que es posible obtener fases habiendo recogido los datos de difracción en un difractómetro con fuente convencional (el valor de f'' es de $7e^-$ para el yodo con la radiación K α del cobre)

Los iones de los lantánidos son muy utilizados por las proteínas que unen Ca. En efecto, el tamaño de sus radios iónicos y su afinidad química son muy similares de tal modo que los iones lantánidos reemplazan fácilmente el calcio de sus sitios de fijación a la proteína. En tal caso, podemos aprovecharnos de la elevada dispersión anómala de los lantánidos bien a partir de su línea de absorción LIII, utilizando radiación sincrotrón, o bien a partir de una fuente convencional de laboratorio (Kahn *et al.*, 1985).

Utilización de gases nobles: xénon y krypton.

Podemos obtener cristales derivados para la obtención de fases por métodos SIRAS o MIRAS, introduciendo gases nobles, xenón o kriptón, en los cristales nativos (Schiltz *et al.*, 2003). Efectivamente, sometiendo los cristales a una atmósfera de gas a presión (1-100 bar) en una

celdilla apropiada. Estos gases son relativamente solubles en agua, difundiendo rápidamente por los canales de solvente del cristal. La unión de los átomos de Xe o de Kr pone en juego las débiles fuerzas de las cavidades hidrofóbicas, de los sitios enzimáticos o de las cavidades intermoleculares. Las ventajas de este método de obtención de cristales derivados son, el isomorfismo elevado de los cristales resultantes de las uniones de baja afinidad, el número elevado de sitios potenciales de fijación y su complementariedad en relación a los sitios de fijación de las sales clásicas, de naturaleza fundamentalmente polar. Por último, podemos variar el número y la ocupación de los sitios variando la presión del gas. La probabilidad de fijación suficiente es elevada, >30-40% (Schiltz *et al.*, 2003), y, cuando no hay fijación, los sitios potenciales pueden ser introducidos por las operaciones de mutagénesis dirigida. La reversibilidad de la fijación permite obtener diferentes derivados a partir de un cristal único. El registro de los datos puede hacerse a temperatura ambiente o a temperatura criogénica. La dispersión anómala del Xe es relativamente elevada para los experimentos en laboratorio ($f''_{\text{CuK}\alpha} = 7.4 \text{ e}^-$).

1.2.5. UTILIZACIÓN DE COMPLEJOS LANTÁNIDOS PARA OBTENER DERIVADOS.

Otra aproximación para introducir los átomos pesados en los cristales de proteína consiste en utilizar los complejos quelatos del elemento elegido. Esta aproximación permite separar las propiedades físicas del átomo pesado, sus propiedades de dispersión de los rayos X, de las propiedades químicas del complejo que impone su comportamiento en el cristal. Las primeras son esencialmente determinadas por el átomo pesado elegido, mientras que las segundas vienen impuestas por el ligando coordinado al metal pesado. Si la constante de asociación del complejo formado por el ligando y el átomo pesado es elevada, esta aproximación puede permitir la fijación de los átomos pesados utilizando la interacción del ligando con la molécula de la proteína. De este modo, podemos concebir una gran variedad de ligandos. No obstante, el tamaño del complejo resultante tiene que estar limitado por el tamaño habitual de los canales de solvente en los cristales de proteína. El complejo resultante debe ser compatible con las condiciones físico-químicas utilizadas para cristalizar las proteínas, y suficientemente estable en esas condiciones.

Por ejemplo, es posible concebir ligandos diseñados para formar interacciones covalentes con ciertos residuos proteicos, tales como los complejos tiol reactivos que se unen a cisteínas libres (Purdy *et al.*, 2002). Tal reacción conduce frecuentemente a un fuerte descenso de la solubilidad de la proteína debido a las fuertes modificaciones de las condiciones de cristalización.

Es igualmente posible concebir complejos cuyas propiedades físico-químicas permitan detectar fácilmente su fijación en el cristal: este sería el caso por ejemplo de los complejos coloreados que permiten estimar su fijación mediante la coloración de los cristales derivados obtenidos (Chesne, 2002).

Particularmente interesante ha resultado ser una clase particular de complejos lantánidos que proporcionan excelentes derivados para resolver estructuras mediante dispersión anómala (Figura 1.2.9) (Girard *et al.*, 2003a; Girard *et al.*, 2003b; Girard *et al.*, 2002). Todos estos ligandos se coordinan a un ion de Gd^{3+} pero pueden también coordinarse a otros iones, Ln^{3+} , de los lantánidos. Los ligandos son derivados de un macrociclo de tetra-azaciclododecano salvo el ligando lineal DTPA (ácido dietilentriaminopentaacético) y su derivado DTPA-BMA, cuyos grupos carboxílicos están desprotonados en los complejos. Por lo tanto, los complejos resultantes son neutros excepto en el caso de Gd-DOTA, Gd-DOTA-BOM y Gd-DOTMA, que poseen una carga global neta negativa, y Gd-DTPA, que posee una doble carga negativa. Los valores de pKa de los complejos se sitúan entorno a 2 (Bianchi *et al.*, 2000) por lo que en las condiciones físico-químicas de cristalización habituales no habrá protonación ninguna. Los complejos son altamente solubles en agua, su solubilidad se encuentra entorno a 1M, salvo para Gd-DOTA-BOM cuya solubilidad es aproximadamente 0.6M.

Las constantes de estabilidad termodinámica de los diferentes complejos son muy elevadas, oscilando entre $\log K \sim 22$ (Gd-DO3A) y $\log K \sim 28$ [Gd-DOTA]⁻ (a título comparativo, para el complejo [Gd-EDTA]⁻, $\log K = 17.35$) (Bianchi *et al.*, 2000)). Dichas constantes nos indican que se trata de complejos muy estables y que no se disocian en las condiciones habituales de cristalización (sólo medios muy ácidos ($pH \leq 2$) provocan su disociación).

1.2.5.1. Presentación de una clase particular de complejos de Gadolinio.

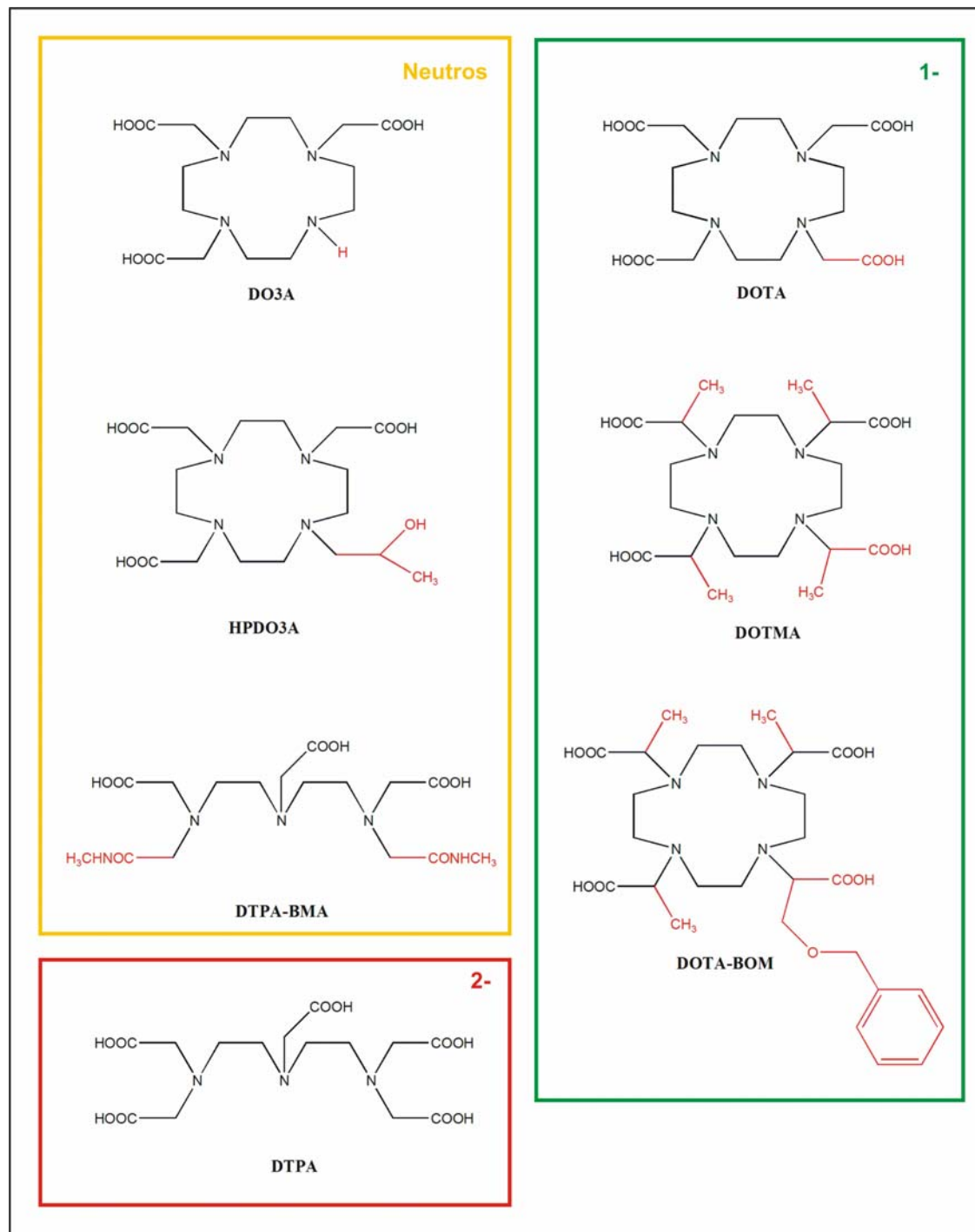


Figura 1.2.9: Ligandos correspondientes a los siete complejos de Gd presentados en este estudio. DO3A, ácido 1,4,7,10-tetraazaciclododecano-1,4,7-triacético; DOTA, ácido 1,4,7,10-tetraazaciclododecano-1,4,7,10-tetraacético; HPDO3A, ácido 10-(2-hidroxipropil)-1,4,7,10-tetraazaciclododecano-1,4,7-triacético; DOTMA, ácido a,a',a''-tetrametil-1,4,7,10-tetraazaciclododecano-1,4,7,10-tetraacético; DTPA-BMA, ácido bismetilamida dietilenetriaminopentaacético; DTPA, ácido dietilenetriaminopentaacético; DOTA-BOM, ácido (fenilmetoxi)metil-1,4,7,10-tetraazaciclododecano-1,4,7,10-tetraacético.

1.2.5.2. Química de coordinación del Gd(III).

El estado de oxidación más estable de los lantánidos es el estado Ln(III) y para el ion Gd^{3+} , el número de coordinación más común es el nueve. Cuando los ligandos indicados en la figura 1.2.10 se coordinan a los iones Ln^{3+} sus cadenas se repliegan en torno al centro metálico para formar una especie de jaula (figura 1.4.2), confiriendo a esos complejos una estabilidad muy elevada. Para dar un orden del tamaño de estos complejos, la distancia entre los dos átomos más alejados del complejo varía entre $\sim 8.5 \text{ \AA}$ (Gd-DO3A) y 10 \AA (Gd-DOTMA, Gd-DTPA-BMA).

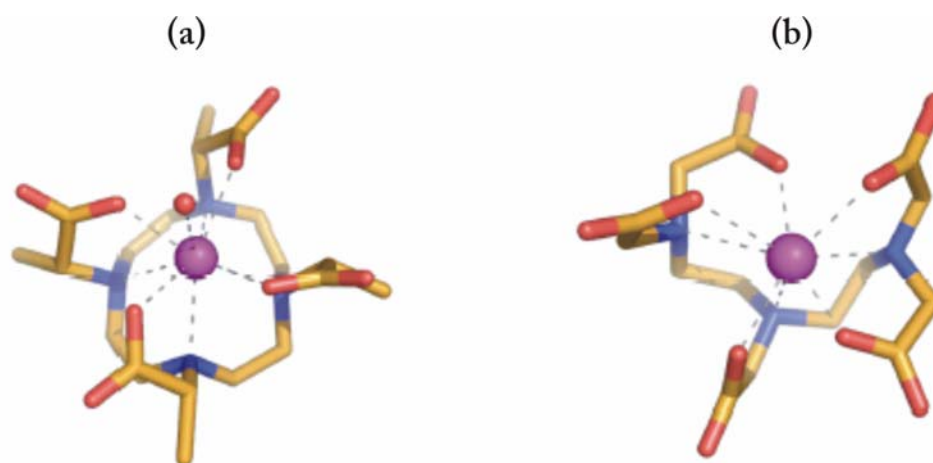


Figura 1.2.10: Estructura tridimensional de los complejos de Gd empleados. a) Gd-DOTMA tiene como ligando en la novena posición de coordinación una molécula de agua. b) En el caso de Gd-DTPA, la novena posición de coordinación la puede aportar un oxígeno carboxílico de un residuo proteico (no aparece representado).

En los complejos utilizados, el ion Gd^{3+} se encuentra unido, por una parte, a los cuatro átomos de nitrógeno del macrociclo o a los tres átomos de nitrógeno del “pseudo-macrociclo” en los ligandos lineales, y de otra parte, a los átomos de oxígenos de los ligandos (figura 1.2.10). La coordinación de estos ligandos octadentados (heptadentado en el caso del Gd-DO3A) deja una vacante en la esfera de coordinación del metal (2 en Gd-DO3A) la cual va a determinar su modo de fijación a las proteínas. Estas posiciones vacantes pueden ser ocupadas por moléculas de agua, por centros donadores de residuos proteicos o por el oxígeno de un ligando de otra molécula de complejo próxima.

1.2.5.3. La dispersión anómala de los lantánidos.

Los lantánidos (Z=57-71) presentan todos una fuerte dispersión anómala de $f'' \sim 28 e^-$ debido a una fuerte resonancia (“línea blanca”) en su pico de absorción LIII, como muestra la Tabla 1.2.2. Los picos están situados a longitudes de onda inferiores a 2 \AA para los lantánidos que van desde el Nd al Lu (tabla 1.2.3). Con la radiación $\text{CuK}\alpha$, los lantánidos que van del La al Gd tienen valores elevados de f'' .

Elemento	La	Ce	Pr	Nd	Pm	Sm	Eu
Z	57	58	59	60	61	62	63
Pico L_{III} [\AA]	2.26	2.16	2.08	1.99	1.92	1.84	1.78
f''_{PicoLIII} [e^-]	28.0	28.0	28.0	28.0	28.0	28.0	28.0
$f''_{\text{CuK}\alpha}$ [e^-]	9.6	10.3	11.0	11.7	12.4	13.3	11.0

Elemento	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
Z	64	65	66	67	68	69	70	71
Pico L_{III} [\AA]	1.71	1.65	1.58	1.54	1.48	1.43	1.39	1.34
f''_{PicoLIII} [e^-]	28.0	28.0	28.0	28.0	28.0	28.0	28.0	28.0
$f''_{\text{CuK}\alpha}$ [e^-]	12.0	8.0	8.0	4.0	4.0	5.0	5.0	5.0

Tabla 1.2.2: Longitudes de onda de los picos de absorción L_{III} (Rieck, 1985) y parte imaginaria f'' de los factores de dispersión atómica a la longitud de onda $\lambda_{\text{CuK}\alpha} = 1.54 \text{ \AA}$ (Ibers *et al.*, 1985) para los diferentes lantánidos.

El gadolinio presenta la ventaja de combinar una señal anómala elevada ($f'' = 12 e^-$) con la radiación $\text{K}\alpha$ del Cu y un pico LIII a la longitud de onda de 1.711 \AA , relativamente accesible.

1.2.5.4. Obtención de cristales derivados a partir de complejos de gadolinio

Inserción de complejos de gadolinio en los cristales

Las propiedades de los complejos descritos en la sección 1.2.5.1, es decir, su estabilidad en la mayoría de las condiciones físico-químicas de cristalización (pH, agente precipitante, tampón), su elevada solubilidad y su ausencia de toxicidad (algunos se emplean como agentes de

contraste en técnicas de diagnóstico por imagen), les convierten en excelentes candidatos para introducirlos como átomos pesados en los cristales de proteína.

Los derivados correspondientes se obtienen empleando soluciones concentradas de complejos lantánidos en las que se sumergen los cristales de proteína (inmersión) o bien se adicionan desde el principio en la propia gota de cristalización (cocrystalización) sin alterar significativamente el proceso de cristalización. Una particularidad de estos complejos es la fácil reversibilidad del proceso de fijación a las moléculas biológicas con las cuales establece uniones de tipo no covalente. Para obtener una fijación suficiente se utilizan concentraciones elevadas del complejo, del orden de 50-100 mM que, sin embargo, toleran las proteínas durante la inmersión o la cocrystalización para proporcionar cristales isomorfos a los nativos.

Cocrystalización

La cocrystalización reduce las manipulaciones del cristal y permite al complejo fijarse poco a poco durante el crecimiento de los cristales, evitando así el posible deterioro asociado a la inmersión, en particular en el caso de cristales con un bajo contenido en solvente.

Las condiciones de obtención de cristales derivados corresponden a las condiciones en las cuales se obtienen los cristales nativos con la adición de en torno a 100 mM de complejo en la gota de cristalización. La adición del complejo puede provocar ligeros cambios de solubilidad, por lo que también es aconsejable variar la concentración del agente precipitante. En caso de fuerte solubilización de la proteína (lo cual es posible por la adición de los complejos Gd-DOTA-BOM y Gd-HPDO3A) es aconsejable bajar la concentración del complejo a 50 mM. En casos excepcionales, la adición del complejo en la gota de cristalización impide la cristalización y provoca la precipitación de la proteína. De este modo podemos suponer que tal cambio en las condiciones de cristalización puede indicar una fijación notable del complejo. En ese caso, hay muchas posibilidades de obtener buenos derivados con los mismos complejos por inmersión de los cristales nativos.

Inmersión

Si disponemos de cristales nativos de la proteína y poca cantidad de la misma, el medio más cómodo para obtener cristales derivados consiste en sumergir los cristales nativos. Si no

podemos añadir proteína en la gota donde practicamos la inmersión, es conveniente aumentar la cantidad de precipitante en la gota de inmersión con el fin de evitar la disolución del cristal. Típicamente, la duración de la inmersión es relativamente corta, variando de segundos a horas. Los complejos son generalmente bien tolerados por los cristales, lo cual puede ser debido a su débil afinidad por las proteínas y su electroneutralidad. Si los cristales no soportan la inmersión, es conveniente bajar la concentración del complejo o reducir la duración de la inmersión.

Para los cristales especialmente sensibles a las manipulaciones, podemos proceder a la inmersión mediante la adición directa de la solución del complejo a la propia gota del cristal: la elevada solubilidad de los complejos permite añadir pequeños volúmenes de solución.

Crioprotección de los cristales derivados

La crioprotección de cristales derivados se realiza de forma similar a la crioprotección de los cristales nativos, es decir, por inmersión en la solución crioprotectora habitual. Sin embargo, en este caso, y debido fundamentalmente a una dilución de la concentración de la solución del átomo pesado, la inmersión en la solución crioprotectora puede provocar una disminución de la afinidad del compuesto de átomo pesado por la proteína revirtiendo el equilibrio de dicha unión. Este fenómeno se le conoce como “back-soaking”. Es decir, se produce la salida del compuesto de átomo pesado de su sitio de fijación en el cristal al solvente disminuyendo la calidad del derivado obtenido. Para minimizar este problema el paso de inmersión en la solución crioprotectora ha de hacerse lo más rápido posible de tal manera que acto seguido congelemos el cristal en nitrógeno líquido evitando así la difusión del átomo pesado al solvente y dejando preparado el cristal para medir. Otro de los métodos adicionales que se pueden emplear para evitar el back-soaking es la inclusión en la solución crioprotectora de una concentración de átomo pesado similar a la del átomo pesado unido a la proteína en el cristal evitando de este modo la difusión por diferencia de concentraciones.



2. OBJETIVOS

2. OBJETIVOS.

Neumococo, responsable de un elevado número de procesos infecciosos e invasivos severos en humanos, es el único microorganismo conocido que presenta un requerimiento nutricional absoluto por colina. Este aminoalcohol es incorporado en forma de fosforilcolina a los ácidos (lipo)teicoicos que forman parte de la envuelta celular y desempeña un papel fundamental en la fisiología y la patogenicidad de neumococo mediante su unión a los componentes séricos del sistema inmune y actuando como punto de anclaje para la familia de proteínas de superficie de unión a colina (CBPs). En este contexto, es interesante el estudio estructural y funcional de la proteína CbpF, máxime teniendo en cuenta que se trata de una de las proteínas más abundantes en la pared celular de neumococo. Además cabe destacar que, a la dificultad de la cristalización de proteínas modulares, debemos sumar la inexistencia de información alguna respecto a la funcionalidad de la proteína objeto de estudio.

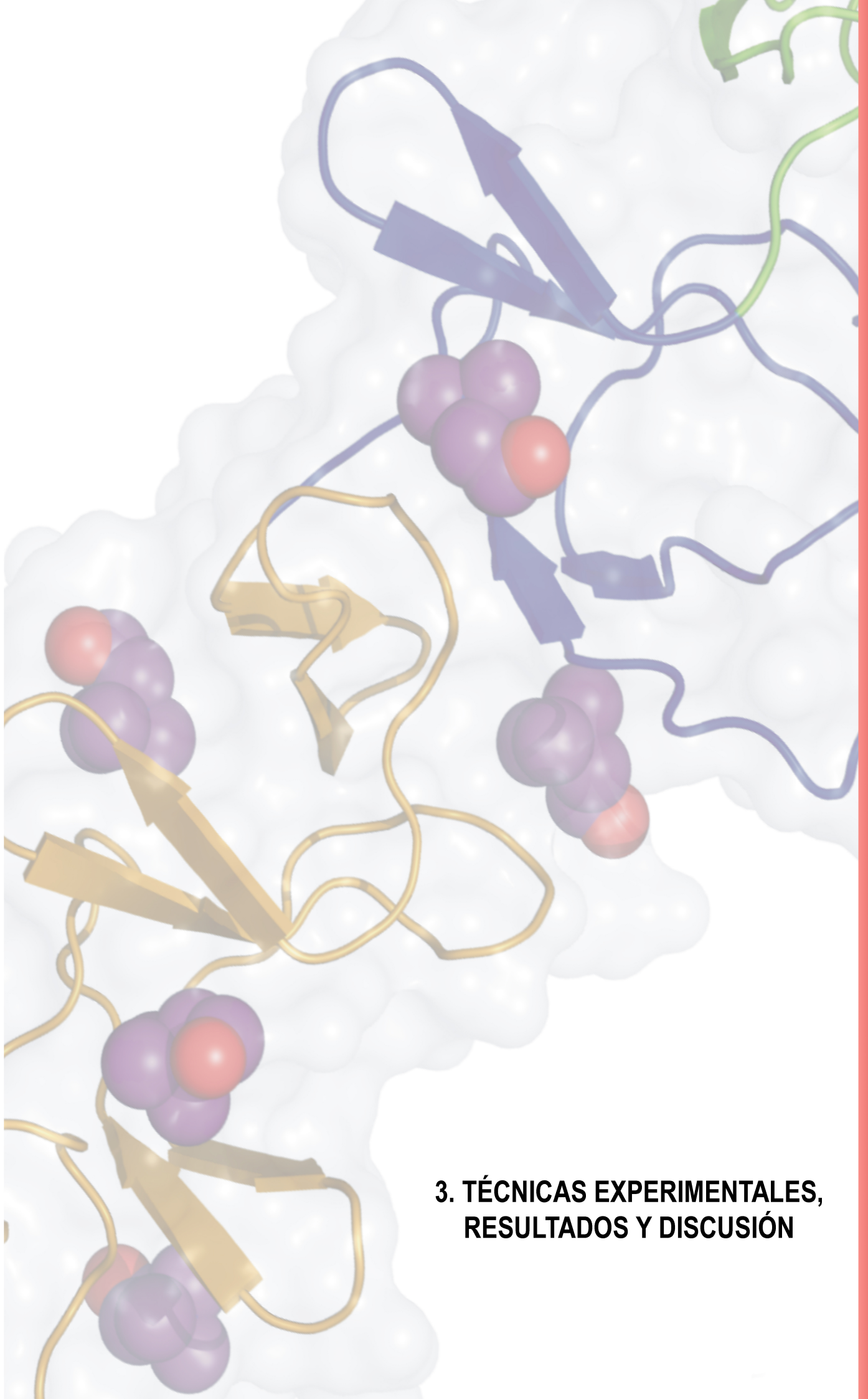
De forma paralela, y empleando a la proteína CbpF como modelo, se abordó un estudio metodológico con una serie de complejos novedosos de Gd. Tras la producción de la proteína pura y estable, de las diferentes etapas que transcurren de la proteína a la estructura, las etapas críticas son la cristalización y la obtención de cristales derivados que permitan determinar las fases experimentales. Es esta última etapa del proceso en la que nos centramos y más concretamente en la obtención de cristales derivados para los métodos de obtención de fases que emplean la dispersión anómala. En definitiva, el objetivo es aportar una herramienta eficaz para obtener cristales derivados como alternativa de otros métodos bien establecidos.

Resumiendo, los objetivos concretos que se proponen en esta tesis pueden sintetizarse en:

- Determinar la estructura y función de CbpF:
 - Expresión y purificación
 - Cristalización
 - Recogida y análisis de datos de difracción de rayos X
 - Resolución estructural
 - Análisis de la estructura
 - Análisis funcional

Objetivos

- Analizar la obtención de fases en cristalografía macromolecular mediante el uso de complejos lantánidos:
 - Análisis comparativo del poder de obtención de fases de los distintos complejos de Gd
 - Análisis comparativo de la cocrystalización frente a la inmersión como método de obtención de derivados
 - Análisis del modo unión de los diferentes complejos de Gd
 - Análisis comparativo de los datos obtenidos por radiación sincrotrón frente al ánodo rotatorio



3. TÉCNICAS EXPERIMENTALES, RESULTADOS Y DISCUSIÓN

3. TÉCNICAS EXPERIMENTALES, RESULTADOS Y DISCUSIÓN.

3.1. ESQUEMA GENERAL EN CRISTALOGRAFÍA DE RAYOS X.

La Cristalografía de rayos X es una de las técnicas físicas más potentes para la obtención de información estructural de la materia ordenada. Nos permite obtener una imagen tridimensional, a escala atómica, del material cristalizado y, aplicado a las proteínas, proporciona información de la estructura tridimensional, plegamiento y detalles de un gran significado físico-químico (distancias, ángulos de enlace o empaquetamiento cristalino) que constituyen información clave para entender la función biológica. La resolución estructural mediante Cristalografía de rayos X puede dividirse en distintas etapas, las cuales se muestran en la figura que aparece en la figura 3.1.1.

3.2. EXPRESIÓN Y PURIFICACIÓN.

3.2.1. INTRODUCCIÓN.

Dadas las características de los estudios cristalográficos, desde el punto de vista preparativo se requiere *a priori* una cantidad abundante de la muestra proteica, así como un elevado grado de pureza y homogeneidad en la misma.

En lo referente al problema de la cantidad se ha resuelto en gran medida con los métodos de ingeniería genética que hacen posible clonar y sobreexpresar genes en células bacterianas o eucariotas obteniendo cantidades importantes de proteína. La producción masiva de una proteína concreta puede llegar a ser de un cuarto del total expresado, pero a altas concentraciones intracelulares las proteínas pueden formar agregados, cuerpos de inclusión o desnaturalizarse. Por tanto, los niveles de producción deben ser optimizados para obtener la mayor proporción de la proteína deseada, funcionalmente activa y en su conformación natural. Sin embargo actualmente, en el caso de la obtención de cuerpos de inclusión en una proteína expresada en *E. coli*, existen cada vez más ejemplos que sugieren el empleo de proteína sometida a procesos de plegamiento para los ensayos de cristalización.

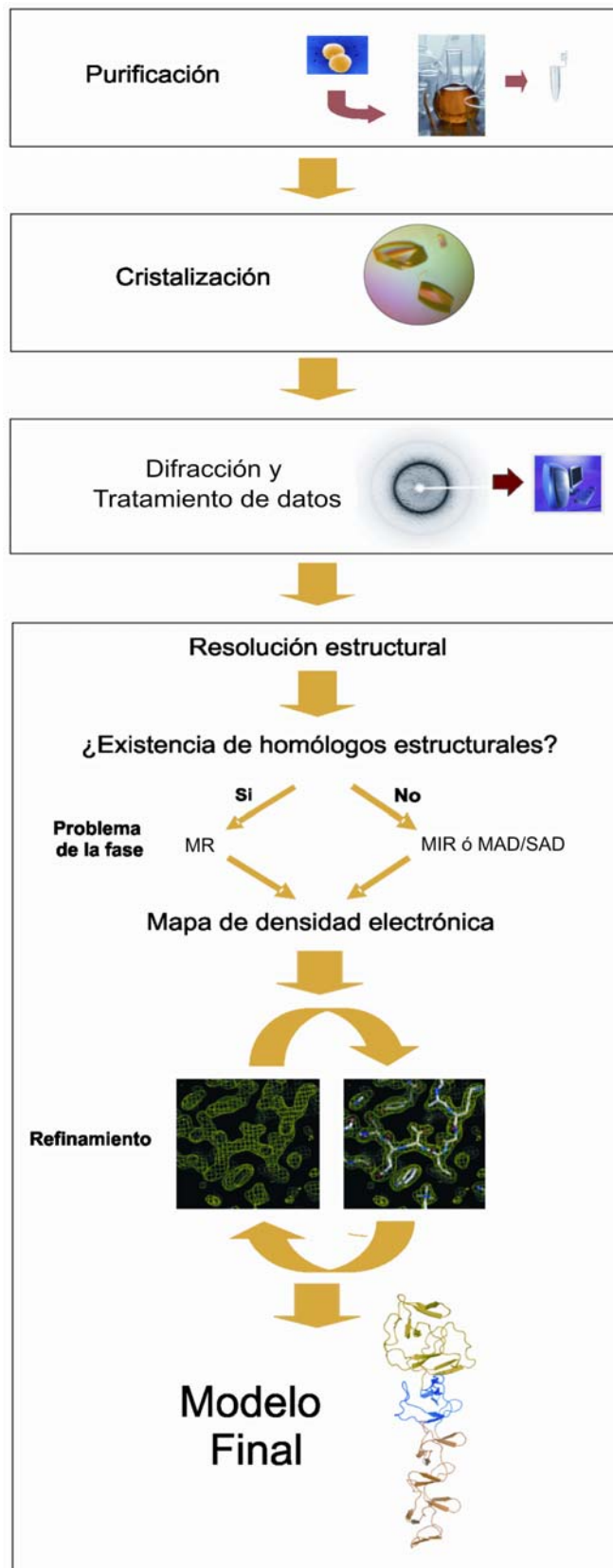


Figura 3.1.1: Esquema de las distintas etapas que configuran un estudio estructural por Cristalografía de rayos X.

La purificación de una proteína tiene como objetivo final obtener el material deseado con el máximo rendimiento y utilizando el proceso más corto y, a ser posible, más económico. Para definir una buena estrategia es recomendable describir previamente el escenario básico de la purificación, es decir, cuál será la aplicación del producto purificado (que determinará el grado de pureza), cuál es el material de partida disponible y cómo debe manejarse, qué pureza se requiere, qué se debe eliminar, qué puede quedar, cuál será la escala final de la purificación y, por supuesto, de qué tipo de equipamiento se dispone o se necesita para la purificación. Obviamente, toda información relativa a las propiedades de la proteína y sus contaminantes será de gran ayuda en todo el proceso de purificación. El punto final es la optimización del esquema de purificación.

Si bien la estrategia de purificación en tres etapas se describió inicialmente como una ayuda para el desarrollo de procesos de purificación de proteínas terapéuticas, se ha demostrado también válida para todo tipo de proteínas y a escala de laboratorio. Las tres fases de las que consta son: captura, purificación y pulido (Figura 3.2.1).

- En la fase inicial de **captura**, los objetivos son aislar, concentrar y estabilizar la proteína a purificar. Para ello se usan columnas cromatográficas con gran cantidad de carga y buenas propiedades de flujo puesto que se trata de la purificación inicial a partir de un material crudo o clarificado.
- Durante la **purificación intermedia**, el objetivo sería la eliminación de la mayor parte de los contaminantes (proteínas, ácidos nucleicos, endotoxinas, virus). En este punto es importante usar columnas de máxima resolución, ya que los contaminantes que quedan tienen propiedades cromatográficas similares a las de la proteína diana. Esto se lleva a cabo combinando técnicas de distinta selectividad que se usan en modo de alta resolución. El flujo y la capacidad de carga tienen que controlarse para que la resolución sea máxima.
- En la etapa de **pulido**, el objetivo es conseguir una alta pureza eliminando cualquier traza de impurezas o productos relacionados con la sustancia de interés que no se hayan conseguido eliminar en la fase intermedia.

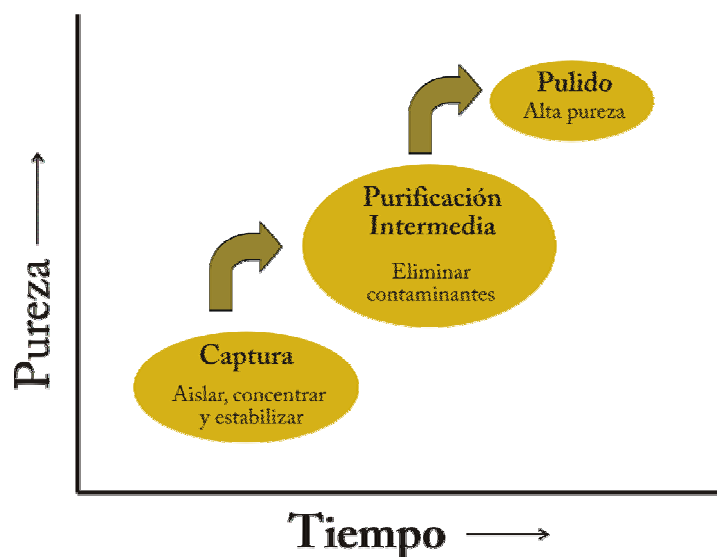


Figura: 3.2.1: Secuencia de la estrategia de purificación en tres fases.

3.2.2. EXPRESIÓN Y PURIFICACIÓN DE CbpF.

Una aproximación bioquímica para identificar proteínas con alta afinidad por la colina que contiene la pared celular de neumococo, usando una cepa de neumococo mutante LytA-, permitió la extracción de cuatro proteínas que permanecían firmemente unidas a la capa de peptidoglicano (García *et al.*, 1999). Tres de ellas resultaron ser LytB, LytC y Pce, que junto con LytA, constituyen el conjunto de mureín hidrolasas bien caracterizadas en neumococo. El análisis N-terminal de la cuarta proteína extraída a partir de la envoltura celular (NTTGGRFVDKDRNK) reveló la existencia de una nueva proteína denominada CbpF.

El gen spr0337 de la cepa R6 de neumococo, que codifica la proteína CbpF, fue clonado dentro del plásmido pIN-III-A3 (Masui *et al.*, 1984). El fragmento PCR correspondiente fue amplificado con el oligonucleótido 5'-**ggaattc**atgAATACCACAGGTGGCCGATTTG-3' y 5'-**cgggatcc**CGCAATCGCTTCTTTCATTATTG-3' complementario a los extremos 5' y 3' de la secuencia de DNA que codifica para la proteína CbpF madura. Las letras en minúscula en los oligonucleótidos indican la extensión para introducir los sitios de restricción adecuados (EcoRI and BamHI). El codón de inicio y los codones de parada complementarios se muestran en negrita. El plásmido recombinante resultante (pAPM40) contiene el gen de CbpF y su secuencia de DNA se confirmó con un secuenciador automático de DNA Abi Prism 3700 (Applied Biosystems). Todos los cebadores se sintetizaron en un sintetizador de oligos Beckman Oligo

1000M. Las enzimas de restricción y otros enzimas que modifican DNA fueron proporcionados por Amersham. Para la purificación de la forma madura, CbpF fue expresada en las cepas DH5 α de *Escherichia coli* transformadas con pAPM40 y crecidas en medio de cultivo Terrific Broth (Sambrook *et al.*, 1989) conteniendo ampicilina (1000 μ g/ml) a 37 °C con aireación. Cuando el cultivo alcanzó una densidad óptica de 0.8 se indujo la expresión en la fase de crecimiento logarítmico con 50 μ M isopropiltio- β -D-galactopiranosido y la incubación se continuó a 25 °C durante 16 horas para minimizar la presencia de cuerpos de inclusión. Las células se recogieron en la fase estacionaria de crecimiento mediante centrifugación a 10.000 x g durante 15 minutos, resuspendiéndose posteriormente en tampón fosfato sódico 20 mM, pH 6,9. A continuación, se lisaron las células utilizando una prensa francesa (1100 psi) y el extracto celular se centrifugó a 15.000 x g durante 15 minutos adicionando sulfato de estreptomycin con el fin de eliminar los restos de ADN por precipitación (Weissborn *et al.*, 1994). CbpF se purificó a partir de la fracción soluble mediante cromatografía de afinidad en columnas de DEAE-celulosa (Sánchez-Puelles *et al.*, 1990). Todo el proceso se llevó a cabo a 4 °C. Tras agregar NaCl a una concentración final de 1M y sembrar el extracto en la columna, CbpF se purificó siguiendo el procedimiento descrito a continuación:

- 1) Lavado exhaustivo de la columna con tampón A (fosfato 20 mM pH 6.9 y NaCl 1.5 M) para eliminar impurezas retenidas en la columna.
- 2) Lavado con tampón B (fosfato 20 mM pH 6.9 y NaCl 0.1 M) hasta conseguir la eliminación total de ADN.
- 3) Elución de CbpF con 4% colina (280 mM) en tampón B, recogiendo fracciones de 1 ml. El exceso de colina provoca la separación de CbpF de la columna dada su mayor afinidad por este ligando.

Registramos las distintas fracciones de elución mediante espectros de absorción barriendo longitudes de onda comprendidas entre 240 y 360 nm comprobando de este modo la pureza y concentración de las mismas. Del mismo modo, la homogeneidad de las muestras se verificó mediante electroforesis en geles desnaturizantes de poliacrilamida (SDS-PAGE) al 12% (Laemli *et al.*, 1970; Fig. 3.2.2) y espectrometría de masas (MALDI-TOF; Fig. 3.3.1). La tinción de las proteínas en el gel se realizó con Coomassie.



Figura 3.2.2: Electroforesis en SDS-PAGE al 12%. Gel correspondiente a las distintas fracciones obtenidas durante la purificación. La movilidad de los marcadores de peso molecular (Mw) se indican por orden creciente: Miosina, β -galactosidasa, fosforilasa b, seroalbúmina bovina, ovoalbúmina, anhidrasa carbónica e inhibidor de la tripsina de soja.

Las fracciones con alta pureza y concentración se mezclaron, dializándose extensivamente frente a un tampón Tris (20 mM, pH = 8.0) que contenía 140 mM de cloruro de colina. En este caso, fue necesario alcanzar esta elevada concentración de colina (140 mM) en el tampón de diálisis debido a que a menores concentraciones la proteína precipitaba rápidamente. La proteína se concentró a 277 K con un concentrador Amicon, YM-10 a aproximadamente 8 mg ml⁻¹. La concentración final de proteína fue determinada por espectrometría asumiendo un coeficiente de extinción molar de 152640 M⁻¹ cm⁻¹ a 280 nm.

3.3. ESPECTROMETRÍA DE MASAS.

3.3.1. CONSIDERACIONES GENERALES.

La espectrometría de masas es una técnica basada en la ionización de las moléculas y en su separación en función de la relación masa/carga (m/z). La ionización de las moléculas puede realizarse mediante diferentes técnicas:

- Electronebulización o electrospray (ESI): La ionización se produce mediante la aplicación de una diferencia de potencial (3-5 kV) al capilar por donde se introduce la disolución al analizador, que actúa como contraelectrodo. Con ayuda de nitrógeno gaseoso se consigue una nube de gotas cargadas que van siendo desolvatadas hasta obtener los correspondientes iones pseudomoleculares.
- Ionización y desorción por láser inducida por matriz (MALDI): La ionización de las moléculas de la muestra se consigue tras mezclar éstas con una matriz formada por una

sustancia capaz de absorber en el UV. La mezcla sólida formada se irradia con un láser cuya energía es absorbida por la matriz, lo que provoca su ionización y la de la proteína. Una vez ionizadas y en estado gas, las moléculas se hacen pasar a un analizador donde se produce la separación en función de la relación masa/carga.

En la actualidad se utilizan dos tipos de analizadores distintos para proteínas:

- Analizador de trampa iónica (IT): Cuando la ionización se ha realizado por ESI se suele utilizar un analizador de trampa iónica (IT). Los iones se confinan durante un cierto tiempo (ms) en el analizador de campos eléctricos de corriente continua y radiofrecuencias. Posteriormente, son expulsados hacia el detector, de forma secuencial, en orden creciente de la relación m/z por la aplicación de una rampa de radiofrecuencias.
- Analizador de tiempo de vuelo (TOF): Es el sistema utilizado cuando se ha ionizado por MALDI. Los iones formados en la cámara de ionización son acelerados y se dejan volar libremente hacia el detector. Su separación se consigue por la distinta velocidad que adquieren debido a la diferencia de masas.

Finalmente, la representación de la corriente iónica detectada frente a la relación masa/carga de los iones, constituye el espectro de masas de la muestra analizada.

3.3.2. DETERMINACIÓN DE ESPECIES MOLECULARES EN MUESTRAS PURIFICADAS DE CbpF.

Los experimentos de MALDI-TOF se realizaron en el Servicio de Espectrometría de Masas del Instituto de Química-Física Rocasolano (CSIC) y se llevaron a cabo en un espectrómetro de masas Voyager DE-PRO (Applied Biosystems) equipado con un láser de nitrógeno ($\lambda = 337$ nm; anchura de pulso = 3 ns; $\nu = 20$ Hz) y una fuente iónica con extracción retardada. Los iones positivos generados por la desorción láser fueron introducidos en el tubo de vuelo (1.3 m de longitud) con un voltaje de aceleración de 25 kV, trabajando en modo lineal. Todos los espectros de masas se obtuvieron recogiendo una media de 5000 disparos, y la matriz utilizada fue ácido sinapínico (10 mg/ml en una mezcla TFA al 0.21% y acetonitrilo al 30%). La calibración necesaria para los espectros de masas fue externa, empleándose apomioglobina de caballo.

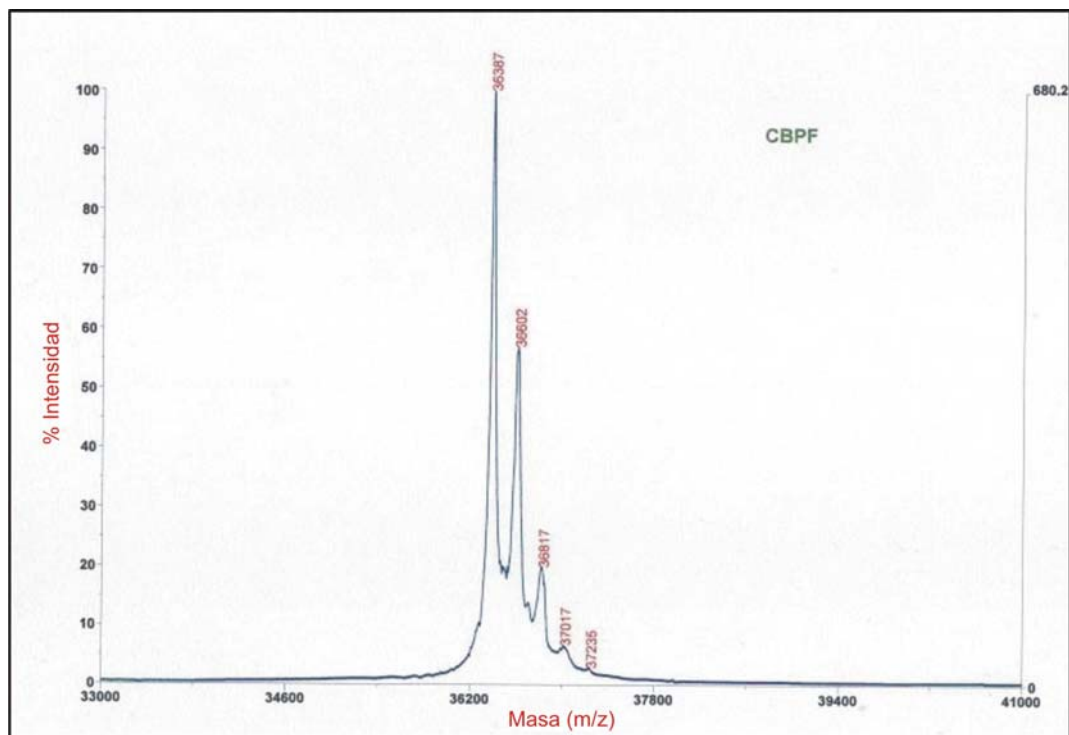


Figura 3.3.1: Registro del experimento de MALDI-TOF. El pico mayoritario registrado a 36387 Da se corresponde al PM de CbpF con una molécula de colina. El intervalo constante de PM (≈ 200 Da) del resto de picos es característico de los diferentes complejos de CbpF con aductos de la matriz y no a otras especies moleculares. Por tanto, la muestra de CbpF contiene una única especie molecular.

3.4. ULTRACENTRIFUGACIÓN ANALÍTICA.

Las técnicas de ultracentrifugación analítica permiten determinar la masa molecular y los parámetros hidrodinámicos de las macromoléculas biológicas en solución, así como juzgar la homogeneidad de una muestra. Además, posibilitan la detección y cuantificación de las interacciones macromoleculares reversibles proteína-proteína o proteína-ligando. Estas interacciones son generalmente débiles y cuando las fuerzas del campo cristalino son equivalentes o de mayor magnitud que ellas, la ultracentrifugación analítica es una de las herramientas más potentes para su caracterización cuantitativa. Hay dos tipos de experimentos básicos en la ultracentrifugación analítica: la velocidad de sedimentación y el equilibrio de sedimentación. Mediante los primeros es posible estimar los coeficientes de sedimentación y difusión de las macromoléculas estudiadas, los cuales están relacionados con su tamaño y forma. Por otro lado, los experimentos de equilibrio de sedimentación permiten determinar la masa molecular promedio de las macromoléculas estudiadas, así como poner de manifiesto la

existencia de equilibrios de asociación y llevar a cabo su caracterización cuantitativa mediante la determinación de las constantes de equilibrio y la estequiometría de los complejos.

En el caso de CbpF, empleamos la ultracentrifugación analítica para determinar su estado de asociación y la masa molecular promedio, por lo que se recurrió a medidas de equilibrio de sedimentación. Teniendo en cuenta la elevada cantidad de residuos aromáticos que nos podrían falsear el resultado, además de los experimentos de equilibrio de sedimentación a distintas concentraciones de colina, se hicieron también experimentos de equilibrio de sedimentación por refractometría, a las mismas concentraciones de colina, pero a mayores concentraciones de proteína.

3.4.1. EXPERIMENTOS DE EQUILIBRIO DE SEDIMENTACIÓN POR ABSORBANCIA.

Los experimentos se realizaron en una ultracentrífuga analítica modelo Optima-XLA (Beckman), equipada con un sistema óptico UV-Vis instalado en el interior de la cámara de la centrífuga que permite la selección de una longitud de onda en el intervalo de 200-700 nm. Se utilizaron celdas con ventanas de cuarzo de doble sector y seis agujeros con un paso óptico de 12 mm, siendo la temperatura de medida de 20 °C. Las muestras de proteína (3.3 μM \approx 0.12 mg/ml) se equilibraron en tampón fosfato 20 mM pH 8.0 y diferentes concentraciones de colina (0, 30, 50, 60, 70, 90, 140, 180, 220 y 280 mM). Con el volumen de muestra utilizado (\cong 80 μl) la altura que se genera en la celda es de unos 0,14 cm, por lo que el equilibrio se alcanza en menos de 12 horas si la selección de la velocidad de sedimentación es adecuada (Chervenka, 1969). En nuestro caso se emplearon 3 velocidades distintas (10500, 12500 y 18000 rpm) con el fin de comprobar la homogeneidad de las masas moleculares. Posteriormente, se restó la línea de base (contribución del material no proteico) realizada tras centrifugar la muestra, una vez alcanzado el equilibrio, a 42000 rpm durante seis horas.

La determinación de las masas moleculares promedio se realizó ajustando a los datos experimentales la ecuación [1], que describe la distribución radial del gradiente de concentración para un soluto ideal tras alcanzarse el equilibrio entre los flujos de las moléculas que sedimentan y difunden en cada posición radial:

$$w_i(r) = w_i(r_0) \exp\left[\left(M_i^* \omega^2 / 2RT\right)(r^2 - r_0^2)\right] \quad [1]$$

donde w_i es la concentración (peso/volumen) de la especie i , r y r_0 son las posiciones radial y de referencia, respectivamente, M_i^* es la masa molecular de flotación de la especie i , ω es la velocidad angular, R es la constante universal de los gases y T es la temperatura absoluta. La masa de flotación de la especie i está relacionada con la masa molecular aparente, M_i , por la expresión:

$$M_i^* = M_i(1 - v_i \rho_o) \quad [2]$$

donde v_i es el volumen específico parcial del soluto y ρ_o es la densidad del disolvente. El ajuste se realizó utilizando el programa HETEROANALYSIS basado en el algoritmo de conservación de la señal (Minton, 1994).

Los valores de M_i^* se convirtieron en las correspondientes masas moleculares aparentes usando un volumen específico parcial de 0.72 ml/g, calculado a partir de la composición de aminoácidos (Laue *et al.*, 1992). La densidad de las disoluciones de colina a 20 °C se determinó con un densitómetro de precisión Antón PAAR DM2. Los cálculos se realizaron con los datos obtenidos a 18000 rpm.

Los resultados mostraron, por un lado, que la masa molecular promedio de CbpF era siempre superior a la del monómero (Figura 3.4.1), y por otro, que ésta era independiente de la concentración de colina a partir de 50 mM, al menos a la concentración de proteína empleada (siempre inferior a 0.2 mg/ml). Es decir, por debajo de 50 mM colina, la proteína agregaba masivamente mientras que a partir de 50 mM colina hasta 280 mM, el estado de asociación de la proteína permanecía inalterado aunque la masa molecular promedio en torno a 1.3 veces la del monómero apuntaba a la existencia de un posible equilibrio de asociación monómero-dímero no dirigido por la concentración de colina. Con el fin de aclarar la existencia o no de un equilibrio de asociación-disociación monómero-dímero, se realizaron experimentos de equilibrio de sedimentación por interferometría.

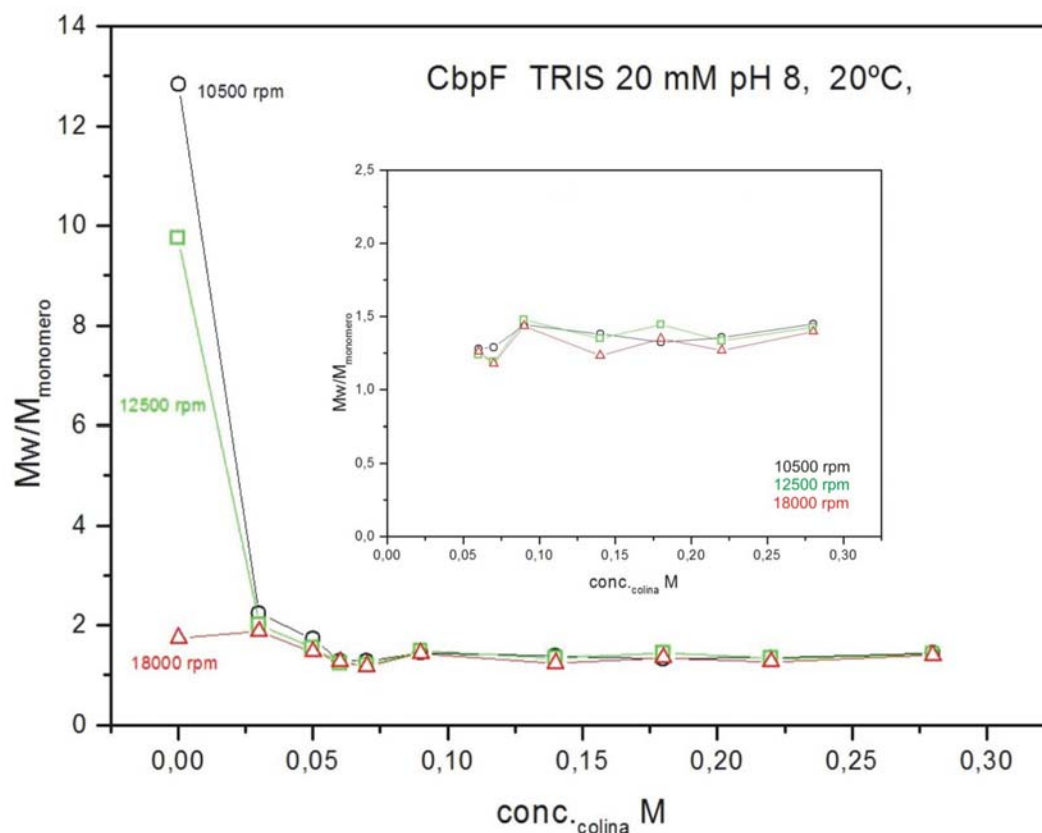


Figura 3.4.1. Ultracentrifugación analítica empleando medidas de absorbancia. Representación de la relación masa molecular experimental: masa teórica del monómero en función de distintas concentraciones de colina (0, 30, 50, 60, 70, 90, 140, 180, 220 y 280 mM) a 3 velocidades diferentes (10500, 12500 y 18000 rpm). En el centro se muestra la misma representación limitando los valores de la relación masa molecular experimental: masa teórica del monómero entre 0 y 2.5 con el fin de analizar la posible existencia de equilibrios de asociación.

3.4.2. EXPERIMENTOS DE EQUILIBRIO DE SEDIMENTACIÓN POR REFRACTOMETRÍA.

Frente a los experimentos de equilibrio de sedimentación por absorbancia, que como su nombre indica está basado en un sistema de análisis óptico de absorción (medidas de absorbancia), existen otros dos sistemas ópticos de análisis basados en este caso en un análisis refractométrico: sistema óptico de Schlieren y sistema óptico interferométrico. De este modo, mientras el sistema óptico de absorción muestra un cambio de absorbancia al pasar de la zona en la que sólo hay disolvente a aquella en la que hay disolución, el sistema interferométrico aprecia en esa misma zona un cambio en el índice de refracción del medio y el sistema óptico de Schlieren convierte ese cambio en una desviación vertical de la luz, apareciendo así un pico,

que, cualitativamente, se podría relacionar con la primera derivada del gráfico proporcionado por el sistema de absorción. Así, a través de estos experimentos la ultracentrifugación analítica permite juzgar la homogeneidad de una muestra. Una muestra heterogénea daría lugar a diferentes frentes de avance bajo la acción del campo centrífugo, lo que a su vez se traduciría en varios picos, o varios cambios de absorción o de índice de refracción, según cual sea el sistema de observación empleado. Todo ello bajo la perspectiva, siempre, de que sería un criterio negativo de homogeneidad: un comportamiento homogéneo no permite asegurar que la muestra analizada lo sea.

Los experimentos de equilibrio de sedimentación empleando sistemas ópticos interferométricos fueron realizados en el laboratorio del Dr. Germán Rivas en el Centro de investigaciones Biológicas. A través de este sistema óptico de análisis pudimos emplear concentraciones de proteína más elevada (0.3, 0.5, 1, 1.2, 2 y 2.2 mg/ml) de las que nos permitía el sistema óptico de absorción. Como hemos comentado anteriormente, la elevada cantidad de residuos aromáticos y la necesidad de mantener la concentración de la proteína de análisis en unos determinados rangos de concentración, hizo necesario la utilización de esta técnica para complementar los datos de absorción pudiendo analizar el comportamiento de la proteína a mayores concentraciones de proteína más cercanas al comportamiento en el cristal.

Tras los resultados obtenidos (Figura 3.4.2) vemos como el comportamiento de la proteína a concentraciones de colina de 140 mM es independiente de la concentración de proteína en un rango de 0.5 a 2 mg/ml apareciendo casi exclusivamente en forma monomérica. Sin embargo, a medida que disminuimos la concentración de colina la proteína empieza a agregar pudiendo sugerir la existencia de un dímero entorno a 70 mM colina. Curiosamente, aunque disminuimos la concentración de proteína es la disminución de la concentración de colina la que dirige el desplazamiento del equilibrio hacia el dímero. A concentraciones de colina en torno a 35 mM, la proteína empieza a agregar masivamente apareciendo formas multimérica. Estos resultados vienen a confirmar y completar los experimentos realizados mediante medidas de absorción.

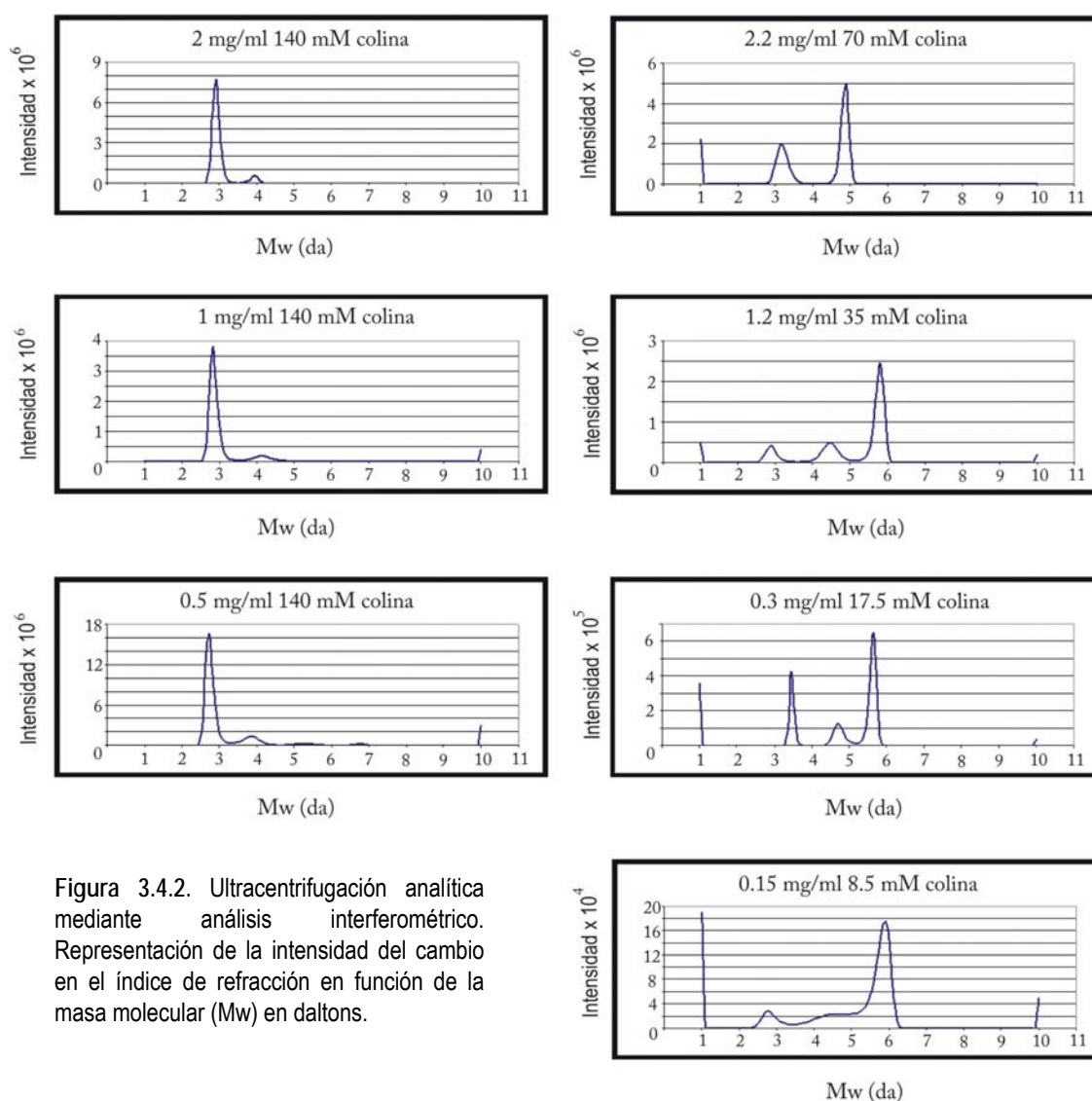


Figura 3.4.2. Ultracentrifugación analítica mediante análisis interferométrico. Representación de la intensidad del cambio en el índice de refracción en función de la masa molecular (Mw) en daltons.

3.5. CRISTALIZACIÓN.

3.5.1. INTRODUCCIÓN.

La determinación de la estructura tridimensional de cualquier molécula por difracción de rayos X depende, en primer lugar, del éxito en la obtención de monocristales con tamaño adecuado para someterlos a difracción (McPherson, 1982).

Un cristal se puede considerar como una red imaginaria de unidades elementales idénticas, denominadas cada una de ellas celdilla unidad. La repetición por translación en las tres direcciones del espacio de este elemento estructural genera el volumen del cristal. Si el cristal está formado por celdillas unidad, la celdilla a su vez presenta en su interior una unidad mínima,

denominada unidad asimétrica. Mediante la aplicación de ciertas operaciones de simetría a la unidad asimétrica, se genera el contenido de la celdilla unidad y, en consecuencia, el de todo el cristal. Dicha celdilla unidad está definida por tres ejes (a , b , c) y tres ángulos interaxiales (α , β , γ) (Figura 3.5.1) que se usan como sistema de referencia para las coordenadas atómicas. La combinación de la celdilla con los elementos de simetría, designa el grupo espacial que describe el patrón de repetición que presenta una o varias moléculas (unidad asimétrica) en el espacio tridimensional.

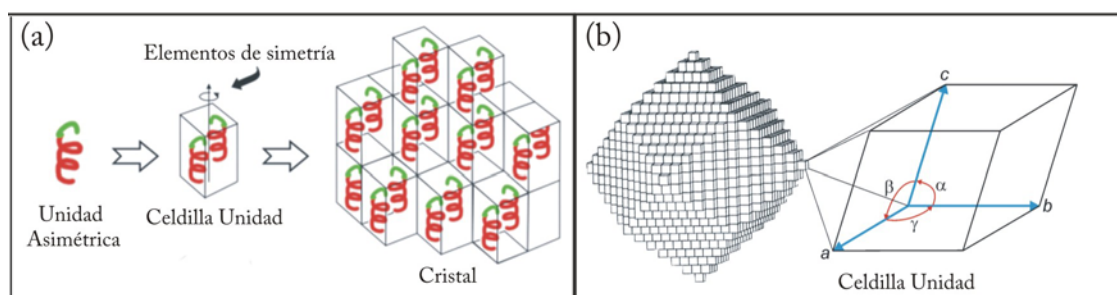


Figura 3.5.1: (a) La unidad asimétrica se repite a sí misma mediante los elementos de simetría, reproduciendo la celdilla unidad (b) El apilamiento de celdillas unidad forma el cristal.

Los cristales de proteína difieren mucho de los cristales convencionales de moléculas pequeñas (como los compuestos orgánicos en general) o sales inorgánicas (como la sal común) ya que las características de los primeros están condicionadas por la propia naturaleza de estas macromoléculas. Así:

- Al estar formados por moléculas quirales, se describen exclusivamente con 65 grupos espaciales (de los 230 posibles), es decir, sólo son posibles aquellos en los que no existen las operaciones de repetición que impliquen centros o planos de simetría, planos de deslizamiento o ejes quirales.
- Debido al tamaño de las proteínas y su esfera de solvatación, el volumen de la celdilla unidad suele ser bastante grande.
- Como las macromoléculas empaquetan con baja densidad, al ordenarse en el cristal dejan grandes canales constituidos por moléculas de agua desordenadas y el contenido en solvente suele ser elevado (50-80%) (Mathews, 1968).

Estos factores hacen que la fuerza de interacción entre las moléculas sea débil, de manera que los cristales son muy frágiles, sensibles a la radiación y difíciles de manipular. Por otra parte, el elevado contenido en solvente desordenado hace que los datos de difracción que se obtienen sean generalmente muy débiles. Este efecto puede ser especialmente acusado cuando en la molécula existen segmentos de alta movilidad, o la cristalinidad de la muestra es baja. Sin embargo, el alto contenido en solvente de los cristales de proteína también tiene aspectos positivos ya que la presencia de una esfera de solvatación permite a la proteína estar correctamente plegada y la presencia de grandes canales posibilita la inclusión en el cristal de átomos pesados, sales y moléculas, importantes (como se verá más adelante) para resolver el problema de la fase o para obtener estructuras de complejos proteína-ligando/inhibidor a partir del cristal de proteína nativa.

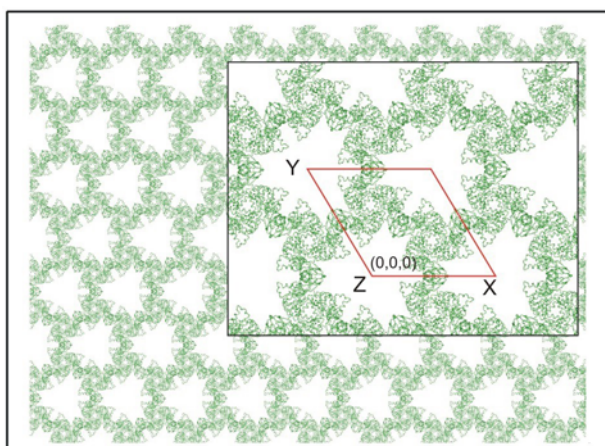


Figura 3.5.2: Estructura cristalina donde destacan los huecos que genera el empaquetamiento molecular. Dichos huecos son los que constituyen los grandes canales de solvente presentes en los cristales de proteína. En rojo aparece la celdilla unidad vista desde el eje c.

Por tanto, el primer objetivo en la resolución de la estructura de una proteína por Cristalografía de rayos X es la obtención de un monocristal que tenga el mayor tamaño posible (mayor número de elementos dispersores) y la menor mosaicidad (parámetro que nos informa del orden interno que posee un cristal). Cuanto mayor desorden interno, mayor será la mosaicidad y peor la calidad del cristal, lo que se traduce en peores datos de difracción.

3.5.1.1. Principios básicos de la cristalización.

Para la producción de monocristales de tamaño adecuado va a ser fundamental disponer de una muestra pura y conformacionalmente homogénea. Durante el proceso de cristalización, las moléculas en solución, en presencia de agentes precipitantes suaves, son conducidas a un estado de sobresaturación (termodinámicamente inestable) que posteriormente deriva en la

aparición de una fase cristalina donde las moléculas son llevadas a un equilibrio entre las fases sólida y líquida (Figura 3.5.3.) (Ducruix & Giegé, 1992).

El proceso de cristalización se puede dividir en dos fases: la fase de nucleación (formación de los primeros agregados ordenados) y la fase de crecimiento del cristal (en donde el conjunto de interacciones atractivas entre las moléculas son máximas, y las repulsivas mínimas).

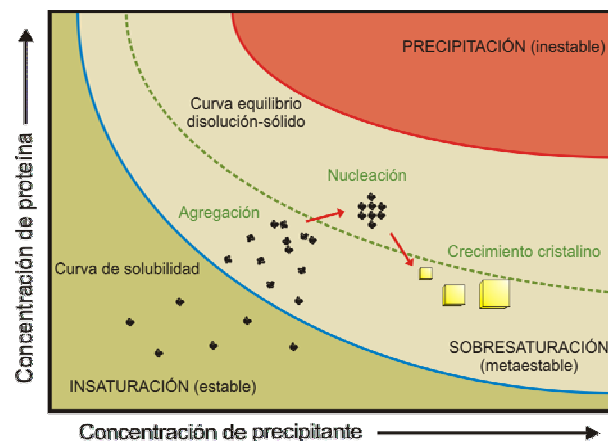


Figura 3.5.3: Descripción esquemática de un diagrama bidimensional de solubilidad, en el que se ilustra el cambio de concentración de proteína con la concentración de agente precipitante.

Durante este proceso hay que superar una barrera energética análogamente a como ocurre en cualquier reacción química (Figura 3.5.4). En el caso de las proteínas, la superación excesiva del punto de saturación, sobre todo si sucede rápidamente, conlleva con gran frecuencia la precipitación de la proteína en forma de agregados amorfos, no ordenados, que no son adecuados para los estudios de difracción. Por el contrario, si la sobresaturación es leve se alcanza un estado metaestable que no da lugar a núcleos cristalinos (a menos que se le aporte energía al sistema) pero que sería la zona ideal para el crecimiento de cristales. Por tanto, la formación de cristales requiere alcanzar lentamente un grado limitado de sobresaturación, para que se formen pocos núcleos, de manera que la masa proteica se invierta en el crecimiento de los cristales y así éstos adquieran un buen tamaño.

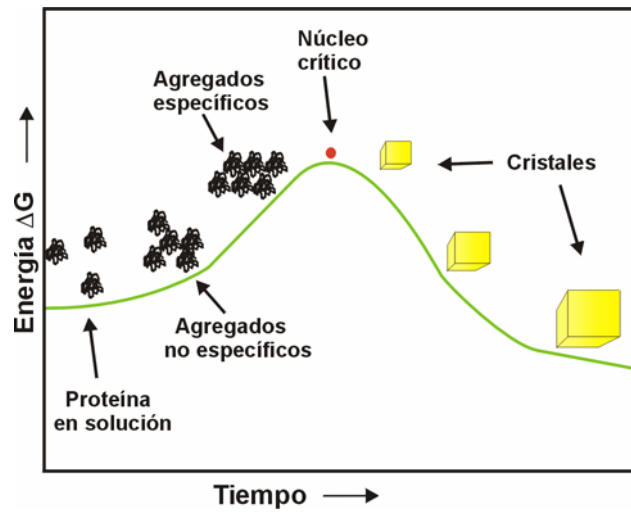


Figura 3.5.4: Diagrama de energía de la cristalización. Imagen extraída de la página web <http://www-structured.cimr.cam.ac.uk/Course/Crystals>.

3.5.1.2. Condiciones ideales de la proteína a cristalizar.

Hay que considerar cinco aspectos principales: concentración, pureza, homogeneidad, estabilidad, y cantidad disponible.

- **Concentración elevada:** dado que hay que alcanzar la sobresaturación, es esencial conseguir que la proteína posea una concentración inicial elevada. Se considera adecuado conseguir concentraciones en el rango 5-20 mg/ml, preferiblemente ≥ 10 mg/ml, aunque no hay que descartar usar concentraciones menores si la solubilidad de la proteína no permite alcanzarlas mayores ya que lo importante es sobresaturar la solución durante el experimento.
- **Pureza de la proteína:** debe ser tan alta como sea posible, aceptándose comúnmente que al menos ha de ser del 95 %. Las impurezas pueden limitar el tamaño de los cristales, ya que su interposición puede abortar el crecimiento del cristal, y/o introducir irregularidades que disminuyan la calidad del cristal obtenido. Por eso, cuando los cristales que se obtienen con una proteína siempre difractan mal o son pequeños, es buena estrategia revisar el grado de pureza e intentar aumentarla. En cualquier caso, no es imposible obtener cristales con purezas relativamente bajas.

- **Homogeneidad de la muestra:** también es importante que la proteína sea homogénea. La heterogeneidad puede ser tanto intrínseca (fosforilación, nitrosilación, grados diversos de glicosilación, formas alternativas de procesamiento, formas isoenzimáticas, etc.), como resultado de la manipulación experimental (grados distintos de oxidación, proteólisis, agregación en grado variable, desnaturalización parcial, etc.). Hay que evitar durante la preparación de la proteína someterla a tratamientos que puedan introducir heterogeneidad, tales como tratamientos térmicos. El uso de espectrometría de masas (MALDI-TOF) para la determinación del peso molecular de las subunidades de la proteína es muy informativa sobre la coexistencia de variantes con diferentes masas (formas proteolizadas, modificaciones postraduccionales). No es recomendable usar la liofilización como modo de preservar las proteínas para cristalizar, ya que muchas veces, al redissolver la proteína liofilizada, una parte de ella agrega, lo que generalmente entorpece el proceso de cristalización. Una forma de detectar esta agregación es mediante técnicas de dispersión óptica. La congelación y descongelación repetida tampoco es recomendable, ya que puede también causar que una fracción de la proteína agregue o se desnaturalice, al menos parcialmente.
- **Estabilidad:** las condiciones de cristalización (pH, temperatura, adición de ligandos y de protectores de grupos –SH) deben preservar al máximo la estabilidad de la proteína. Ello afecta a la conservación de la proteína tras su purificación, así como a la solución en la que se disuelve para intentar la cristalización. Salvo que la proteína se desnaturalice y precipite muy rápidamente, el saber que la proteína tiene una estabilidad limitada no debe ser obstáculo para iniciar pruebas de cristalización, siempre que la proteína sea estable en la solución inicial en que se concentra. En todo caso, por lo general lo mejor es usar, cuando hay alternativa, la forma más estable posible. Además, debemos tener en cuenta que las proteínas y los agentes precipitantes orgánicos son buenos nutrientes para el crecimiento de bacterias u hongos por lo que se recomienda usar una técnica estéril, o al menos aséptica, o añadir azida sódica al 0.02% como biocida, aunque con ello se aumenta el contenido de sal de la solución de proteína.
- **Cantidad disponible:** desde el punto de vista logístico es importante disponer de abundante proteína purificada. El consejo general es que para hacer un primer rastreo

son precisos unos pocos mgs de proteína altamente purificada, y si el rastreo debe incluir la presencia o ausencia de ligandos de la proteína, y condiciones a 20 °C y 4 °C, la cantidad mínima aumenta. Actualmente, la introducción de robots de cristalización está reduciendo estas cantidades en factores de 2-5 veces. Por otro lado, la producción de proteínas recombinantes ha facilitado mucho la generación de cantidades importantes de proteína y a su vez hace posible la incorporación de selenometionina en vez de metionina, lo que abre las puertas a la determinación de las fases mediante técnicas de difracción anómala.

3.5.1.3. Proceso experimental de cristalización.

La cristalización es un proceso que depende de muchos factores que aún hoy en día no se han llegado a racionalizar y controlar totalmente, por lo que se puede considerar que este proceso es fundamentalmente empírico.

Como se ha visto previamente y de forma general, la estrategia que se sigue para alcanzar la sobresaturación consiste en empezar en una concentración relativamente elevada de proteína (alrededor de 10 mg/ml), y aumentar la concentración por eliminación gradual de agua, introduciendo además ciertas variables que afecten a la solubilidad de la proteína, como son:

- Naturaleza y concentración de diferentes agentes precipitantes, como por ejemplo polímeros (los más utilizados son los polietilenglicoles), sales, compuestos orgánicos volátiles y no volátiles,... etc.
- pH
- Temperatura
- Naturaleza y concentración de aditivos, como por ejemplo sales, ligandos comunes de proteínas,...etc.

Esta búsqueda de las condiciones de cristalización se desarrolla en dos pasos, primero se realiza una búsqueda de condiciones iniciales, seguida, en segundo lugar, de una optimización sistemática de dichas condiciones.

- **Búsqueda de condiciones iniciales:** debido a que el número de factores a controlar es excesivo, es imposible realizar una búsqueda sistemática en la que se exploren todos los factores posibles. Por este motivo se han diseñado estrategias de muestreo, que utilizan matrices que exploran el espacio de cristalización. La matriz más utilizada actualmente fue desarrollada por Jancarik y Kim (1991) y consiste en el rastreo de condiciones que han dado en el pasado condiciones favorables. Sin embargo, también existen programas que generan matrices que exploran al azar el espacio de cristalización, basándose únicamente en incompatibilidades de reactivos.
- **Optimización:** a partir de las condiciones iniciales, se hace un rastreo barriendo la concentración de precipitante, de proteína, la temperatura o el método de cristalización. Una vez optimizadas esas condiciones, si es necesario, se prueban ciertos aditivos, como compuestos orgánicos, cationes, aniones, detergentes,...etc. Además, también se recurre a las técnicas de macro/microsembrado y al crecimiento de cristales en geles de agarosa.

3.5.1.4. Métodos de cristalización.

El número de técnicas descritas es muy elevado aunque sólo unas pocas se usan generalmente. La más común se basa en alcanzar la sobresaturación mediante difusión de vapor (Figura 3.5.5.a). Una pequeña gota que contiene la solución de proteína y la de cristalización, se equilibra contra un reservorio de mucho mayor volumen y que contiene una solución de cristalización a una concentración superior que en la gota. El equilibrio se consigue por difusión de las especies volátiles (agua o solvente orgánico), hasta que las concentraciones de precipitantes en gota y reservorio son iguales. Existen principalmente dos modalidades de esta técnica. En la modalidad de la gota colgante (Figura 3.5.5.a) la gota se deposita sobre un cubreobjetos siliconizado, que luego se invierte y se utiliza para sellar un pocillo de una placa multipocillo, de modo que la gota quede dentro del recinto sellado, separada de la solución de cristalización presente en el pocillo. En la segunda alternativa (Figura 3.5.5.a) la gota se asienta sobre una elevación plana o cóncava situada por encima del nivel de la solución de cristalización. Esta última aproximación, combinada con el uso de placas de 96 pocillos que se sellan con una lámina flexible de plástico adhesivo y transparente, puede simplificar la manipulación y ahorrar soluciones de cristalización y de proteína.

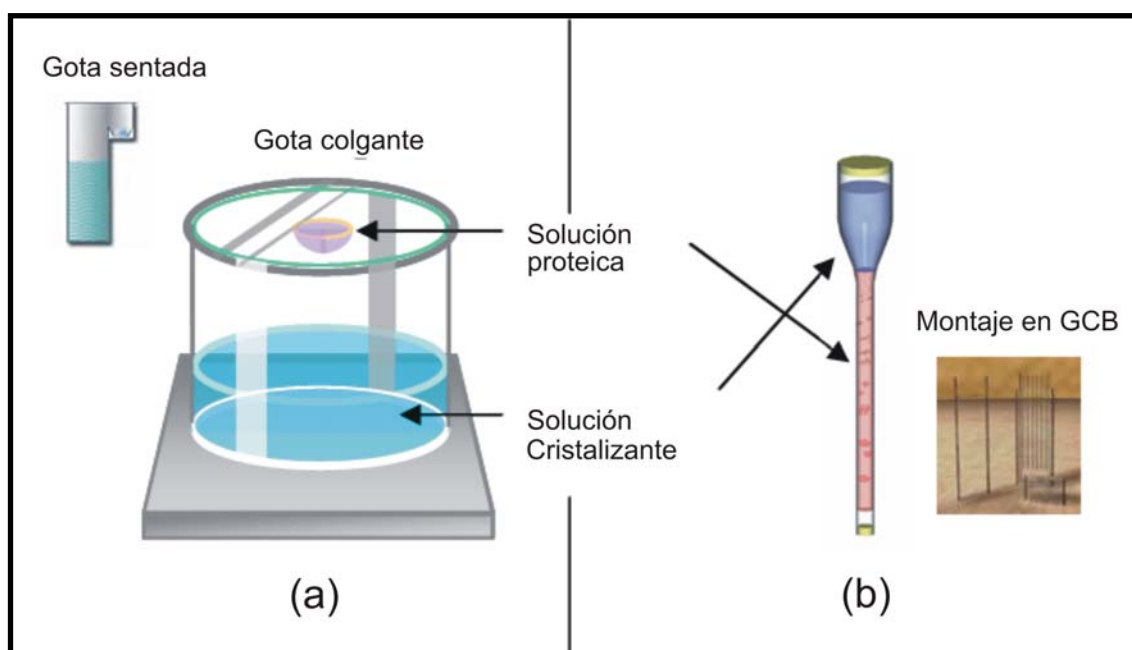


Figura 3.5.5: Esquema de los montajes experimentales de cristalización mediante (a) la técnica de difusión de vapor en gota sentada o colgante y (b) la de contradifusión en capilar. El montaje en contradifusión se ha extendido gracias a la comercialización de las GCB (*Granada Crystallization Box*).

Otra estrategia es la basada en el método de contradifusión (J. M. García-Ruiz, 2003). Aunque fundamentalmente se usa para la optimización de cristales a partir de condiciones de cristalización conocidas, actualmente se están desarrollando kits comerciales, como Triana CCD (*Capillary Counterdiffusion de Triana Science & Technology*. Granada, España), para la búsqueda de condiciones iniciales de cristalización. Como se muestra en la figura 3.5.5.b, la solución de proteína, en contacto con la solución de cristalización a través de una interfase líquido-líquido o líquido-gel, se equilibra por difusión de la segunda, que es más densa, a través de la primera. El experimento se realiza en un capilar de pequeño tamaño (0.1 - 0.3 mm), para minimizar el transporte de masa por convección, y sólo se produce transporte por difusión, con lo que, en principio, existe un aporte de proteína al cristal más regular, que debe redundar en un crecimiento cristalino más ordenado y en cristales con menores problemas mecánicos. Los cristales aparecen a lo largo del capilar debido al gradiente de sobresaturación que se genera, obteniéndose al principio precipitados amorfos y, a lo largo del capilar, cristales separados unos de otros (Gavira *et al.*, 2002). Esta técnica presenta dos ventajas destacables: la poca cantidad de proteína necesaria para los ensayos, y, sobre todo, que una vez producidos los cristales no

hay que manipularlos para recoger datos de difracción ya que se pueden obtener directamente a partir del montaje en el propio capilar.

3.5.2. CRISTALIZACIÓN DE CBPF Y SUS DERIVADOS LANTÁNIDOS.

3.5.2.1. Cristalización de CbpF.

Las condiciones preliminares de cristalización se establecieron empleando el método de difusión de vapor en gota colgante a 291 K (ver sección 3.5.1.4) y usando las soluciones de cristalización “Crystal Screen I”, “Crystal Screen II” e “Index”. (Hampton Research, CA, USA), que están basadas en la matriz de rastreo de condiciones ideada por Jancarik y Kim (1991). Así, se mezclaron 1 μl de la solución de proteína CbpF (7.8 mg ml^{-1} en 20 mM Tris-HCl pH = 8.0, 140 mM cloruro de colina) con 1 μl de solución de cristalización y las gotas se equilibraron frente a 500 μl de solución de cristalización. Los cristales crecieron con dos hábitos cristalinos diferentes a los que denominamos como hábito cristalino I y hábito cristalino II. Los cristales pertenecientes al hábito cristalino I crecieron a 291 K en 0.2 M sulfato amónico, 30% PEG 8000 y tampón 0.1M cacodilato sódico pH = 6.5 (Figura 3.5.6.a) mientras que los cristales pertenecientes al hábito cristalino II crecieron a la misma temperatura en 0.01M NiCl_2 , 20% PEGMME 2000, 0.1M Tris pH 8.5 (Figura 3.5.6.b). Los cristales pertenecientes al hábito II eran los que mejor difractaban (ver apartado 3.6.2.1), por ello sus condiciones de cristalización fueron optimizadas mezclando 3 μl de la solución proteica (3.9 mg ml^{-1} en 20 mM Tris-HCl pH = 8.0, 140 mM cloruro de colina) y 1 μl de la solución de cristalización, produciendo en tres días cristales con unas dimensiones máximas de 0.7 x 0.4 x 0.3 mm que difractaban hasta 1.67 Å de resolución.

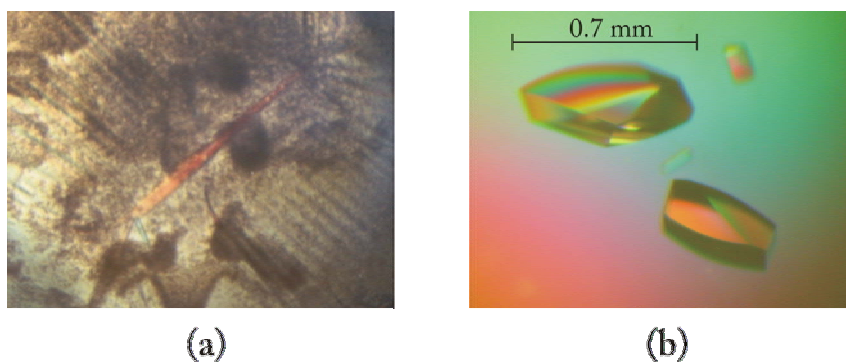


Figura 3.5.6: Hábitos cristalinos obtenidos tras los ensayos de cristalización con CbpF. (a) Hábito cristalino I. (b) Hábito cristalino II.

3.5.2.2. Cristalización de derivados de CbpF. Derivados lantánidos.

Generalidades.

La obtención de cristales derivados por el método de la gota colgante se puede realizar mediante inmersión o cocrystalización. La inmersión consiste en sumergir los cristales nativos en soluciones del complejo de átomo pesado y la cocrystalización en la cristalización de la proteína objeto de estudio en soluciones madre** que contienen dichos complejos. Aunque, al igual que en el proceso de cristalización de cristales nativos, no controlamos todos los factores que intervienen en dicho proceso, existen una serie de pautas para elegir unos complejos de metales pesados frente a otros en función de las características físico-químicas de la proteína y del medio donde cristaliza la forma nativa (ver sección 1.2.4.1). Además, debemos estar seguros de elegir átomos pesados que nos proporcionen una diferencia en la amplitud de difracción considerable de tal modo que minimicemos la importancia de la contribución de los errores en la recogida de datos. Como el número atómico del átomo pesado y el número de sitios derivatizados requeridos para una obtención de fases exitosa es proporcional al peso molecular de la macromolécula, para macromoléculas grandes necesitaremos átomos pesados o agregados de mayor número atómico y más de un átomo pesado por molécula.

Cristalización de derivados de CbpF.

En cuanto a la preparación de los cristales derivados de átomo pesado se llevaron a cabo multitud de ensayos tanto para aplicarlos a métodos de reemplazo isomorfo múltiple (MIR) como a métodos de dispersión anómala (MAD/SAD). Inicialmente se usaron todo tipo de complejos de metales pesados convencionales (Hampton Research, CA, USA) barriendo condiciones diferentes de concentración (1 - 20 mM), tiempo de cocrystalización o inmersión (10 sg – 3 días). Posteriormente, la elección de las soluciones de átomo pesado no fue al azar, empleándose fundamentalmente soluciones de Pt y Au ya que éstas eran las que *a priori* podrían darnos mejores resultado dado la composición aminoacídica de CbpF y sus condiciones de cristalización. Tras analizar la secuencia de aminoácidos de CbpF, comprobamos que si bien no poseía cisteínas libres (diana de los complejos mercuriales), si poseía metioninas por lo que

** solución de cristalización optimizada donde crecen los cristales

los complejos de Pt podrían interaccionar con dichas metioninas mediante unión covalente (Blundell *et al.*, 1976; Petsko *et al.*, 1985; Rould *et al.*, 1997). Por otro lado, a la hora de seleccionar una solución de átomo pesado para obtener derivados debemos tener en cuenta que dicho complejo de átomo pesado ha de ser soluble en la solución de cristalización. Soluciones de átomo pesado que cumplieran al menos una de estas características aparecían dentro de las “siete sales mágicas” (ver sección 1.2.5.4 “Introducción de átomos pesados para inmersión o cocrystalización”) donde K_2PtCl_4 cumplía ambas y $KAu(CN)_2$ era perfectamente soluble a pH = 8.5. Tras explorar finamente las condiciones de concentración y el tiempo de cocrystalización / inmersión con ambos complejos, no obtuvimos ningún derivado que nos permitiera obtener fases cristalográficas.

También se recurrió al empleo de haluros (ver sección 1.2.5.4 “Introducción de átomos pesados para inmersión o cocrystalización”) empleando concentraciones entre 0.25 y 1M del haluro dentro de las soluciones crioprotectoras (experimentos de inmersión) y concentraciones de 0.2M para los experimentos de cocrystalización. Al igual que en el caso del empleo soluciones de las soluciones de átomo pesado convencional, no obtuvimos ningún derivado que nos permitiera obtener las fases cristalográficas.

Derivados lantánidos de CbpF.

La cristalización de los derivados lantánidos de CbpF (ver sección 1.2.5.4) se obtuvo mediante la técnica de difusión de vapor en gota colgante a partir de las condiciones de cristalización de la proteína nativa (Molina *et al.*, 2007) (ver apartado 3.5.2.1). Los cristales derivados se prepararon mediante cocrystalización o mediante inmersión empleando las condiciones de cristalización nativas (0.01M $NiCl_2$, 20% PEG MME 2K, 0.1M Tris, pH 8.5. Hábito II) añadiendo 50 mM de cada complejo de Gd (Molina *et al.*, en prensa). Los mejores cristales se obtuvieron a partir de gotas que contenían 3 μ l de solución proteica y 1 μ l de solución de átomo pesado. En el caso de los derivados obtenidos por inmersión, el tiempo de inmersión fue de en torno a 10 segundos (tiempo de inmersión común que preservaba la calidad cristalina de todos los complejos), el cual es significativamente menor que tiempos de inmersión documentados anteriormente (Girard *et al.*, 2003a). Todos los cristales derivados fueron isomorfos a los cristales nativos ($P2_12_12$).

3.6. TÉCNICAS CRISTALOGRÁFICAS Y RESOLUCIÓN ESTRUCTURAL

3.6.1. CONCEPTOS GENERALES.

La biología molecular tiene como fin último comprender los procesos biológicos en términos de la física y la química de las macromoléculas que participan en ellos. Una de las diferencias esenciales entre la química de los sistemas vivos y los inertes es la gran complejidad estructural de las macromoléculas biológicas. Por tanto, la química de la vida no se podrá desvelar sin antes conocer a resolución atómica, o cercana a la atómica, la estructura de las macromoléculas biológicas, especialmente las proteínas (Branden & Tooze, 1991), que son las verdaderas máquinas de la vida. Así, comprender la estructura de cada proteína es comprender en que consiste la vida (Max Perutz, Premio Nobel de Química).

Como se ha comentado anteriormente, la cristalografía de rayos X es una de las técnicas experimentales más potentes para la determinación de estructuras moleculares ya que frente a otras técnicas encaminadas al mismo fin, presenta la gran ventaja de proporcionar una imagen tridimensional, a escala atómica, del material cristalizado.

3.6.1.1. Concepto de difracción

La difracción de rayos X es el fenómeno físico fundamental a través del cual se manifiesta la interacción de los rayos X con la materia cristalina. Cuando un haz de rayos X incide sobre un cristal, éste interacciona con los electrones de los átomos que componen el cristal, haciéndolos vibrar acopladamente con las variaciones periódicas de su campo eléctrico. Los electrones oscilantes se convierten, de esta manera, en focos de nueva radiación X que se emite en todas las direcciones, fenómeno que se denomina “dispersión”. Estos rayos X dispersados por los electrones interfieren entre sí, pudiendo dar interferencias destructivas. Sin embargo, como los átomos en un cristal están ordenados de un modo regular y periódico, también ocurre que en determinadas direcciones, las ondas interfieren constructivamente, reforzándose, dando lugar al fenómeno de dispersión cooperativa o difracción. Es en este caso de interferencia constructiva en el que se cumple la Ley de Bragg:

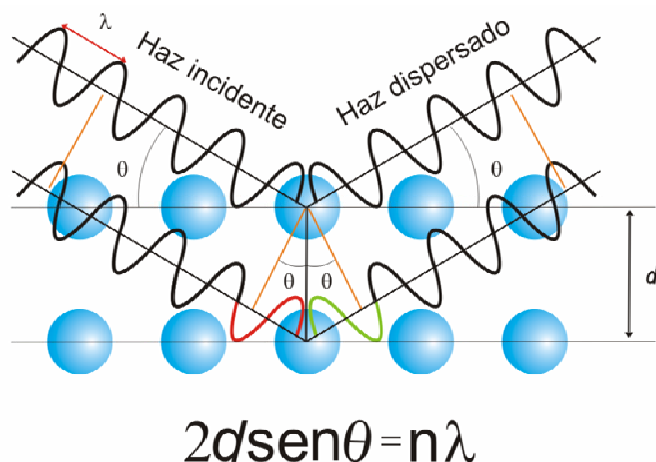


Figura 3.6.1: Ley de Bragg. La onda que incide en el plano inferior se retrasa un número entero de longitudes de onda y sale en fase con la onda superior, reforzándose la señal.

Por otro lado, el hecho de utilizar radiación X se debe a que el fenómeno de difracción requiere que la longitud de onda de la radiación electromagnética sea del mismo orden de las distancias entre los focos dispersores, los átomos de cristal.

El modo en que la cristalografía de rayos X llega al conocimiento de la estructura cristalina de cualquier material puede comprenderse mejor si se compara con el símil que representa la observación de un objeto a través de un microscopio óptico. En éste, la radiación dispersada por el objeto en estudio se recombina de nuevo, a través de un sistema de lentes, para dar lugar a la imagen aumentada del objeto dispersor. Sin embargo, no existe ninguna lente que sea capaz de focalizar los rayos X dispersados por los átomos. Para ello, las técnicas cristalográficas hacen uso de cálculos matemáticos denominados “síntesis de Fourier”, con los que se consigue “focalizar” los rayos X.

3.6.1.2. Toma de datos de difracción

En los experimentos de difracción se hace incidir un haz monocromático de rayos X (con una longitud de onda alrededor de 1 \AA) sobre un monocristal de la muestra en estudio.

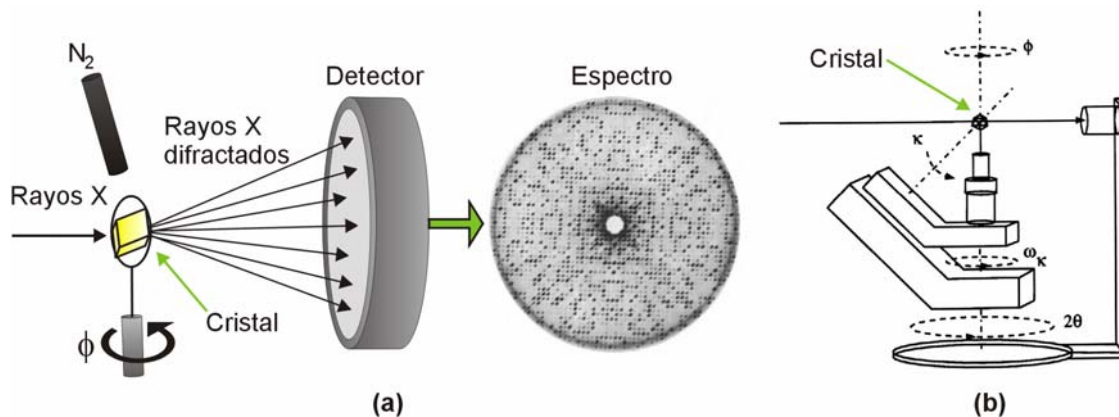


Figura 3.6.2: (a) Esquema de un experimento de difracción con geometría Phi. (b) Esquema de un goniómetro con geometría Kappa.

El espectro de difracción de rayos X contiene un patrón de intensidades correspondientes a los haces difractados por el cristal y la recogida de datos de difracción consiste en la medida de dichas intensidades y su posición espacial respecto al sistema de referencia del laboratorio. El diseño experimental implica la toma de imágenes sucesivas en un detector plano situado en la perpendicular del haz de rayos X, para distintas posiciones angulares del cristal, hasta obtener un conjunto de datos completo. Esto se consigue mediante oscilaciones simples del cristal (giros en Phi) o bien mediante una combinación de giros más complejos (Figura 3.6.2)

3.6.1.3. Fuentes de radiación y equipos de difracción.

La estrategia de recogida de los datos de difracción y la elección de la fuente de radiación dependerá del tipo de datos de difracción que sean necesarios para poder abordar la resolución estructural, pudiéndose realizar en un difractómetro con fuente convencional o en una estación de radiación sincrotrón. En nuestro caso, se utilizaron ambas fuentes para llevar a cabo la medida de los datos de difracción descritos en este trabajo.

Generador de ánodo rotatorio.

Los difractómetros utilizados normalmente en cristalografía de proteínas usan como fuente de rayos X un generador de ánodo rotatorio (Figura 3.6.3). Este generador permite suministrar un elevado voltaje al tubo de alto vacío de rayos X donde los electrones de un filamento incandescente se desprenden e inciden sobre el ánodo metálico. De esta forma, electrones de

los átomos del ánodo se excitan pasando a niveles superiores de energía que a continuación abandonan para volver al estado fundamental liberando el exceso de energía en forma de radiación X. Convencionalmente, se utiliza un ánodo de cobre, el cual genera radiación X de 1.5418 \AA de longitud de onda, correspondiente a la línea K α del cobre.

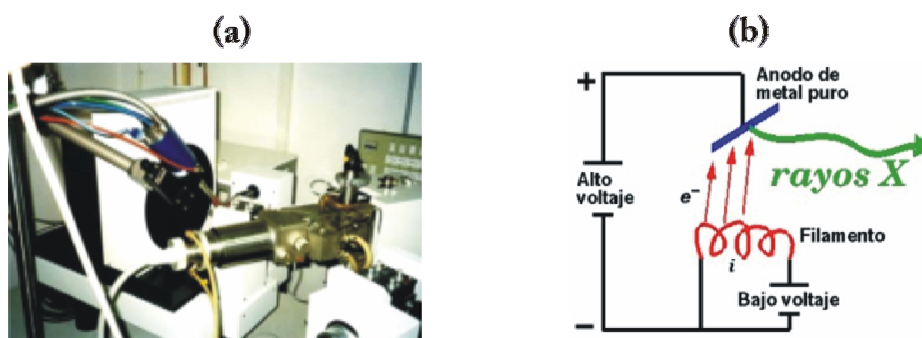


Figura 3.6.3: Generador de ánodo rotatorio. (a) Generador ENRAF-NONIUS FR571 e “Imaging Plate” pertenecientes al GCMBE (Grupo de Cristalografía Macromolecular y Biología Estructural) del departamento de Química-Física Rocasolano (CSIC). (b) Esquema de la generación de rayos X.

Para este trabajo se utilizó uno de los equipos del grupo de Cristalografía Macromolecular y Biología Estructural del departamento de Química-Física Rocasolano (CSIC) consistente en un generador de rayos X de ánodo rotatorio ENRAF-NONIUS FR571, con monocromador de grafito y ánodo de cobre. Este equipo presenta acoplado un detector Imaging Plate MAR345 y un sistema criogénico de Oxford Cryosystems que permite enfriar la muestra a 120 K. La cabeza goniométrica presenta geometría Phi lo que permite realizar oscilaciones simples del cristal sobre un solo eje. Gracias a este equipo se recogieron datos tanto de la proteína nativa como de los distintos intentos de obtención de derivados.

Radiación sincrotrón.

La radiación sincrotrón es la fuente de radiación X más intensa que existe. El sincrotrón consta de un anillo de cientos de metros de longitud por el que se hacen circular electrones o positrones a velocidades relativistas. Cuando a este haz de electrones se le obliga a cambiar la dirección de su trayectoria mediante imanes, se produce la emisión de una radiación muy energética tangencial a la dirección seguida por los electrones denominada radiación sincrotrón. La radiación emitida muestra un espectro continuo de longitud de onda muy amplio, en el que se encuentran incluidos los rayos X.

Para este trabajo se utilizó la radiación sincrotrón generada en diferentes líneas del ESRF (*European Synchrotron Radiation Facility*) de Grenoble, Francia. Debido al tipo de información que requeríamos, recurrimos a las líneas policromáticas BM30A y fundamentalmente ID29 donde se utilizó la longitud de onda correspondiente a la línea L_{III} del gadolinio ($\lambda = 1.7110 \text{ \AA}$), determinada a partir de espectros de fluorescencia de los cristales. Como sistemas de detección se utilizó un detector MARCCD (línea BM30A) y un detector ADSC Quantum 210 (línea ID29). En ambos casos la temperatura de trabajo fue de 100 K.

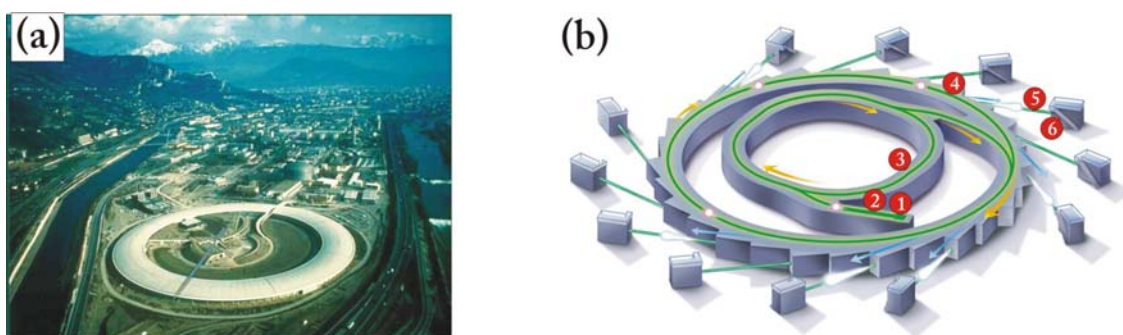


Figura 3.6.4: Radiación sincrotrón. (a) Vista aérea del sincrotrón ESRF (Grenoble). (b) Partes de una instalación sincrotrón. 1. Cañón de electrones. 2. Acelerador lineal. 3. Anillo acelerador. 4. Anillo de almacenamiento. 5. Haz de rayos X. 6. Cabañas de medida.

Ventajas y desventajas de las distintas fuentes de radiación.

Las características técnicas de una fuente de radiación convencional le confieren una gran versatilidad y un bajo coste a la hora de diseñar y realizar una estrategia de recogida de datos, mientras que la radiación sincrotrón tiene como ventajas fundamentales su mayor brillo (mayor intensidad y más concentrada) y la posibilidad de sintonizar la longitud de onda de los rayos X. Dado que el fenómeno de difracción es muy débil en el caso de los cristales de proteína, el mayor brillo de la radiación sincrotrón permite obtener datos de mayor calidad. En cuanto a la sintonización de la radiación, posible en las instalaciones de radiación sincrotrón, es muy útil a la hora de resolver estructuras mediante el uso de la difracción anómala.

3.6.1.4. Preparación de la muestra.

Los cristales de proteína se deterioran rápidamente al ser expuestos a la radiación. La longitud de onda penetrante y la intensidad del haz, unidas a prolongados tiempos de exposición,

provocan fenómenos de absorción que hacen que la muestra se caliente, generándose radicales libres en las moléculas de proteína y solvente, que, con el tiempo, dan lugar a daños irreparables en el cristal que redundan en la pérdida de orden interno. Por este motivo es conveniente enfriar el cristal durante el proceso de difracción (Garman y Schneider, 1997), mediante un flujo laminar de nitrógeno (120 - 100K) (Figura 3.6.2). Para que las temperaturas utilizadas en criocrystalografía ($\cong 100$ K) no dañen el cristal por la formación de cristales de hielo, es necesario utilizar sustancias crioprotectoras (McFerrin & Snell, 2002) como el glicerol, metilpentanodiol, etilenglicol, etanol, polietilenglicol, sacarosa, etc, para que las moléculas de agua no estructurales se reemplacen por las crioprotectoras generando un estado vítreo alrededor del cristal. Sin embargo, estas sustancias pueden alterar también el cristal, dañando el orden interno, o disolviéndolo en el crioprotector. Por este motivo, hay que elegir apropiadamente el agente crioprotector, su concentración y el tiempo de permanencia del cristal en la solución crioprotectora. El procedimiento seguido en nuestro caso fue realizar durante unos segundos la inmersión del cristal en la solución crioprotectora antes de colocarlo, a través de un lazo de nylon que se monta mediante un imán en la cabeza goniométrica, bajo el flujo laminar de nitrógeno a baja temperatura. La solución crioprotectora utilizada fue la solución madre con un 17% de glicerol y la temperatura de trabajo fue de 100 K.

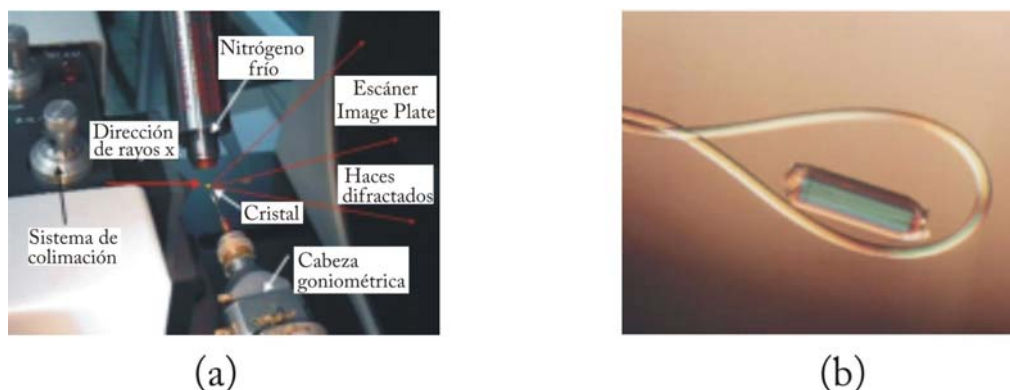


Figura 3.6.5: Montaje de la muestra en el experimento de difracción. (a) Detalles del sistema de detección, cabeza goniométrica, cristal, sistema de enfriamiento y detector. (b) Cristal de proteína montado en un lazo e inmerso en solución crioprotectora.

Una vez se ha montado el cristal, éste es irradiado con rayos X mientras se le hace girar en intervalos pequeños. La medida realizada en los sucesivos intervalos nos proporciona un conjunto de imágenes de difracción en la que se recoge la totalidad de los haces difractados por

el cristal. Estos haces constituyen el patrón de intensidades y aparecen en la imagen de difracción en forma de pequeños puntos o “spots”.

3.6.1.5. Resolución de la estructura.

Como se ha mencionado anteriormente, un espectro de difracción de rayos X consiste en un patrón de intensidades. La información estructural está concentrada en dicho patrón de difracción (Figura 3.6.6) y en forma de los denominados factores de estructura, asociados a cada intensidad, cuya fase se pierde en el experimento de difracción.

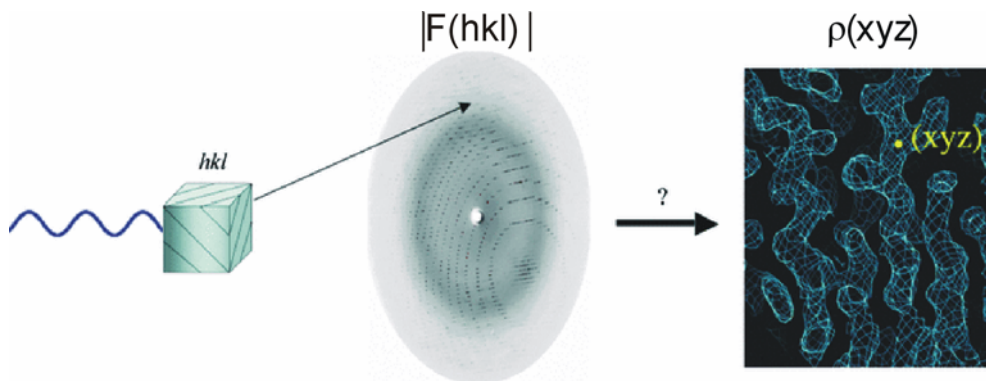


Figura 3.6.6: Proceso de resolución estructural

Haciendo uso de la teoría de interferencias de ondas es fácil demostrar que la geometría del espectro da cuenta de la celdilla unidad y la simetría de las intensidades informa de la simetría cristalina. Además, se puede establecer una relación directa entre la estructura cristalina y el experimento, mediante la transformada de Fourier (Figura 3.6.7).

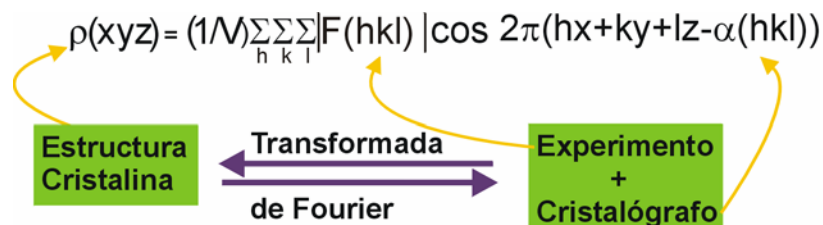


Figura 3.6.7: Función de densidad electrónica. V es el volumen de la celdilla unidad y $\alpha(hkl)$ es la fase asociada al módulo del factor de estructura, $|F(hkl)|$. Es interesante resaltar el hecho de que la densidad electrónica en un punto viene determinada por la información de todo el espectro. Inversamente, el valor de un factor de estructura (módulo y fase) viene determinado por las posiciones de todos los átomos que forman la estructura, $\rho(xyz)$

Esta expresión (Figura 3.6.7) relaciona la densidad electrónica, $\rho(xyz)$, en cada punto (xyz) del cristal, es decir la estructura, con el experimento de difracción, donde $|F(hkl)|$ es proporcional a la raíz cuadrada de la intensidad de cada punto (hkl) del espectro de difracción. Para poder resolver la ecuación es necesario disponer no sólo de los módulos, $|F(hkl)|$, sino también de las fases, $\alpha(hkl)$, asociadas a los factores de estructura. Esta información, sin embargo, se pierde durante el experimento. Esto constituye el denominado “problema de la fase” y el trabajo del cristalógrafo consiste en averiguar el valor de $\alpha(hkl)$ mediante las diferentes metodologías expuestas en el apartado 1.2.2. (Determinación estructural de macromoléculas por difracción de rayos X).

Metodologías para recuperar las fases

Brevemente recordamos que en lo que a cristales de proteína se refiere, el problema de la fase puede resolverse mediante tres métodos: (i) mediante el Reemplazo Molecular (MR, del inglés *Molecular Replacement*), haciendo uso de un modelo estructural de una proteína homóloga, previamente determinada, (ii) introduciendo átomos altamente dispersores, o método del Reemplazo Isomorfo Múltiple (MIR, del inglés, *Multiple Isomorphous Replacement*) o (iii) introduciendo átomos dispersores anómalos, o método de Dispersión Anómala Simple o Múltiple (SAD, del inglés *Single-wavelength Anomalous Diffraction* o MAD, del inglés *Multi-wavelength Anomalous Diffraction*). Con cualquiera de estos tres métodos podemos calcular unas fases iniciales, $\alpha_c(hkl)$ que, junto con las amplitudes experimentales, $|F_{\text{exp}}(hkl)|$, nos permitirán calcular una función de densidad electrónica (ecuación de la figura 3.6.7) aproximada, sobre la cual podremos construir nuestro modelo estructural.

Metodologías para mejorar las fases

En ocasiones, los valores de las fases obtenidas inicialmente son sólo aproximados, y en estas circunstancias, los mapas de densidad electrónica que se calculan a partir de éstas no son completamente interpretables. Por tanto, han de ser mejoradas haciendo uso de distintos métodos de modificación de densidad, que están basados en algún tipo de conocimiento estructural previo (Read & Kleywegt, 2001; Read, 2003).

Las técnicas de modificación de densidad más empleadas son: el aplanado de solvente (*solvent flattening*), el promediado por simetría no cristalográfica (*non-crystallographic symmetry averaging*), el ajuste a histograma (*histogram matching*) y la extensión de fases (*phase extension*).

- El aplanado de solvente parte del conocimiento de la fracción que ocupan solvente y proteína en el cristal. Asume que la densidad electrónica no puede ser negativa y fija unos valores promedios para la densidad de la proteína y del solvente.
- La técnica de promediado por simetría no cristalográfica es también una técnica muy potente y consiste en imponer equivalencia en los valores de densidad electrónica cuando existen varias copias en la unidad asimétrica relacionadas por simetría no cristalográfica. La densidad correspondiente a cada copia se promedia de manera que aumenta la relación señal/ruido. El promediado es más efectivo cuanto mayor sea el número de copias independientes en la unidad asimétrica y cuanto mayor sea el contenido de solvente.
- La técnica de ajuste a histograma es una mejora de la técnica de aplanado de solvente. Asume que la densidad electrónica de la región proteica sigue una distribución de densidad electrónica que no es constante.
- La técnica de extensión de fases se emplea para propagar la información que se tiene de las fases desde baja a alta resolución. De esta manera, si se tienen las fases para un conjunto de datos a baja resolución, se puede hacer uso de éstas sobre un conjunto de datos a alta resolución para obtener fases más allá de la resolución inicialmente obtenida. En este caso se asume que la distribución de densidad electrónica existente en los datos a baja resolución se va a conservar para los datos de alta resolución.

La mejora de fases es un proceso cíclico que termina cuando se alcanza la convergencia y que conlleva la aplicación alternada de dos procesos: la modificación del mapa y combinación de fases, de manera que con un mapa mejorado se recalculan fases, que luego se vuelven a utilizar para obtener un nuevo mapa mejorado. Todas estas técnicas pueden llevarse a la práctica con programas como DM (Cowtan & Main, 1998), SOLOMON (Leslie, 1987a; Abrahams & Leslie, 1996), RESOLVE (Terwillinger, 2000) o CNS (Brünger *et al.*, 1998).

3.6.1.6. Construcción del modelo y refinamiento.

Una vez que se dispone de un mapa de densidad electrónica interpretable, se procede al trazado de la cadena polipeptídica, proceso que se puede automatizar si se dispone de datos a alta resolución (mejores de 2.5 Å) (Perrakis *et al.*, 1999). Si el mapa no es de la suficiente calidad, el trazado tiene que realizarse manualmente. El modelo estructural inicial no suele explicar demasiado bien las observaciones experimentales (desacuerdo entre $|F_o|$ y $|F_c|$), mostrando un elevado factor de desacuerdo R (Ecuación 2.3.1). Para mejorar el modelo (minimizar R hasta un 20-25%) se realizan procedimientos de refinamiento matemático, seguidos del cálculo de mapas de densidad electrónica y de una reconstrucción manual, necesaria cuando existen errores serios en el modelo que el refinamiento matemático no puede corregir.

$$R(\%) = 100 \left(\frac{\sum_{hkl} |F_o| - K \sum_{hkl} |F_c|}{\sum_{hkl} |F_o|} \right)$$

Ecuación 3.6.1: Factor de desacuerdo.

Los parámetros atómicos que se refinan son fundamentalmente las coordenadas (x,y,z) de cada átomo y el factor térmico B isotrópico, que da cuenta de la vibración térmica o movilidad del átomo. Esto quiere decir que en una macromolécula, el número de parámetros por refinar es muy superior al número de datos experimentales, lo que obliga a introducir en el refinamiento una serie de restricciones geométricas (longitudes y ángulos de enlace, interacciones de van der Waals, ángulos de torsión, etc.), que se recogen en diccionarios.

Existen distintos tipos de refinamiento que se aplican dependiendo del estado en que se encuentre el modelo. En las primeras etapas del refinamiento se aplican los métodos de templado simulado (*simulated annealing*) y refinamiento de cuerpo rígido. Posteriormente se aplican métodos probabilísticos de máxima verosimilitud (*maximum likelihood*). Un valor bajo del factor de desacuerdo R no corresponde necesariamente a un buen modelo final. Para solucionar este problema, se ha introducido el método estadístico de validación cruzada (Kleywegt & Brünger, 1996). Así, los datos de difracción se dividen en dos grupos: un grupo que contiene el 90 - 95% de los datos y que se utiliza en el proceso de refinamiento (*working set*), y un grupo que contiene el 10 - 5% restante de datos de difracción (*test set*), que no se usan en el

refinamiento. Con este último conjunto de datos se calcula otro factor de desacuerdo R denominado R -free. El R -free mide el grado en que el modelo se ajusta a los datos de difracción que fueron excluidos del proceso de refinamiento y modelado. Por tanto, cambios en el modelo que no mejoren el ajuste con este 10% de datos de difracción harán que el R -free no baje. Así pues, el R -free está correlacionado con la exactitud del modelo atómico. Los programas que se suelen utilizar para refinar las estructuras son REFMAC (Murshudov *et al.*, 1997) y CNS (Brünger *et al.*, 1998) y para la reconstrucción manual O (Jones *et al.*, 1991)

El modelo final puede ser considerado aceptable cuando presenta unos buenos valores para el factor de desacuerdo ($R \approx 20$ -25% y R -free < 30%), se ajusta al mapa de densidad electrónica y presenta la mayoría de torsiones de enlace dentro de los valores permitidos (Ramachandran *et al.*, 1963).

3.6.2. RESOLUCIÓN ESTRUCTURAL DE CBPF.

3.6.2.1. Toma de datos de difracción.

Se recogieron datos de difracción de los dos hábitos cristalinos (I y II) obtenidos en diferentes condiciones de cristalización (apartado 3.5.2.1). Los conjuntos de datos se recogieron por medio de un detector tipo Imaging Plate empleando la radiación $\text{Cu}\alpha$ ($\lambda = 1.5418 \text{ \AA}$) proporcionada por una fuente de ánodo rotatorio Enraf-Nonius (Figura 3.6.3.a) equipado con un sistema de focalización de doble espejo y operando a 40 kV y 90 mA. Los cristales fueron crioprotectados mediante una inmersión rápida en sus respectivas soluciones de cristalización y 10% de glicerol para el hábito cristalino I y el 17% para el hábito cristalino II.

Los datos de difracción se procesaron empleando MOSFLM, SORTMTZ, REINDEX, SCALA y TRUNCATE pertenecientes al *paquete de programas de CCP4 (Collaborative Computational Project, Number 4, 1994)*. Las estadísticas del procesado de datos se detallan en la tabla 3.6.1. Ambos hábitos cristalinos pertenecían al grupo ortorrómbico $P2_12_12$, pero los parámetros de la celdilla unidad son diferentes. Mientras los cristales con hábito cristalino I ($a = 52.40$, $b = 115.79$, $c = 72.98 \text{ \AA}$) difractaron hasta una resolución de 1.95 \AA (Figura 3.6.8.a), los cristales pertenecientes al hábito cristalino II ($a = 49.13$, $b = 114.94$, $c = 75.69 \text{ \AA}$) difractaron hasta 1.67 \AA de resolución (Figura 3.6.8.b).

	Cristal nativo. Hábito I	Cristal nativo. Hábito II
Recogida de datos		
Grupo espacial	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2
Celdilla unidad		
<i>a</i> (Å)	52.40	49.15
<i>b</i> (Å)	115.79	115.01
<i>c</i> (Å)	72.98	75.79
Temperatura (K)	120	120
Longitud de onda (Å)	1.5418	1.5418
Resolución (Å)	26.9 (2.2)-2.1	21.1 (1.8)-1.7
Reflexiones totales	102.654	625.729
Reflexiones únicas	25.934	50.379
Multiplicidad	3.9 (3.9)	12.4 (11.5)
Completitud (%)	99.9 (100)	99.1(96.7)
<i>I</i> / σ <i>I</i>	11.0 (2.3)	33.1 (7.0)
<i>R</i> _{sym} *	0.10 (0.56)	0.05 (0.34)

$$R_{\text{sym}}^* = \frac{\sum_{hkl} |I_{hkl} - \langle I_{hkl} \rangle|}{\sum_{hkl} I_{hkl}} \text{ para todo el rango de resolución.}$$

Tabla 3.6.1: Estadísticas de procesamiento para los 2 hábitos cristalinos. Los valores entre paréntesis son para la concha de alta resolución.

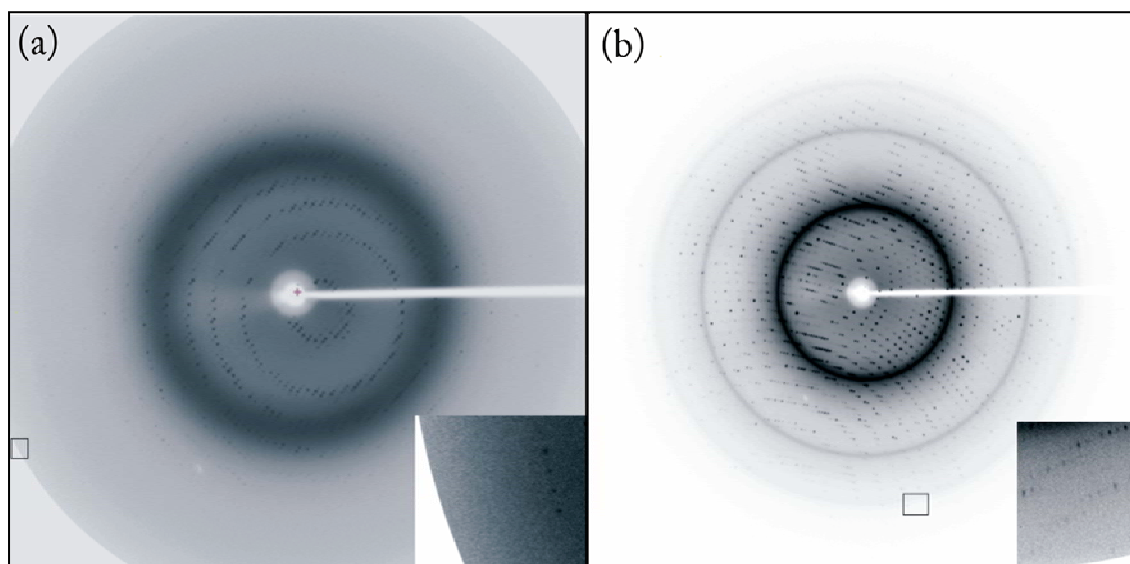


Figura 3.6.8: Patrones de difracción de los cristales de CbpF. (a) Hábito cristalino I (grado de oscilación 1°). (b) Hábito cristalino II (grado de oscilación 1°). El pequeño cuadrado en (a) muestra reflexiones a 2.0 Å de resolución; en (b) las reflexiones alcanzan el límite del detector, las cuales corresponden a 1.67 Å de resolución.

Considerando la mayor resolución obtenida en el patrón de difracción correspondiente al hábito cristalino II, se decidió emplear estas condiciones de cristalización para las posteriores etapas. Teniendo en cuenta la masa molecular de CbpF y el volumen de la celdilla unidad, obtuvimos un coeficiente de Mathews de $2.94 \text{ \AA}^3 \text{ Da}^{-1}$ (Mathews, 1968) para un monómero en la unidad asimétrica y un contenido de solvente del 58%.

3.6.2.2. Obtención de fases con soluciones de átomo pesado convencionales. Obtención de fases con yoduro.

Para resolver la estructura de CbpF se abordó la preparación de derivados de átomo pesado mediante técnicas de reemplazo isomorfo (SIR/MIR) o bien mediante dispersión anómala (SAD/MAD). La información acerca de la química de los átomos pesados y de los ligandos con la que éstos interactúan, se tomó del banco de datos de átomos pesados accesible (*Heavy Atom Databank*, <http://www.bmm.icnet.uk/had/heavyatom.html>). Esta página proporciona información acerca de la estrategia inicial de derivatización, de la manipulación física posterior en la producción de derivados y aspectos prácticos mucho menos documentados. De la escasa documentación existente acerca de la solución de átomo pesado a elegir para obtener derivados de éxito, hay que resaltar que dicha elección se ha de hacer teniendo en cuenta la secuencia aminoacídica de la proteína a resolver, las condiciones de cristalización en la que ésta cristaliza en su forma nativa (criterio químico) y el método de obtención de fases empleado (criterio de dispersión) (ver sección 1.2.4).

A partir de la composición aminoacídica de CbpF y las condiciones de cristalización, se determinó que las soluciones de átomo pesado más apropiadas para la obtención de derivados eran las sales de Pt y Au y más en concreto las soluciones de K_2PtCl_4 y $\text{KAu}(\text{CN})_2$, respectivamente. Se analizaron distintas condiciones de concentración de solución de átomo pesado (1-20 mM), a distintos pHs, por cocrystalización e inmersión y a tiempo variable entre pocos segundos y una semana. Gracias a la información proporcionada por las secciones de los mapas de Patterson y las estadísticas de obtención de fases, en ninguno de los casos pudimos certificar la obtención de derivados. Si bien la falta de isomorfismo de los cristales derivados nos impidió aplicar la técnica MIR para la resolución estructural, la baja intensidad de señal de los picos observados en los mapas de Patterson (3σ como máximo) no permitía distinguir los

átomos pesados por lo que también se descartó su posible aplicación para técnicas de dispersión anómala.

Dado que el empleo de soluciones de átomo pesado convencionales no nos proporcionó derivados de suficiente calidad para abordar la obtención de fases, se decidió emplear la metodología de uso de haluros (I, Br) como átomos pesados (Dauter *et al.*, 2000; Dauter & Dauter, 2001; Dauter *et al.*, 2001). Teniendo en cuenta lo descrito en la sección 1.2.4.4. (Introducción de átomos pesados para inmersión o cocrystalización), se abordó tanto la cocrystalización como la inmersión de CbpF con NaI a distintas concentraciones (1M, 0.5M y 0.1M) con el fin de obtener un derivado adecuado para SAD.

A pesar de que las secciones de Harker de los mapas de Patterson (Figura 3.6.9) así como las estadísticas de obtención de fases nos indicaban *a priori* un derivado más prometedor que los obtenidos anteriormente, los mapas de densidad electrónica obtenidos no tenían la suficiente calidad para abordar la construcción del modelo inicial.

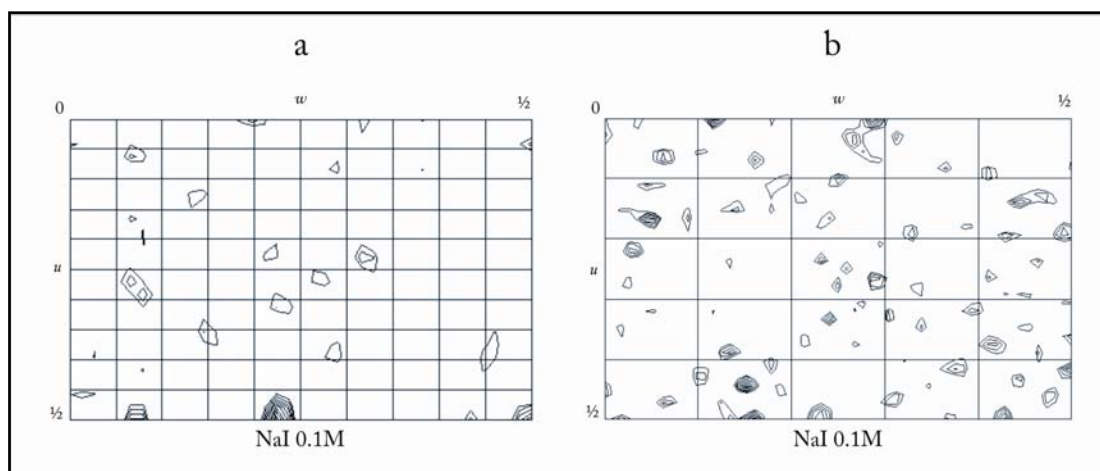


Figura 3.6.9: Secciones de Harker obtenidas en los experimentos de derivatización por cocrystalización con una solución 0.1M de NaI. Los niveles de intensidad de los picos están representados en pasos de 0.5σ comenzando a partir de 1.5σ . a. Sección de Harker del mapa de Patterson diferencia isomorfo $v = \frac{1}{2}$. b. Sección de Harker del mapa de Patterson anómalo.

3.6.2.3. Obtención de fases con complejos lantánidos.

Ejemplos de obtención de fases de estructuras macromoleculares cristalinas basadas en dispersión anómala a una única longitud de onda (SAD) han demostrado que esta aproximación tiene una aplicación general en biología estructural. Gracias a la mejora de las instalaciones de

recogida de datos y técnicas criogénicas, combinado con un procesado de datos más potente, programas de obtención de fases y modificación de densidad, la aproximación SAD prueba que en determinados casos la obtención de fases es más sencilla que obtener fases a partir de medidas a múltiples longitudes de onda (MAD) reduciendo el problema del daño por radiación y el tiempo de recogida de los datos. Se puede presentar a cualquier longitud de onda donde la dispersión anómala pueda ser observada, en muchos casos usando fuentes de rayos X de laboratorio. (Dauter *et al.*, 2002). Sin embargo, aunque a energías muy altas o bajas es posible obtener dispersión anómala, otros problemas pueden provocar que la recogida de datos resulte muy difícil.

El actual énfasis sobre la determinación estructural extensiva ha conducido al desarrollo de métodos de preparación de derivados de átomo pesado que sean más factibles y menos destructivos que los tradicionales ensayos de inmersión o cocrystalización. El método MAD a la longitud de onda de la línea de absorción K del Se, combinada con la incorporación de Se en las proteínas reemplazando metionina con selenometionina, ha revolucionado la cristalografía macromolecular (Hendrickson & Ogata, 1997). La incorporación de la selenometionina (Doublíe, 1997; Hendrickson *et al.*, 1990), también se puede llevar a cabo en células de insectos (Bellizi *et al.*, 1999; Chen & Bahl, 1991; McWhirter *et al.*, 1999) y células de mamíferos (Lustbader *et al.*, 1995), aunque la optimización de estos sistemas eucarióticos no facilita la incorporación de dicha selenometionina. Además, muchas proteínas eucariotas no pueden ser sobreexpresadas y plegadas adecuadamente en sistemas procariontas siendo necesarias modificaciones postraduccionales tales como la glicosilación o la formación de puentes disulfuro. Incluso cuando la proteína contiene selenometionina quizá no cristalice del mismo modo que la nativa.

Por lo tanto, métodos alternativos para obtener derivados para experimentos de dispersión anómala tienen que ser propuestos, incluyendo el uso de haluros (Dauter & Dauter, 1999, 2001; Dauter *et al.*, 2000, 2001; Evans & Bricogne, 2002) y de los gases nobles en los sitios de unión nativos (Vitali *et al.*, 1991; Schiltz *et al.*, 1994; Cohen *et al.*, 2001) o en sitios de unión creados por mutagénesis directa (Quillin & Mathews, 2002). La señal anómala a partir de los átomos de S intrínsecos ha sido usada de manera creciente para resolver la estructura por el método SAD (Hendrickson & Teeter, 1981; Wang, 1985; Dauter *et al.*, 1999; Liu *et al.*, 2000; Bond *et al.*, 2001; Gordon *et al.*, 2001; Micossi *et al.*, 2002). También, el método de etiquetaje basado en un

quelato lantánido con un grupo tiol reactivo ha sido propuesto por Purdy *et al.*, (2002) y la tradicional preparación de átomos pesados ha sido revisada usando inmersiones cortas (Sun & Radaev, 2002; Sun *et al.*, 2002). Técnicas tales como la electroforesis en geles nativos (Boggon & Shapiro, 2000) o la espectrometría de masas (Cohen *et al.*, 2000; Sun & Hammer, 2000) han sido usadas para la búsqueda rápida de derivados.

El exitoso uso de una nueva clase de complejos de Gd (Figura 1.2.9) para obtener derivados de átomo pesado que explotan la alta señal anómala en su línea de absorción LIII ($\lambda = 1.7106 \text{ \AA}$, $f'' = 28 e^-$), pone en evidencia la utilidad de estos derivados (Girard *et al.*, 2003a; Girard *et al.*, 2003b). En este sentido, debido a sus intensas “líneas blancas” y a sus elevados valores de f'' , los átomos lantánidos son de gran interés para resolver estructuras de macromoléculas biológicas usando métodos SAD o MAD. Los derivados son obtenidos usando complejos lantánidos altamente solubles que pueden ser fácilmente introducidos dentro de cristales de proteína a alta concentración bien por inmersión o cocrystalización (Girard *et al.*, 2002; Girard *et al.*, 2003a; Girard *et al.*, 2003b). Esta capacidad ha sido ya demostrada incluso para obtener derivados de átomo pesado en fases cúbicas lipídicas para los ensayos de proteínas de membrana (Girard *et al.*, 2004). Otros complejos lantánidos como Yb-HPDO3A y Lu-HPDO3A proporcionaron también de modo similar al complejo Gd-HPDO3A buenos cristales derivados para experimentos de obtención de fases (Girard *et al.*, 2003c). En estos casos, lutecio e yterbio, con líneas de absorción LIII a menores longitudes de onda que la correspondiente al Gd, forman complejos lantánidos que son de especial interés para experimentos con radiación sincrotrón (Girard *et al.*, 2003c) aunque el átomo de gadolinio puede proporcionar la mejor señal anómala en las fuentes de ánodo rotatorio convencional de los laboratorios.

Recogida de datos y procesamiento.

Para eliminar cualquier resto de complejo de Gd del líquido que rodea los cristales derivados, los cristales de CbpF fueron sumergidos unos pocos segundos en soluciones nativas que contenían 17% de glicerol como crioprotector. Antes de la recogida de datos, todos los cristales fueron enfriados a 100K (Oxford Cryosystems).

Con el fin de emplear el mejor derivado para la resolución estructural de CbpF por la técnica SAD, se recogieron datos con todos los derivados lantánidos ensayados (Figura 1.2.9) tanto por

cocrystalización como por inmersión (ver apartado 3.5.5). Los datos del derivado lantánido de mayor calidad (CbpF: Gd-HPDO3A obtenido por cocrystalización en 50mM de complejo) fueron recogidos en la línea ID29 del ESRF (Grenoble) a la longitud de onda de la línea de absorción LIII del Gd ($\lambda=1.711 \text{ \AA}$). Los datos fueron procesados usando los programas XDS (Kabsch, 1988) y MOSFLM (Leslie, 1987), y escalados con el programa SCALA del paquete CCP4 (Collaborative Computational Project, Number 4, 1994). Los datos relativos a la recogida y las estadísticas del procesado se muestran en la tabla 3.6.2.

Recogida de datos	CbpF:HPDO3A
Grupo espacial	P2 ₁ 2 ₁ 2
Celdilla unidad	
<i>a</i> (Å)	49.35
<i>b</i> (Å)	114.52
<i>c</i> (Å)	76.23
Temperatura (K)	100
Longitud de onda (Å)	1.7106
Resolución (Å)	100-2.23 (2.36-2.23)
Reflexiones totales	143.495
Reflexiones únicas	21.579
Multiplicidad	6.6 (5.5)
Complejidad (%)	98.9 (95.0)
<i>I</i> / $\sigma(I)$	10.4 (7.7)
<i>R</i> _{sym} (%)*	4.7 (8.9)

Tabla 3.6.2: Recogida de datos y estadísticas de procesado.

Cálculo de la subestructura de átomos pesados y obtención del mapa de densidad electrónica

El cálculo de la subestructura de átomos pesados se realizó tanto de forma manual a partir de las secciones de Harker de los mapas de Patterson anómalos como *ab initio* empleando el programa SHELXD (Schneider & Sheldrick, 2002). El programa SHELXD es muy efectivo y robusto a la hora de encontrar la subestructura, sin embargo, hay que tener en cuenta ciertos parámetros críticos para este fin: la resolución a la que truncan los datos, que se evalúa muy bien con XPREP y SHELXC, el número de sitios que buscar, que debe ser al menos un 20%

más del número real, y en el caso de inmersión del cristal en una solución de átomo pesado, permitir que la búsqueda sea posible en posiciones especiales. SHELXD encontró 4 átomos de Gd por cada molécula de CbpF. Una vez encontrada la subestructura de átomos pesados, dichas coordenadas fueron usadas para calcular las fases con SHARP (De la Fortelle and Bricogne, 1997), obteniendo fases iniciales a una resolución de 2.23 Å. Las fases fueron seguidamente refinadas mediante programas de mejora de fases como DM integrado dentro de la interfase de SHARP o SHELXE integrado dentro de la interfase HKL2MAP (Schneider *et al.*, 2004) y asumiendo un 58% de contenido en solvente tal y como estimaba el programa TRUNCATE del paquete CCP4 (Cowtan & Main, 1996). SHELXE no es un programa tan eficaz como DM para la mejora de fases pero aporta información muy fiable acerca del enantiómero correcto o de si la estructura se puede resolver con los datos recogidos.

Tras obtener el enantiómero correcto, SHARP nos proporcionó finalmente mapas de densidad electrónica experimentales (2fofc y fofc) de excelente calidad así como un fichero con las coordenadas de los átomos pesados. Dichos mapas de densidad electrónica experimentales fueron visualizados con el programa O (Jones *et al.*, 1991), iniciando de este modo el trazado *de novo* del modelo.

3.6.2.4. Construcción y refinamiento de la estructura del complejo CbpF-colina.

A partir del mapa de densidad electrónica experimental (resolución = 2.23 Å), el trazado de la cadena polipeptídica del derivado CbpF: Gd-HPDO3A, se realizó manualmente con el programa O (Jones *et al.*, 1991) para la mayoría de los residuos excepto para dos pequeños fragmentos de dicha cadena (residuos 169-177 y 147-150) donde la densidad electrónica no estaba bien definida. El refinamiento del modelo se llevó a cabo con el programa CNS (Brünger *et al.*, 1998) que realiza una dinámica molecular de templado simulado para aliviar las tensiones que pudieran existir (se llegó hasta 3000K, límites de resolución 100 - 2.23 Å). Con el fin de resolver la estructura completa y mejorar la resolución del modelo, se recurrió al procedimiento del reemplazo molecular mediante MOLREP (perteneciente al conjunto de programas CCP4) empleando los datos nativos de los dos hábitos cristalinos obtenidos y el modelo de CbpF construido previamente. Posteriormente se llevó a cabo, de nuevo, el refinamiento de ambos modelos empleando el programa CNS (Brünger *et al.*, 1998) en las etapas iniciales y el

programa REFMAC (Murshudov *et al.*, 1997) en las etapas finales. A partir del reemplazo con los datos nativos correspondientes al hábito cristalino II (resolución = 1.67 Å), obtuvimos un modelo de excelente calidad pero no proporcionaba mapas de densidad electrónica que permitieran resolver los dos fragmentos mal definidos y completar el modelo. Sin embargo, a partir del reemplazo con los datos correspondientes al hábito cristalino I (resolución = 2.1 Å), obtuvimos un modelo de la proteína completa. Tras el proceso cíclico de modelado y refinamiento se terminó de construir la cadena y se añadieron 7 moléculas de colina. Finalmente se modelaron las moléculas de agua.

Las estadísticas finales del refinamiento y las características de los modelos (hábitos I y II) se muestran en la tabla 3.6.3. La estereoquímica de los modelos se verificó con el programa PROCHECK (Laskowski *et al.*, 1993). En la figura 3.6.10 se representa el diagrama de Ramachandran, donde se presentan los ángulos de torsión Ψ y ϕ de los residuos de la cadena polipeptídica. Para el hábito I, el 86.7% de los residuos se encuentran en las zonas más favorables, el 13.3% en zonas permitidas y ninguno en zonas prohibidas, mientras que para el hábito II, el 90.8% de los residuos se encuentran en las zonas más favorables, el 9.2% en zonas permitidas y ninguno en zonas prohibidas.

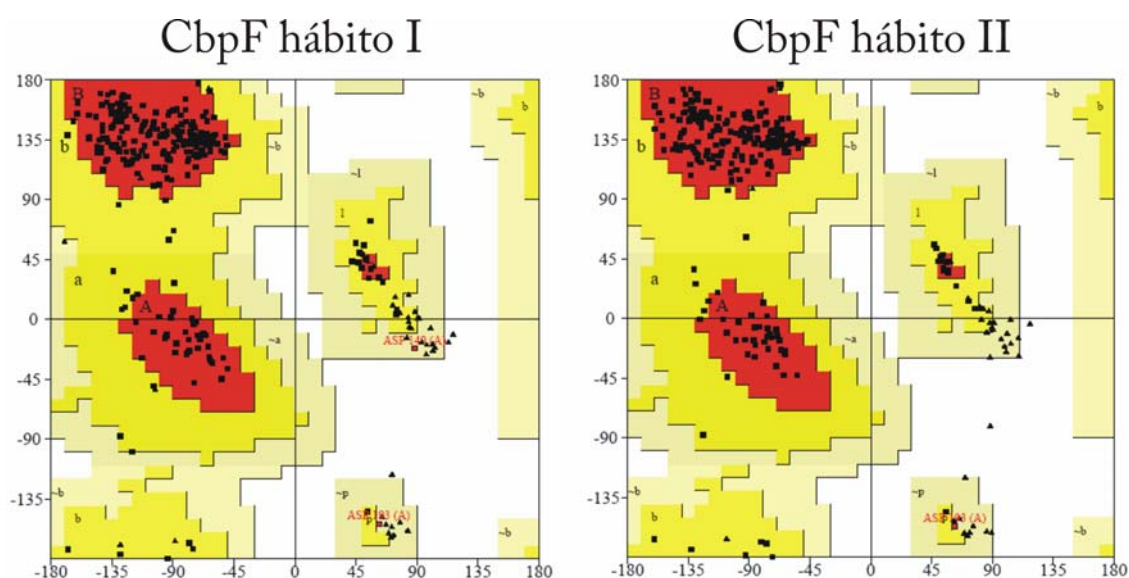


Figura 3.6.10: Diagramas de Ramachandran de los modelos estructurales correspondientes al hábito cristalino I y II (Laskowski *et al.*, 1993).

	Cristal Nativo. Hábito I	Cristal Nativo. Hábito II
Refinamiento		
Resolución (Å)	26.9-2.1	21.1-1.7
R (%)	20.9	19.9
R-free (%)	24.0	21.1
Modelo		
Moléculas de proteína por unidad asimétrica	1	1
Aminoácidos	311	298
Moléculas de colina	7	7
Moléculas de agua	311	494
RMSD		
Longitud de enlace (Å)	0.010	0.006
Angulo de enlace (°)	1.294	1.089

Tabla 3.6.3: Estadísticas del refinamiento de CbpF y modelos depositados (PDBs: 2V05 y 2V04).

3.6.3. ANÁLISIS DE LA ESTRUCTURA TRIDIMENSIONAL DEL COMPLEJO CBPF-COLINA

3.6.3.1. Organización y plegamiento general.

La estructura de CbpF está plegada en dos módulos bien definidos (Figura 3.6.11): el módulo N-terminal (residuos 1-128), que presenta una conformación en forma de disco (37 x 34 x 24 Å), y el módulo de unión a colina (residuos 129-311) organizado a su vez en dos dominios denominados CI y CII (Figura 3.6.11) que presentaban el plegamiento superhelicoidal encontrado previamente en otras CBPs. El dominio CII (residuos 193-311) contiene cinco repeticiones de unión a colina ($p1-p5$), con el motivo consenso (GWXK-X₄₋₅-WYY-hidrofóbico-X₃₋₅-GXMX₂₋₃), y un extremo C-terminal de 16 aminoácidos (295-311) que presenta el mismo plegamiento superhelicoidal que el resto del CBM. Cada repetición está compuesta por una horquilla β simétrica seguida de un lazo y una región sin estructura secundaria regular, y están estrictamente dispuestas siguiendo un plegamiento superhelicoidal a izquierdas, similar al observado en C-LytA (Fernández-Tornero *et al.*, 2001), el dominio CI de Cpl-1 (Hermoso *et al.*, 2003) y el módulo de anclaje de Pce (Hermoso *et al.*, 2005). Las horquillas se extienden perpendicularmente al eje de la superhélice y cada repetición está localizada a 120° en sentido antihorario. El dominio estructural CI (residuos 129-192) presenta dos repeticiones modificadas ($dp5$ y $dp6$), mostrando inserciones/adiciones de aminoácidos dentro de la horquilla β de la

repetición *dp6* y también en el bucle posterior a la horquilla β de la repetición *dp5* (representadas en rojo y azul en la figura 3.6.12a). Este dominio CI define una región de conexión entre el modulo N-terminal y el CBM. El dominio de unión a colina puede ser descrito a grandes rasgos como un prisma triangular de unos 76 Å de altura con caras de 26 Å, con los sitios de unión a colina situados a lo largo de las tres caras laterales.

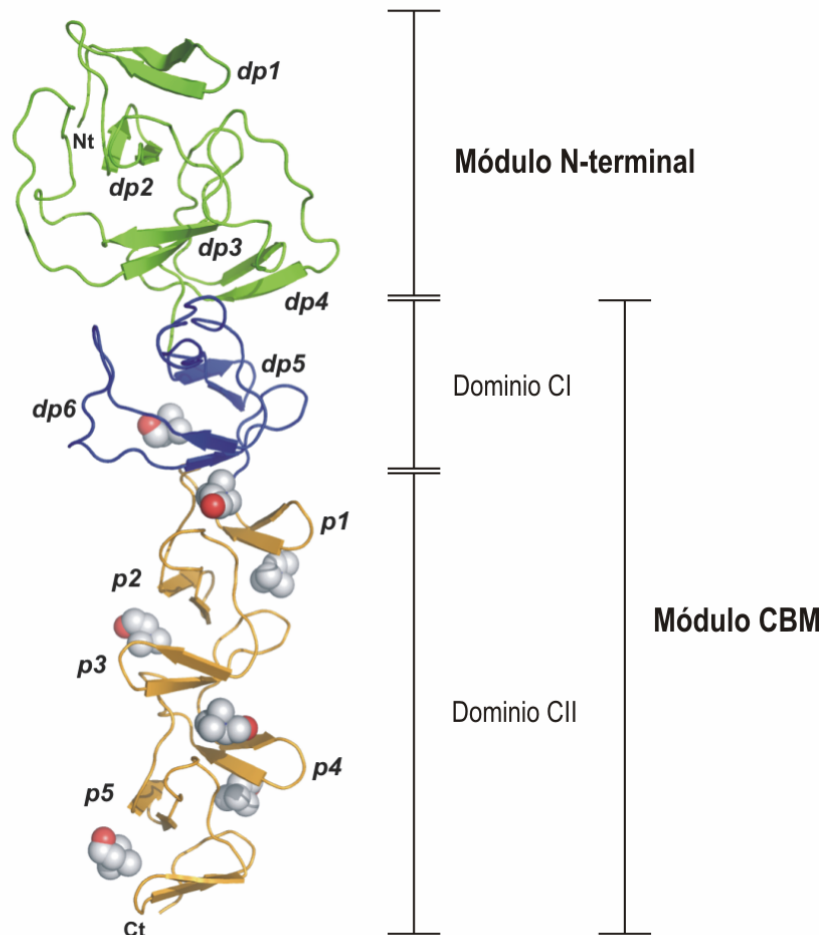


Figura 3.6.11: Estructura tridimensional de CbpF en complejo con colina. La estructura de CbpF aparece coloreada en función de las regiones estructurales diferenciadas: en verde, región N-terminal; en azul, región interdominio; en naranja, región C-terminal. Las moléculas de colina unidas a los sitios de unión a colina están representadas por esferas.

CbpF presenta una característica única con respecto a otros miembros de las proteínas de unión a colina: la secuencia de CbpF está constituida enteramente por repeticiones de unión a

colina, unas de ellas siguiendo estrictamente el motivo consenso (región CII, repeticiones $p1$ - $p5$), otras aparecen como repeticiones parcialmente modificadas por inserciones de aminoácidos (región CI, $dp5$ - $dp6$) y otras altamente modificadas, tanto por inserciones de aminoácidos como por mutaciones sobre diferentes posiciones del motivo consenso (módulo N-terminal, $dp1$ - $dp4$) (Figura 3.6.12a). La presencia de este esquema de repeticiones de unión a colina a lo largo de toda la secuencia primaria produce la forma alargada de la estructura de CbpF (más de 117 Å de longitud) de tal modo que las repeticiones canónicas y modificadas son perfectamente superponibles entre sí (Figura 3.6.12b).

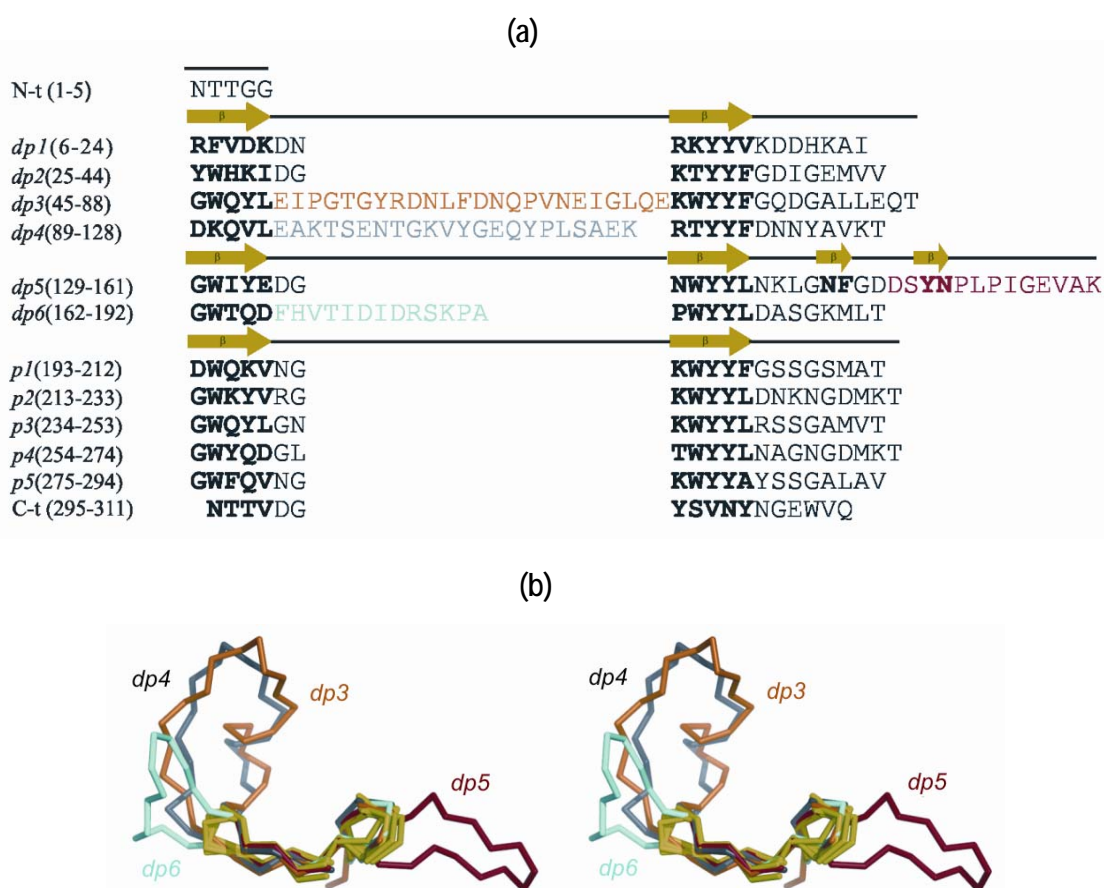


Figura 3.6.12: Análisis secuencial y estructural de CbpF (a). Comparación de la secuencia aminoacídica de los sitios de unión a colina. Las repeticiones canónicas son representadas con la letra p y las repeticiones modificadas como dp. Los aminoácidos implicados en hebras beta aparecen resaltados en negrita. (b) Representación estereográfica de la superposición de carbonos α de todas las repeticiones de unión a colina. En verde, se representan las repeticiones canónicas de unión a colina $p1$, $p2$, $p3$, $p4$, $p5$ así como las repeticiones modificadas $dp1$ y $dp2$. En naranja se representa $dp3$, en gris $dp4$, en azul claro $dp6$, y en rojo $dp5$. El intervalo de valores de la desviación cuadrática media (r.m.s.d) de las repeticiones $p1$ a $p5$ va de 0.30 a 0.70Å; mayor diferencia se observa entre las repeticiones canónicas y repeticiones modificadas (los valores r.m.s.d oscilan entre 0.44 y 1.02 Å para la superposición de la repetición $p1$ con las $dp1$ - $dp6$).

La estructura cristalina de CbpF en complejo con colina muestra que hasta siete sitios están disponibles (Figura 3.6.11). Estas siete moléculas de colina están todas localizadas en el CBM en sitios de unión a colina canónicos de tal modo que, al menos, tres residuos aromáticos estructuralmente conservados (dos Trp y una Tyr) más un residuo hidrofóbico (Met o Leu) forman una cavidad en la que se sitúa el grupo metilo de la colina (Figura 3.6.13). Este patrón para los sitios de unión se observa tanto en el dominio CI como en el CII, mientras que en el módulo N-terminal, constituido por cuatro repeticiones de unión a colina altamente modificadas (*dp1-dp4*), no se cumplen los requerimientos para la unión de colina (como prueba la ausencia de dicha molécula en el N-terminal de CbpF). En las repeticiones *dp1*, *dp2* y *dp4* al menos uno de los residuos críticos de Trp es reemplazado por Thr o Lys (Figura 3.6.12a) de modo que se produce una reducción en la capacidad de unión a colina y, además en el caso de las mutaciones por Lys posibilitan la formación de interacciones por puente salino críticas en la disposición modular de CbpF. Particularmente interesante resulta el potencial sitio de unión a colina formado entre las repeticiones *dp3* y *dp4* que presenta *a priori* todos los aminoácidos necesarios, pero que no une colina (Figura 3.6.14) debido a la presencia de un residuo de Lys (Lys 96) en el interior de la cavidad. Por otro lado existe un sitio de unión a colina funcional formado por las repeticiones *dp5* y *dp6* del dominio CI que están modificadas respecto a las repeticiones canónicas. Las repeticiones *dp5* y *dp6* tienen inserciones de 13 residuos cada una y, a pesar de esta gran alteración, los aminoácidos que participan en el reconocimiento de colina (Trp130, Trp137, Tyr182 y Met190) están conservados, cumpliéndose de este modo los requisitos básicos para la unión.

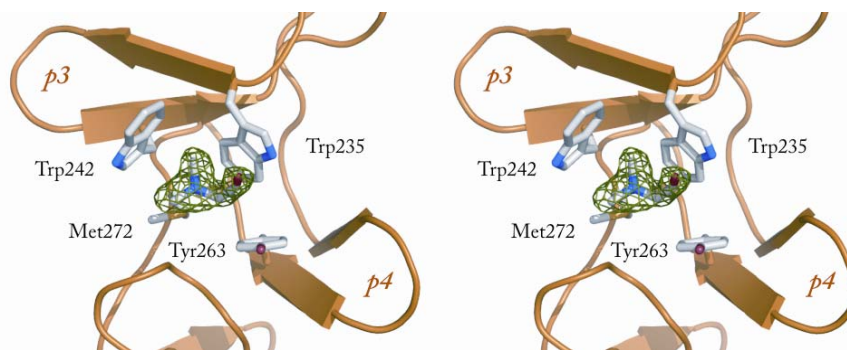
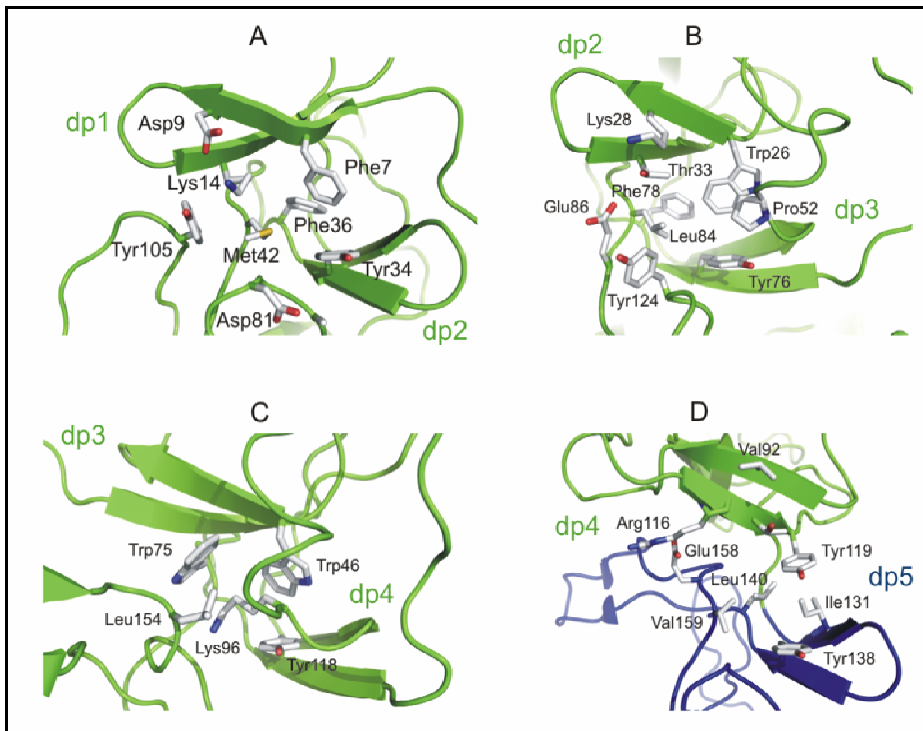


Figura 3.6.13: Representación estereográfica de la molécula de unión a colina unida al sitio de unión a colina, mostrando su mapa de densidad electrónica $F_o - F_c$ representado a una relación señal / ruido 1σ (verde).

Repeticiones de unión a colina en el módulo N-terminal



Repeticiones de unión a colina en el CBM

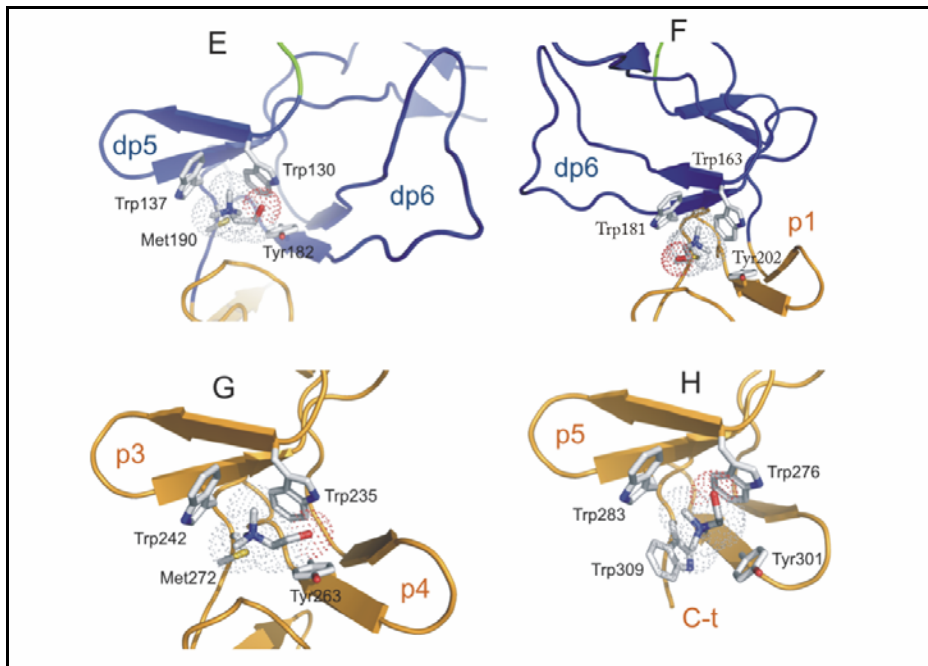


Figura 3.6.14: Repeticiones de unión a colina en CbpF tanto en el dominio funcional (A-D) como en el dominio de unión a colina (E-H).

3.6.3.2. Divergencia secuencial en las repeticiones de unión a colina.

La divergencia secuencial de las repeticiones de unión a colina observada en CbpF es generada por mutaciones sobre los aminoácidos consenso de la repetición, por inserción de aminoácidos en el giro de la horquilla β o en el lazo, o por combinación de ambos (Figura 3.6.12a). Como se ha descrito anteriormente, la alta divergencia de las primeras cuatro repeticiones (*dp1-dp4*) es responsable de la incapacidad de unión de colina en la región N-terminal (Figura 3.6.14). Además, modificaciones en la naturaleza y composición de las repeticiones de unión a colina alteran de forma muy crítica otras características de CbpF. Inserciones de aminoácidos observadas en las repeticiones *dp3*, *dp4*, *dp5* y *dp6* (Figura 3.6.12a) originan repeticiones más largas con 43, 39, 32 y 30 aminoácidos, respectivamente, frente a los 20 aminoácidos de las repeticiones de unión a colina canónicas. Aquellas inserciones de aminoácidos están principalmente compuestas por residuos hidrofóbicos y cargados negativamente (29 y 13 aminoácidos, respectivamente, de 72 aminoácidos insertados en las repeticiones *dp3*, *dp4*, *dp5* y *dp6*). La naturaleza hidrofóbica y cargada de los residuos insertados producen tanto un módulo N-terminal compacto como una elevada carga global electronegativa sobre una de sus caras, en contraposición al carácter electropositivo del módulo de unión a colina (Figura 3.6.15)

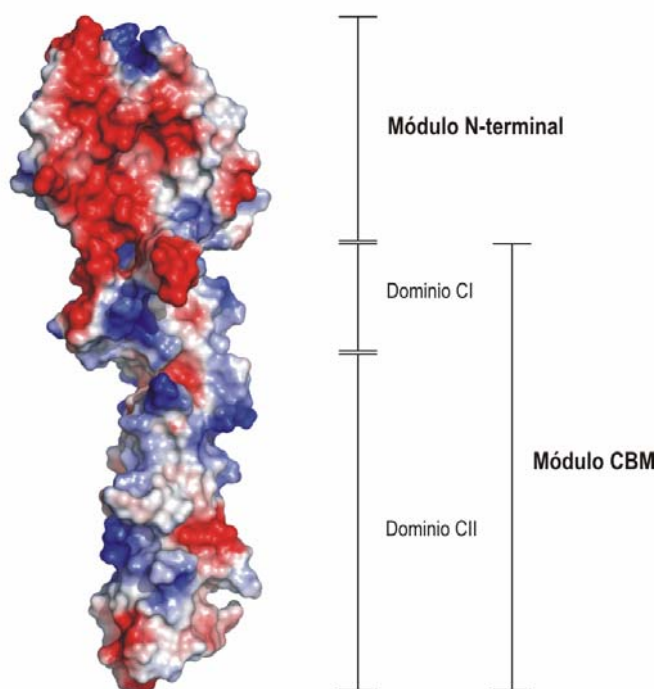


Figura 3.6.15: Superficie molecular de CbpF representada en función de su potencial electrostático (potencial negativo coloreado en rojo y positivo coloreado en azul).

Las variaciones en la secuencia de las repeticiones de unión a colina son también responsables de la disposición modular en CbpF. Las interacciones intermodulares se producen a través del módulo N-terminal y el dominio CI. La Lys90 (que reemplaza un residuo de Trp en dp4) se encuentra a distancia de puente salino respecto a los residuos de Asp89 (3.37 Å) y Glu133 (4.14 Å), y la Lys96 está implicada en una red de interacciones por puente salino conectando el módulo N-terminal y el módulo de unión a colina (Lys96-Glu158-Arg116-Glu114). La estabilización de esta red está asegurada por la fijación de la posición del residuo de Lys96 a través de interacciones tanto con el posible sitio de unión a colina formado entre las repeticiones *dp3* y *dp4* y con la Leu 154 (Figura 3.6.16b). Estas inserciones de aminoácidos y las nuevas interacciones en la que están implicadas son responsables de la modificación de la disposición de las repeticiones produciendo un ángulo de alrededor de 140° entre módulos (Figura 3.6.16a).

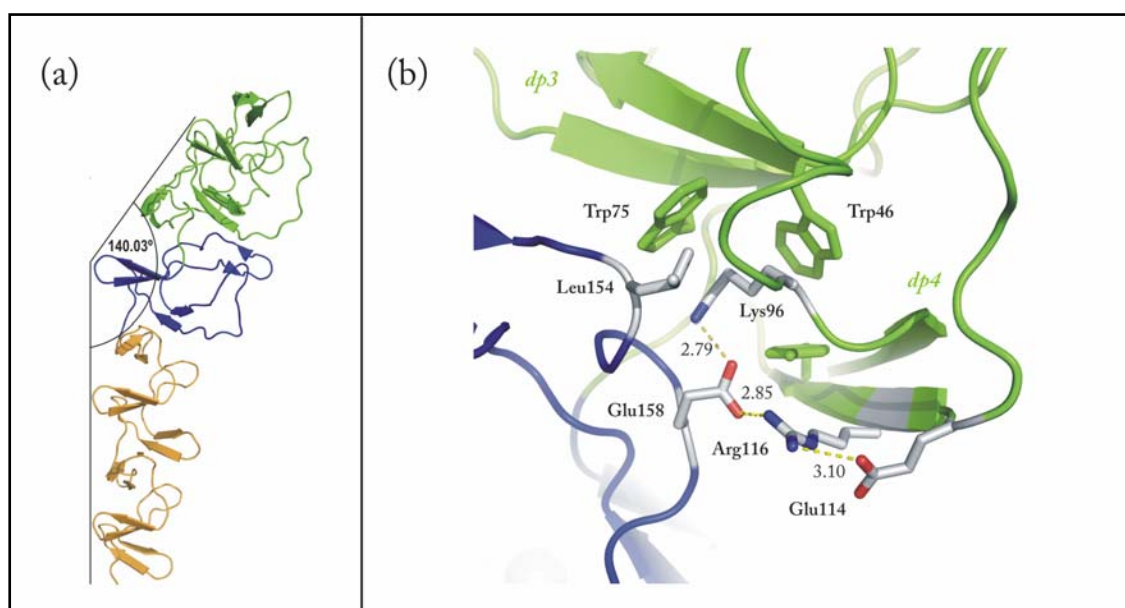


Figura 3.6.16: (a) Disposición relativa entre módulos tomando como referencia el eje helicoidal formado por el módulo de unión a colina. (b) Interacciones intermodulares que intervienen en la estabilidad del modelo.

Bisagra molecular en CbpF.

La obtención de dos estructuras cristalinas diferentes para CbpF (ver sección 3.6.2.5), nos reveló la existencia de una bisagra molecular entre el módulo N-terminal y el dominio CI del CBM (Ver figura 3.6.17). Esta bisagra molecular se dedujo a partir de que las diferencias

estructurales principalmente concernían a la posición relativa de los módulos (ver Figura 3.6.18) debido a una redistribución de las interacciones entre los aminoácidos del módulo N-terminal y el dominio CI del CBM. Esto quedaba reflejado en cambios en la conformación de los aminoácidos y en la calidad del mapa de densidad electrónica de los dos grandes bucles pertenecientes a las repeticiones *dp5* y *dp6* del dominio CI (residuos 141-161 y 167-179). De este modo, cuando todos los residuos de ambos bucles están claramente definidos en el mapa de densidad electrónica del hábito cristalino I, trece residuos de la parte más externa de estos bucles no están definidos en el hábito cristalino II (Figura 3.6.17)

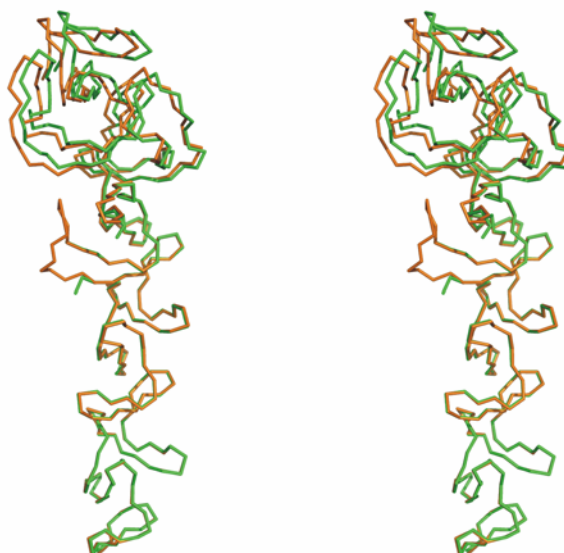


Figura 3.6.17: Representación estereográfica de la superposición del modelo cristalográfico del hábito cristalino I (naranja) frente al hábito cristalino II (verde)

3.6.3.3. Modelo de unión de CbpF a la cadena de teicoico. Modelo de interacción de CbpF a la pared de neumococo.

Para la construcción del modelo CbpF-teicoico se utilizó el fragmento pentamérico de ácido teicoico resuelto por RMN (Klein *et al.*, 1996). Con el fin de mantener las conformaciones originales de los anillos de azúcar de los ácidos teicoicos, solamente se realizaron torsiones, siempre permitidas, para los residuos de fosforilcolina y la molécula de ribitol. Tomando como modelo de interacción de una proteína de unión a colina con la cadena de teicoico el modelo de interacción de la fosfocolinesterasa de neumococo (Pce) con dicho teicoico (Hermoso *et al.*,

2005), superpusimos mediante el programa O (Jones *et al.*, 1991) los respectivos módulos canónicos de unión a colina obteniendo un modelo preliminar de interacción CbpF-teicoico. La presencia de colina en los sitios de unión a colina en la forma cristalizada nos sirvió de guía para la generación de un modelo más fiable. Finalmente, el modelo se minimizó energéticamente empleando el programa CNS (Brünger *et al.*, 1998).

Hasta tres unidades repetidas de polisacárido pudieron ser situadas a lo largo de la superficie del módulo de unión colina. Curiosamente al situar una cuarta unidad de polisacárido, se produce un giro en la cadena de ácido teicoico impidiendo la interacción con la cavidad electronegativa del módulo N-terminal (Figura 3.6.18).

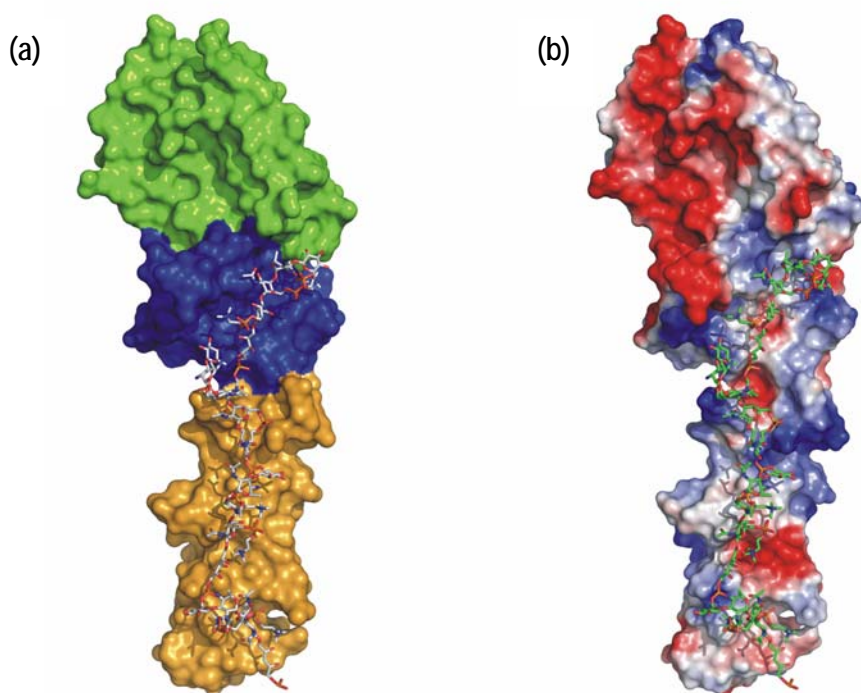


Figura 3.6.18: Modelo de interacción de la superficie de CbpF con la cadena de teicoico. (a) Verde (región N-terminal del módulo funcional), azul (región de conexión intermodular) y naranja (módulo de unión a colina). (b) Superficie de potencial electrostático. En rojo se representan las regiones cargadas negativamente y en azul las cargadas positivamente.

La obtención de la estructura del peptidoglicano por RMN (Meroueh *et al.*, 2006) proporcionó una conformación helicoidal a derechas de la cadena polisacáridica produciendo un eje de simetría 3 en la molécula. Este resultado condujo al modelo panel de abeja propuesto para la pared celular. El modelo de CbpF podría ser directamente acomodado dentro de este patrón (Figura 3.6.19)

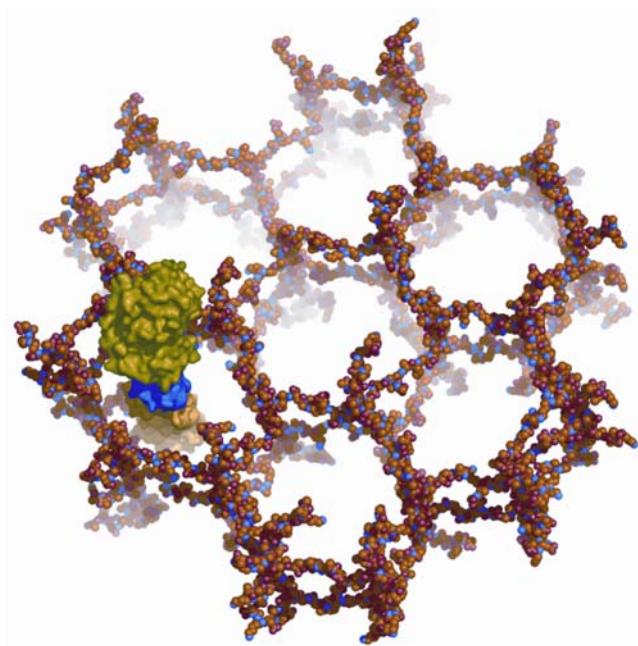


Figura 3.6.19: Modelo de unión de CbpF a la pared de neumococo.

3.6.4. BÚSQUEDA DE LA FUNCIÓN DE CBPF.

3.6.4.1. Ensayos de unión con diferentes análogos de la pared y análogos de teicoico.

Dada la ausencia de información acerca de la posible función de CbpF, se trató de conseguir información funcional a partir del análisis secuencial-estructural de la proteína y de la información relativa a otras CBPs de neumococo. A pesar de estar compuesta exclusivamente por motivos de unión a colina (canónicos y modificados), CbpF cumple la organización bimodular característica de otras CBPs de neumococo, donde el módulo de unión a colina se encuentra situado en el extremo C-t cumpliendo su correspondiente papel de anclaje a la pared celular (ver sección 1.1.4). Sin embargo, la función de CbpF que debería residir en el extremo N-t (módulo funcional) no se pudo extraer por analogía con otros módulos funcionales ya que su plegamiento y secuencia se asemejan más a un dominio de unión a colina compuesto por repeticiones altamente modificadas que a cualquier otro módulo funcional descrito hasta la fecha. Este novedoso plegamiento en el módulo funcional de CbpF sin función catalítica conocida indicaba *a priori* que podría tratarse más bien de una proteína con una función de reconocimiento más que con función catalítica. Otro de los aspectos que también apuntaban en

la dirección de la función de reconocimiento de CbpF fue obtenido a partir del análisis de la superficie de potencial electrostático de la misma. Este análisis reveló la existencia de una cavidad en el módulo funcional con una elevada cantidad de residuos ácidos (10 aminoácidos ácidos), así como la presencia de varias Asn, Gln y residuos aromáticos, sugiriendo que esta región pudiera estar implicada en el reconocimiento molecular. De hecho, este tipo de residuos es típico de regiones de unión de carbohidratos (Malik & Sandra, 2007) por lo que es factible pensar en una función de este tipo para el módulo N-terminal de CbpF. Además, la existencia y localización de dos largos bucles extendidos hacia el solvente en la región interdominio, donde la densidad electrónica es de peor calidad, podría indicarnos que probablemente emplee dicha región en el reconocimiento del sustrato estabilizando dicha región.

Con el fin de probar esta posible función de reconocimiento de carbohidratos y teniendo en cuenta el entorno molecular en el que se encuentra CbpF, se abordaron ensayos de unión de CbpF tanto con azúcares análogos de la pared celular como con azúcares análogos del ácido teicoico (ver tabla 3.6.4). Los ensayos de unión se realizaron mediante técnicas de cristalización por inmersión y cocristalización barriendo concentraciones de azúcares entre 100 y 10 mM. En ninguno de los casos los mapas de densidad electrónica obtenidos nos revelaron la presencia del azúcar ensayado.

Análogos de la pared celular	Análogos del ácido teicoico
N-acetilmurámico	N-acetil-D-galactosamina-6-sulfato
Glucosamina	N-acetil-D-galactosa-6-sulfato
N-acetil-glucosamina (NAG)	Ribitol
Chitobiosa (NAG) ₂	D-ribulosa-5-fosfato
Chitotetraosa (NAG) ₄	L-α-glicerofosforilcolina
Chitohexaosa (NAG) ₆	
Cellotetraosa	

Tabla 3.6.4: Análogos de la pared celular y del ácido teicoico empleados para ensayos de unión con CbpF mediante cristalización.

Dada la especificidad de unión de las proteínas por determinados azúcares como ha sido puesto de manifiesto en el caso de Cpl-1 (Pérez-Dorado *et al.*, 2007), no se descartó el hecho de que pudiéramos no haber encontrado el correcto. Tras la obtención de una serie de fragmentos sintéticos de peptidoglicano (Meroueh *et al.*, 2006) y con ello de la aparición de un conjunto de

nuevos análogos de la pared celular (Figura 3.6.20), decidimos emplearlos para probarlos como posibles sustratos de CbpF.

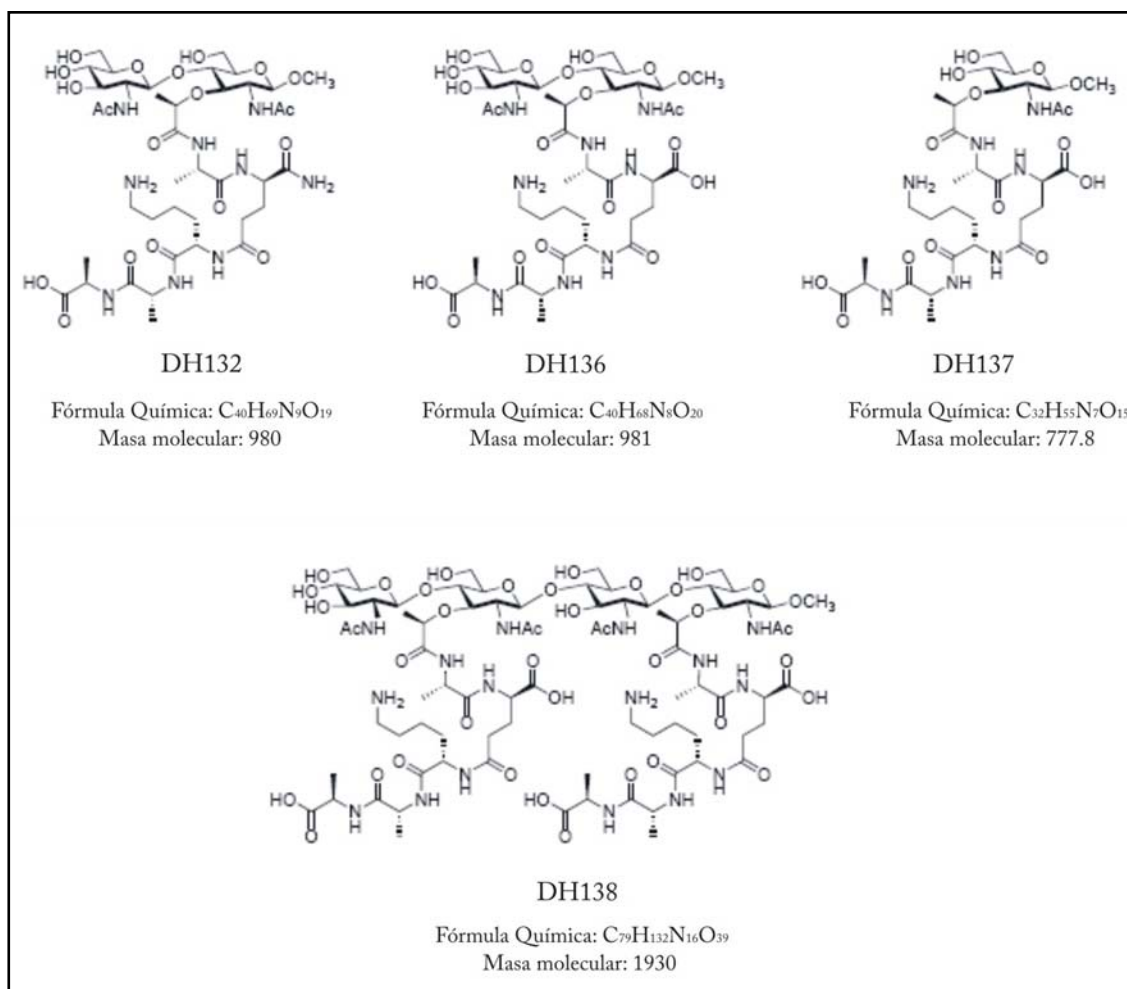


Figura 3.6.20: Fragmentos sintéticos de la cadena de peptidoglicano.

De modo análogo al procedimiento seguido con los anteriores posibles sustratos, se llevaron a cabo ensayos de unión por cocristalización e inmersión barriendo concentraciones de ligando entre 10 y 50 mM. En el caso del empleo del ligando $(2S5P)_2$ (denominado DH138) se produce un cambio en los parámetros de celdilla (Tabla 3.6.5) que se traduce en un drástico giro de más de 10° del módulo funcional respecto del módulo de anclaje, indicando la posibilidad de algún tipo de interacción con dicho ligando (Figura 3.6.21). Si bien no se obtuvo la densidad electrónica del análogo del peptidoglicano que proporcionara una prueba directa de interacción, estos indicios sentaron las bases para realizar estudios de acoplamiento y resonancia de

Plasmón Superficie con las paredes de la bacteria.

Recogida de datos	Cristal nativo. Hábito I	Complejo CbpF: (2S5P) ₂
Grupo espacial	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2
Celdilla unidad a, b, c (Å)	52.40, 115.79, 72.98	45.06, 112.73, 79.50
Longitud de onda (Å)	1.5418	1.5418
Resolución (Å)	26.9 (2.2)-2.1	45.1 (2.7)-2.4
Multiplicidad	3.9 (3.9)	5.5 (5.5)
Completitud (%)	99.9 (100)	87.0 (84.0)
<i>I</i> / σ /	11.0 (2.3)	9.7 (7.9)
<i>R</i> _{sym} *	0.10 (0.56)	0.19(0.56)
Refinamiento		
Resolución	26.9-2-1	45.1-2.4
Nº de reflexiones	102654	152424
<i>R</i> _{work} / <i>R</i> _{free}	20.9/24.0	19.2/27.2
Nº de átomos		
Proteínas	2581	2492
Ligando	49	49
Agua	311	285
Factores B		
Proteína	34.82	15.9
Ligando/ion	53.89	49.7
Agua	41.92	23.4
Desviaciones RMS		
Longitud de enlace (Å)	0.010	0.006
Angulo de enlace (°)	1.294	1.147

Tabla 3.6.5: Análogos de la pared celular y del ácido teicoico empleados para ensayos de unión con CbpF mediante cristalización.

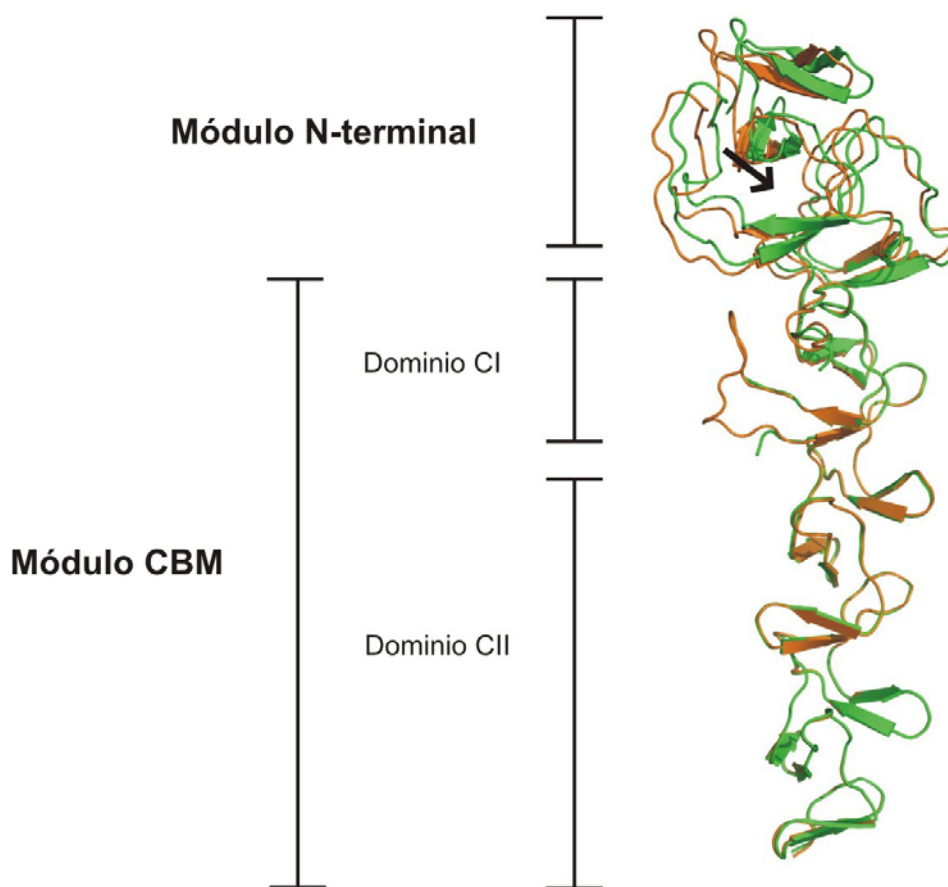


Figura 3.6.21: Superposición estructural de las estructuras cristalinas de CbpF nativa (naranja) y CbpF en presencia del análogo de peptidoglicano (2S5P)₂. La estructura cristalina de CbpF obtenida por inmersión con el análogo del peptidoglicano presenta un dominio CII similar a la proteína nativa, pero el dominio CI presenta una elevada flexibilidad en los bucles y el módulo N-terminal está rotado 10° respecto al de la proteína nativa. El posible sitio de unión del peptidoglicano obtenido por simulaciones de Docking está indicado con una flecha.

3.6.4.2. Estudios de acoplamiento.

Los estudios de acoplamiento se llevaron a cabo con el módulo FlexiDock del conjunto de programas SYBYL 6.9 (Tripos Inc., St. Louis. Mo Tripos, USA). Los estudios de acoplamiento revelaron que el peptidoglicano de neumococo puede estar perfectamente acomodado en la cavidad del extremo N-terminal de CbpF (Figura 3.6.22), reforzando la hipótesis de que CbpF quizá actúe bloqueando el acceso de LytC a su sustrato. Un fragmento de peptidoglicano ([GlcNAc-MurNAc-(L-Ala-γD-Glu-L-Lys-D-Ala-DAla)]₂), interacciona con los residuos aromáticos e hidrofóbicos de la hendidura (Tyr35, Tyr48, Trp75, Pro153 and Leu154) y con los residuos polares (Asp38, Glu50 y Asn152. Ver Tabla 3.6.6). Los ensayos de cristalización fallaron a la

hora de obtener información directa de la unión del peptidoglicano sobre el módulo N-terminal de CbpF. Los cristales nativos se resquebrajaban cuando se ponían en contacto con el fragmento sintético de pared celular y sólo en un caso mediante una inmersión rápida pudimos recoger un conjunto de datos (Tabla 3.6.5). A pesar de la ausencia de densidad electrónica para el peptidoglicano, la presencia del ligando causó una significativa alteración de la estructura de CbpF en el dominio CII y en el módulo N-terminal (Figura 3.6.21). Los bucles alargados del dominio CI de la estructura del complejo están altamente desordenados y no fue posible trazar ni siquiera la cadena principal entre los residuos 168 y 178. Particularmente interesante resultó el hecho de que el modulo N-terminal en presencia del análogo del peptidoglicano rotaba 10° alrededor de una bisagra molecular entre el modulo N-terminal y la región de conexión conduciendo a fuertes diferencias en las posiciones relativas entre módulos, apoyando una interacción directa entre el peptidoglicano y el módulo N-terminal.

Átomo del ligando			Átomo de la proteína				Distancia
N	Nombre	Clase	Residuo	Nombre	Clase		
8	O6	I	THR 101	OG1	I	5.1	
8	O6	I	ASN 152	OD1	II	5.8	
21	O5	II	SER 150	OG	I	4.4	
23	O6	I	SER 150	OG	I	4.5	
23	O6	I	ASP 149	O	II	5.1	
47	O7	II	ASN 152	ND2	III	5.5	
51	O	II	TYR 35	OH	I	5.7	
55	N2	I	GLU 73	OE1	II	3.9	
65	O2N	II	TYR 35	OH	I	5.6	
70	N	I	GLU 73	OE1	II	4.9	
70	N	I	LEU 71	O	II	5.0	
73	O1	I	GLU 50	OE1	II	2.7	
73	O1	I	ILE 69	O	II	3.1	
73	O1	I	LEU 71	N	III	3.7	
73	O1	I	LEU 71	O	II	4.2	
74	O2	II	ARG 57	NH2	III	3.1	
86	NZ	I	ASP 172	OD1	II	3.6	

Tabla 3.6.6: Lista de posibles puentes de hidrógeno entre el análogo de peptidoglicano (2S5P)₂ y CbpF.

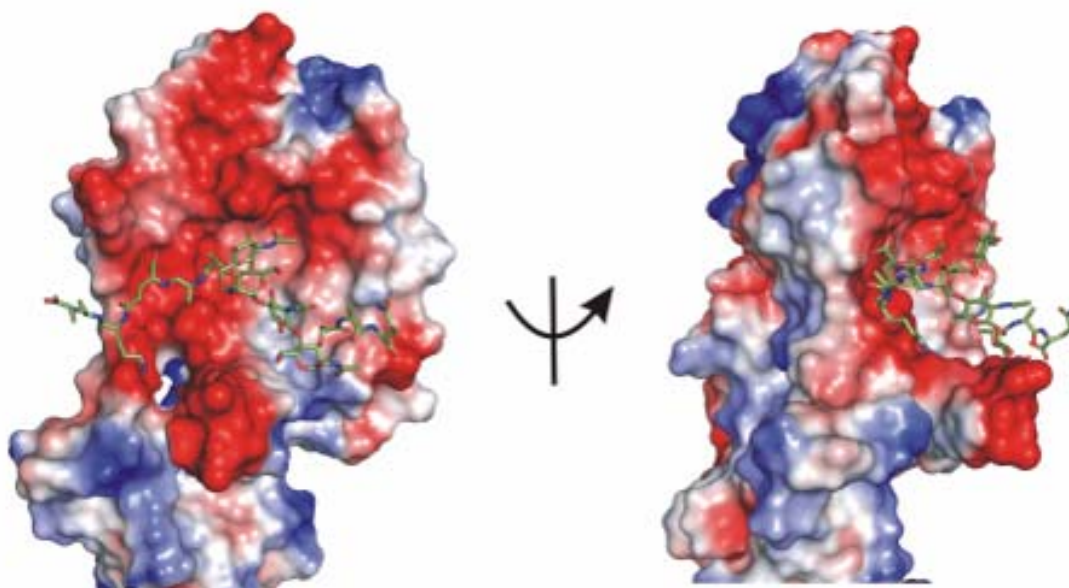


Figura 3.6.22: Modelo acoplado del peptidoglicano dentro del módulo N-terminal de CbpF. El potencial electrostático es representado sobre la superficie molecular de CbpF. La fracción del peptidoglicano sintético $(2S5P)_2$ que interacciona con el módulo N-terminal está representada como sticks con los átomos de carbono en verde, átomos de nitrógeno en azul y, átomos de oxígeno en rojo.

3.6.4.3. Ensayos *in vitro* e *in vivo* de actividad lítica de murein hidrolasas en presencia y ausencia de CbpF.

CbpF modula la actividad autolítica de LytC^{††}

Para detectar actividades líticas sobre paredes celulares de neumococo, se emplearon paredes celulares marcadas radiactivamente. Mediante este ensayo no se pudo detectar ninguna actividad lítica de la pared celular por parte de CbpF, por lo que es muy improbable que CbpF pudiera poseer tal actividad enzimática. También se analizó la influencia de CbpF sobre la actividad de murein hidrolasas conocidas (LytA, LytB, LytC y Pce) asumiendo que CbpF podría modular la actividad hidrolítica de la pared celular de estas enzimas por competición con los sitios de unión a colina o con su sustrato. Para probar esta hipótesis, se determinó *in vitro* la actividad lítica de la pared celular de diferentes murein hidrolasas en ausencia/presencia de

^{††} Estudios llevados a cabo por el grupo de Genética Bacteriana del departamento de Microbiología Molecular del Centro de Investigaciones Biológicas. Director: Pedro A. García López.

CbpF o de un módulo de unión a colina estándar (C-LytA) (éste último a modo de control). Los resultados revelaron que las actividades de LytA, LytB, y Pce no eran afectadas significativamente por la adición de CbpF, por el contrario, la actividad lítica de LytC estaba fuertemente reducida en presencia de 200 nM CbpF (Figura 3.6.23). Para comprobar la contribución a este efecto inhibitorio de los módulos aislados de CbpF se usaron proteínas truncadas, el módulo N-terminal N-CbpF (aminoácidos 1-192) o el dominio CII C-CbpF (residuos, 193-311), empleando el mismo método radiactivo *in vitro* utilizado anteriormente. Los resultados indican que C-CbpF no altera la actividad enzimática de LytC, sugiriendo que el dominio N-terminal de CbpF juega un papel crítico en esta inhibición específica. Sin embargo, la presencia del módulo de anclaje de CbpF es necesaria para realizar su función inhibitoria ya que el dominio N-terminal aislado de CbpF era incapaz de inhibir la actividad lítica de LytC (Figura 3.6.24A). Estas observaciones excluyen la posibilidad de que el efecto inhibitorio de CbpF sobre LytC pueda ser simplemente atribuido a una competición por la unión de residuos de colina sobre la superficie celular. Aunque el dominio N-terminal es capaz de unir la pared celular (véase más adelante), esta interacción aparece demasiado débil para ser capaz de prevenir el acceso de LytC al sustrato de modo que sería necesaria la presencia de la proteína modular completa para inhibir LytC. Este resultado concuerda con el comportamiento de otras CBPs donde la ausencia del dominio de unión a colina reduce su actividad funcional a niveles prácticamente indetectables (Sanz *et al.*, 1992).

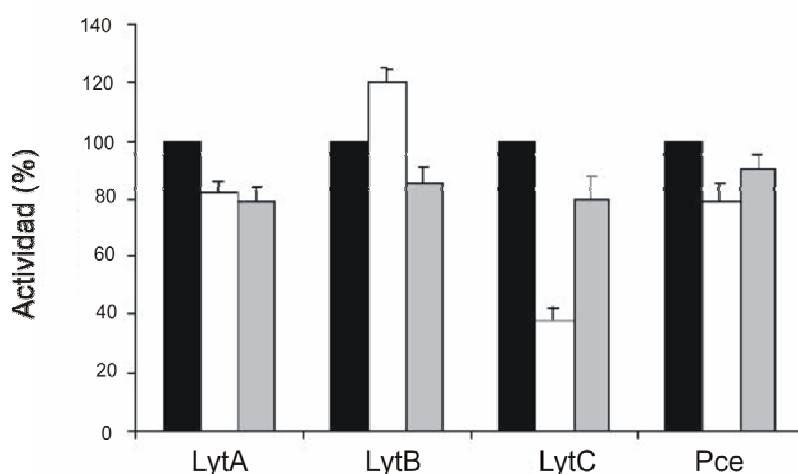


Figura 3.6.23: Efecto inhibitorio de CbpF sobre la actividad hidrolítica de murein hidrolasas de neumococo. Actividad lítica de la pared celular *in vitro* de las murein hidrolasas LytA, LytB, LytC, y Pce sin CbpF (barras negras), con 0.2 µM de CbpF (barras blancas), o 0.2 µM de C-LytA (barras grises). La barra de error se estima como el triple de la desviación estándar (n = 5 experimentos).

Para comprobar si el efecto inhibitorio de CbpF sobre LytC podría ocurrir también con otras lisozimas de neumococo, se repitieron los experimentos con las lisozimas fágicas Cpl-1 y Cpl-7 (todas ellas pertenecientes a la familia 25 de glicosilhidrolasas al igual que LytC). Cabe mencionar que los módulos catalíticos de Cpl-1 y Cpl-7 son virtualmente idénticos, pero mientras Cpl-1 presenta un CBM, Cpl-7 no (García *et al.*, 1990). Los resultados obtenidos indican (Figura 3.6.24A) que ambas enzimas son inhibidas por CbpF mostrando por una lado, que el efecto inhibitorio no está limitado a la lisozima LytC sino también a otras glicosilhidrolasas de la misma familia y por otro que el efecto no es debido a la competición por moléculas de colina (Cpl-7 no une colina). Las diferencias estructurales entre Cpl-1 y Cpl-7 quizá podrían explicar por qué la inhibición de Cpl-1 por CbpF fue mayor que en el caso de Cpl-7. Mientras que la actividad de Cpl-7 es independiente de colina, en Cpl-1 depende de la unión de colina. Este resultado sugiere que el CBM de CbpF refuerza el efecto inhibitorio del módulo N-terminal sobre aquellas enzimas donde el anclaje a colina es requerido para alcanzar la máxima actividad. Por otro lado, aunque no se puede descartar que el efecto inhibitorio sea debido a una interacción directa CbpF-LytC, el hecho de que CbpF inhiba diferentes lisozimas sugiere que la inhibición se debe a un bloqueo del sustrato común.

Como LytC está implicada en la autólisis de los cultivos de neumococo a 30°C (García *et al.*, 1999), empleamos esta propiedad para investigar el papel de CbpF *in vivo*. Por lo tanto, analizamos sobre R6lytAcbpF de *S. pneumoniae* (un mutante doble sin LytA y CbpF), el efecto sobre la autólisis en ausencia de CbpF o por adición externa de esta proteína. El mutante era deficiente de LytA para evitar el efecto de la autolisina mayoritaria que podría enmascarar el papel autolítico de LytC a 30°C. La lenta curva de autólisis que muestra R6lytA (López *et al.*, 1986), un mutante sencillo sin LytA, a 30°C se aceleró en el caso de la cepa R6lytAcbpF (doble mutante, Figura 3.6.24B), sugiriendo que la ausencia de CbpF facilita el proceso autolítico de LytC. Este resultado concuerda con el efecto de represión de LytC por CbpF observado *in vitro*, indicando que CbpF juega el mismo papel regulador *in vivo*. Además, cuando analizamos el crecimiento de la cepa R6lytA en presencia de diferentes cantidades de CbpF añadidas durante el crecimiento bacteriano, observamos que la adición de 1.2 µM de CbpF eliminaba la autólisis inducida por LytC a 30°C, mientras la adición de proteínas homólogas C-LytA o C-CbpF, usados como controles, no alteraban la curva autolítica (Figura 3.6.24C). El cultivo que contiene CbpF externo mostró la presencia de células R6lytA sin lisar con una morfología celular formada por

cadena cortas bajo microscopía de contraste de fase, mientras muchas células lisadas se observaban en el cultivo sin CbpF, indicando que CbpF protege la pared celular del daño lítico.

Por otro lado, comprobamos el comportamiento de *R6lytA-cbpF* en presencia de LytC o de LytC y CbpF añadida al medio (Figura 3.6.24D). Como se esperaba, las células mutantes se autolisaban cuando recibían una cantidad determinada de LytC (García *et al.*, 1999), mientras la autólisis, se inhibía cuando la misma cantidad de LytC era añadida en presencia de 1.2 μ M de CbpF (Fig 3.6.24D). Cabe resaltar que la inhibición de este proceso autolítico no tuvo lugar en presencia de cantidades similares de C-CbpF o de otras proteínas homólogas como LytB.

El balance de todos estos resultados, apunta a que CbpF modula selectivamente la función autolítica de la lisozima LytC y que la divergencia de las repeticiones de unión a colina que constituye el dominio N-terminal juega un papel crítico en este efecto específico. Está bien establecido que las autolisinas juegan papeles diferentes y fundamentales en la fisiología y patología de la bacteria (López *et al.*, 2004; Gosink *et al.*, 2000). Estas enzimas son capaces de cambiar hidrolíticamente la integridad de la pared celular de la bacteria. El efecto de reacciones incontroladas catalizadas por estas enzimas podría de hecho ser bastante perjudicial y como tal actividad enzimática debe estar estrictamente regulada. La naturaleza de este proceso regulatorio es desconocido hasta la fecha, aunque, en este sentido, se ha atribuido al contenido de ácido lipoteicoico en la membrana celular un papel regulador (Höltje & Tomasz, 1975). Recientemente, se ha sugerido la implicación de pequeños RNA no codificados en el control de la autólisis en la fase estacionaria (Halfmann *et al.*, 2007). Nuestros resultados demuestran que CbpF puede inhibir la actividad de LytC tanto *in vitro* como *in vivo* y por lo tanto, concluimos que CbpF juega un papel clave en la regulación de la actividad de LytC.

Interacción entre el módulo N-terminal y la pared celular^{††}

Por otro lado, los experimentos de resonancia de Plasmón Superficie (SPR) y de entrecruzamiento llevados a cabo entre CbpF y LytC han descartado la posibilidad de una interacción directa entre ambas proteínas como la causa de la inhibición. Por lo tanto, el mecanismo inhibitorio podría deberse a interacciones específicas de CbpF y el peptidoglicano que impedirían el acceso de la lisozima a su sustrato. Para estudiar esta hipótesis se analizaron estas interacciones mediante SPR empleando diferentes paredes celulares (paredes celulares

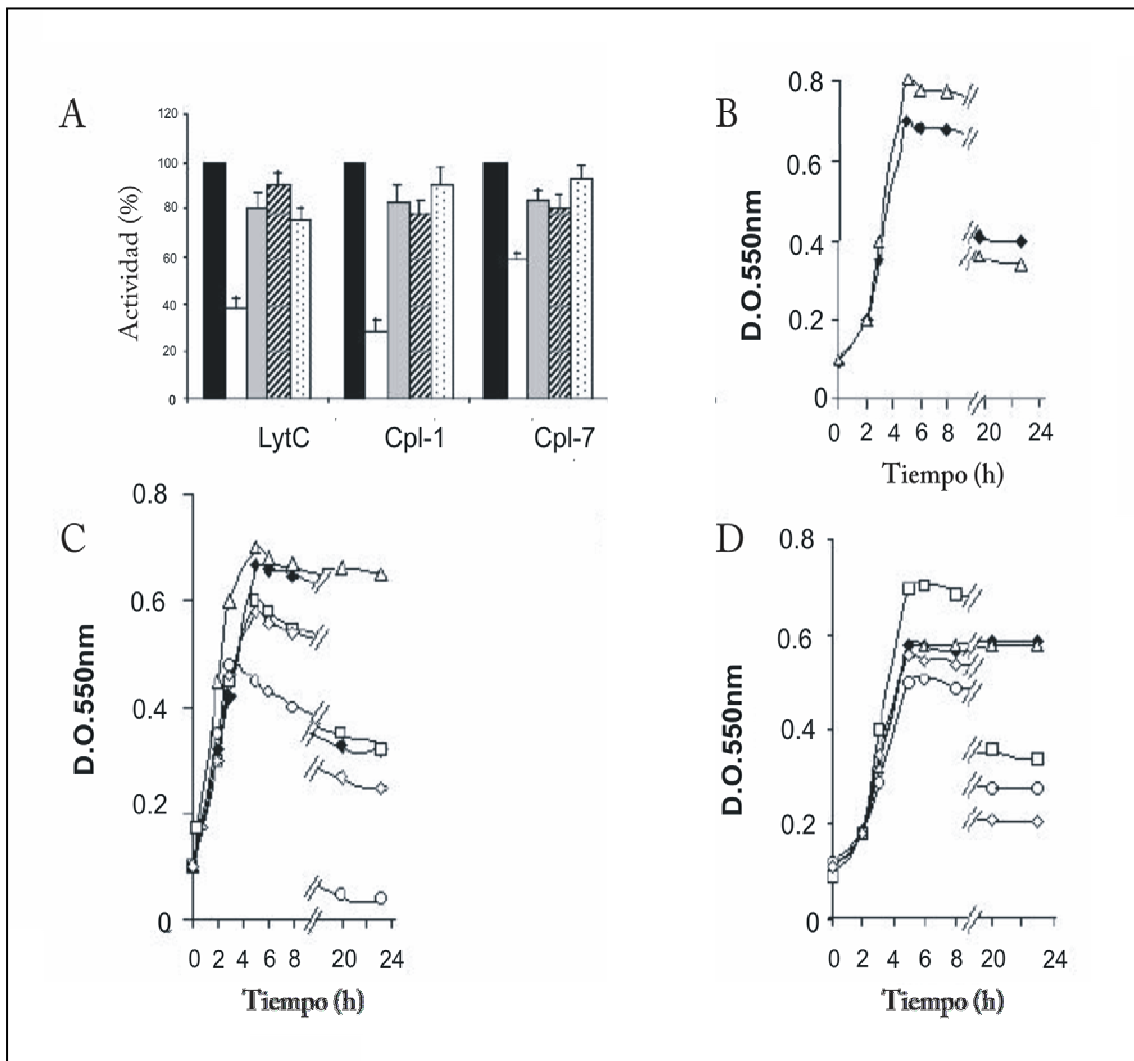


Figura 3.6.24: Efecto inhibitorio de CbpF sobre la actividad hidrolítica de LytC. (A) Actividad lítica de la pared celular *in vitro* de las enzimas LytC, Cpl-1 y Cpl-7 sin CbpF (barras negras), con 0.2 μ M de CbpF (barras blancas), 0.2 μ M de C-LytA (barras grises), 0.2 μ M de C-CbpF (barras ralladas), o 0.2 μ M de N-CbpF (barras punteadas). La barra de error se estima como el triple de la desviación estándar ($n=5$ experimentos) (B) Curvas de crecimiento de las cepas R6lytA (diamantes negros) y R6lytAcbpF (triángulos blancos) a 30 °C. (C) Efecto de la adición externa de proteínas purificadas añadidas, a una concentración final de 1.2 μ M sobre cultivos crecidos exponencialmente de la cepa R6lytA, a 30 °C: con CbpF (triángulos blancos), con C-CbpF (círculos blancos), con C-LytA (cuadrados blancos), con LytB (rombos blancos). También se muestra la curva de crecimiento de un cultivo sin tratar de R6lytA (rombos negros). (D) Efecto de diferentes proteínas sobre la actividad de LytC ensayada en un cultivo de la cepa R6lytAlytC mediante proteínas purificadas añadidas externamente, a una concentración final de 1.2 μ M: con LytC (rombos blancos), con LytC + CbpF (triángulos blancos), con LytC + C-CbpF (círculos blancos), con LytC + LytB (cuadrados blancos). También se muestra la curva de crecimiento del cultivo sin tratar R6lytAlytC (rombos negros).

de *Micrococcus lysodeikticus* y de *S. pneumoniae* con y sin colina). Como era de esperar, CbpF fue capaz de unirse fuertemente a las paredes celulares de neumococo que contenían colina. Particularmente interesante fue el hecho de que CbpF también se uniera intensamente a las

paredes celulares de *S. pneumoniae* sin colina e incluso, aunque en menor extensión, a paredes celulares de *Micrococcus lysodeikticus* (con un peptidoglicano bacteriano diferente del de neumococo y que no contiene colina) (Figura 3.6.25). Los experimentos con los módulos de anclaje, C-CbpF y C-LytA fueron sólo capaces de unir paredes celulares de neumococo que contenían colina, sugiriendo que su capacidad de unión estaba restringida a su interacción específica con este aminoalcohol. Sin embargo, el módulo N-terminal de CbpF era capaz de unir ambos tipos de paredes celulares de neumococo (con y sin colina) sugiriendo que el módulo N-terminal es capaz de reconocer específicamente las paredes celulares de neumococo y que los módulos N- y C-terminal de CbpF se unen a diferentes componentes del peptidoglicano.

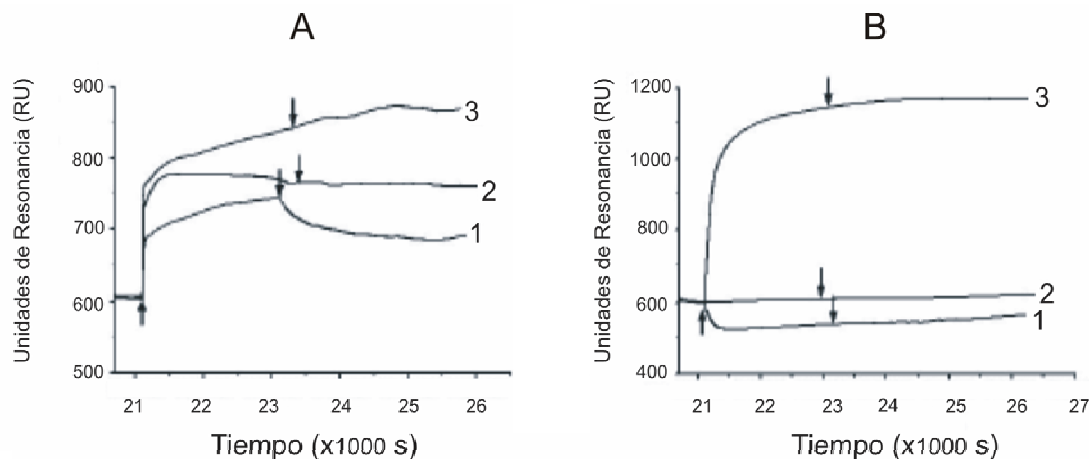


Figura 3.6.25: Análisis de unión mediante Resonancia de Plasmón Superficie (SPR). (A) Curvas de unión de CbpF a diferentes muestras purificadas inmovilizadas de paredes celulares: *M. lysodeikticus* (1), *S. pneumoniae* sin colina, (2) y *S. pneumoniae* con colina (3). (B) Curvas de unión de C-LytA (1), C-CbpF (2), y N-CbpF (3) a células inmovilizadas sin colina de *S. pneumoniae*.

CbpF encabeza una familia de proteínas del neumococo con función desconocida

Un estudio acerca de los ortólogos de CbpF en otras cepas neumocócicas ha revelado la existencia de las proteínas CbpC y CbpJ en la cepa TIGR4 (Figuras 3.6.26 y 3.6.27). El análisis comparativo de estos genes demuestra que *CbpF* quizá haya sido el resultado de la recombinación de *CbpC* y *CbpJ*. La coexistencia de dos genes altamente similares en puntos de TIGR4 que se corresponde con las proteínas similares a CbpF codificadas juegan un papel fisiológico crítico y sugiere que esta duplicación pueda favorecer la adquisición de propiedades reguladoras más selectivas.

Posteriores análisis sobre genomas neumocócicos conocidos revelan la existencia de varias proteínas que parecen mostrar la misma arquitectura que CbpF. Estas proteínas quizá constituyan una nueva subfamilia del tipo CbpF dentro de la extensa familia de las CBPs que posean un módulo CBM típico y una región N-terminal formada por repeticiones no canónicas. A pesar de las pequeñas diferencias en la composición de aminoácidos a lo largo de todos los miembros de la subfamilia CbpF, se mantienen la mayoría de los residuos implicados tanto en el

armazón estructural de la región N-terminal como en las interacciones intermodulares, sugiriendo que estas proteínas podrían exhibir funciones reguladoras similares a CbpF.

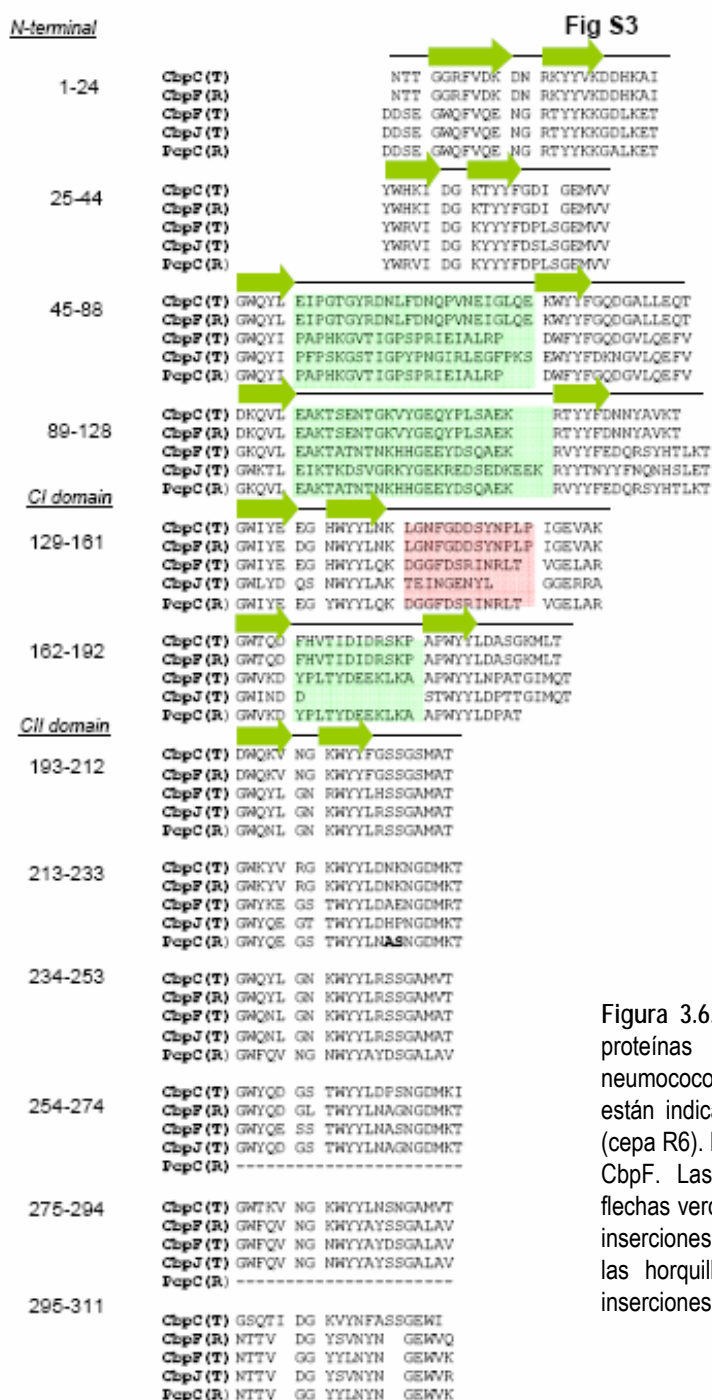


Figura 3.6.26: Alineamiento de secuencia de proteínas similares a CbpF en cepas de neumococo. Los orígenes de las proteínas están indicadas por (T) (cepa TIGR) y por R (cepa R6). Los números indican los residuos en CbpF. Las horquillas β se muestran como flechas verdes. Los marcos verdes representan inserciones de aminoácidos en las regiones de las horquillas y los marcos rojos muestran inserciones tras el giro β .

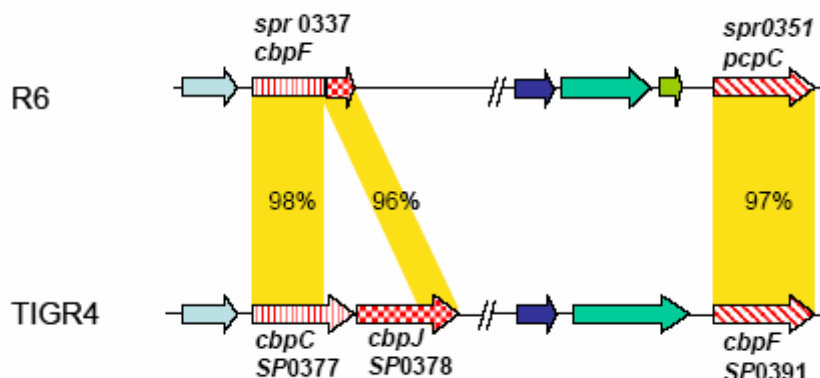


Figura 3.6.27: Representación esquemática de la localización cromosómica de los homólogos de CbpF en las cepas neumocócicas de R6 y TIGR. Colores similares indican genes homólogos. Los genes que codifican para CbpF están coloreados en rojo con diferentes marcos. Las líneas truncadas indican que las regiones cromosómicas no están continuas. El cromosoma de la cepa R6 contiene 2 marcos de lectura abiertos que codifican para proteínas similares a CbpF referenciadas como *spr0337* (CbpF) y *spr0351* (pcpC). El cromosoma de la cepa TIGR contiene 3 marcos de lectura abiertos que codifican para proteínas similares a CbpF referenciadas como SP0377 (CbpC), SP0378 (CbpJ) y SP0391 (CbpF). Las similitudes de la secuencia se indican en porcentajes. Una comparación detallada del gen CbpF de la cepa R6 y los genes CbpC y CbpJ de TIGR apunta claramente a que CbpF resultó de la recombinación de CbpC y CbpJ.

3.6.5. ANÁLISIS DE LA OBTENCIÓN DE FASES MEDIANTE COMPLEJOS LANTÁNIDOS DE GD EN CRISTALOGRAFÍA MACROMOLECULAR.

Se emplearon siete complejos de Gd (Figura 1.2.9) en la preparación de los derivados de átomo pesado para resolver mediante la técnica SAD la estructura de la proteína de unión a colina F (CbpF), una proteína de superficie de 36 KDa de *Streptococcus pneumoniae*. Además, se utilizó la proteína CbpF como sistema modelo para analizar la capacidad de obtención de fases de cada uno de estos derivados. Se analizaron sistemáticamente tres aspectos diferentes: la eficacia de la cocrystalización frente a la inmersión en la unión de los diferentes complejos de Gd, su modo de interacción y el estudio comparativo de la obtención de fases por SAD empleando radiación sincrotrón frente a los rayos X obtenidos a partir de un generador ánodo rotatorio. Este estudio revela el elevado potencial de estos complejos para obtener fases mediante SAD empleando ánodos rotatorios convencionales y refuerza su relevancia para la cristalografía macromolecular extensiva.

3.6.5.1. Consideraciones generales.

El método de difracción anómala a una única longitud de onda (SAD) es actualmente una aproximación bien establecida y altamente efectiva en biología estructural. Avances recientes en instrumentación y programas de análisis de datos (especialmente en programas de obtención de fases y de modificación de la densidad) combinados con una creciente conciencia del daño por radiación de las fuentes de tercera generación han conducido a un marcado incremento en la popularidad y eficiencia del método SAD. Los experimentos SAD generalmente implican un único conjunto de datos recogido en la línea de absorción o por encima de la energía de la línea de absorción de un dispersor anómalo incorporado en la muestra. Los experimentos SAD generalmente requieren menos datos que los experimentos MAD, dependiendo del tipo de átomo pesado, su unión a la proteína y la intensidad de su señal anómala, proporcionando de este modo un método más rápido de obtención de fases y un potencial de daño por radiación limitado. Además, la obtención de fases por SAD ha sido documentado a longitudes de onda remotas por encima de la energía de la línea de absorción (Leonard *et al.*, 2005) e incluso con las fuentes de rayos X convencionales de los laboratorios (Dauter *et al.*, 2002).

Un obstáculo asociado con la obtención de fases basado en átomos pesados es el proceso de obtener un derivado de átomo pesado útil. Hoy en día, la mayoría de los experimentos de difracción anómala son llevados a cabo empleando proteínas en las cuales los residuos de metionina son sustituidos por selenometionina, un método aplicable a proteínas recombinantes expresadas en sistemas bacterianos o células de insectos (Doublé, 1997; Hendrickson *et al.*, 1990).

Se han propuesto métodos alternativos para obtener derivados para proteínas que no pueden ser producidas de forma recombinante, o donde el material sustituido por la selenometionina es o bien insoluble o bien impide la cristalización, incluyendo el uso de haluros (Dauter & Dauter, 1999; Dauter *et al.*, 2000, 2001; Evans & Brice, 2002) y de gases nobles introducidos bien en la proteína nativa (Vitali *et al.*, 1991), o en los sitios de unión creados por mutagénesis dirigida (Quillin & Matthews, 2002). La señal anómala de los átomos de S intrínsecos se ha ido empleando cada vez con mayor frecuencia para resolver estructuras mediante el método SAD (Hendrickson & Teeter, 1981; Micossi *et al.*, 2002). Sin embargo, debido a la débil señal anómala del azufre a longitudes de onda accesibles a la mayoría de las líneas de haces

empleadas en cristalografía de proteínas, el método es realmente adecuado únicamente para cristales que difracten muy bien. Los tradicionales protocolos de inmersión y cocrystalización de átomos pesados han sido revisados recientemente, siguiendo la actual tendencia de aplicar tiempos de inmersión menores (Sun & Radaev, 2002; Sun *et al.*, 2002). A pesar de los intentos para desarrollar aproximaciones racionales para la derivatización mediante átomos pesados empleando técnicas como la electroforesis en geles nativos (Boggon & Shapiro, 2000) o espectrometría de masas (Cohen *et al.*, 2000; Sun & Hammer, 2000), el procedimiento sigue siendo esencialmente experimental.

Una nueva clase de complejos de Gd han demostrado ser una herramienta útil para obtener cristales derivados de átomo pesado adecuados para obtener fases experimentalmente, explotando la intensa señal anómala en el pico de absorción L_{III} del Gd ($f'' = 28 e^-$, $\lambda = 1.711 \text{ \AA}$) y para la radiación Cu K α ($f'' = 12 e^-$) (Girard *et al.*, 2003a; Girard *et al.*, 2004). Cabe destacar que el último valor es mayor que la señal anómala obtenida a partir del pico de absorción K del Se ($f'' = 6 - 10 e^-$).

Obtención de los complejos de gadolinio empleados.

Los complejos de gadolinio de los ligandos HPDO3A, DOTMA, DO3A y DOTA-BOM (estos dos últimos no son comerciales) nos fueron proporcionados en forma de polvo por Bracco Imaging spa, Milán, Italia. Los complejos Gd-DTPA, Gd-DTPA-BMA y Gd-DOTA fueron obtenidos en forma de soluciones acuosas de los complejos comercialmente disponibles para soluciones médicas (MAGNEVISTTM, Schering, Alemania, OMNISCANTM, Nycomed Inc., Noruega y DOTAREMTM, Guerbet, Francia, respectivamente) que son proporcionados a una concentración de 0.5 M.

3.6.5.2. Recogida de datos y procesamiento.

Todos los cristales fueron sumergidos durante unos pocos segundos en las soluciones de cristalización nativas complementadas con un 17% de glicerol como crioprotector para eliminar cualquier resto de complejo de Gd no unido a los cristales. Todos los cristales fueron congelados mediante corriente de nitrógeno líquido a 100K (Oxford Cryosystems).

Los datos de difracción fueron recogidos en el ESRF: en la línea ID29 del ESRF para todos los cristales derivados obtenidos por cocrystalización y en la línea BM30 para los cristales derivados obtenidos por inmersión (derivados de Gd-HPDO3A, Gd-DOTMA y Gd-DTPA). La recogida de los datos fue llevada a cabo a la longitud de onda correspondiente a la "línea blanca" ($\lambda=1.711$ Å) del pico de absorción L_{III} del Gd como se determinó a partir de los espectros de fluorescencia obtenidos (Figura 1.2.6). La distancia cristal-detector fue fijada a 110 mm en ambas líneas de medida. Los cristales difractaban habitualmente hasta 1.9 Å de resolución, pero la máxima resolución estaba limitada debido al tamaño del detector a 2.23 Å en ID29, y a 2.68 Å en BM30A. Los datos fueron recogidos barriendo un rango de 180° con el fin de incrementar la multiplicidad y de este modo mejorar la precisión de la medida de los datos.

Para evaluar el potencial de obtención de fases mediante dispersión anómala de estos compuestos cuando empleamos las fuentes de rayos X convencionales usadas en los laboratorios, se llevó a cabo una recogida de datos empleando dicha radiación convencional sobre cristales derivados CbpF: Gd-HPDO3A. Los conjuntos de datos obtenidos a partir de una fuente de radiación convencional fueron recogidos sobre un detector MAR345 tipo Imaging Plate empleando la radiación Cu K α proporcionada por un generador de ánodo rotatorio Enraf-Nonius FR571.

Todos los conjuntos de datos recogidos en el ESRF fueron procesados usando el programa XDS (Kabsch, 1988) y escalados con el programa SCALA (Evans, 1993) del paquete CCP4 (Collaborative Computational Project, Number 4, 1994). Los conjuntos de datos recogidos con la radiación Cu K α sobre una fuente convencional de rayos X fueron integrados usando MOSFLM (Leslie, 1987) y escalados con SCALA. Un resumen de los parámetros de la recogida de datos y estadísticas de procesado se muestra en las tablas 3.6.7, 3.6.8 y 3.6.9.

3.6.5.3. Obtención de fases por SAD.

Mediante cocrystalización se produjeron siete diferentes derivados isomorfos de Gd empleando las condiciones de cristalización nativas y añadiendo el correspondiente complejo a una concentración de 50 mM (se probaron concentraciones diferentes de 20, 50 y 100 mM siendo la concentración de 50 mM la más exitosa). Todos los conjuntos de datos se recogieron en la misma línea de medida (ID29, ESRF). Las posiciones de los átomos pesados fueron

determinadas empleando el programa *SHELXD* (Schneider & Sheldrick, 2002) para cada derivado y sus coordenadas se introdujeron directamente en el protocolo de obtención de fases implementado en *SHARP* (De la Fortelle and Bricogne, 1997). A continuación, se mejoraron las fases mediante aplanado de solvente y ajuste a histograma empleando el programa *DM* (Cowtan & Main, 1996), asumiendo un 58 % de contenido en solvente como estimó el programa *TRUNCATE* del paquete *CCP4*. El número de sitios de Gd, así como las estadísticas de obtención de fases para los cristales derivados obtenidos con los siete complejos de Gd diferentes, se muestran en la tabla 3.6.7. Para estudiar las interacciones proteína-ligando, los sitios de unión se visualizaron con el programa gráfico *O* (Jones *et al.*, 1991). El refinamiento de las posiciones del ligando se llevaron a cabo con *CNS* (Brünger *et al.*, 1998).

Para tres de los siete complejos (Gd-HPDO3A, Gd-DOTMA y Gd-DTPA) se obtuvieron cristales derivados tanto por cocrystalización como mediante inmersión empleando las mismas condiciones de cristalización y la misma concentración de complejo (50 mM). El procedimiento para identificar posiciones de átomo pesado y para el cálculo y modificación de fase fue el mismo que el descrito anteriormente. El número final de sitios de Gd y las estadísticas de obtención de fases para derivados de CbpF obtenidos por cocrystalización y por inmersión se muestran en la tabla 3.6.8.

Las estadísticas de obtención de fases para los datos de difracción del cristal derivado de Gd-HPDO3A recogido con radiación sincrotrón y con radiación CuK α se muestran en la tabla 3.6.9.

La estructura de CbpF fue modelada manualmente con el programa *O* (Jones *et al.*, 1991) y refinada con *CNS* (Brünger *et al.*, 1998). Los coeficientes de correlación entre los mapas de densidad electrónica experimentales y los calculados a partir de los modelos finales refinados se determinaron con el programa *OVERLAPMAP*. Las estadísticas de refinamiento para todos los derivados de Gd obtenidos se muestran en las tablas 3.6.10, 3.6.11 y 3.6.12.

3.6.5.4. Estudio comparativo del poder de obtención de fases de los distintos complejos de Gd.

La derivatización de CbpF mediante protocolos de cocrystalización o inmersión fue llevada a cabo de forma extensiva y no requirió significativos cambios respecto a las condiciones de cristalización originales. La elevada solubilidad de los complejos permite concentraciones

considerablemente superiores a las que generalmente se emplean para preparar derivados de átomo pesado. A pesar de esta alta concentración, se obtuvieron cristales derivados con buena calidad de difracción y, en algunos casos, difrataron incluso a una resolución superior a la de los cristales nativos. De hecho, mientras el cristal nativo de CbpF de mayor calidad difractó hasta una resolución máxima de 1.67 Å (Molina *et al.*, 2007), un cristal derivado proporcionó datos de buena calidad hasta una resolución de 1.4 Å. En estos casos se debían a la aparición de nuevas interacciones, mediadas por los complejos de Gd, estabilizando el empaquetado cristalino.

El éxito de cualquier obtención de fases con Gd depende del número de sitios de unión y de su ocupación. La elevada dispersión anómala de los átomos de Gd, combinado con la alta ocupación de los sitios de unión, conduce a intensos efectos anómalos que pueden ser directamente atestiguados por los altos valores de $\Delta F/\sigma$ (XDS) y R_{ano} (SCALA), por la altura de los picos en los mapas de Patterson anómalos, así como por los parámetros de obtención de fases (R_{cullis} , poder de obtención de fases y figura de mérito). Para todos los cristales derivados, los valores de $\Delta F/\sigma$ y R_{ano} se encontraban en rangos de 1.30-4.25 y 4.6-12.1%, respectivamente (Tabla 3.6.7). Los valores más bajos corresponden al derivado cocrystalizado Gd-DTPA-BMA. En todos los casos la obtención de fases fue exitosa, demostrando que todos los complejos de Gd se unían a CbpF, como confirmaban los mapas de Patterson anómalos (figura 3.6.28).

Todos los derivados proporcionaron fases de suficiente buena calidad para permitir la construcción del modelo (Tabla 3.6.7). El mejor derivado, Gd-HPDO3A, conduce a la figura de mérito más alta (0.588) y el factor R-cullis más bajo (0.298) después de la obtención de fases con *SHARP*. Tras la modificación de densidad, la figura de mérito para este derivado fue de 0.860. La calidad de las fases obtenidas con este cristal derivado permitió trazar el modelo usando el mapa experimental a 2.23 Å de resolución. Los derivados Gd-DO3A y Gd-DTPA también proporcionaron excelentes estadísticas de obtención de fases con figuras de mérito superiores a 0.45 antes de la modificación de la densidad. Incluso para los cristales derivados Gd-DTPA-BMA y Gd-DOTMA, los valores de la figura de mérito antes de la modificación de la densidad fueron de 0.37 y 0.39, respectivamente, y condujeron a valores superiores a 0.75 tras la modificación de densidad permitiendo la fácil construcción del modelo.

Recogida de datos	Gd-HPDO3A	Gd-DOTMA	Gd-DTPA	Gd-DO3A	Gd-DTPA-BMA	Gd-DOTA-BOM	Gd-DOTA
Parámetros de celdilla							
<i>a</i> (Å)	49.35	49.18	48.42	48.98	48.71	49.03	49.19
<i>b</i> (Å)	114.52	115.00	113.49	114.92	114.00	114.37	114.24
<i>c</i> (Å)	76.23	75.80	75.38	76.25	76.29	76.07	75.66
Rango de resolución (Å)	100-2.23 (2.36-2.23)	100-2.23 (2.36-2.23)	100-2.23 (2.35-2.23)	100-2.29 (2.42-2.29)	100-2.98 (3.14-2.98)	100-2.23 (2.35-2.23)	100-2.30 (2.42-2.30)
No. de reflexiones observadas	143,495	144,217	144,177	77081	60,403	118,066	134,747
No. de reflexiones únicas	21,579	21,372	20,726	19934	9,076	20,962	19,830
Acéntricas	18,835	18,571	18,108	17,349	7,528	18,338	17,202
Céntricas	2,744	2,801	2,618	2585	1,548	2,624	2,628
Relación señal-ruido $I/\sigma(I)$	10.4(7.7)	10.5(4.6)	3.7(3.6)	11.3(5.4)	8.3(3.1)	7.5(1.3)	9.2(7.4)
Redundancia	6.6(5.5)	6.7(5.8)	7.0(7.0)	3.9(3.7)	6.7(5.1)	5.6(3.6)	6.8(6.7)
Complejidad (%)	98.9(95.0)	99.7(98.4)	98.7(95.5)	98.6(93.9)	98.9(92.8)	97.2(96.1)	99.4(97.7)
R_{sym} (%)	4.7(8.9)	5.4(15.0)	10.9(17.1)	4.8(12.4)	7.1(19.7)	4.9(13.7)	5.4(10.0)
R_{ano} (%)	10.6(5.9)	4.5(5.4)	9.2(11.7)	10.1(12.5)	4.3(7.3)	5.6(9.0)	5.4(6.6)
$\Delta F/\sigma^*$	4.249	1.541	1.760	2.788	1.307	2.164	1.852
Estadísticas de obtención de fases							
Número de sitios encontrados por SHELXD	5	1	2	5	4	2	2
Número de sitios finalmente modelados***	4	1	1	3	2	1	1
Ocupación de los sitios **	0.81, 0.68, 0.47, 0.38.	0.41	0.84	0.74, 0.64, 0.61	0.34, 0.18	0.82	0.68
$\langle \Delta F \rangle / F$	0.100	0.037	0.069	0.095	0.032	0.067	0.056
$\Delta F/\sigma^*$	5.496	1.997	1.865	3.629	1.307	2.711	2.246
FOM antes de DM**	0.588	0.371	0.456	0.537	0.392	0.416	0.391
R_{cullis} para reflexiones acéntricas**	0.298	0.752	0.549	0.436	0.632	0.695	0.737
Poder de obtención de fases para reflexiones acéntricas**	4.905	1.371	2.340	3.231	1.949	1.854	1.441
FOM tras DM**	0.860	0.765	0.821	0.844	0.802	0.812	0.778
Factor R_{free} en el espacio real space**	0.093	0.104	0.088	0.100	0.101	0.104	0.107
Correlación de mapa con OVERLAPMAP	0.724	0.584	0.672	0.711	0.626	0.654	0.605

Tabla 3.6.7: Recogida de datos, estadísticas de refinamiento y obtención de fases para los siete derivados de Gd obtenidos por co cristalización. Los datos de difracción fueron recogidos empleando radiación sincrotron a la longitud de onda de absorción de la línea LIII del Gd a partir de los cristales derivados de CbpF obtenidos mediante co cristalización con los siete complejos de Gd. Los valores entre paréntesis se refieren a la concha de mayor resolución. *Datos proporcionados por XDS. ** Datos proporcionado por SHARP. *** Datos proporcionados por O. La resolución de las estadísticas de faseado se limitó a 2.98Å en todos los casos

Capacidad de obtención de fases de los derivados de Gd

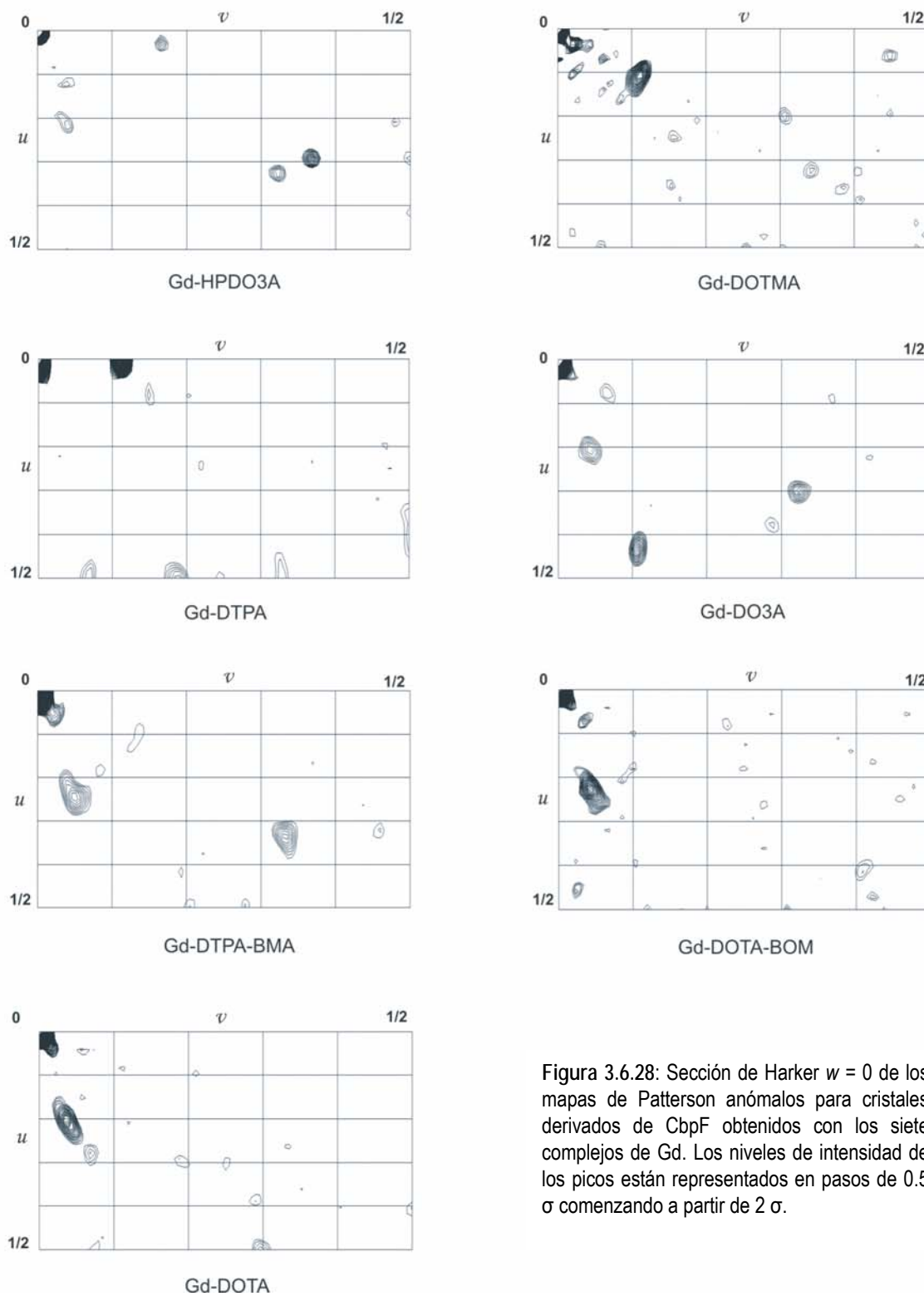


Figura 3.6.28: Sección de Harker $w = 0$ de los mapas de Patterson anómalos para cristales derivados de CbpF obtenidos con los siete complejos de Gd. Los niveles de intensidad de los picos están representados en pasos de 0.5σ comenzando a partir de 2σ .

3.6.5.5. Estudio comparativo de la cocrystalización frente a la inmersión como método de obtención de derivados.

Se usaron tres de los siete complejos de Gd para preparar cristales derivados de CbpF por inmersión además de los obtenidos por cocrystalización. En todos los casos XDS proporcionó elevados valores de $\Delta F/\sigma$ (superiores a 1.5) indicando una señal anómala significativa en cada caso. Las estadísticas de obtención de fases correspondientes a los cristales sumergidos (Tabla 3.6.8) confirmaron la unión efectiva para los tres complejos. Para tres de los derivados analizados, las figura de mérito antes del proceso de modificación de densidad fueron superiores a 0.49 y los factores R_{free} en el espacio real menores a 0.12, reflejando la buena calidad de las fases iniciales. Además, se obtuvieron elevados valores del poder de obtención de fases (superiores a 2.4), bajos valores de R_{cullis} (menores a 0.55) y figuras de mérito tras el proceso de modificación de densidad, superiores a 0.77. Por lo tanto, los protocolos tanto de cocrystalización y de inmersión facilitaron la derivatización de los cristales y condujeron a la exitosa obtención de fases de la estructura.

Los mejores cristales derivados, obtenidos con Gd-HPDO3A, se caracterizaron por la presencia de varios sitios de unión con alta ocupación (Tabla 3.6.8) empleando cualquier protocolo. Cuatro moléculas de Gd-HPDO3A se unieron a la proteína, con mayor ocupación para los derivados cocrystalizados (0.81, 0.68, 0.47 y 0.38) que para los obtenidos por inmersión (0.41, 0.24, 0.22 y 0.21), explicándose de este modo los mejores resultados de obtención de fases obtenidos con los cristales derivados cocrystalizados. Por el contrario, para los otros dos complejos de Gd, Gd-DOTMA y Gd-DTPA, las estadísticas de obtención de fases para los derivados obtenidos por inmersión fueron mejores que las obtenidas para las muestras cocrystalizadas. Para el derivado cocrystalizado Gd-DOTMA, sólo una molécula se unió a la proteína con una ocupación del sitio de unión de 0.41, mientras que en el derivado obtenido por inmersión se unieron dos moléculas a la proteína con ocupaciones de 0.51 y 0.41, respectivamente. Del mismo modo, para los derivados Gd-DTPA, los cristales derivados obtenidos por cocrystalización presentaron un único sitio de unión altamente ocupado (0.74) mientras el derivado obtenido por inmersión reveló dos sitios de unión con bajas ocupaciones.

Con cada uno de los complejos analizados, el número de sitios de unión fue igual o superior para los derivados obtenidos por inmersión que para los obtenidos por cocrystalización.

Método de derivatización	Inmersión			Cocrystalización		
	Gd-HPDO3A	Gd-DOTMA	Gd-DTPA	Gd-HPDO3A	Gd-DOTMA	Gd-DTPA
Recogida de datos						
Parámetros de celdilla						
a (Å)	48.97	48.89	48.87	49.35	49.18	48.42
b (Å)	114.93	114.86	114.06	114.52	115.00	113.49
c (Å)	76.17	76.16	75.76	76.23	75.80	75.38
Rango de resolución (Å)	100-2.70 (2.85-2.70)	100-2.70 (2.85-2.70)	100-2.70 (2.85-2.70)	100-2.23 (2.36-2.23)	100-2.23 (2.36-2.23)	100-2.23 (2.35-2.23)
No. de reflexiones observadas	78,896	76,577	78,780	143,495	144,217	144,177
No. de reflexiones únicas	11,791	11,598	11,879	21,579	21,372	20,726
Acéntricas	9,999	9,761	9,995	18,835	18,571	18,108
Céntricas	1,792	1,837	1,884	2,744	2,801	2,618
Relación señal-ruido $I/\sigma(I)$	18.6(3.0)	10.3(1.5)	8.6(1.6)	10.4(7.7)	10.5(4.6)	3.7(3.6)
Redundancia	6.7(4.8)	6.6(4.5)	6.6(4.7)	6.6(5.5)	6.7(5.8)	7.0(7.0)
Complejidad (%)	99.3(98.4)	96.3(96.1)	100(99.9)	98.9(95.0)	99.7(98.4)	98.7(95.5)
R_{sym} (%)	3.0(14.0)	4.5(21.9)	3.8(21.0)	4.7(8.9)	5.4(15.0)	10.9(17.1)
R_{ano} (%)	5.3(10.3)	6.9(18.7)	4.3(15.8)	10.6(5.9)	4.5(5.4)	9.2(11.7)
$\Delta F/\sigma^*$	3.614	2.951	2.370	4.249	1.541	1.760
Estadísticas de obtención de fases						
Número de sitios encontrados por SHELXD	4	4	4	5	1	2
Número de sitios finalmente modelados***	4	2	3	4	1	1
Ocupación de los sitios **	0.41, 0.24, 0.22, 0.21	0.51, 0.41	0.42, 0.19, 0.12	0.55, 0.53, 0.40, 0.31	0.39	0.74
$\langle \Delta F \rangle / F$	0.046	0.054	0.038	0.075	0.032	0.061
$\Delta F/\sigma^*$	4.081	3.331	2.539	5.136	1.855	1.833
FOM antes de DM**	0.518	0.509	0.491	0.567	0.346	0.441
R_{cullis} para reflexiones acéntricas**	0.424	0.455	0.552	0.313	0.773	0.586
Poder de obtención de fases para reflexiones acéntricas**	3.256	2.990	2.405	4.745	1.257	2.194
FOM tras DM**	0.861	0.837	0.802	0.845	0.741	0.801
Factor R_{free} en el espacio real space**	0.113	0.104	0.088	0.101	0.107	0.096
Correlación de mapa con OVERLAPMAP	0.701	0.686	0.667	0.718	0.566	0.641

Tabla 3.6.8: Recogida de datos, estadísticas de refinamiento y obtención de fases comparando el método de inmersión frente al método de cocrystalización. Los datos de difracción fueron recogidos empleando radiación sincrotrón a la longitud de onda de absorción de la línea LIII del Gd a partir de los cristales derivados de CbpF obtenidos mediante cocrystalización o mediante inmersión con tres de los siete complejos de Gd empleados. Los valores entre paréntesis se refieren a la concha de mayor resolución. *Datos proporcionados por XDS. ** Datos proporcionados por SHARP. *** Datos proporcionados por O. La resolución de las estadísticas de faseado se limitó a 2.70Å en todos los casos.

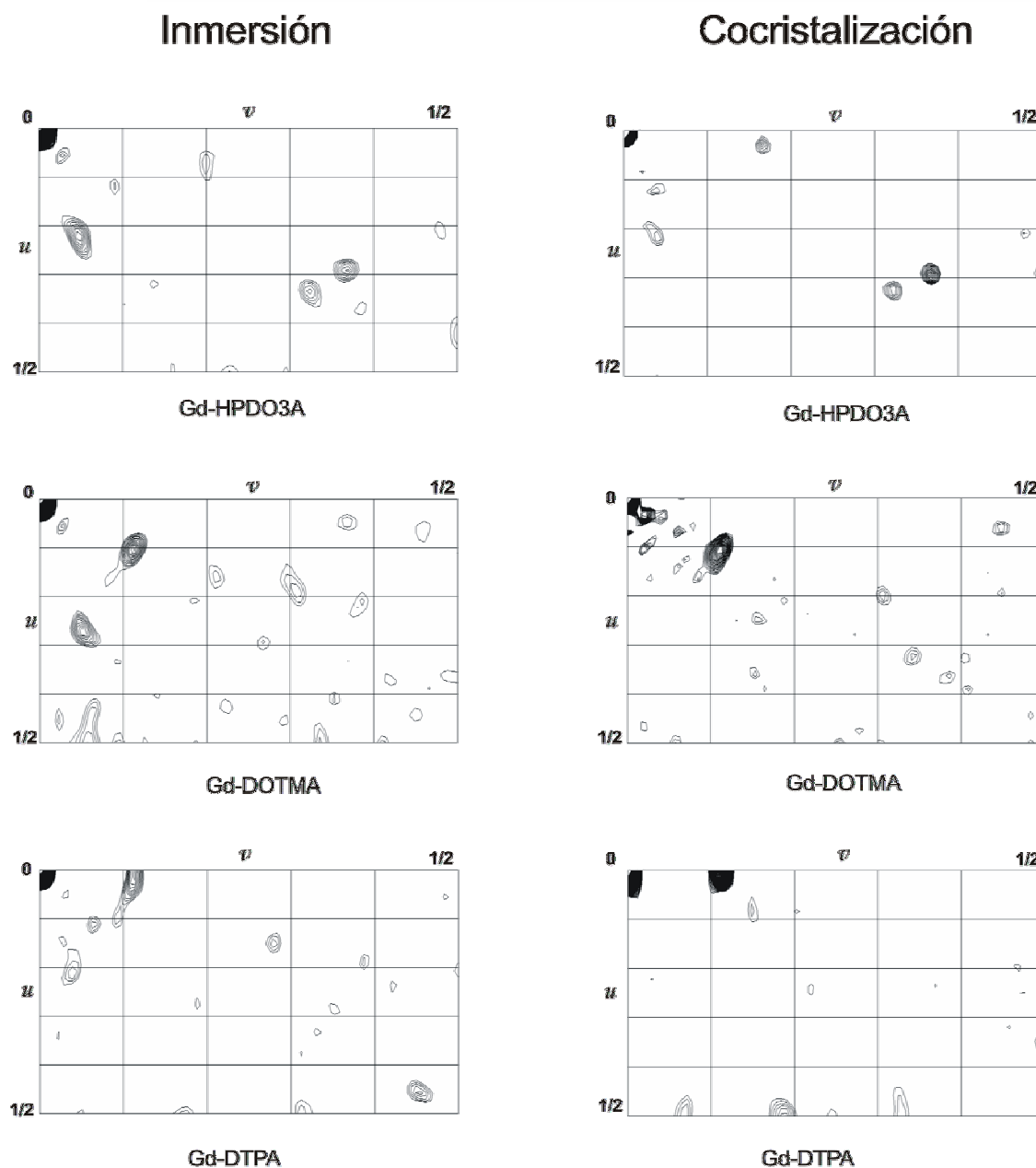


Figura 3.6.29: Comparación de las secciones de Harker $w = 0$ de los mapas de Patterson anómalos entre derivados obtenidos por cocrystalización y por inmersión con tres de los siete complejos empleados. Los niveles de intensidad de los picos están representados pasos de 0.5σ comenzando a partir de 2σ .

Contrariamente, las ocupaciones de los sitios de unión de los cristales derivados obtenidos por cocrystalización eran generalmente superiores a los correspondientes a los obtenidos por inmersión, como también aparece ilustrado por la intensidad de los picos en los mapas de Patterson anómalos (Figura 3.6.29). Hay que resaltar que aunque los tiempos de inmersión eran muy cortos (no más de diez segundos), originaban la unión efectiva del complejo generando

buenos resultados de obtención de fases. Las pequeñas diferencias observadas entre los métodos de inmersión y de cocrystalización podrían estar relacionadas con la alta accesibilidad de los sitios de unión en este caso. De acuerdo con esto, se ha documentado que la cocrystalización era el único camino para preparar cristales derivados cuando los sitios de unión eran inaccesibles a los complejos de Gd en el cristal (Girard *et al.*, 2003b).

3.6.5.6. Modo de unión de los diferentes complejos de Gd empleados.

El número y la localización de los diferentes sitios de unión del ligando fueron analizados para los siete complejos empleados^{##} (Figura 3.6.30). Todos los sitios de unión fueron localizados en el módulo C-terminal de CbpF. Las diferencias estereoquímicas y de carga global neta entre los complejos originaron sitios de unión diferentes en localización y número. Las extensiones del carboxilato de la molécula de ligando (Figura 1.2.9) se pliegan hacia el ion de gadolinio en todos los complejos, formando una jaula alrededor del ion cargado (Figura 3.6.31). Se observaron principalmente tres modos de interacción: (i) Interacciones hidrofóbicas CH- π (Brandl *et al.*, 2001) fueron las interacciones observadas para los complejos eléctricamente neutros (Gd-HPDO3A, Gd-DO3A, Gd-DTPA-BMA) y para dos de los complejos con una única carga global negativa (Gd-DOTA and Gd-DOTA-BOM), (ii) Interacciones polares que se detectaron en la unión del ligando Gd-DOTMA y (iii) Interacciones electrostáticas y polares implicadas en la unión del ligando Gd-DTPA.

Las interacciones CH- π juegan el papel principal en la unión de cinco de los siete complejos analizados y generan el mayor número de sitios de unión del ligando (oscilando de 1 a 4 dependiendo del ligando). Este tipo de interacción se observó en varios sitios diferentes, el más común de ellos (observado en todos los cinco ligandos construidos alrededor del macrociclo de tetraazaciclododecano) corresponde al bolsillo hidrofóbico más grande (sitio1) formado por la cadena lateral de cuatro residuos aromáticos (tres residuos de Trp y un residuo de Tyr) (Figura 3.6.31a). En este caso el macrociclo está completamente encapsulado dentro de una cavidad aproximadamente cuadrada (aproximadamente 8 Å de lado). Los restantes sitios de unión hidrofóbicos, sitios 2, 4, 6 y 7, (Figura 3.6.31b, d, f y g) estabilizan el ligando por medio de

^{##} Para estudiar las interacciones proteína-ligando, la localización de los sitios de unión fue visualizado con el programa O y el refinamiento de las posiciones del ligando fueron llevadas a cabo con CNS.

interacciones con los tres residuos aromáticos (dos Trp estructuralmente conservados y una Tyr) que están implicados en la unión de colina (Molina *et al.*, 2009). En estos bolsillos aromáticos más pequeños el macrociclo está orientado hacia uno de los residuos de Trp. La unión del ligando se alcanza principalmente a través de interacciones CH- π , pero también se observaron interacciones polares con un residuo de Asn, mediado por una molécula de agua (Figura 3.6.31b, d y f). Particularmente interesante resulta el hecho de que el complejo lineal Gd-DTPA-BMA adopta una conformación similar a los complejos basados en el macrociclo e interacciona con la proteína de modo similar. Los ligandos de Gd más pequeños como Gd-DO3A y Gd-HPDO3A manifiestan el mayor número de sitios (3 y 4 respectivamente) presumiblemente debido a que se pueden acomodar más fácilmente dentro de los bolsillos hidrofóbicos de tamaños variables respecto a ligandos más grandes. Sin embargo, la ocupación más alta (0.82) dentro de los ligandos que interaccionan a través de interacciones CH- π es alcanzada por Gd-DOTA-BOM (Tabla 3.6.7). Este resultado es presumiblemente debido a la presencia de un anillo fenilo extra permitiendo a DOTA-BOM interaccionar con los residuos aromáticos de la proteína a través del macrociclo y del anillo fenilo.

Otro modo de unión fue observado para el complejo Gd-DOTMA (sitio 5). Este complejo, que presenta una única carga global negativa, se une a la proteína a través de enlaces de hidrógeno con las amidas de la cadena principal (e.g. Gly 239) vía las dos cadenas de carboxilato del ligando y a través de otros dos enlaces de hidrógeno mediados por moléculas de agua (Figura 3.6.31e). Además, dos grupos metilo de DOTMA también contactan con el residuo de Leu238 reforzando la estabilización del complejo. No se observó interacción de este ligando y sólo un sitio fue ocupado (ocupación de 0.41), indicando un débil modo de interacción.

La unión del complejo Gd-DTPA a CbpF implica la coordinación del ion Gd³⁺ mediante el oxígeno del grupo amido de la Asn266 y la formación de un fuerte puente salino (2.04 Å) entre la Lys 273 y uno de los grupos carboxílicos del ligando (Figura 3.6.31c). Esta interacción está reforzada a través de una fuerte red de siete enlaces de hidrógeno entre los átomos de oxígeno de los grupos carboxílicos del ligando y las cadenas laterales de Asn266, Tyr264 y dos amidas de la cadena principal (Ala267, Gly268) (Figura 3.6.31c). En contraste al escenario observado para otros complejos, los grupos carboxílicos de Gd-DTPA están orientados hacia la proteína y no hacia el solvente. Este modo de interacción se observa para el único ligando que presenta

dos cargas negativas. La estructura lineal de DTPA probablemente induce una elevada plasticidad del ligando, maximizando así el número de potenciales interacciones con la proteína. Su elevada ocupación (0.84), la mayor observada para todos los complejos, refleja el poder de esta interacción. Para el complejo CbpF:Gd-DTPA la densidad electrónica era suficientemente buena para permitir la construcción y el refinamiento de la estructura del propio complejo de Gd, cuya estructura no había sido documentada anteriormente. En las estructuras refinadas la esfera de coordinación es normalmente completada por una molécula de agua aunque las cadenas laterales de los aminoácidos también están implicadas en este papel.

En resumen, se observan principalmente tres modos de unión para los siete complejos empleados con la misma proteína. Mientras que la interacción CH- π es la más frecuente, interacciones electrostáticas también conducen a elevadas ocupaciones de los sitios de unión. Esta versatilidad aumenta el potencial de estos complejos como agentes para la preparación de derivados de átomo pesado con diferentes proteínas.

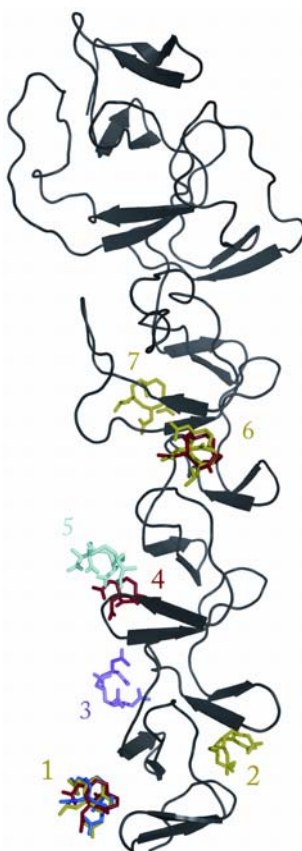


Figura 3.6.30: Localización de los sitios de unión a colina de los diferentes complejos numerados de 1 a 7. Los ligandos de los complejos están representados en verde, (Gd-HPDO3A), en rojo, (DO3A), en cian (DOTMA), en magenta, (DTPA) y en azul (DOTA). Los ligandos DOTA-BOM y DTPA-BMA no estaban bien definidos en el mapa de densidad electrónica pero su átomo Gd fue localizado en el sitio número 1 en ambos casos.

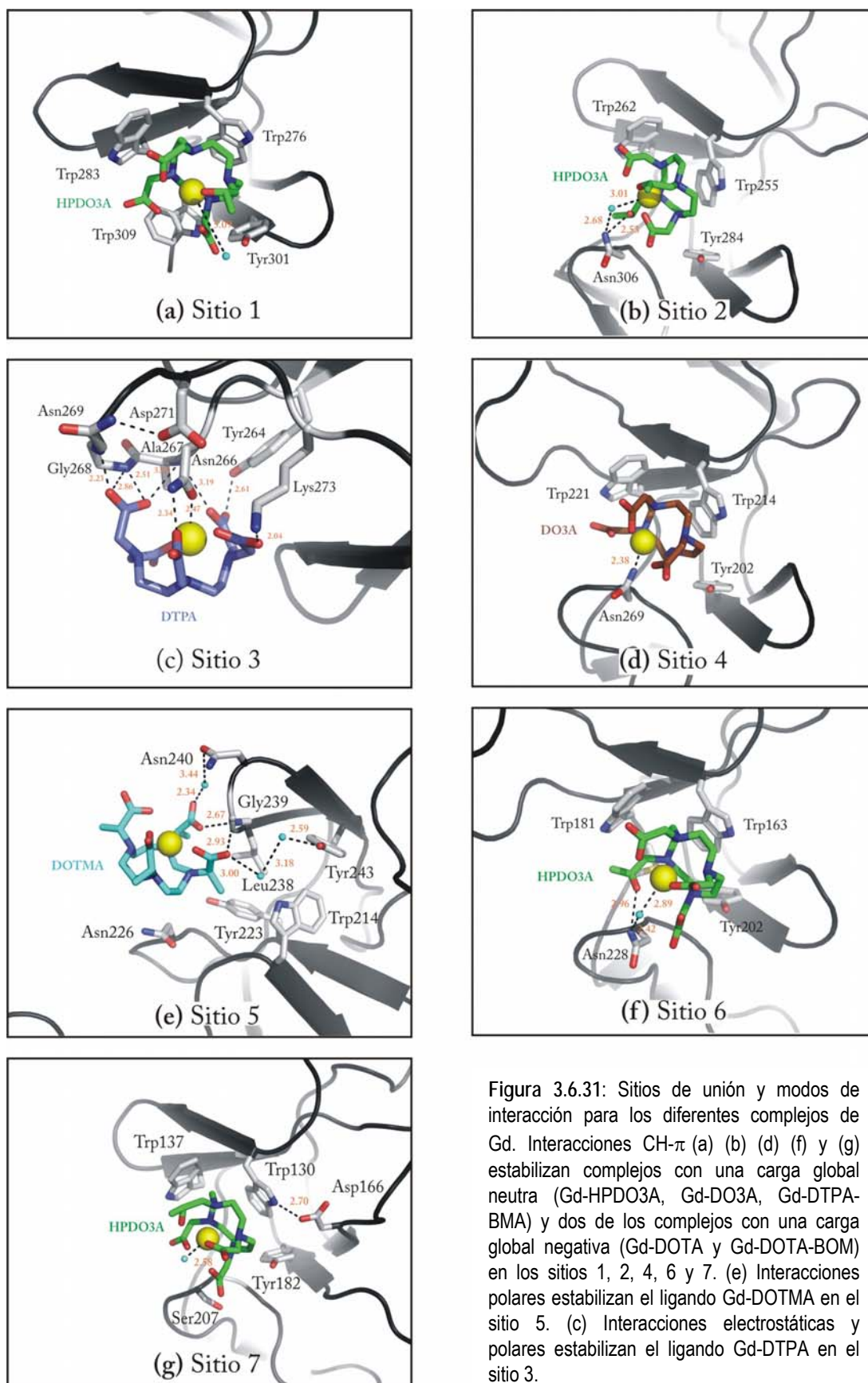


Figura 3.6.31: Sitios de unión y modos de interacción para los diferentes complejos de Gd. Interacciones CH- π (a) (b) (d) (f) y (g) estabilizan complejos con una carga global neutra (Gd-HPDO3A, Gd-DO3A, Gd-DTPA-BMA) y dos de los complejos con una carga global negativa (Gd-DOTA y Gd-DOTA-BOM) en los sitios 1, 2, 4, 6 y 7. (e) Interacciones polares estabilizan el ligando Gd-DOTMA en el sitio 5. (c) Interacciones electrostáticas y polares estabilizan el ligando Gd-DTPA en el sitio 3.

3.6.5.7. Estudio comparativo de los datos obtenidos por radiación sincrotrón frente al ánodo rotatorio.

Los datos de difracción del cristal derivado Gd-HPDO3A recogidos en un difractómetro convencional de laboratorio, fueron comparados con datos recogidos empleando radiación sincrotrón con el fin de probar el potencial de obtención de fases de estos compuestos para emplearlos con fuentes de radiación convencional (Tabla 3.6.9). Elevados valores para R_{ano} y $\Delta F/\sigma$ (4.5% y 1.66, respectivamente) indicaron una señal anómala significativa. Las secciones de Harker de los mapas de Patterson anómalos, claramente revelaron la presencia de dispersores anómalos (Figura 3.6.32). Cuatro sitios de Gd fueron identificados, los cuales se correspondían a los sitios encontrados a partir de los datos obtenidos en el sincrotrón (estadísticas de obtención de fases se muestran en la tabla 3.6.9). Los valores de la figura de mérito estaban alrededor de 0.45 antes de la modificación de densidad. Se obtuvieron valores de poder de obtención de fases de 1.74, R-cullis de 0.64 y una figura de mérito de 0.84 tras el programa de modificación de densidad. La obtención de un mapa de densidad electrónica experimental muy bien definido nos permitió trazar fácilmente el modelo. Estos resultados confirmaron la capacidad de estos compuestos para la obtención de fases por SAD empleando la fuente de rayos X convencional de los laboratorios.

Como era de esperar, los resultados de la obtención de fases para los datos recogidos con la radiación sincrotrón a la línea de absorción LIII del Gd eran mejores que aquellos obtenidos con un ánodo rotatorio convencional propiedad del Instituto Rocasolano. Sin embargo, la señal anómala y las fases obtenidas empleando una fuente de radiación convencional condujeron a mapas de densidad electrónica experimentales muy bien definidos y de suficiente calidad para permitir la construcción del modelo. A partir del cristal derivado Gd-HPDO3A obtenido por cocrystalización, se recogieron datos con radiación sincrotrón a la línea de absorción LIII del Gd proporcionando una excelente señal anómala hasta una resolución de 2.23 Å, la cual, combinada con extensión de fases nos permitió obtener un mapa experimental inicial a 1.91 Å de resolución (ver figura 3.6.32d). La buena calidad del mapa permitió trazar de forma sencilla incluso la mayoría de las cadenas laterales. De modo análogo, a partir del cristal derivado Gd-HPDO3A obtenido por cocrystalización, los datos recogidos empleando una fuente de rayos X convencional permitió la determinación de fases a 2.93 Å de resolución conduciendo a un mapa

de densidad electrónica experimental bien definido de 1.86 Å de resolución tras el aplanado de solvente y la modificación de la densidad (Figura 3.6.32c). Aunque el mapa obtenido a partir de la fuente de radiación convencional mostraba menos información detallada que la obtenida con radiación sincrotrón, la diferencia entre ambos mapas era insignificante como se puede observar para el sitio de unión a colina.

Recogida de datos	Ánodo Rotatorio	Sincrotrón (ESRF)
Parámetros de celdilla		
a (Å)	49.04	49.35
b (Å)	114.86	114.52
c (Å)	76.05	76.23
Rango de resolución (Å)	100-2.39	100-2.23
	(2.51-2.39)	(2.36-2.23)
No. de reflexiones observadas	115,647	143,495
No. de reflexiones únicas	16,575	21,579
Acéntricas	14,195	18,835
Céntricas	2,380	2,744
Relación señal-ruido $I / \sigma(I)$	13.7(2.3)	10.4(7.7)
Redundancia	7.0(6.9)	6.6(5.5)
Completitud (%)	99.6(97.1)	98.9(95.0)
R_{sym} (%)	4.9(33.9)	4.7(8.9)
R_{ano} (%)	4.5(15.6)	10.6(5.9)
$\Delta F/\sigma$ *	1.661	4.249
Estadísticas de obtención de fases		
No. de sitios encontrados por SHELXD	4	5
No. de sitios modelados***	4	4
Ocupación de los sitios	0.51, 0.28, 0.20, 0.19	0.64, 0.63, 0.49, 0.38.
$\langle \Delta F \rangle / F$	0.023	0.090
$\Delta F/\sigma$ *	1.661	4.654
FOM tras refinamiento ML**	0.426	0.521
R-Cullis para reflexiones acéntricas**	0.719	0.352
Poder de obtención de fases para reflexiones acéntricas**	1.581	4.546
FOM tras DM**	0.802	0.823
Factor R Free en el espacio real**	0.096	0.103
Correlación de mapas con <i>OVERLAPMAP</i>	0.096	0.103

Tabla 3.6.10: Recogida de datos, estadísticas de refinamiento y obtención de fases comparando datos obtenidos mediante radiación sincrotrón y mediante ánodo rotatorio. Datos recogidos a partir de cristales derivados de CbpF obtenidos mediante cocrystalización con el complejo Gd-HPDO3A y medidos con radiación CuK α del ánodo rotatorio convencional y con radiación sincrotrón. Los valores entre paréntesis se refieren a la concha de mayor resolución. * Datos proporcionados por XDS. ** Datos proporcionados por SHARP. *** Datos proporcionados por O. La resolución en las estadísticas de obtención de fases se limitó a 2.39 Å en ambos casos

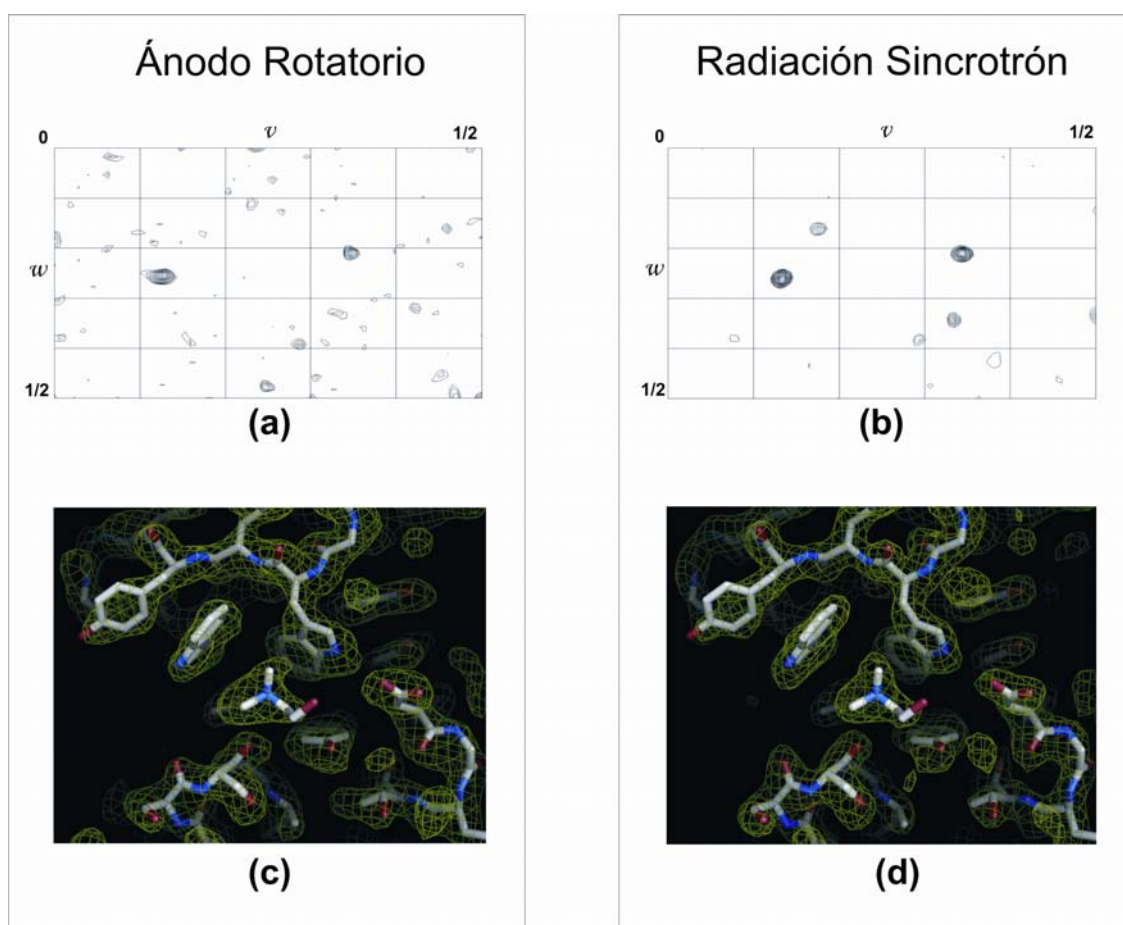


Figura 3.6.32: Comparación entre los datos obtenidos mediante radiación sincrotrón y mediante ánodo rotatorio. Secciones de Harker $u = 1/2$ del mapa de Patterson anómalo para el cristal derivado obtenido con el complejo Gd-HPDO3A y medido con radiación sincrotrón (b) y datos recogidos usando un ánodo rotatorio convencional (a). Los niveles de intensidad de los picos están representados pasos de 0.5σ comenzando a partir de 2σ . Los mapas de densidad electrónica experimentales obtenidos a partir de cristales derivados Gd-HPDO3A medidos empleando la radiación Cu K α de una fuente de radiación convencional (c) y con radiación sincrotrón a la longitud de onda del pico de absorción LIII del Gd (d). Las imágenes representan los mapas de densidad electrónica experimentales en la región del sitio de unión de colina a 2 Å de resolución.

Refinamiento	Gd-HPDO3A	Gd-DOTMA	Gd-DTPA	Gd-DO3A	Gd-DTPA-BMA	Gd-DOTA-BOM	Gd-DOTA
Rango de resolución (Å)	46.13-2.23	45.64-2.23	45.175-2.23	45.92-2.29	45.14-2.98	45.81-2.23	45.91-2.30
R (%)	20.65	20.95	21.57	21.78	23.55	22.14	21.39
R _{free} (%)	22.44	22.96	23.42	23.56	25.91	25.28	24.33
Ocupación de los sitios*	0.87, 0.66, 0.77, 0.60	0.65	0.99	0.83, 0.66, 0.63	0.34, 0.28	0.95	0.80
Factores B (Å ²)*	33, 36, 31, 46	44	37	41, 41, 50	62, 62	38	53
Modelo							
Moléculas de proteína por unidad asimétrica	1	1	1	1	1	1	1
Aminoácidos	298	298	298	298	298	298	298
Moléculas de colina	3	6	6	4	5	6	6
Moléculas de complejo de Gd	4	1	1	3	0	0	1
Moléculas de agua	296	297	289	293	215	295	269
Desviación RMS del ideal:							
Distancia de enlace (Å)	0.014	0.013	0.013	0.015	0.023	0.014	0.013
Ángulos de enlace (grados)	1.344	1.339	1.361	1.348	2.141	1.403	1.297
Observaciones	Bucles C147-C150 y C169-C177 no están bien definidos						

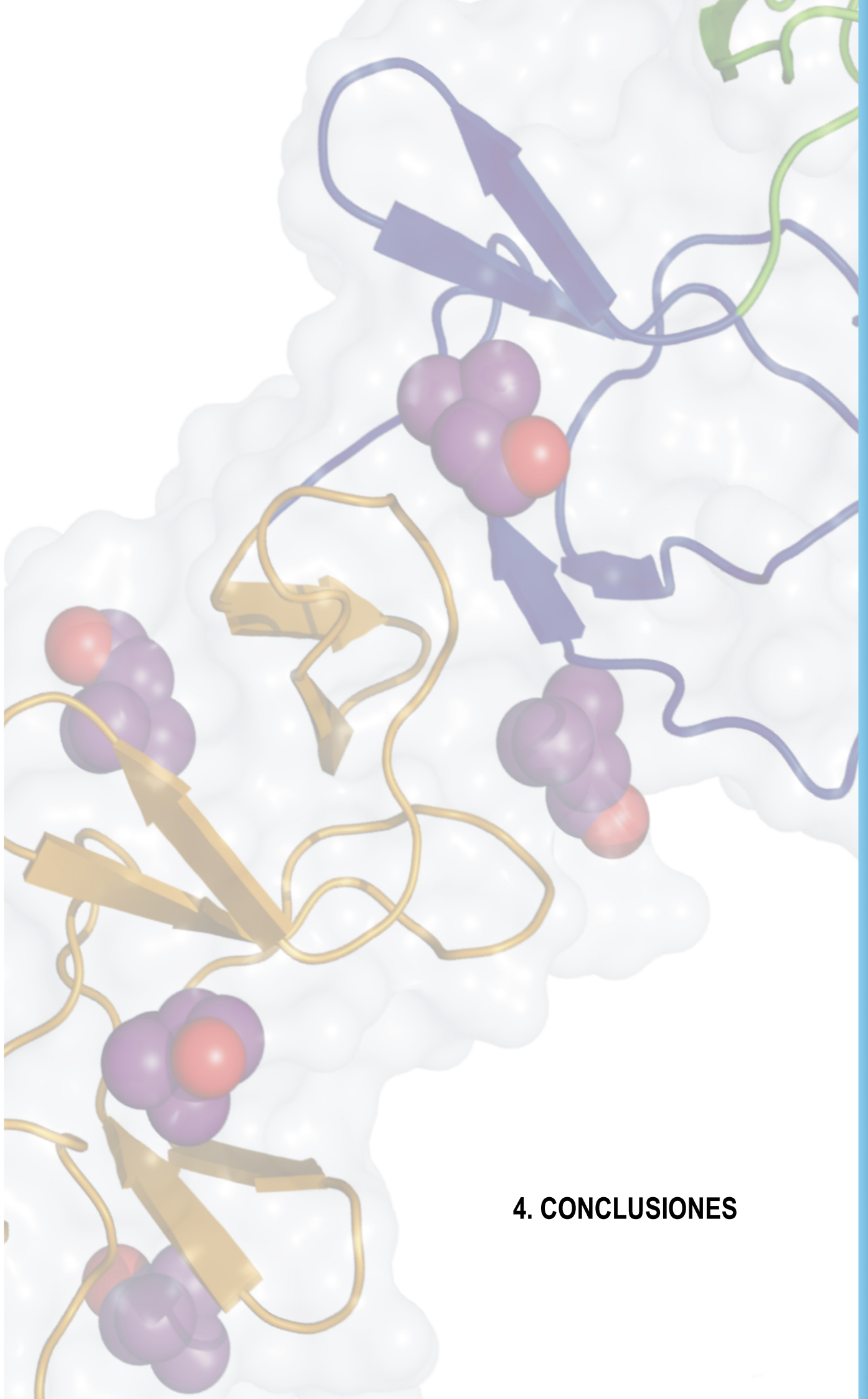
Tabla 3.6.10: Estadísticas de las estructuras refinadas correspondientes a los siete derivados de Gd obtenidos por cocrystalización. * Datos proporcionados por CNS.

Refinamiento	Gd-HPDO3A	Gd-DOTMA	Gd-DTPA
Rango de resolución (Å)	45.36-2.70	45.14-2.70	46.94-2.70
<i>R</i> (%)	22.95	23.15	23.52
<i>R</i> _{free} (%)	25.04	25.36	26.21
Ocupación de los sitios*	0.56, 0.41, 0.37, 0.34	0.51, 0.39	0.52, 0.26, 0.21
Factores B (Å ²)*	41, 44, 40, 42	45, 49	42, 48, 51
Modelo			
Moléculas de proteína por unidad asimétrica	1	1	1
Aminoácidos	298	298	298
Moléculas de colina	3	5	4
Moléculas de complejo de Gd	4	2	3
Moléculas de agua	296	297	289
Desviación RMS del ideal:			
Distancia de enlace (Å)	0.014	0.013	0.013
Ángulos de enlace (grados)	1.344	1.339	1.361
Observaciones	Bucles C147-C150 y C169-C177 no están bien definidos		

Tabla 3.6.11: Estadísticas de las estructuras refinadas correspondientes a los derivados de Gd obtenidos mediante inmersión. * Datos proporcionados por CNS.

Refinamiento	Gd-HPDO3A
Rango de resolución (Å)	45.36-2.39
<i>R</i> (%)	22.28
<i>R</i> _{free} (%)	25.16
Ocupación de los sitios*	0.71, 0.63, 0.44, 0.37
Factores B (Å ²)*	39, 42, 37, 49
Model	
Moléculas de proteína por unidad asimétrica	1
Aminoácidos	298
Moléculas de colina	3
Moléculas de complejo de Gd	4
Moléculas de agua	296
Desviación RMS del ideal:	
Distancia de enlace (Å)	0.019
Ángulos de enlace (grados)	1.511
Observaciones	Bucles C147-C150 y C169-C177 no están bien definidos

Tabla 3.6.12: Estadísticas de la estructura refinada para el derivado Gd-HPDO3A obtenido mediante ánodo rotatorio.



4. CONCLUSIONES

4. CONCLUSIONES.

1. CbpF presenta una estructura modular compuesta enteramente por repeticiones de unión a colina canónicas y no canónicas, distribuidas a lo largo de su secuencia, lo cual altera dramáticamente su forma, distribución de carga y capacidad de unión, organizando la proteína en dos módulos bien definidos. Mientras el módulo N-terminal está constituido exclusivamente por repeticiones no canónicas incapaces de unir colina, en el módulo C-terminal o de unión a colina se concentran las repeticiones canónicas (subdominio CII), aunque no exclusivamente ya que también presenta repeticiones no canónicas (subdominio CI) con capacidad de unir colina.
2. Como proteína perteneciente a la familia de proteínas de unión a colina (CBPs), uno de sus módulos, en este caso el C-terminal, se erigió como el implicado en la unión a la pared celular, mientras que el módulo N-terminal se reveló como crítico para la inhibición de la acción autolítica de la muramidasa LytC, proporcionando una función reguladora para la autólisis de neumococo.
3. Los ensayos de unión con diferentes análogos de la pared y análogos de teicoico no proporcionaron una prueba directa de interacción pero en el caso del análogo $(2S5P)_2$ el cambio en los parámetros de la celdilla apuntó a la interacción con dicho ligando. Mediante estudios de docking se profundizó en el análisis de la interacción validando la posibilidad de que se produjera dicha interacción ya que se acomodaba el fragmento perfectamente.
4. Los ensayos de inhibición *in vivo* e *in vitro* probaron que CbpF modula selectivamente la función autolítica de la lisozima LytC y que la divergencia de las repeticiones de unión a colina que constituye el dominio N-terminal juega un papel crítico en este efecto específico. Es decir, nuestros resultados demuestran que CbpF puede inhibir la actividad de LytC tanto *in vivo* como *in vitro* y por lo tanto, aseveramos que CbpF juega un papel clave en la regulación de la actividad de LytC.
5. Los experimentos de SPR y de entrecruzamiento llevados a cabo entre CbpF y LytC han descartado la hipótesis de que dicha inhibición se produjera debido a interacciones directas proteína-proteína. El módulo N-terminal de CbpF era capaz de unir tanto paredes celulares de neumococo con y sin colina, sugiriendo que el módulo N-terminal es capaz de reconocer

específicamente las paredes celulares de neumococo y que los módulos N- y C-terminal de CbpF se unen a diferentes componentes del peptidoglicano.

6. Un estudio sobre los ortólogos de CbpF en otras cepas de neumococo reveló la existencia de las proteínas CbpC y CbpJ en la cepa TIGR4. De hecho, un análisis comparativo demostró que CbpF puede ser el resultado de la recombinación de CbpC y CbpJ. La coexistencia de dos genes muy similares en TIGR4 apunta al hecho de que CbpF juega un papel fisiológico crítico y sugiere que esta duplicación puede que favorezca la adquisición de propiedades reguladoras selectivas. Posteriores análisis de genomas de neumococo revelaron la existencia de varias proteínas que parecen mostrar la misma arquitectura que CbpF. Todo esto nos conduce a pensar que estas proteínas puedan constituir una nueva subfamilia tipo CbpF dentro de la amplia familia de proteínas CBP poseyendo un CBM típico y una región N-terminal constituida por una serie de repeticiones no consenso. A pesar de las pequeñas diferencias en la composición aminoacídica dentro de todos los componentes de la subfamilia tipo CbpF, se conservan la mayoría de los residuos implicados tanto en la estructura N-terminal y las interacciones intermodulares, sugiriendo que esta proteínas pueden exhibir funciones reguladoras similares a CbpF.

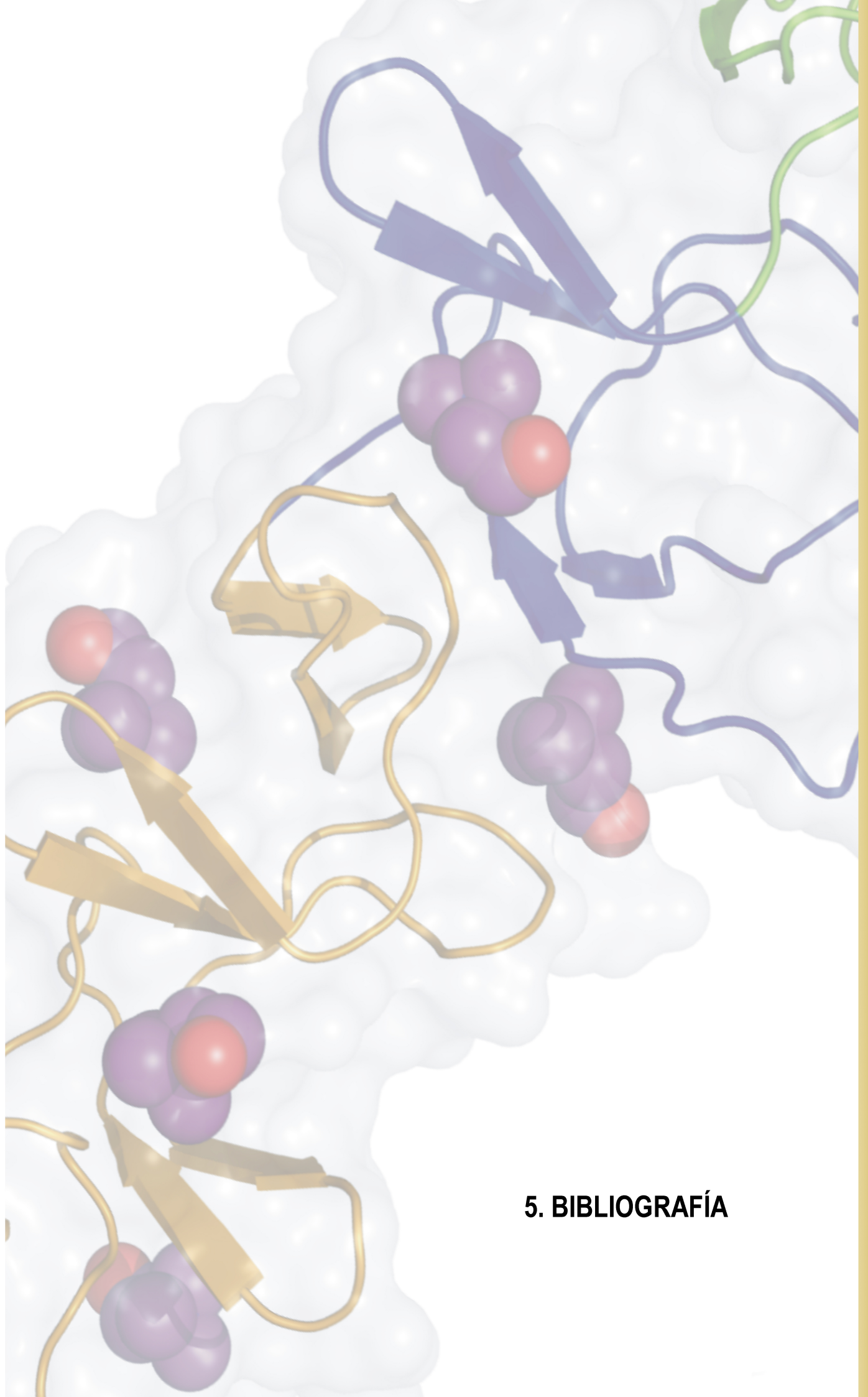
7. Los siete complejos de gadolinio presentados en este estudio son excelentes candidatos para métodos de obtención de fases experimentales basados en dispersión anómala. Pueden ser usados a elevadas concentraciones para obtener cristales derivados mediante inmersión o cocrystalización con nuestra proteína modelo y los experimentos SAD condujeron a mapas de densidad electrónica que permitieron la construcción inequívoca del modelo proteico en todos los casos.

8. Los métodos de inmersión y cocrystalización mostraron ser igualmente adecuados para proporcionar cristales derivados de buena calidad. Mientras el número de los sitios de unión del complejo era igual o mayor para los cristales sumergidos, las ocupaciones de los sitios de unión eran generalmente superiores para los derivados cocrystalizados.

9. Estos complejos de gadolinio también ofrecen la posibilidad de obtener buenas fases empleando datos de difracción a partir de una fuente de rayos X proporcionada por un ánodo CuK α , haciendo uso de sus 12 e⁻ del factor de dispersión f' a esta longitud de onda. La introducción de estos complejos dentro de los cristales no perjudicó su calidad de difracción

a diferencia de lo que ocurre habitualmente con la mayoría de los compuestos de átomo pesado convencionales.

10. Se observaron tres modos de interacción diferentes para nuestra proteína modelo dependiendo de la naturaleza química del complejo. Debido a que para una proteína dada, los sitios de unión probablemente sean diferentes dependiendo del ligando, estos complejos podrían ser usados como una mezcla para tomar la ventaja de sus diferentes modos de unión. La versatilidad de los complejos de gadolinio tetraazaciclododecano, junto con su elevado poder de obtención de fases incluso empleando fuentes de radiación de rayos X convencionales, podría convertirlos en una herramienta útil en cristalografía macromolecular.



5. BIBLIOGRAFÍA

5. BIBLIOGRAFÍA.

- Albert, A., Martinez-Ripoll, M., Espinosa-Ruiz, A., Yenush, L., Cullanez-Macia, F.A. and Serrano, R. (2000) The X-ray structure of the FMN-binding protein AtHal3 provides the structural basis for the activity of a regulatory subunit involved in signal transduction. *Structure*, **8**, 961-969.
- Austrian, R. (1953) Morphologic variation in pneumococcus. I. An analysis of the bases for morphologic variation in pneumococcus and description of a hitherto undefined morphologic variant. *J Exp Med*, **98**, 21-40.
- Barnickel, G., Nauman, D., Bradaczek, H., Labischinski, H., y Giesbrecht, P. (1983). The target of Penicillin (Hakenbeck, R., Höltje, J.V., y Labischinski, H., eds.), pp 61-66. Berlin: W. de Gruyter.
- Behr, T., Fischer, W., Peter-Katalinic, J. and Egge, H. (1992) The structure of pneumococcal lipoteichoic acid. Improved preparation, chemical and mass spectrometric studies. *Eur J Biochem*, **207**, 1063-1075.
- Bergmann, S. and Hammerschmidt, S. (2006) Versatility of pneumococcal surface proteins. *Microbiology*, **152**, 295-303.
- Berry, A.M., Lock, R.A., Hansman, D. and Paton, J.C. (1989) Contribution of autolysin to virulence of *Streptococcus pneumoniae*. *Infect Immun*, **57**, 2324-2330.
- Bethe, G., Nau, R., Wellmer, A., Hakenbeck, R., Reinert, R.R., Heinz, H.P. and Zysk, G. (2001) The cell wall-associated serine protease PrtA: a highly conserved virulence factor of *Streptococcus pneumoniae*. *FEMS Microbiol Lett*, **205**, 99-104.
- Beveridge, T.J. (1981) Ultrastructure, chemistry, and function of the bacterial wall. *Int Rev Cytol*, **72**, 229-317.
- Bianchi, A., Calabi, L., Corana, F., Fontana, S., Losi, P., Maiocchi, A., Paleari, L., Valtancoli, B. (2000). "Thermodynamic and structural properties of Gd(III) complexes with polyaminopolycarboxylic ligands: basic compounds for the development of MRI contrast agents." *Coord. Chem. Rev.* **204**, 309-393.
- Bierbaum, G. and Sahl, H.G. (1985) Induction of autolysis of staphylococci by the basic peptide antibiotics Pep 5 and nisin and their influence on the activity of autolytic enzymes. *Arch Microbiol*, **141**, 249-254.
- Blue, C.E., Paterson, G.K., Kerr, A.R., Berge, M., Claverys, J.P., y Mitchell, T.J. (2003). ZmpB, a novel virulence factor of *Streptococcus pneumoniae* that induces tumor necrosis factor alpha production in the respiratory tract. *Infect. Immun.* **71**: 4925-4935.

- Boggon, T.J., Helliwell, J.R., Judge, R.A., Olczak, A., Siddons, D.P., Snell, E.H. and Stojanoff, V. (2000) Synchrotron X-ray reciprocal-space mapping, topography and diffraction resolution studies of macromolecular crystal quality. *Acta Crystallogr D Biol Crystallogr*, **56**, 868-880.
- Boggon, T.J. and Shapiro, L. (2000) Screening for phasing atoms in protein crystallography. *Structure*, **8**, R143-149.
- Bond, C.S., Shaw, M.P., Alphey, M.S. and Hunter, W.N. (2001) Structure of the macrocycle thioestrepton solved using the anomalous dispersion contribution of sulfur. *Acta Crystallogr D Biol Crystallogr*, **57**, 755-758.
- Bracha, R., Chang, M., Fiedler, F. and Glaser, L. (1978) Biosynthesis of teichoic acids. *Methods Enzymol*, **50**, 387-340.
- Brandl, M., Weiss, M.S., Jabs, A., Suhnel, J. and Hilgenfeld, R. (2001) C-H...pi-interactions in proteins. *J Mol Biol*, **307**, 357-377.
- Braun, J.S., Sublett, J.E., Freyer, D., Mitchell, T.J., Cleveland, J.L., Tuomanen, E.I. and Weber, J.R. (2002) Pneumococcal pneumolysin and H₂O₂ mediate brain cell apoptosis during meningitis. *J Clin Invest*, **109**, 19-27.
- Bricogne, G. (1988). A Bayesian statistical theory of the phase problem. I. A multichannelmaximum-entropy formalism for constructing generalized joint probability distributions of structure factors. *Acta Cryst. A*, **44**, 517-545.
- Briles, D.E., Hollingshead, S.K., Swiatlo, E., Brooks-Walter, A., Szalai, A., Virolainen, A., McDaniel, L.S., Benton, K.A., White, P., Prellner, K., Hermansson, A., Aerts, P.C., Van Dijk, H. and Crain, M.J. (1997) PspA and PspC: their potential for use as pneumococcal vaccines. *Microb Drug Resist*, **3**, 401-408.
- Brooks-Walter, A., Briles, D.E. and Hollingshead, S.K. (1999) The *pspC* gene of *Streptococcus pneumoniae* encodes a polymorphic protein, PspC, which elicits cross-reactive antibodies to PspA and provides immunity to pneumococcal bacteremia. *Infect Immun*, **67**, 6533-6542.
- Brown, M., Peierls, R., Stern, E. (1997). "White lines in x-ray absorption." *Physical Review B*. **15**, **2**, 738-744.
- Brundish, D.E. and Baddiley, J. (1968) Pneumococcal C-substance, a ribitol teichoic acid containing choline phosphate. *Biochem J*, **110**, 573-582.
- Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*, **54**, 905-921.

- Cartwright, K.A. (2002) Epidemiology of meningococcal disease. *Hosp Med*, **63**, 264-267.
- Chen, W. and Bahl, O.P. (1991) Recombinant carbohydrate and selenomethionyl variants of human choriogonadotropin. *J Biol Chem*, **266**, 8192-8197.
- Claverys, J.P. and Havarstein, L.S. (2007) Cannibalism and fratricide: mechanisms and raisons d'etre. *Nat Rev Microbiol*, **5**, 219-229.
- Chesne, M. L. (2002). "Utilisation de la diffusion anormale de l'uranium pour déterminer la structure de macromolécules biologiques - des rayons X durs aux rayons X mous." *Thèse de doctorat, Université Joseph Fourier, Grenoble*.
- Chiavolini, D., Memmi, G., Maggi, T., Iannelli, F., Pozzi, G. and Oggioni, M.R. (2003) The three extra-cellular zinc metalloproteinases of *Streptococcus pneumoniae* have a different impact on virulence in mice. *BMC Microbiol*, **3**, 14.
- Cohen, S.L., Padovan, J.C. and Chait, B.T. (2000) Mass spectrometric analysis of mercury incorporation into proteins for X-ray diffraction phase determination. *Anal Chem*, **72**, 574-579.
- Cohen, A., Ellis, P., Kresge, N. and Soltis, S.M. (2001) MAD phasing with krypton. *Acta Crystallogr D Biol Crystallogr*, **57**, 233-238.
- Colloborative Computational Project, N. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*, **50**, 760-763.
- Cowtan, K.D. and Main, P. (1996) Phase combination and cross validation in iterated density-modification calculations. *Acta Crystallogr D Biol Crystallogr*, **52**, 43-48.
- Cundell, D., Masure, H.R. and Tuomanen, E.I. (1995) The molecular basis of pneumococcal infection: a hypothesis. *Clin Infect Dis*, **21 Suppl 3**, S204-211.
- Cundell, D.R., Gerard, N.P., Gerard, C., Idanpaan-Heikkila, I. and Tuomanen, E.I. (1995) *Streptococcus pneumoniae* anchor to activated human cells by the receptor for platelet-activating factor. *Nature*, **377**, 435-438.
- Dauter, Z. and Dauter, M. (1999) Anomalous signal of solvent bromides used for phasing of lysozyme. *J Mol Biol*, **289**, 93-101.
- Dauter, Z., Dauter, M. and Rajashankar, K.R. (2000) Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Crystallogr D Biol Crystallogr*, **56**, 232-237.
- Dauter, Z., Li, M. and Wlodawer, A. (2001) Practical experience with the use of halides for phasing macromolecular structures: a powerful tool for structural genomics. *Acta Crystallogr D Biol Crystallogr*, **57**, 239-249.
- Dauter, Z. (2002) New approaches to high-throughput phasing. *Curr Opin Struct Biol*, **12**, 674-678.

Bibliografía

- Dauter, Z., Dauter, M. and Dodson, E. (2002) Jolly SAD. *Acta Crystallogr D Biol Crystallogr*, **58**, 494-506.
- Dauter, Z., Nagem, R. (2002). "Direct way to anomalous scatterers." *Z. Kristallogr.* **217**, 694-702
- Dawson, M.H. (1933). *Proc.Soc. Exp. Biol. And. Med.* **30**: 806.
- De las Rivas, B., Garcia, J.L., Lopez, R. and Garcia, P. (2001) Molecular characterization of the pneumococcal teichoic acid phosphorylcholine esterase. *Microb Drug Resist*, **7**, 213-222.
- De Las Rivas, B., Garcia, J.L., Lopez, R. and Garcia, P. (2002) Purification and polar localization of pneumococcal LytB, a putative endo-beta-N-acetylglucosaminidase: the chain-dispersing murein hydrolase. *J Bacteriol*, **184**, 4988-5000.
- Double, S. (1997) Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol*, **276**, 523-530.
- Drenth, J. (1999). Principles of Protein X-ray Crystallography, 2nd ed. Berlin: Springer-Verlag.
- Evans, P R. (1993). *Proceedings of the CCP4 Study Weekend, 29-30 Enero 1993*, editado por L. Sawyer, N. Isaacs & S. Bailey, pp. 114-123. Warrington: Daresbury Laboratory
- Evans, G., Pettifer, R. F. (2001) CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. *Journal Applied of Crystallography*, **34**, 82-86.
- Evans, G. and Bricogne, G. (2002) Triiodide derivatization and combinatorial counter-ion replacement: two methods for enhancing phasing signal using laboratory Cu Kalpha X-ray equipment. *Acta Crystallogr D Biol Crystallogr*, **58**, 976-991.
- Fernández-Tornero, C., Lopez, R., Garcia, E., Gimenez-Gallego, G. and Romero, A. (2001) A novel solenoid fold in the cell wall anchoring domain of the pneumococcal virulence factor LytA. *Nat Struct Biol*, **8**, 1020-1024.
- Fernández-Tornero, C., Garcia, E., Lopez, R., Gimenez-Gallego, G. and Romero, A. (2002) Two new crystal forms of the choline-binding domain of the major pneumococcal autolysin: insights into the dynamics of the active homodimer. *J Mol Biol*, **321**, 163-173.
- Fischer, W. (1988) Physiology of lipoteichoic acids in bacteria. *Adv Microb Physiol*, **29**, 233-302.
- Fischer, W., Behr, T., Hartmann, R., Peter-Katalinic, J. and Egge, H. (1993) Teichoic acid and lipoteichoic acid of *Streptococcus pneumoniae* possess identical chain structures. A reinvestigation of teichoid acid (C polysaccharide). *Eur J Biochem*, **215**, 851-857.
- Fischer, W. (2000) Phosphocholine of pneumococcal teichoic acids: role in bacterial physiology and pneumococcal infection. *Res Microbiol*, **151**, 421-427.
- Fischer, H. and Tomasz, A. (1985) Peptidoglycan cross-linking and teichoic acid attachment in *Streptococcus pneumoniae*. *J Bacteriol*, **163**, 46-54.

- Formanek, H. (1983). The target of penicillin. (Hakenbeck, R., Höltje, J.V., y Labischinski, H., eds.), pp 55-60. Berlin: W. de Gruyter.
- Garcia-Bustos, J.F., Chait, B.T. and Tomasz, A. (1987) Structure of the peptide network of pneumococcal peptidoglycan. *J Biol Chem*, **262**, 15400-15405.
- Garcia-Bustos, J. and Tomasz, A. (1990) A biological price of antibiotic resistance: major changes in the peptidoglycan structure of penicillin-resistant pneumococci. *Proc Natl Acad Sci U S A*, **87**, 5415-5419.
- Garcia, E., Garcia, J.L., Ronda, C., Garcia, P. and Lopez, R. (1985) Cloning and expression of the pneumococcal autolysin gene in *Escherichia coli*. *Mol Gen Genet*, **201**, 225-230.
- Garcia, J.L., Garcia, E. and Lopez, R. (1987a) Overproduction and rapid purification of the amidase of *Streptococcus pneumoniae*. *Arch Microbiol*, **149**, 52-56.
- Garcia, J.L., Garcia, E., Arraras, A., Garcia, P., Ronda, C. and Lopez, R. (1987b) Cloning, purification, and biochemical characterization of the pneumococcal bacteriophage Cp-1 lysin. *J Virol*, **61**, 2573-2580.
- García, J. L., García, E., Sánchez-Puelles, J. M. y López, R. (1988a). Identification of a lytic enzyme of *Clostridium acetobutylicum* that degrades choline-containing pneumococcal cell walls. *FEMS Microbiol. Lett.* **52**: 133-138.
- Garcia, P., Garcia, J.L., Garcia, E., Sanchez-Puelles, J.M. and Lopez, R. (1990) Modular organization of the lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. *Gene*, **86**, 81-88.
- Garcia, P., Gonzalez, M.P., Garcia, E., Lopez, R. and Garcia, J.L. (1999a) LytB, a novel pneumococcal murein hydrolase essential for cell separation. *Mol Microbiol*, **31**, 1275-1281.
- Garcia, P., Paz Gonzalez, M., Garcia, E., Garcia, J.L. and Lopez, R. (1999b) The molecular characterization of the first autolytic lysozyme of *Streptococcus pneumoniae* reveals evolutionary mobile domains. *Mol Microbiol*, **33**, 128-138.
- García, J. L., Sánchez-Beato, A., Medrano, F.J., y López, R. (2000). Versatility of choline-binding domain. En *Streptococcus pneumoniae* (Tomasz, A., ed.), pp 231-244, Mary Ann Liebert Inc., Larchmont, NY.
- Garman, E. and Murray, J.W. (2003) Heavy-atom derivatization. *Acta Crystallogr D Biol Crystallogr*, **59**, 1903-1913.
- Garman, E. and Nave, C. (2002) Radiation damage to crystalline biological molecules: current view. *J Synchrotron Radiat*, **9**, 327-328.
- Girard, E., Chantalat, L., Vicat, J. and Kahn, R. (2002) Gd-HPDO3A, a complex to obtain high-phasing-power heavy-atom derivatives for SAD and MAD experiments: results with tetragonal hen egg-white lysozyme. *Acta Crystallogr D Biol Crystallogr*, **58**, 1-9.

- Girard, E., Stelter, M., Anelli, P.L., Vicat, J. and Kahn, R. (2003a) A new class of gadolinium complexes employed to obtain high-phasing-power heavy-atom derivatives: results from SAD experiments with hen egg-white lysozyme and urate oxidase from *Aspergillus flavus*. *Acta Crystallogr D Biol Crystallogr*, **59**, 118-126.
- Girard, E., Stelter, M., Vicat, J. and Kahn, R. (2003b) A new class of lanthanide complexes to obtain high-phasing-power heavy-atom derivatives for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, **59**, 1914-1922.
- Girard, E., Anelli, P.L., Vicat, J. and Kahn, R. (2003c) High-phasing-power lanthanide derivatives: taking advantage of ytterbium and lutetium for optimized anomalous diffraction experiments using synchrotron radiation. *Acta Crystallogr D Biol Crystallogr*, **59**, 1877-1880.
- Girard, E., Pebay-Peyroula, E., Vicat, J. and Kahn, R. (2004) Heavy-atom derivatives in lipidic cubic phases: results on hen egg-white lysozyme tetragonal derivative crystals with Gd-HPDO3A complex. *Acta Crystallogr D Biol Crystallogr*, **60**, 1506-1508.
- Glaser, L. and Lindsay, B. (1974) The synthesis of lipoteichoic acid carrier. *Biochem Biophys Res Commun*, **59**, 1131-1136.
- Gordon, E.J., Leonard, G.A., McSweeney, S. and Zagalsky, P.F. (2001) The C1 subunit of alpha-crustacyanin: the de novo phasing of the crystal structure of a 40 kDa homodimeric protein using the anomalous scattering from S atoms combined with direct methods. *Acta Crystallogr D Biol Crystallogr*, **57**, 1230-1237.
- Gosink, K.K., Mann, E.R., Guglielmo, C., Tuomanen, E.I. and Masure, H.R. (2000) Role of novel choline binding proteins in virulence of *Streptococcus pneumoniae*. *Infect Immun*, **68**, 5690-5695.
- Greco, A., Ho, J. G. S., Shuang-Jun Lin, Palcic, M. M., Rupnik, M., K-S Ng, K. (2006) Carbohydrate recognition by *Clostridium difficile* toxin A. *Nature Structural & Molecular Biology*, **13**, 460 - 461.
- Green, D., Ingram, V., Perutz, M.(1954). "The structure of haemoglobin IV. Sign of determination by isomorphous replacement method." *Proc. R. Soc. London SER. A* **225**, 287-307.
- Green, G.A., Schue, V. and Monteil, H. (1995) Cloning and characterization of the cytotoxin L-encoding gene of *Clostridium sordellii*: homology with *Clostridium difficile* cytotoxin B. *Gene*, **161**, 57-61.
- Griffith, F. (1928). The significance of pneumococcal types. *J. Hyg.* **27**: 113-159.
- Halfmann, A., Kovacs, M., Hakenbeck, R. and Bruckner, R. (2007) Identification of the genes directly controlled by the response regulator CiaR in *Streptococcus pneumoniae*: five out of 15 promoters drive expression of small non-coding RNAs. *Mol Microbiol*, **66**, 110-126.

- Hammerschmidt, S., Talay, S.R., Brandtzaeg, P. and Chhatwal, G.S. (1997) SpsA, a novel pneumococcal surface protein with specific binding to secretory immunoglobulin A and secretory component. *Mol Microbiol*, **25**, 1113-1124.
- Hammerschmidt, S., Bethe, G., Remane, P.H. and Chhatwal, G.S. (1999) Identification of pneumococcal surface protein A as a lactoferrin-binding protein of *Streptococcus pneumoniae*. *Infect Immun*, **67**, 1683-1687.
- Hendrickson, W.A., Teeter, M. M. (1981) Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. *Nature*, **290**, 107-113.
- Hendrickson, W.A., Horton, J.R. and LeMaster, D.M. (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *Embo J*, **9**, 1665-1672.
- Hendrickson, W.A., Ogata, C. M. (1997) *Methods Enzymology*, **276**, 494-523.
- Hendrickson, W. (1999). Maturation of MAD phasing for the determination of macromolecular structures. *J. Synchrotron Rad.* **6**, 845-851.
- Hermoso, J.A., Monterroso, B., Albert, A., Galan, B., Ahrazem, O., Garcia, P., Martinez-Ripoll, M., Garcia, J.L. and Menendez, M. (2003) Structural basis for selective recognition of pneumococcal cell wall by modular endolysin from phage Cp-1. *Structure*, **11**, 1239-1249.
- Hermoso, J.A., Lagartera, L., Gonzalez, A., Stelter, M., Garcia, P., Martinez-Ripoll, M., Garcia, J.L. and Menendez, M. (2005) Insights into pneumococcal pathogenesis from the crystal structure of the modular teichoic acid phosphorylcholine esterase Pce. *Nat Struct Mol Biol*, **12**, 533-538.
- Hermoso, J.A., Garcia, J.L. and Garcia, P. (2007) Taking aim on bacterial pathogens: from phage therapy to enzybiotics. *Curr Opin Microbiol*, **10**, 461-472.
- Hirschfield, G.M. and Pepys, M.B. (2003) C-reactive protein and cardiovascular disease: new insights from an old molecule. *Qjm*, **96**, 793-807.
- Holmes, A.R., McNab, R., Millsap, K.W., Rohde, M., Hammerschmidt, S., Mawdsley, J.L. and Jenkinson, H.F. (2001) The pavA gene of *Streptococcus pneumoniae* encodes a fibronectin-binding protein that is essential for virulence. *Mol Microbiol*, **41**, 1395-1408.
- Holtje, J.V. and Tomasz, A. (1974). Teichoic acid phosphorylcholine esterase. A novel enzyme activity in pneumococcus. *J Biol Chem*, **249**, 7032-7034.
- Höltje, J.V. and Tomasz, A. (1975a) Lipoteichoic acid: a specific inhibitor of autolysin activity in *Pneumococcus*. *Proc Natl Acad Sci U S A*, **72**, 1690-1694.
- Holtje, J.V. and Tomasz, A. (1975b) Lipoteichoic acid: a specific inhibitor of autolysin activity in *Pneumococcus*. *Proc Natl Acad Sci U S A*, **72**, 1690-1694.

- Holtje, J.V. and Tomasz, A. (1976) Purification of the pneumococcal N-acetylmuramyl-L-alanine amidase to biochemical homogeneity. *J Biol Chem*, **251**, 4199-4207.
- Horne, D. and Tomasz, A. (1985) Pneumococcal Forssman antigen: enrichment in mesosomal membranes and specific binding to the autolytic enzyme of *Streptococcus pneumoniae*. *J Bacteriol*, **161**, 18-24.
- Hoskins, J., Alborn, W.E., Jr., Arnold, J., Blaszczyk, L.C., Burgett, S., DeHoff, B.S., Estrem, S.T., Fritz, L., Fu, D.J., Fuller, W., Geringer, C., Gilmour, R., Glass, J.S., Khoja, H., Kraft, A.R., Lagace, R.E., LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P., McAhren, S.M., McHenney, M., McLeaster, K., Mundy, C.W., Nicas, T.I., Norris, F.H., O'Gara, M., Peery, R.B., Robertson, G.T., Rockey, P., Sun, P.M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C.A., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R., Jr., Skatrud, P.L. and Glass, J.I. (2001) Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J Bacteriol*, **183**, 5709-5717.
- Houldsworth, S., Andrew, P.W. and Mitchell, T.J. (1994) Pneumolysin stimulates production of tumor necrosis factor alpha and interleukin-1 beta by human mononuclear phagocytes. *Infect Immun*, **62**, 1501-1503.
- Jancarik, J., Kim, S-H. (1991) Sparse matrix sampling: a screening method for crystallization of proteins. *Journal Applied Crystallography*, **24**, 409-411.
- Janecek, S., Svensson, B. and Russell, R.R. (2000) Location of repeat elements in glucansucrases of *Leuconostoc* and *Streptococcus* species. *FEMS Microbiol Lett*, **192**, 53-57.
- Jedrzejewski, M.J. (2001) Pneumococcal virulence factors: structure and function. *Microbiol Mol Biol Rev*, **65**, 187-207.
- Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A*, **47 (Pt 2)**, 110-119.
- Jonsson, S., Musher, D.M., Chapman, A., Goree, A. and Lawrence, E.C. (1985) Phagocytosis and killing of common bacterial pathogens of the lung by human alveolar macrophages. *J Infect Dis*, **152**, 4-13.
- Kabsch, W. (1988) Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J. Appl. Cryst.*, **21**, 916-924.
- Kahn, R., Fourme, R., Bosshard, R., Chiadmi, M., Risler, J.L., Dideberg, O. and Wery, J.P. (1985) Crystal structure study of Opsanus tau parvalbumin by multiwavelength anomalous diffraction. *FEBS Lett*, **179**, 133-137.
- Kamerling, J.P. (2000). Pneumococcal polysaccharides. A chemical view. Molecular Biology and mechanisms of disease. En *Streptococcus pneumoniae*. (Tomasz, A., ed.), pp 81-84. Mary Ann Liebert Inc., Larchmont, NY.

- Karle, J. & Hauptman, H. (1956). A theory of phase determination for the four types of noncentrosymmetric space groups 1P222, 2P22, 3P12, 3P22. *Acta Cryst.* **9**: 635-651.
- Karlsson, C., Jansson, Per-Erik., Sorensen, U. B. S. (1999) The pneumococcal common antigen C-polysaccharide occurs in different forms. *European journal of biochemistry / FEBS*, **265**, 1091-1097.
- Katayama, S., Dupuy, B., Garnier, T. and Cole, S.T. (1995) Rapid expansion of the physical and genetic map of the chromosome of *Clostridium perfringens* CPN50. *J Bacteriol*, **177**, 5680-5685.
- Kausmally, L., Johnsborg, O., Lunde, M., Knutsen, E. and Havarstein, L.S. (2005) Choline-binding protein D (CbpD) in *Streptococcus pneumoniae* is essential for competence-induced cell lysis. *J Bacteriol*, **187**, 4338-4345.
- Kim, J.O. and Weiser, J.N. (1998) Association of intrastrain phase variation in quantity of capsular polysaccharide and teichoic acid with the virulence of *Streptococcus pneumoniae*. *J Infect Dis*, **177**, 368-377.
- Kobe, B. and Kajava, A.V. (2000) When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends Biochem Sci*, **25**, 509-515.
- La Fortelle, E.d.l., Bricogne, G. (1997) Maximum-Likelihood Heavy-Atom Parameter Refinement for Multiple Isomorphous Replacement And Multiwavelength Anomalous Diffraction Methods. *Methods in Enzymology*, **276**, 472-494.
- Laskowski, R.A., MacArthur, M. W., Moss, D. S., Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *Journal Applied Crystallography*, **26**, 283-291.
- Lazarevic, V. and Karamata, D. (1995) The tagGH operon of *Bacillus subtilis* 168 encodes a two-component ABC transporter involved in the metabolism of two wall teichoic acids. *Mol Microbiol*, **16**, 345-355.
- Leonard, G.A., Sainz, G., de Backer, M.M. and McSweeney, S. (2005) Automatic structure determination based on the single-wavelength anomalous diffraction technique away from an absorption edge. *Acta Crystallogr D Biol Crystallogr*, **61**, 388-396.
- Leslie, A.G.W. (1987). *Proceedings of the CCP4 study Weekend. Computational Aspects of Protein Crystal Data Analysis*, editado por J. R. Helliwell, P. A. Machin & M. Z. Papiz, pp. 39-50. Warrington: Daresbury Laboratory.
- Leslie, A., Duke, E., Brown, D., Naismith, J., Nave, C., Phillips, S., Stuart, D., Thompson, A., Walsh, M., Evans, G. (2004). "Diamond Beamline Proposal 044 - A Tuneable Beamline for Macromolecular Crystallography Optimised for Long Wavelengths (1.5-2.5Å)." Doc. No: BLS-044-PRO-0003.

Bibliografía

- Liu, Z.J., Vysotski, E.S., Chen, C.J., Rose, J.P., Lee, J. and Wang, B.C. (2000) Structure of the Ca²⁺-regulated photoprotein obelin at 1.7 Å resolution determined directly from its sulfur substructure. *Protein Sci*, **9**, 2085-2093.
- López, R., García, E., García, P., Ronda, C. and Tomasz, A. (1982) Choline-containing bacteriophage receptors in *Streptococcus pneumoniae*. *J Bacteriol*, **151**, 1581-1590.
- Lopez, R., Sanchez-Puelles, J.M., García, E., García, J.L., Ronda, C. and García, P. (1986) Isolation, characterization and physiological properties of an autolytic-deficient mutant of *Streptococcus pneumoniae*. *Mol Gen Genet*, **204**, 237-242.
- López, R., García, E., García, P., García, J. L. (2004) Cell wall hydrolases. In Tuomanen, E.I., Mitchell, T. J., Morrison, D. A., Spratt, B. G. (ed.), *The Pneumococcus*. ASM Press, Washington, DC, pp. 75-88.
- López, R., García, E. (2004) Recent trends on the molecular biology of pneumococcal capsules, lytic enzymes, and bacteriophage. *FEMS Microbiology Reviews*, **28**, 553-580.
- Luo, R., Mann, B., Lewis, W.S., Rowe, A., Heath, R., Stewart, M.L., Hamburger, A.E., Sivakolundu, S., Lacy, E.R., Bjorkman, P.J., Tuomanen, E. and Kriwacki, R.W. (2005) Solution structure of choline binding protein A, the major adhesin of *Streptococcus pneumoniae*. *Embo J*, **24**, 34-43.
- Lustbader, J.W., Wu, H., Birken, S., Pollack, S., Gawinowicz Kolks, M. A., Pound, A. M., Austen, D., Hendrickson, W. A., Canfield, R. E. . (1995) *Endocrinology*. **136**, 640-650.
- Machin, P. (1985). Editor. *Molecular Replacement*. Proceedings of the Daresbury Study Weekend, 15-16 Feb. 1985. SERC Daresbury Laboratory, Warrington, England.
- Malik, A. and Ahmad, S. (2007) Sequence and structural features of carbohydrate binding in proteins and assessment of predictability using a neural network. *BMC Struct Biol*, **7**, 1.
- Mann, B., Orihuela, C., Antikainen, J., Gao, G., Sublett, J., Korhonen, T.K. and Tuomanen, E. (2006) Multifunctional role of choline binding protein G in pneumococcal pathogenesis. *Infect Immun*, **74**, 821-829.
- Masui, Y., Mizuno, T., Inouye, M. (1984) Novel High-level Expression Cloning Vehicles: 104-fold Amplification of *Escherichia coli* Minor Protein. *Bio/Technology*, **2**, 81-85.
- Matthews, B.W. (1968) Solvent content of protein crystals. *J Mol Biol*, **33**, 491-497.
- McCullers, J.A. and Tuomanen, E.I. (2001) Molecular pathogenesis of pneumococcal pneumonia. *Front Biosci*, **6**, D877-889.
- Meroueh, S.O., Bencze, K.Z., Heseck, D., Lee, M., Fisher, J.F., Stemmler, T.L. and Mobashery, S. (2006) Three-dimensional structure of the bacterial cell wall peptidoglycan. *Proc Natl Acad Sci U S A*, **103**, 4404-4409.

- Micossi, E., Hunter, W.N. y Leonard, G.A. (2002) De novo phasing of two crystal forms of trypanothione II using the anomalous scattering from S atoms: a combination of small signal and medium resolution reveals this to be a general tool for solving protein crystal structures. *Acta Crystallogr D Biol Crystallogr*, **58**, 21-28.
- Mitchell, T.J. (2003) The pathogenesis of streptococcal infections: from tooth decay to meningitis. *Nat Rev Microbiol*, **1**, 219-230.
- Mold, C., Rodgers, C.P., Kaplan, R.L. and Gewurz, H. (1982) Binding of human C-reactive protein to bacteria. *Infect Immun*, **38**, 392-395.
- Molina, R., Gonzalez, A., Moscoso, M., Garcia, P., Stelter, M., Kahn, R. and Hermoso, J.A. (2007) Crystallization and preliminary X-ray diffraction studies of choline-binding protein F from *Streptococcus pneumoniae*. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, **63**, 742-745.
- Molina, R., Gonzalez, A., Stelter, M., Perez-Dorado, I., Kahn, R., Morales, M., Campuzano, S., Campillo, N.E., Mobashery, S., Garcia, J.L., Garcia, P. and Hermoso, J.A. (2009) Crystal structure of CbpF, a bifunctional choline-binding protein and autolysis regulator from *Streptococcus pneumoniae*. *EMBO Rep*, **10**, 246-251.
- Molina, R., A., Stelter, M., Kahn, R., and Hermoso, J.A. (En prensa) Characterization of Gadolinium Complexes for SAD Phasing in Macromolecular Crystallography: Application to CbpF. *Acta Crystallogr D Biol Crystallogr*, **65**.
- Müller, J., Heinemann, U. (2005). "Use of anomalous scattering for solving the phase problem in protein crystallography." dans : *Encyclopedic Reference of Genomics and Proteomics* (Oschkinat, H. & Schmieder, P. eds) Springer, Heidelberg.
- Musher, D.M. (1992) Infections caused by *Streptococcus pneumoniae*: clinical spectrum, pathogenesis, immunity, and treatment. *Clin Infect Dis*, **14**, 801-807.
- Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*, **53**, 240-255.
- Navaza, J. (1994) A MoRe: an Automated Package for Molecular Replacement. *Acta Crystallographica Section A*, **50**, 157-163.
- Neeleman, C., Geelen, S.P., Aerts, P.C., Daha, M.R., Mollnes, T.E., Roord, J.J., Posthuma, G., van Dijk, H. and Fleer, A. (1999) Resistance to both complement activation and phagocytosis in type 3 pneumococci is mediated by the binding of complement regulatory protein factor H. *Infect Immun*, **67**, 4517-4524.
- Novak, R. and Tuomanen, E. (1999) Pathogenesis of pneumococcal pneumonia. *Semin Respir Infect*, **14**, 209-217.

Bibliografía

- Oggioni, M.R., Memmi, G., Maggi, T., Chiavolini, D., Iannelli, F. and Pozzi, G. (2003) Pneumococcal zinc metalloproteinase ZmpC cleaves human matrix metalloproteinase 9 and is a virulence factor in experimental pneumonia. *Mol Microbiol*, **49**, 795-805.
- OMS. (2007). Pneumococcal conjugate vaccine for childhood immunization – WHO position paper. *Geneva: World Health Organization*.
- Paton, J.C. and Ferrante, A. (1983) Inhibition of human polymorphonuclear leukocyte respiratory burst, bactericidal activity, and migration by pneumolysin. *Infect Immun*, **41**, 1212-1216.
- Paton, J.C., y Morona, J.K. (2000). *Streptococcus pneumoniae* capsular polysaccharide. In: Gram-positive pathogens. (Fischetti, V.A., Novick, R.P., Ferreti, J.J., Portnoy, D.A., y Rood, J.I., eds.) ASM press, Washington D.C.
- Pelzer, H. (1983). En The target of Penicillin. (Hakenbeck, R., Höltje, J.V., y Labischinski. H., eds.), pp 105-111. Berlin: W. de Gruyter.
- Pepis, M.B. and Hirschfield, G.M. (2003) C-reactive protein: a critical update. *J Clin Invest*, **111**, 1805-1812.
- Port, M., Idee, J.M., Medina, C., Robic, C., Sabatou, M. and Corot, C. (2008) Efficiency, thermodynamic and kinetic stability of marketed gadolinium chelates and their possible clinical consequences: a critical review. *Biomaterials*, **21**, 469-490.
- Purdy, M.D., Ge, P., Chen, J., Selvin, P.R. and Wiener, M.C. (2002) Thiol-reactive lanthanide chelates for phasing protein X-ray diffraction data. *Acta Crystallogr D Biol Crystallogr*, **58**, 1111-1117.
- Quillin, M.L. and Matthews, B.W. (2002) Generation of noble-gas binding sites for crystallographic phasing using site-directed mutagenesis. *Acta Crystallogr D Biol Crystallogr*, **58**, 97-103.
- Ramagopal, U.A., Dauter, M. and Dauter, Z. (2003b) Phasing on anomalous signal of sulfurs: what is the limit? *Acta Crystallogr D Biol Crystallogr*, **59**, 1020-1027.
- Riesenfeld-Orn, I., Wolpe, S., Garcia-Bustos, J.F., Hoffmann, M.K. and Tuomanen, E. (1989) Production of interleukin-1 but not tumor necrosis factor by human monocytes stimulated with pneumococcal cell surface components. *Infect Immun*, **57**, 1890-1893.
- Rogers, H. J., y Thurman, P. (1985). The effect of celerenin on the morphogenesis and autolytic activity of *Bacillus subtilis*. *J. Gen. Microbiol.* **131**: 591-599.
- Rosenow, C., Ryan, P., Weiser, J.N., Johnson, S., Fontan, P., Ortqvist, A. and Masure, H.R. (1997) Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of *Streptococcus pneumoniae*. *Mol Microbiol*, **25**, 819-829.
- Rossmann, M. (1972). Editor. *The Molecular Replacement Method*. New York: Gordon and Breach.

- Rubins, J.B., Charboneau, D., Fasching, C., Berry, A.M., Paton, J.C., Alexander, J.E., Andrew, P.W., Mitchell, T.J. and Janoff, E.N. (1996) Distinct roles for pneumolysin's cytotoxic and complement activities in the pathogenesis of pneumococcal pneumonia. *Am J Respir Crit Care Med*, **153**, 1339-1346.
- Salton, M.R.J. (1994). The bacterial cell envelope. A historical perspective. En Bacterial cell wall. (Ghuysen, J.M., y Hakenbeck, R., eds.) pp 1-22. Elsevier Science Publishing BV, Amsterdam, The Netherlands.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. , 18.82-18.84. (1989) *Molecular Cloning: A Laboratory Manual*.
- Sánchez-Beato, A. R. (1995a). Tesis Doctoral. Caracterización molecular de genes de *Clostridium acetobutylicum* y *Streptococcus pneumoniae* que codifican para proteínas con dominios de unión a colina. En: *Microbiología II*. Universidad Complutense, Madrid.
- Sánchez-Beato, A.R., Ronda, C. and Garcia, J.L. (1995b) Tracking the evolution of the bacterial choline-binding domain: molecular characterization of the *Clostridium acetobutylicum* NCIB 8052 cspA gene. *J Bacteriol*, **177**, 1098-1103.
- Sanchez-Beato, A.R., Lopez, R. and Garcia, J.L. (1998) Molecular characterization of PcpA: a novel choline-binding protein of *Streptococcus pneumoniae*. *FEMS Microbiol Lett*, **164**, 207-214.
- Sanchez-Puelles, J.M., Sanz, J.M., Garcia, J.L. and Garcia, E. (1990) Cloning and expression of gene fragments encoding the choline-binding domain of pneumococcal murein hydrolases. *Gene*, **89**, 69-75.
- Sayre, D. (1952) The squaring method: a new method for phase determination. *Acta Cryst*, **5**: 60-65.
- Sanz, J.M., Diaz, E. and Garcia, J.L. (1992) Studies on the structure and function of the N-terminal domain of the pneumococcal murein hydrolases. *Mol Microbiol*, **6**, 921-931.
- Schiltz, M., Prangé, T., Fourme, R. (1994) On the preparation and X-ray data collection of isomorphous xenon derivatives. *J. Appl. Cryst.*, **27**, 950-960.
- Schiltz, M., Shepard, W., Fourme, R., Prange, T., de la Fortelle, E. and Bricogne, G. (1997) High-pressure krypton gas and statistical heavy-atom refinement: a successful combination of tools for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*, **53**, 78-92.
- Schiltz, M., Fourme, R., Prangé, T. (2003). "Use of noble gases xenon and krypton as heavy atoms in protein structure determination." *Methods Enzymol*. **374**, 83-119.
- Schleifer, K.H. and Kandler, O. (1972) Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol Rev*, **36**, 407-477.

Bibliografía

- Schneider, T.R. and Sheldrick, G.M. (2002) Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr*, **58**, 1772-1779.
- Severin, A. and Tomasz, A. (1996) Naturally occurring peptidoglycan variants of *Streptococcus pneumoniae*. *J Bacteriol*, **178**, 168-174.
- Severin, A., Horne, D. and Tomasz, A. (1997) Autolysis and cell wall degradation in a choline-independent strain of *Streptococcus pneumoniae*. *Microb Drug Resist*, **3**, 391-400.
- Shankar, K. and Bard, R.C. (1952) The effect of metallic ions on the growth and morphology of *Clostridium perfringens*. *J Bacteriol*, **63**, 279-290.
- Shann, F. (1990) Modern vaccines. Pneumococcus and influenza. *Lancet*, **335**, 898-901.
- Sheldrick, G.M. (1998) *Directs Methods for Solving Macromolecular Structures*, editado por S. Fortier, pp. 401-411. Kluwer Academic Publishers.
- Skov Sørensen, U.B., Blom, J., Birch-Andersen, A. and Henrichsen, J. (1988) Ultrastructural localization of capsules, cell wall polysaccharide, cell wall proteins, and F antigen in pneumococci. *Infect Immun*, **56**, 1890-1896.
- Smith, B.L. and Hostetter, M.K. (2000) C3 as substrate for adhesion of *Streptococcus pneumoniae*. *J Infect Dis*, **182**, 497-508.
- Seo, H.S., Cartee, R.T., Pritchard, D.G. and Nahm, M.H. (2008) A new model of pneumococcal lipoteichoic acid structure resolves biochemical, biosynthetic, and serologic inconsistencies of the current model. *J Bacteriol*, **190**, 2379-2387.
- Sun, P.D. and Hammer, C.H. (2000) Mass-spectrometry assisted heavy-atom derivative screening of human Fc gamma RIII crystals. *Acta Crystallogr D Biol Crystallogr*, **56**, 161-168.
- Sun, P.D., Radaev, S., Kattah, M. (2002) Generating isomorphous heavy-atom derivatives by a quick-soak method. Part I: test cases. *Acta Crystallographica Section D*, **58**, 1092-1098.
- Sun, P.D., Radaev, S. (2002) Generating isomorphous heavy-atom derivatives by a quick-soak method. Part II: phasing of new structures. *Acta Crystallographica Section D*, **58**, 1099-1103.
- Sun, P.D. and Hammer, C.H. (2000) Mass-spectrometry assisted heavy-atom derivative screening of human Fc gamma RIII crystals. *Acta Crystallogr D Biol Crystallogr*, **56**, 161-168.
- Swiatlo, E., Champlin, F.R., Holman, S.C., Wilson, W.W. and Watt, J.M. (2002) Contribution of choline-binding proteins to cell surface properties of *Streptococcus pneumoniae*. *Infect Immun*, **70**, 412-415.

- Szalai, A.J., Briles, D.E. and Volanakis, J.E. (1996) Role of complement in C-reactive-protein-mediated protection of mice from *Streptococcus pneumoniae*. *Infect Immun*, **64**, 4850-4853.
- Teeter, M.M., Mazer, J.A. and L'Italien, J.J. (1981) Primary structure of the hydrophobic plant protein crambin. *Biochemistry*, **20**, 5437-5443.
- Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Umayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T., Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and Fraser, C.M. (2001) Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science*, **293**, 498-506.
- Tilled, W.S., Goebel, W.F., y Avery, O.T. (1930). Chemical and immunological properties of a species-specific carbohydrate of pneumococci. *J. Exp. Med.* **52**: 895-900.
- Tilley, S.J., Orlova, E.V., Gilbert, R.J., Andrew, P.W. and Saibil, H.R. (2005) Structural basis of pore formation by the bacterial toxin pneumolysin. *Cell*, **121**, 247-256.
- Tomasz, A. (1967) Choline in the cell wall of a bacterium: novel type of polymer-linked choline in *Pneumococcus*. *Science*, **157**, 694-697.
- Tomasz, A. (1968) Biological consequences of the replacement of choline by ethanolamine in the cell wall of *Pneumococcus*: chanin formation, loss of transformability, and loss of autolysis. *Proc Natl Acad Sci U S A*, **59**, 86-93.
- Tomasz, A. (1970) Cellular metabolism in genetic transformation of pneumococci: requirement for protein synthesis during induction of competence. *J Bacteriol*, **101**, 860-871.
- Tomasz, A., Westphal, M., Briles, E.B. and Fletcher, P. (1975) On the physiological functions of teichoic acids. *J Supramol Struct*, **3**, 1-16.
- Tseng, H.J., McEwan, A.G., Paton, J.C. and Jennings, M.P. (2002) Virulence of *Streptococcus pneumoniae*: PsaA mutants are hypersensitive to oxidative stress. *Infect Immun*, **70**, 1635-1639.
- Tsuchido, T., Hiraoka, T., Takano, M. and Shibasaki, I. (1985) Involvement of autolysin in cellular lysis of *Bacillus subtilis* induced by short- and medium-chain fatty acids. *J Bacteriol*, **162**, 42-46.
- Tu, A.H., Fulgham, R.L., McCrory, M.A., Briles, D.E. and Szalai, A.J. (1999) Pneumococcal surface protein A inhibits complement activation by *Streptococcus pneumoniae*. *Infect Immun*, **67**, 4720-4724.

Bibliografía

- Tuomanen, E., Tomasz, A., Hengstler, B. and Zak, O. (1985a) The relative role of bacterial cell wall and capsule in the induction of inflammation in pneumococcal meningitis. *J Infect Dis*, **151**, 535-540.
- Tuomanen, E., Liu, H., Hengstler, B., Zak, O. and Tomasz, A. (1985b) The induction of meningeal inflammation by components of the pneumococcal cell wall. *J Infect Dis*, **151**, 859-868.
- Tuomanen, E., Rich, R. and Zak, O. (1987) Induction of pulmonary inflammation by components of the pneumococcal cell surface. *Am Rev Respir Dis*, **135**, 869-874.
- Umeda, A., Yokoyama, S., Arizono, T. and Amako, K. (1992) Location of peptidoglycan and teichoic acid on the cell wall surface of *Staphylococcus aureus* as determined by immunoelectron microscopy. *J Electron Microsc (Tokyo)*, **41**, 46-52.
- Usobiaga, P., Medrano, F.J., Gasset, M., Garcia, J.L., Saiz, J.L., Rivas, G., Laynez, J. and Menendez, M. (1996) Structural organization of the major autolysin from *Streptococcus pneumoniae*. *J Biol Chem*, **271**, 6832-6838.
- Usón, I. and Sheldrick, G.M. (1999) Advances in direct methods for protein crystallography. *Curr Opin Struct Biol*, **9**, 643-648.
- Varea, J., Saiz, J.L., Lopez-Zumel, C., Monterroso, B., Medrano, F.J., Arrondo, J.L., Iloro, I., Laynez, J., Garcia, J.L. and Menendez, M. (2000) Do sequence repeats play an equivalent role in the choline-binding module of pneumococcal LytA amidase? *J Biol Chem*, **275**, 26842-26855.
- Vitali, J., Robbins, A. H., Almo, S. C., Tilton, R. F. (1991) Using Xenon as a Heavy Atom for Determining Phases in Sperm Whale Metmyoglobin. *Journal Applied Crystallography*, **24**, 931-935.
- Vollmer, W. and Tomasz, A. (2000) The *pgdA* gene encodes for a peptidoglycan N-acetylglucosamine deacetylase in *Streptococcus pneumoniae*. *J Biol Chem*, **275**, 20496-20501.
- Wang, B.C. (1985) *Methods Enzimology*, **115**, 90-111.
- Webb, M. (1951) The influence of magnesium on cell division. The effect of magnesium on the growth of bacteria in chemically-defined media of varying complexity. *J Gen Microbiol*, **5**, 485-495.
- Weiser, J.N., Austrian, R., Sreenivasan, P.K., y Masure, H.R. (1994). Phase variation in pneumococcal opacity: relationship between colonial morphology and nasopharyngeal colonization. *Infect. Immun.* **62**: 2582-2589.
- Weissborn, A.C., Liu, Q., Rumley, M.K. and Kennedy, E.P. (1994) UTP: alpha-D-glucose-1-phosphate uridylyltransferase of *Escherichia coli*: isolation and DNA sequence of the *galU* gene and purification of the enzyme. *J Bacteriol*, **176**, 2611-2618.

-
- Wren, B.W. (1991) A family of clostridial and streptococcal ligand-binding proteins with conserved C-terminal repeat sequences. *Mol Microbiol*, **5**, 797-803.
- Yeung, M.K., y Mattingly, S.J. (1986). Covalent linkage of the type- and group- specific antigens to the peptide moiety of the peptidoglycan of serotype III group B *Streptococcus*. *Curr. Microbiol.* **14**: 205-211.
- Yother, J., Handsome, G. L., Briles, D. E. (1992) Truncated Forms of PspA That Are Secreted from *Streptococcus pneumoniae* and Their Use in Functional Studies and Cloning of the *pspA* Gene. *Journal of Bacteriology*, **174**, 610-618.
- Yother, J., White, J. M. (1994) Novel surface attachment mechanism of the *Streptococcus pneumoniae* protein PspA. *Journal of Bacteriology*, **176**, 2976-2985.
- Yother, J., Leopold, K., White, J. and Fischer, W. (1998) Generation and properties of a *Streptococcus pneumoniae* mutant which does not require choline or analogs for growth. *J Bacteriol*, **180**, 2093-2101.
- Yother, J. (2004). Capsules. En *The pneumococcus*. (Tuomanen, E., Mitchell, T., Morrison, D.A., y Spratt, B.G., eds.), pp. 30-48. Washington: ASM Press.
- Zhang, J.R., Mostov, K.E., Lamm, M.E., Nanno, M., Shimida, S., Ohwaki, M. and Tuomanen, E. (2000) The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. *Cell*, **102**, 827-837.

A 3D molecular model of a protein structure. The protein backbone is shown as a blue ribbon, with several alpha-helices and beta-strands. A gold ribbon structure is also visible, possibly representing a different conformation or a different part of the protein. Several clusters of atoms are shown as spheres: purple spheres and one red sphere. The entire structure is set against a light blue, semi-transparent surface that represents the protein's surface or a specific binding pocket. The text "6. ARTÍCULOS" is located in the bottom right corner of the image.

6. ARTÍCULOS

Crystallization and preliminary X-ray diffraction studies of choline-binding protein F from *Streptococcus pneumoniae*.

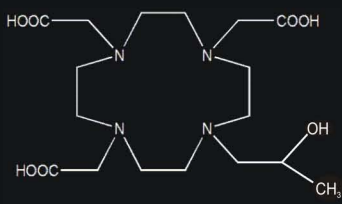
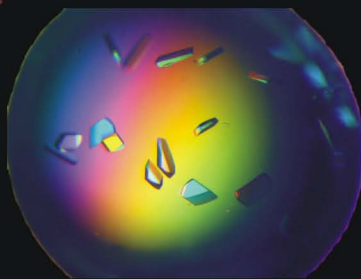
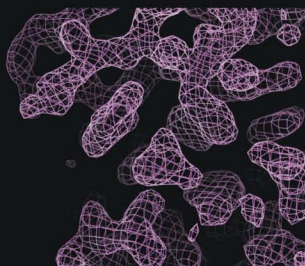
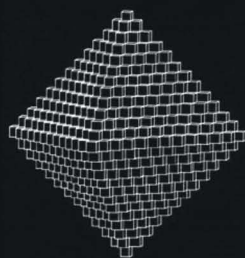
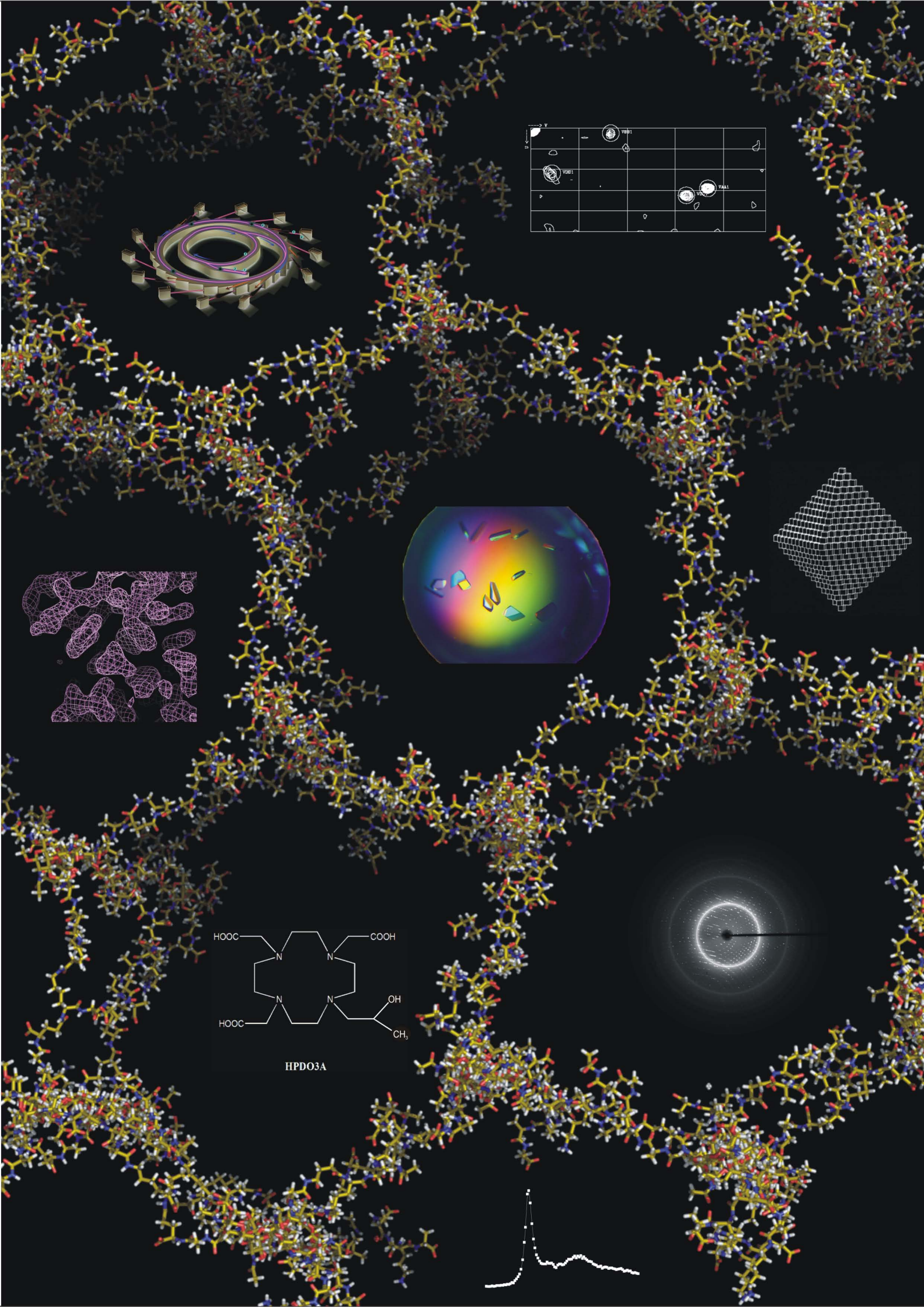
Molina, R., Gonzalez, A., Moscoso, M., Garcia, P., Stelter, M., Kahn, R. and Hermoso, J.A. (2007). *Acta Crystallogr Sect F Struct Biol Cryst Commun*, **63**, 742-745.

Crystal structure of CbpF, a bifunctional choline-binding protein and autolysis regulator from *Streptococcus pneumoniae*.

Molina, R., Gonzalez, A., Stelter, M., Perez-Dorado, I., Kahn, R., Morales, M., Campuzano, S., Campillo, N.E., Mobashery, S., Garcia, J.L., Garcia, P. and Hermoso, J.A. (2009). *EMBO Rep*, **10**, 246-251.

Characterization of Gadolinium Complexes for SAD Phasing in Macromolecular Crystallography: Application to CbpF.

Molina, R., A., Stelter, M., Kahn, R., and Hermoso, J.A. (2009). *Acta Crystallogr D Biol Crystallogr*, **65**, 823-831.



HPDO3A

