



UNIVERSITAT AUTÒNOMA DE BARCELONA
DEPARTAMENT DE FÍSICA

Estructura Tridimensional del Connector del Bacteriòfag $\Phi 29$

Joan Pous i Ramos
IBMB-CSIC Barcelona

MEMÒRIA PRESENTADA PER OPTAR AL
TÍTOL DE DOCTOR EN CIÈNCIES FÍSiques

8 de febrer de 2002

UNIVERSITAT AUTÒNOMA DE BARCELONA
DEPARTAMENT DE FÍSICA

Estructura Tridimensional del Connector del Bacteriòfag $\Phi 29$

MEMÒRIA PRESENTADA PER
JOAN POUS I RAMOS,
IBMB-CSIC BARCELONA
PER OPTAR AL TÍTOL DE
DOCTOR EN CIÈNCIES FÍSiques

DIRECTOR
Dr. Miquel Coll Capella

TUTOR
Dr. Fernando López Aguilar

8 de febrer de 2002

A Search for Structure

If we want to understand the human body we must know its structure, in terms of the cells that make it up. If we want to understand cells we must know their structure, in terms of molecules.

If we want to know molecules, we must know their structure, in terms of atoms. And to understand atoms we must know their structure, in terms of electrons and nuclei... To understand nuclei, we must learn their structure in terms of protons and neutrons....

Linus Pauling

Presentació

En aquest treball es descriu l'estructura afinada del connector del bacteriòfag $\Phi 29$, resolta per cristal·lografia de raigs X i es proposa un mecanisme de funcionament d'aquesta proteïna.

Els bacteriòfags són virus que ataquen els bacteris i el connector és una proteïna de grans dimensions situada en un vèrtex de la càpsida d'aquests virus que juga un paper central en la formació del bacteriòfag, en l'empaquetament del seu DNA i en la infecció dels bacteris.

El connector és part d'una bomba que transporta DNA. Té un mecanisme basat en asimetries dels components d'aquesta bomba que faciliten la rotació del connector amb un consum molt baix d'energia, i a la vegada una gran eficiència en el seu propòsit. La seva estructura no s'assembla a cap altra proteïna coneguda de funcions similars i, per tant en aquest treball no només es presenta una estructura important, per mida i funció, sinó que s'aporten idees noves de com s'estructuren les proteïnes per a complir les seves funcions.

Si volem entendre els virus, haurem de conèixer la seva estructura en termes dels àtoms i de les molècules que els formen.

Organització d'aquest treball

Aquest treball està dividit en les següents parts:

- **Introducció: El connector del fag $\Phi 29$** [plana 3] Descripció general dels estudis que s'han fet sobre l'estructura i mecanisme de virus bacteriòfags, enfocat al tema de la seva maduració, empaquetament del DNA i infecció de les cèl·lules, i al paper que té el connector en tots aquests processos.
- **Objectius** [plana 17] Objectius proposats alhora d'afrontar aquest treball.
- **Materials i Mètodes: La Cristal·lografia de proteïnes** [plana 21] Descripció dels mètodes utilitzats per a resoldre l'estructura. En aquest treball es va intentar afrontar la resolució de l'estructura de diverses maneres. No totes van resultar favorables, per la qual cosa s'ha fet una descripció de com es va intentar, junt amb els resultats que es van obtenir. Altres estan encara en desenvolupament i es descriu l'estat actual en aquest capítol.
- **Resultats:** [plana 75] Articles publicats i acceptats. Altres treballs en preparació no s'han inclòs aquí ja que es pretén continuar desenvolupant el tema en un futur. Aquests resultats preliminars estan descrits a Materials i Mètodes.

Aquesta tesi està basada en els següents treballs:

- Purification, crystallization and preliminary X-ray diffraction studies of the bacteriophage $\phi 29$ connector particle.
A. Guasch, A. Párraga, J. Pous, J.M. Valpuesta, J.L. Carrascosa, and M. Coll.
FEBS Lett, 430:283–287, 1998. [plana 77]
- Crystallographic analysis reveals the 12-fold symmetry of the bacteriophage $\phi 29$ connector particle.
A. Guasch, J. Pous, A. Párraga, J.M. Valpuesta, J.L. Carrascosa, and M. Coll.
J Mol Biol, 281(2):219–225, 1998. [plana 83]
- Detailed architecture of a DNA translocating machine: The high-resolution structure of the bacteriophage $\Phi 29$ connector particle.
J. Pous, A. Guasch, B. Ibarra, F.X. Gomis-Rüth, J.M. Valpuesta,

N. Sousa, J.L. Carrascosa and M. Coll.
J Mol Biol 315, 663–676, 2002. [plana 91]

- **Discussió** [plana 107] Resum de de l'estructura i mecanisme proposat per l'empaquetament del DNA pel connector. Discussió de les possibilitats d'aquest model de mecanisme respecte altres proposats.
- **Conclusions** [plana 115] Conclusions que responen als objectius plantejats al començar el treball i a altres que s'han anat plantejant durant el procés de resolució de l'estructura i d'interpretació del mecanisme. Conclusions independents de cada article.

Índex

Presentació	i
Organització d'aquest treball	iii
I Introducció	1
1 El connector del fag Φ29	3
1.1 El fag Φ 29	3
1.2 El connector i la proteïna p10	5
1.2.1 Estructura del connector	5
1.2.2 El connector en la formació del fag	9
1.2.3 La infecció	12
1.3 Les simetries: cinc, sis, deu i dotze (o tretze)	12
2 Objectius	17
II Materials i Mètodes	19
3 La Cristal·lografia de proteïnes	21
3.1 Definicions	21
3.1.1 Els cristalls	21
3.1.2 Els cristalls de proteïna	21
3.2 Obtenció de la proteïna, dels cristalls i dels derivats	24
3.2.1 Obtenció de la proteïna p10	24
3.2.2 Els cristalls	24
3.2.3 Els derivats	25
3.3 Raigs X: generació, òptica, interacció amb els cristalls i detecció	26
3.3.1 Generació	26
3.3.2 Òptica	27
3.3.3 Interacció amb els cristalls	28
3.3.4 Detectores	30

3.4	L'experiment	31
3.4.1	Congelació i muntatge dels cristalls	31
3.4.2	Recollida i processament de les dades	33
3.5	Solució de l'estructura tridimensional	42
3.5.1	MIR i MAD	42
3.5.2	Reemplaçament Molecular	46
3.5.3	NCS	49
3.5.4	Màscares	50
3.5.5	L'extensió de fases	51
3.5.6	El traçat de l'estructura	54
3.6	Estructura final de la proteïna	55
3.6.1	L'afinament de l'estructura	55
3.6.2	Màxim de versemblança i Mínims quadrats	60
3.7	Macles	62
3.7.1	Macles hemièdriques	63
3.7.2	Procediment pràctic	66
3.8	Microscòpia electrònica	69
3.8.1	Reconstrucció de la M.E.	69
3.8.2	Ajust de la microscòpia i del model afinat	70
3.9	El PDB	72
III Resultats		73
4	Resultats	75
Purification, crystallization and preliminary X-ray diffraction studies of the bacteriophage Φ29 connector particle.		77
Crystallographic analysis reveals the 12-fold symmetry of the bacteriophage Φ29 connector particle.		83
Detailed architecture of a DNA translocating machine: The high-resolution structure of the bacteriophage Φ29 connector particle.		91
IV Discussió i Conclusions		105
5	Discussió	107
5.1	Estructura del connector	107
5.2	Mecanisme del connector	110

6	Conclusions	115
6.1	Conclusions globals	115
6.2	Conclusions dels articles	117
	Referències	119

Índex de figures

1.1	Primera micrografia de fags	3
1.2	Bacteriòfag T4	4
1.3	Perspectives del coll del fag $\Phi 29$	6
1.4	Cristall bidimensional	6
1.5	Representació tridimensional del connector a 9Å	6
1.6	Encaix del connector en la càpsida	7
1.7	El connector amb anticossos	8
1.8	DNA superenrotllat al connector	9
1.9	Connectors visualitzats amb AFM	9
1.10	Model d'AFM del cristall 2D	10
1.11	Acoblament del fag	10
1.12	Relació entre l'energia d'interacció i el nombre de subunitats	14
1.13	Simetria del DNA	14
3.1	Simetries dels cristalls	22
3.2	El codi genètic	23
3.3	Obtenció de la proteïna	24
3.4	Efecte dels raigs X sobre els cristalls	32
3.5	Wilson Plot	35
3.6	Fluorescència TaBr	44
3.7	Factors de difusió atòmics	45
3.8	Extensió de fases I	52
3.9	Extensió de fases II	53
3.10	Molècula del criotampó	55
3.11	Angles peptídics	57
3.12	Hèlix α i fulla β	58
3.13	Ramachandran del connector	59
3.14	Difracció amb microfocus	62
3.15	Difracció d'una macla	63
3.16	Distribució acumulada	64
3.17	Mapa de Patterson	65
3.18	Eixos de la macla	67

3.19	Macla reconstruïda	68
3.20	Seccions M.E.	70
3.21	Reconstrucció de la M.E.	70
3.22	Ajust de la microscòpia i del model afinat	71
5.1	Monòmer	108
5.2	Secció lateral	109
5.3	Un model per a l'empaquetament	111
5.4	El nostre model per a l'empaquetament	112

Índex de taules

1.1	Bacteriòfags	4
3.1	Aminoàcids i Bases	23
3.2	Llista dels cristalls analitzats	26
3.3	Línies de difracció	28
3.4	Capes de reflexions	34
3.5	Recollida i processat dels cristalls de I422	36
3.6	Recollida i processat dels cristalls de I422Ag	37
3.7	Recollida i processat dels cristalls de I422Au	38
3.8	Recollida i processat dels cristalls de I422Pt	39
3.9	Recollida i processat dels cristalls de I422TaBr	40
3.10	Reemplaçament Molecular	48

Part I
Introducció

Capítol 1

El connector del fag $\Phi 29$

1.1 El fag $\Phi 29$

Bacteriòfag i fag són els noms que es donen als virus que ataquen els bacteris. Van ser descoberts a principis del segle XX quasi simultàniament per Twort (1915) i per d'Herelle (1917) [7]. La simplicitat del sistema i la seva manipulació fàcil va fer dels fags un dels primers objectius a estudiar als inicis de la Biologia Molecular, als anys 40 per Max Delbrück i Salvador Luria. Se sospitava que els fags eren una mena de *gens despallats*. Així, la millor manera de conèixer com era un gen i com es duplicava seria l'estudi de les propietats d'aquests virus [57]. Simultàniament van començar els esforços per a visualitzar la seva estructura. La primera imatge de microscòpia electrònica, obtinguda per Ernst i Helmut Ruska, va néixer amb els primers microscopis electrònics comercials el 1940 (fig. 1.1) [61].

Figura 1.1: Primera micrografia de fags. El 1940, H. Ruska va publicar la primera micrografia d'un bacteri infectat per bacteriòfags. Aquesta va ser presa a una magnificació de $15000\times$ amb el *Siemens-Halske Übermikroskop*, construït per E. Ruska i col·laboradors (magnificació màxima de $30000\times$ i 13 nm de resolució) instal·lat en el laboratori de la fàbrica a Berlín. L'autor menciona que les partícules esfèriques observades al voltant del bacteri es destrueixen després d'una exposició allargada al feix d'electrons [61].



Com que els bacteris no tenen una membrana citoesquelètica tan sofisticada com la dels eucariotes, els seus virus no poden entrar per endocitosi.

Necessiten un sistema més complex per introduir-hi el DNA [38]. A grans trets, un fag està format de dues grans parts: una càpsida que conté els seus gens i un gran apèndix de forma més o menys tubular que s'encarrega de conduir el DNA dins la cèl·lula en la infecció (fig. 1.2).

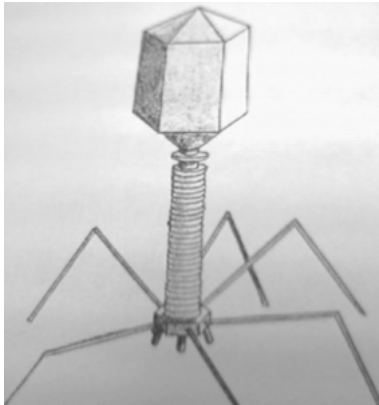


Figura 1.2: Bacteriòfag T4. Esquema del bacteriòfag T4, semblant al $\Phi 29$. Es poden veure les parts principals: la càpsida, la cua, la base i les fibres de la base

El fag $\Phi 29$ és un virus de DNA de doble cadena que infecta *Bacillus subtilis*. Conté uns 20 gens en una càpsida de forma icosaèdrica allargada (*prolate*) al voltant d'un eix 5 [61]. Aquesta càpsida té uns 540Å de llargada i uns 420Å d'amplada. En un dels extrems d'aquest eix es troba el connector (o portal), d'uns 80Å de llargada, a continuació la cua, d'uns 500Å, i la resta del mecanisme d'infecció. Tot i ser un virus relativament petit, té la mateixa estructura que altres de més grans, com el λ i el T4. Les diferències estan principalment en el nombre de proteïnes que formen la càpsida i la cua, però la forma global es conserva. Aquestes proteïnes s'ordenen idènticament, simplement formant cossos de mida diferent [38], per repetició. El $\Phi 29$ és, per tant, un bon model per conèixer l'estructura i mecanisme d'una família de virus similars molt extensa (taula 1.1).

Nom	bacteri	Bases (Kpb)
T2, T4, T6	<i>Escherichia coli</i>	170
T7	<i>Escherichia coli</i>	40
λ	<i>Escherichia coli</i>	48.5
$\Phi 29$	<i>Bacillus subtilis</i>	19
P22	<i>Salmonella typhimurium</i>	43.5
T3	<i>Escherichia coli</i>	40
SPP1	<i>Bacillus subtilis</i>	46
μ	<i>Escherichia coli</i>	39

Taula 1.1: Bacteriòfags. Alguns dels bacteriòfags similars al $\Phi 29$ [37, 38, 61, 55, 17].

1.2 El connector i la proteïna p10

El bacteriòfag $\Phi 29$ està compost de només 6 proteïnes estructurals. Dues d'aquestes proteïnes formen el coll, la p10 i la p11 [8]. La proteïna p12 forma uns apèndixs al coll, i la p9 la cua [30].

El connector del fag $\Phi 29$ (*portal protein, head-tail connector, upper collar*) està format per 12 còpies de la proteïna p10 de 36 kDa cadascuna [25, 3]. Està situat en un dels dos vèrtexs d'ordre 5 que hi ha als extrems allargats de la càpsida. Forma part d'un motor rotatiu que empaqueta el DNA genòmic, de doble cadena.

Aquest motor està format pel connector, un RNA de 174 bases i una altra proteïna (p11), aquests dos últims amb simetria 6 segons les reconstruccions de criomicroscòpia [47].

Les 12 proteïnes es disposen en forma troncocònica amb un canal d'un diàmetre suficient per tal que passi DNA de doble cadena, que el travessa al llarg del seu eix.

Tots els bacteriòfags disposen de connector, amb simetries no totalment determinades, però sempre 12 o 13 [17, 10]. Pel que se sap fins ara, usant criomicroscòpia, tots tenen unes dimensions globals similars, fet consistent amb que tots tenen la mateixa funció de conduir el DNA.

1.2.1 Estructura del connector

Els estudis de l'estructura tridimensional del connector es van començar amb difracció d'electrons de cristalls bidimensionals [11]. En aquests estudis del coll sencer (proteïnes p10 i p11) s'observava ja la seva forma de tronc de con amb un canal central. La part més ampla estava composta de 12 lòbuls. A partir d'aquests lòbuls l'estructura de cada subunitat s'estretava, com un con, aparellant-se cadascuna amb la veïna, alhora que es recargolava cap al cim de la següent subunitat. A la part estreta, doncs, s'observaven només 6 subunitats, probablement en la zona corresponent a la proteïna p11 (fig. 1.3) [9].

Els millors estudis de l'estructura del connector s'han fet a partir de cristalls bidimensionals [54]. Les tècniques usades eren la criomicroscòpia [54, 53] o bé la microscòpia de força atòmica [39].

La resolució útil en criomicroscòpia era d'uns 9 Å. Els cristalls bidimensionals tenien una cel·la de 165 Å × 165 Å en el grup espacial $P42_12$ (fig. 1.4), on els connectors es disposen amb l'eix 12 al llarg del quaternari, i alternats, un cap amunt, un altre cap envall, segons obliga l'eix helicoidal. Es continua veient la vorticitat de la seva estructura, però la gran diferència és que les 12 subunitats estan separades perfectament i ja no hi ha una transició de

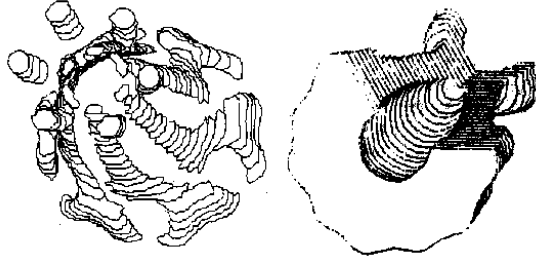


Figura 1.3: Perspectives del coll del fag $\Phi 29$. A la dreta, vista d'un tall de l'estructura des de la part més ampla, la corresponent a la proteïna p10. A l'esquerra, reconstrucció a partir dels valors més alts de la densitat, i en direcció contrària. La forta vorticitat de l'estructura és apreciable, així com el canvi de simetria 12 en la part ampla, a 6 en la més estreta. [9].

simetria 12 a 6. Aquests estudis estan fets només amb el connector, sense la proteïna p11, per la qual cosa és molt possible que allò observat amb simetria 6 representés aquesta part. (Comparar fig. 1.3 i fig. 1.5)

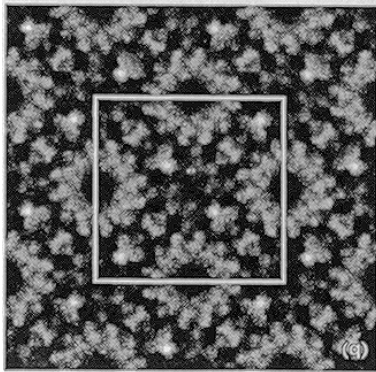


Figura 1.4: Cristall bidimensional. Promitjat de 5 imatges, amb la simetria $P42_12$ imposada. La cel·la de $165\text{\AA} \times 165\text{\AA}$ està indicada en blanc. [53].

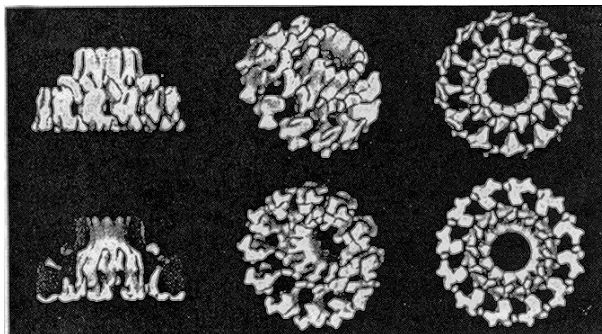


Figura 1.5: Representació tridimensional del connector a 9\AA . Representació tridimensional del connector a 9\AA , en varies vistes.

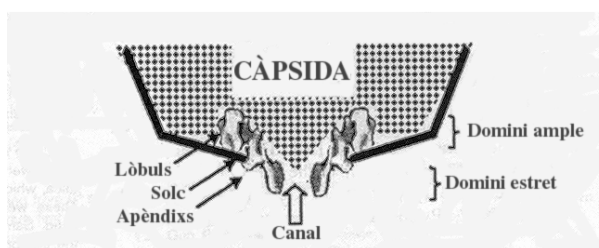
Tot i que la resolució en el pla perpendicular a l'eix 12 és prou alta, no és així en profunditat, ja que en inclinar les mostres per obtenir dades en

aquesta direcció es perd resolució. Les resolucions finals es redueixen a uns 10Å en el pla horitzontal i a uns 17Å en el vertical.

El connector té una alçada de 75Å i conté dos dominis: un d'ample (135Å d'amplada i 40 d'alçada) que s'inserta a dins de la procàpsida, i un de d'estret (62Å d'amplada per 24 d'alçada) que interacciona amb les altres proteïnes que formen el coll i que connecta amb la cua. El domini ample té forma d'anell de 80Å de diàmetre i està envoltat de 12 lòbuls de 27Å cadascun. A l'interior també hi ha 12 petits lòbuls de 15Å que connecten amb els anteriors. El domini estret conté, per la part de fora i a mitja alçada, 12 petits apèndixs. Se suposa que la vall que es forma entre aquests apèndixs i els lòbuls formen un encaix on s'hi posen les proteïnes que formen un dels vèrtexs de simetria 5 de la càpsida, impedit d'aquesta manera moviments longitudinals del connector, però no de giratoris. Ambdues parts envolten un canal que comença amb 58Å de diàmetre en el domini ample i acaba amb 36Å en el domini estret (fig. 1.6) [53].

Figura 1.6: Encaix del connector en la càpsida.

Situació hipotètica del connector dins del vèrtex de la càpsida. Aquesta s'encaixa en el solc entre els lòbuls i els apèndixs, de manera que podria rotar, però ni entrar en la càpsida, ni sortir-ne. [53].



Connectors complexats amb anticossos específics a diferents zones de la seva seqüència d'aminoàcids van ser estudiats amb microscòpia electrònica (*immunoelctrònica*) i promitjat d'imatge [52] (fig. 1.7). Aquests anticossos, quan es lligaven als connectors, mostraven les parts de la seqüència proteïca exposades a l'exterior. Segons els autors del treball, la seqüència del connector comença en els apèndixs de la part central, per dirigir-se després a l'extrem del domini estret. D'alguna manera no especificada, a causa de que no es van poder addicionar anticossos, la seqüència continua per dins del canal fins als lòbuls del domini ample, per acabar un altre cop a prop dels apèndixs de l'inici. Aquesta disposició és coherent amb altres dades conegudes, com que la part inicial de la seqüència és un domini amb afinitat pels àcids nucleics [16, 15, 51]. En determinar l'estructura per difracció de raigs X vàrem veure que aquesta interpretació no és la correcta (fig. 5.1, plana 108). Només els anticossos relacionats amb la zona d'unió amb el pRNA i l'extrem N-terminal van determinar dominis correctes (fig. 1.7, **d** i **e**).

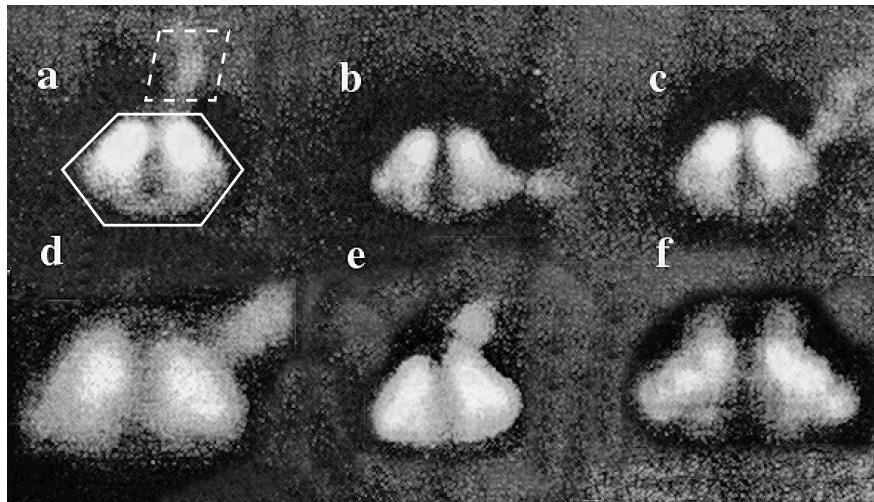
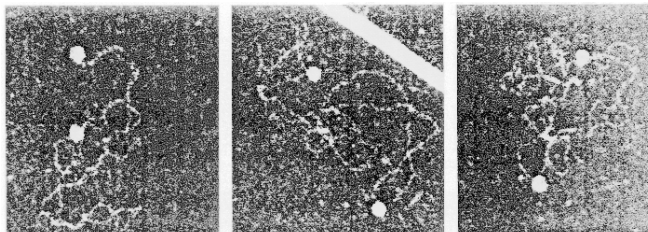


Figura 1.7: El connector amb anticossos. Microscòpia electrònica del connector complexat amb anticossos. Vistes laterals promitjades del connector amb alguns anticossos [52]. A la imatge **a** la línia contínua delimita la projecció lateral promitjada del connector, i la discontinua, la d'un anticòs. L'anticòs de la imatge **a** s'unia a la seqüència d'aminoàcids 78-92. El de la **b** a 167-190. El de la **c** a 275-292. El de la **d** a 1-14. El de l'**e** a la zona d'unió amb el pRNA. La imatge **f** és la de control. Veure el text per una discussió sobre aquests resultats.

L'afinitat pel DNA de la part externa del connector s'ha demostrat mitjançant microscòpia electrònica i electroforesi [51]. DNA de longitud coneguda es posava en presència de connectors aïllats. En observar amb microscòpia electrònica la longitud del DNA (fig. 1.8), es podia veure que aquest era més curt, però a més a més es veien petites zones amb superenrotllament. Si es mesurava la massa d'aquest DNA per electroforesi, resultava ser la del DNA més la de tants connectors com nusos o zones superenrotllades es veien en les imatges. La longitud resultant del DNA implicava que la quantitat de DNA superenrotllat era massa gran com per cabre dins del canal del connector, de manera que l'única explicació és que el DNA estigués embolicant el connector, mostrant així l'elevada afinitat que aquest té pel DNA en la seva zona externa.

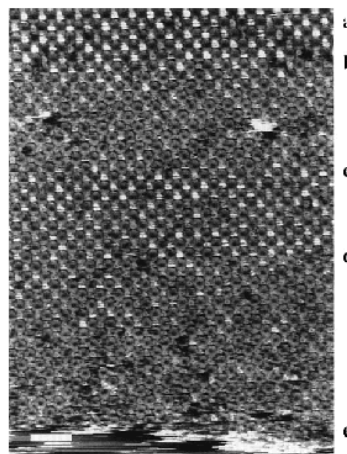
La microscòpia de força atòmica (*AFM*) dels mateixos cristalls bidimensionals havia mostrat altres característiques del connector, com la seva sensibilitat a la força aplicada (fig. 1.9). Un resultat molt espectacular d'aquest estudi és que augmentant i decreixent la força, dins d'uns certs límits, el

Figura 1.8: DNA superenrotllat al connector. Complexos de connector-DNA a una magnificació de $\times 160000$. Les boles blanques són connectors amb el DNA superenrotllat al voltant seu. [51].



connector recuperava la seva forma [39]. Tot i que l'AFM pot donar lloc a artefactes a causa d'aquesta força, també és possible que el connector sigui elàstic per ell mateix, un tema que encara està per determinar.

Figura 1.9: Connectors visualitzats amb AFM. Topografia dependent de la força dels connectors del $\Phi 29$ adquirida amb AFM. Entre **a** i **b** la força aplicada era entre 50 i 100 pN. A **b** es va incrementar a uns 150 pN. La part que estava estesa, corresponent a l'extrem estret del connector quedava enfonsada a causa de la força aplicada. Aquest procés, reversible, es repeteix en la mateixa imatge: entre **c** i **d** la força era de 50-100 pN i 150 pN entre **d** i **e**. A **e** la força es va incrementar a 300 pN i l'ordre cristal·lí es va trencar [39].



Recentment s'ha publicat un model a partir de difracció de raigs X, sense afinar les coordenades [48], i afinades [47]. En ell ja s'hi podent observar algunes de les característiques del connector a nivell atòmic. Aquests estudis han estat fets en paral·lel als nostres i han donat resultats molt similars, tot i que a una resolució menor i amb algunes diferències estructurals causades probablement per aquesta resolució menor. En la discussió es farà una comparació amb el nostre model (pàg. 107).

1.2.2 El connector en la formació del fag

Les parts que formen un fag es construeixen independentment, formant les peces principals per separat [60, 7], que després s'acoblen. A la figura 1.11 es mostra el camí típic per molts fags d'aquest acoblament.

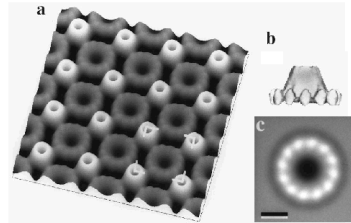


Figura 1.10: Model d'AFM del cristall 2D. **a** Perspectiva del model d'AFM promitjat per correlació i amb simetria 4 aplicada dels connectors de la figura 1.9. Les marques blanques representen la cel·la unitat. **b** Vista lateral del connector combinant dades prèvies amb les d'aquesta reconstrucció. **c** Promitjat rotacional amb simetria 12. La barra representa 50\AA , i els grisos representen una escala de 30\AA . Les mides del connector són $76 \pm 0.5\text{\AA}$ d'alçada, $64 \pm 4\text{\AA}$ de diàmetre en la part estreta i $133 \pm 3\text{\AA}$ en l'ample. El canal té un diàmetre de $12 \pm 0.3\text{\AA}$ a $37 \pm 4\text{\AA}$ en les parts estreta i ampla, respectivament [39].

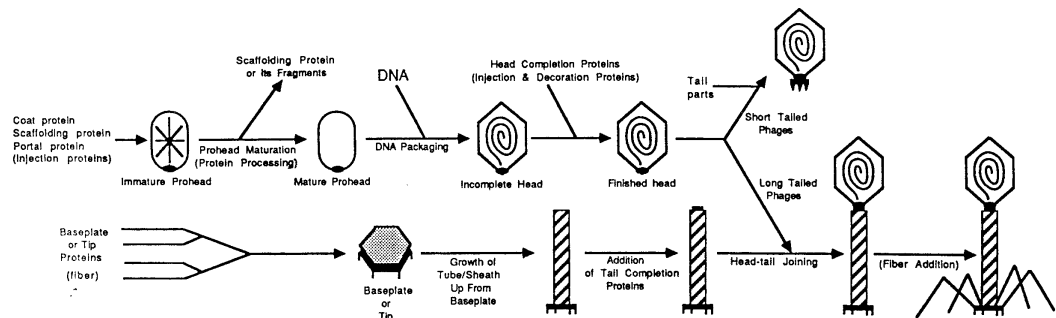


Figura 1.11: Acoblament del fag. Etapes en l'acoblament dels fags de doble cadena de DNA. El connector està representat per un cercle petit negre en la base de la càpsida. Es creu que aquesta creix a partir del connector i que el creixement s'acaba després de que s'hagi introduït el DNA. A continuació es junta amb la part inferior, que s'ha ensamblat apart. [7].

El connector té un paper central en la formació de la càpsida. La complicada geometria de la càpsida requereix d'ajuda externa per formar-se, ja que les proteïnes s'han de disposar en una simetria icosaèdrica exacte, i un petit error en la curvatura pot donar lloc a un objecte de forma diferent, que pot ser no es tancaria mai, i que no podria contenir el DNA. És molt comú obtenir ordenaments tubulars amb les proteïnes de la càpsida quan se les intenta unir sense presència del connector [38].

En el cas de les càpsides allargades (*prolate*), com la del $\Phi 29$, la forma

icosaèdrica només està present en els extrems, en les tapes, essent la part central un prisma de llargària diferent segons de quin fag es tracti. Es coneix que el virus pot suportar petits errors en l'estructura d'aquesta part central, un possible origen de la diversitat d'aquests virus, però no en les parts icosaèdriques, ja que han de contenir el mecanisme d'injecció [38]. Per això, també, el creixement de la càpsida ha de començar en el vèrtex on se situa el connector.

Es creu que la terminasa, que es troba l'inici del DNA que s'ha d'empaquetar, s'uneix al connector. La terminasa és una ATPasa i per tant, segurament la iniciadora de tot el procés energètic en el transport del DNA [27].

L'afinitat del DNA per la part exterior del connector i la seva facilitat per embolicar-se al seu voltant (plana 9 i fig. 1.8) va portar a un model [51] en que el DNA entraria a dins de la càpsida embolicant-se al voltant del connector. Aquest mecanisme s'ha deixat de suposar, ja que totes les altres evidències porten cap a un DNA passant a través del canal. El fet que el connector no estigués junt amb la càpsida segurament va ser el que va provocar aquest efecte.

El DNA entra en el fag després de la formació de la càpsida, i abans de la unió de la cua i de la base. Els models inicials apostaven per un motor unidireccional *brownià*: El connector podria girar en una sola direcció, per motius tèrmics, i quedava travat en cas d'intentar moure's en la direcció contrària. En cada moviment favorable faria avançar el DNA en el seu interior linealment cap a l'interior de la càpsida. L'entrada del DNA no costaria energia, però, en canvi, el DNA adquiriria energia en comprimir-se dins de la càpsida. Aquest motor tindria la propietat de crear energia del no res, contradient la segona llei de la termodinàmica [38].

Els estudis d'empaquetament *in vitro* han mostrat que calen entre una i dues molècules d'ATP per cada 2 parells de bases empaquetats [10]. Aquest és un consum similar a altres sistemes de *translocació* (transport). Tot i que aquests altres sistemes necessiten consumir energia, per exemple, per trencar ponts d'hidrogen, el consum es pot considerar baix, si tenim en compte la gran pressió dins la càpsida del DNA que va entrant i que empeny per sortir.

El DNA s'ha de traslladar per l'interior del connector, cap a dins de la càpsida. Moltes molècules processives se suposa que actuen avançant a través de la hèlix del DNA. El resultat és que el van voltant. Un mecanisme equivalent implicaria un connector, en relació, quiet i un DNA que aniria girant al voltant del seu propi eix. Un mecanisme així faria entrar el DNA girant dins de la càpsida. Allà dins, s'ha de plegar i embolicar, però és força difícil, a causa del poc espai que hi ha pel DNA dins de la càpsida, i a la torsió extra que es generaria, que s'hauria de compensar amb un consum molt

més elevat d'energia. Contràriament, un connector giratori permetria que el DNA avancés sense girar [28]. Un connector d'aquestes característiques, lliure, podria impedir la formació de la càpsida, però s'ha postulat que en els primers moments de la maduració de la càpsida, el connector podria estar-hi travat fins arribar a la maduració de la càpsida. En aquest moment s'alliberaria el connector i entraria en marxa el mecanisme per fer entrar el DNA, alhora que la càpsida s'aniria expandint per la pressió d'aquest a l'entrar (plana 12).

Aquesta seqüència d'acoblament ha estat demostrada, per exemple pel fag T4 [31, 32]. Una càpsida totalment madura també és capaç d'empaquetar DNA [32].

1.2.3 La infecció

Un cop formats connector i càpsida, la cua i la base s'uneixen per formar el fag complet. Les cues tenen longituds molt diverses. La del $\Phi 29$ és de les més curtes. Generalment es creu que el connector o alguna altra proteïna del coll tanca el canal, i pot ser una segona en la base també, si la cua és de les llargues [38].

En el moment de la infecció s'ha de transmetre un senyal per tal que el bloqueig desaparegui. En cues curtes, una modificació en la proteïna que bloqueja (o en el connector) pot ser possible a causa de la proximitat entre aquesta part del fag i la membrana cel·lular, però no és possible en cues llargues, ja que la distància és molt gran. Les cues poden tenir longituds més grans que la pròpia càpsida, i en aquests casos, són flexibles.

En aquest segon cas, en unir-se el connector amb la cua, se suposa que el bloqueig a l'alçada del connector s'allibera, la cua queda plena de DNA i és la base qui l'atura fins que algun canvi en el moment de penetrar la cèl·lula desbloqueja el pas.

Experimentalment s'ha vist que les càpsides dels fags λ es poden unir a cues que ja estan unides a la cèl·lula [14]. En aquest cas el DNA hauria d'estar travat per alguna proteïna del coll, fins que s'unís la cua. Aquest alliberaria el DNA, que penetraria a la cèl·lula automàticament, ja que el bloqueig de la base ja hauria estat alliberat prèviament.

1.3 Les simetries: cinc, sis, deu i dotze (o tretze)

Com ja s'ha vist, la càpsida d'un fag té forma d'icosàedre allargat. Per tant té simetria 5 en els seus vèrtexs. Tots els connectors estudiats, ja sigui per

microscòpia o bé per raigs X tenen simetria 12 o 13. Per exemple, el del $\Phi 29$ té simetria 12 [8, 25, 3], així com el del T3 [55].

L'únic que per ara s'ha demostrat que té simetria 13 és el de l'SPP1 [17, 33]. Aquest resultat sempre ha estat una mica controvertit. Hi ha indicis que la sobreexpressió dels connectors en sistemes heteròlegs podria donar lloc a ordenacions diferents. Per exemple en el fag T7 s'han observat poblacions barrejades de connectors amb simetria 12 i 13. En connectors de T3 i $\Phi 29$ extrets *in vivo* de les seves càpsides, mai no s'ha observat cap simetria diferent de la 12. No es coneix cap estudi d'aquestes característiques dels connectors de l'SPP1 [10].

De totes maneres, si l'SPP1 és una excepció en els fags, però en canvi, l'estructura del connector és idèntica a la dels altres fags, exceptuant-ne la seva simetria, voldrà dir que el mecanisme dels connectors no pot estar basat en cap lligam estructural entre ell i el DNA.

Un estudi a resolució atòmica de mostres que haguessin estat confirmades com a funcionals (no artefactes de la sobreexpressió) es feia necessari, i aquest va ser el primer pas un cop obtingudes les primeres dades de difracció dels cristalls (plana 83 i [25]).

La relació de simetries entre el connector i la càpsida és important per entendre el mecanisme de bombeig del DNA en ambdues direccions. Si la relació és un número enter, (p.e. 10:5, 15:5), podria implicar que proteïnes veïnes trobarien facilitat en lligar-se o interaccionar entre elles [28]. El resultat seria una estructura probablement massa rígida com per permetre una rotació ràpida del connector (fig. 1.12). Per tant s'hauria de suposar un mecanisme de bombeig del DNA a partir d'algun altre element que es trobés en aquella zona. Hi ha estudis que mostren la necessitat de RNA en l'acoblament del DNA dins de la càpsida, una estructura rígida seria possible en aquest cas.

Si la relació no és entera (p.e. 11:5, 12:5, 13:5), no es pot trobar cap posició relativa privilegiada entre el connector i la càpsida. Una orientació que encari un parell d'interfícies entre el connector i la càpsida, allunyarà les d'un sector llunyà. D'aquesta manera, o bé es deforma considerablement l'estructura o bé simplement queda centrada al voltant de l'eix 5 de la càpsida i el connector quedarà lliure per rotar en qualsevol direcció.

Per microscòpia electrònica s'ha observat que el canal interior del connector, per on passa el DNA, té un diàmetre adequat per tal que passi una cadena doble de DNA estirada o molt poc deformada. Una doble hèlix de DNA en la conformació típica té una simetria 10_1 que en projecció axial és de 10 (fig. 1.13). Un altre cop es pot tornar a considerar la mateixa hipòtesi: la relació entre simetries del connector i del DNA no permet una orientació estable relativa entre ells. En aquest cas també es podria considerar la idea

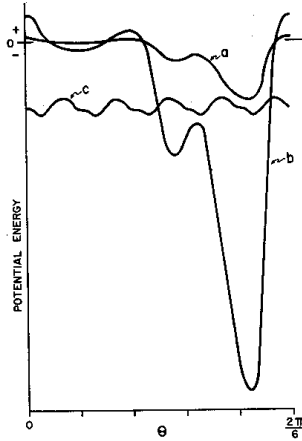


Figura 1.12: Relació entre l'energia d'interacció i el nombre de subunitats. L'energia potencial positiva significa repulsió i la negativa atracció. Situació hipotètica entre un connector situat concèntricament dins del vèrtex d'una càpsida. **a** Interacció entre un dels lòbuls i una de les proteïnes de la càpsida. **b** Interacció entre tots els lòbuls i la càpsida en un cas de simetria entera. En aquest cas, la relació de simetries és 6:6. Hi ha un mínim que podria aturar la rotació. **c** Quan la relació és 6:5, la funció és periòdica ($2\pi/30$), i de molt poca amplitud. La interacció és molt més fluïxa, de manera que no hi ha barreres prou fortes com per aturar el gir. El connector pot rotar amb poca energia. [28].

que el connector no rota i és el DNA el que ho fa, però això provocaria un superenrotllament impossible a l'interior de la càpsida, ja que en anar girant, intentaria fer girar alhora tota la part ja encapsulada. De manera que la hipòtesi més simple torna a ser la d'un connector giratori i amb la nova peça d'un DNA que va avançant linealment a través d'ell.

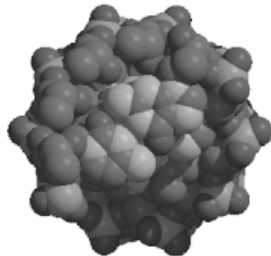


Figura 1.13: Simetria del DNA. Projecció vertical de DNA per mostrar que, en aquesta direcció, té simetria 10.

Altres parts importants en el mecanisme són el pRNA i la proteïna p16. Aquest parell es creu que formen part del sistema que fa entrar el DNA a la càpsida, però que un cop entrat se separa del virus.

El pRNA (*packaging RNA*) té 174 bases repartides en una estructura anular amb uns apèndixs interiors i tot ell amb simetria 6. Està col·locat probablement al voltant de la part més estreta del connector [12, 26]. La proteïna p16 és l'ATPasa que proveeix d'energia tot el sistema [23]. Es creu que l'ATPasa provoca canvis de forma als apèndixs del pRNA que actuen de *rems* fent girar el connector. El pRNA a més reconeix l'extrem del DNA víric que s'ha d'empaquetar i el condueix a l'entrada del connector. La relació de

simetries entre connector i el pRNA és de 12:6, una relació entera. Si l'impuls que fa girar el connector prové del apèndix del pRNA, és important que les simetries casin, ja que d'aquesta manera la força impulsora dels 6 braços se sumaran. L'efecte seria similar al cas **b** de la figura 1.12.

El resultat de totes aquestes peces és, doncs, un motor on p16 proveeix d'energia al pRNA i que aquest la fa servir per fer girar el connector gràcies a un *encaix* de simetries com si es tractés d'una clau que fa girar un cargol. Per l'altre extrem, la càpsida i el connector no queden fixats gràcies al *desencaix* de simetries i es permet una rotació lliure com un coixinet. El pRNA llavors reconeix la proteïna terminasa del DNA víric i l'acosta al connector, que el conduirà cap a dins de la càpsida, ajudat per una altra asimetria entre DNA i connector que a més obligaria al DNA a avançar només en una direcció i sense recargar-se.

Un cop empaquetat tot el DNA, el pRNA i la p16 en deixar de ser útils desapareixen, i a la part estreta s'uneixen la p11 i tota la resta de la cua i la base. La unió és estable, ja que p11 té simetria 6 i pot encaixar bé amb la part estreta del connector i formar el coll del fag tal com s'ha vist en els estudis de microscòpia electrònica (fig. 1.3, plana 6).

L'estructura externa del connector és molt important per conèixer com es produeix aquest efecte de *rem* entre el pRNA i el connector, però la interna, la més desconeguda a la microscòpia electrònica, és imprescindible per entendre com el DNA pot lliscar unidireccionalment al llarg del connector cap a dins de la càpsida.

Capítol 2

Objectius

Tots els estudis previs que s'han fet sobre l'estructura dels connectors eren a partir de microscòpia electrònica o de força atòmica. Aquestes proporcionen una idea general de l'envolupant de la proteïna, però no permet conèixer-ne l'estructura i les interaccions a nivell atòmic que són indispensables per al coneixement del seu mecanisme.

Els objectius plantejats en començar aquesta tesi van ser:

1. Obtenció de l'estructura tridimensional a alta resolució del connector del fag $\Phi 29$ mitjançant les tècniques de la cristal·lografia de proteïnes.

Com que l'únic que es coneixia eren models de baixa resolució obtinguts a partir de microscòpia electrònica, això obligava a:

2. Elaborar un mètode de reemplaçament molecular i extensió de fases per a la resolució d'estructures d'alta simetria a partir de models de molt baixa resolució.

Paral·lelament s'intentaria resoldre l'estructura per tècniques convencionals:

3. Resoldre l'estructura a partir de cristalls derivats fent servir les tècniques del MIR i del MAD.

Per últim, l'estructura a nivell atòmic permetrà:

4. Proposar un mecanisme que expliqui l'empaquetament del DNA a l'interior de la càpsida dels virus bacteriòfags.

Part II
Materials i Mètodes

Capítol 3

La Cristal·lografia de proteïnes

3.1 Definicions

3.1.1 Els cristalls

Un cristall és un ordenament periòdic de molècules. La unitat que es repeteix, per translació, és la cel·la unitat. La manera d'ordenar-se les molècules en una cel·la unitat es diu empaquetament. A causa de la diversitat d'empaquetaments possibles, una cel·la unitat té simetries (rotacions, translacions, inversions, etc.) caracteritzades en 230 grups espacials. Aquestes simetries divideixen la cel·la unitat en unitats asimètriques. Totes aquestes simetries són globals per al cristall, i el defineixen com a tal, però a més, en una unitat asimètrica, hi poden haver altres simetries locals, que només són vàlides dins d'ella. Aquestes simetries s'anomenen *simetries no cristal·logràfiques* (NCS). Generalment hi són per característiques pròpies de la molècula. És la unitat d'empaquetament que ha format el cristall (fig. 3.1).

3.1.2 Els cristalls de proteïna

Fins l'any 1934 es creia que les proteïnes eren objectes globulars sense estructura interna. Les primeres difraccions, eren de molt baixa resolució i suggerien un ordenament en fibra. Aquest any J.D. Bernal i D.Crowfoot varen poder obtenir imatges de difracció de raigs X de prou resolució com per demostrar que una proteïna tenia una estructura interna [41].

Les proteïnes són compostos formats per una seqüència d'aminoàcids (fig. 3.2 i taula 3.1) [20] que es pleguen formant una estructura secundària, principalment hèlixs α i fulles β . Aquesta estructura secundària es replega en l'anomenada terciària que és el que dóna a les proteïnes la seva forma i funció característica. Moltes vegades una proteïna necessita estar unida amb altres

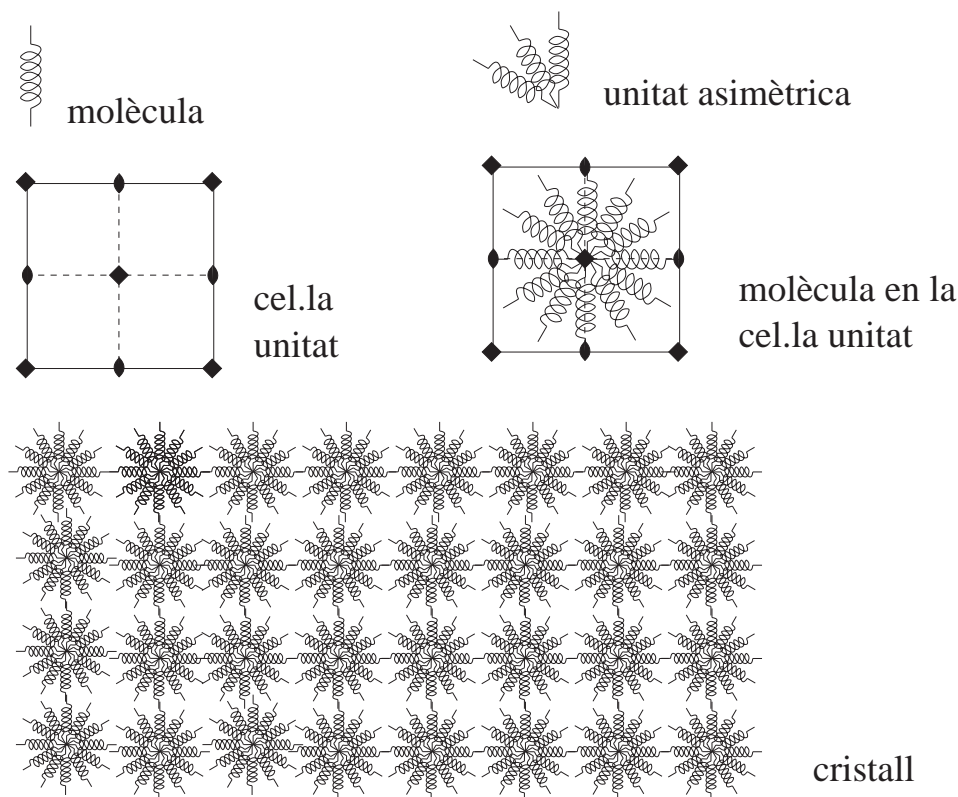
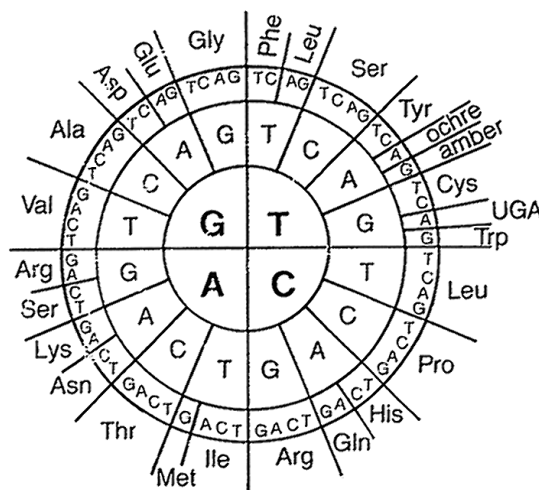


Figura 3.1: Simetries dels cristalls.

per fer la seva funció. Aquesta es la seva estructura quaternària. Veurem que per al cas del connector del $\Phi 29$, són necessàries 12 còpies de la proteïna p10 per formar la unitat funcional.

Un cristall de proteïna està format, a més del material proteic, per una quantitat gran de solvent, generalment de l'ordre del 50%, tot i que és molt variable. Segons les condicions de cristal·lització i segons el mètode emprat, també s'hi poden trobar agents cristal·litzadors (MPD, glicerol, etc. fig. 3.10, plana 55), i àtoms pesants afegits per ajudar a resoldre l'estructura, o bé que formen part del complex actiu. En altres casos, òbviament també hi poden haver presents altres compostos amb els que es vulguin estudiar les interaccions amb la proteïna, com fàrmacs, DNA o RNA. Quan s'introdueixen àtoms pesants per ajudar a resoldre una estructura, els cristalls resultants s'anomenen derivats, respecte del cristall format només per la proteïna pura, que s'anomena natiu. Aquesta gran quantitat de solvent provoca moltes vegades que un cristall de proteïna sigui més aviat esponjós. Es degrada

Figura 3.2: El codi genètic. A la perifèria del cercle hi ha els noms dels aminoàcids, al centre el de les bases (seqüència del DNA). El DNA es transcriu directament a RNA missatger, que està format pel mateix conjunt de bases, però amb U en lloc de T. El Ribosoma és una gran proteïna que té per funció traduir l'RNA missatger a proteïna. El codi genètic es llegeix radialment des del centre del cercle i cada grup 3 bases dóna lloc a un aminoàcid. UGA, ochre i amber són combinacions de bases que no corresponen a cap aminoàcid i marquen el final del codi d'una proteïna.



aminoàcids			bases				
Alanina	Ala	A	Leucina	Leu	L	Adenina	A
Arginina	Arg	R	Lisina	Lys	K	Tirosina	T
Asparagina	Asn	N	Metionina	Met	M	Guanidina	G
Aspàrtic	Asp	D	Fenilalanina	Phe	F	Citosina	C
Cisteïna	Cys	C	Prolina	Pro	P	Uracil	U
Glutamina	Gln	Q	Serina	Ser	S		
Glutàmic	Glu	E	Treonina	Thr	T		
Glicina	Gly	G	Triptòfan	Trp	W		
Histidina	His	H	Tirosina	Tyr	Y		
Isoleucina	Ile	I	Valina	Val	V		

Taula 3.1: Aminoàcids i Bases. Esquerra: Nom dels aminoàcids, junt amb les abreviacions de tres lletres i d'una lletra. Dreta: Nom de les bases del DNA i de l'RNA i el seu codi d'una lletra. El DNA està format per ATGC i l'RNA per AUGC. Les tirosines es transcriuen directament a uracils.

molt ràpidament si no es troba en la solució on ha crescut, o bé congelat.

El nombre de grups espacials permesos en un cristall de proteïna és més reduït que en un cristall convencional, ja que no són permeses operacions d'inversió. Dels 230 grups espacials, només en són possibles 65.

3.2 Obtenció de la proteïna, dels cristalls i dels derivats

3.2.1 Obtenció de la proteïna p10

La proteïna es pot obtenir aïllant-la directament del seu organisme [11] o bé per sobreexpressió en organismes que porten un plasmidi recombinant [30] (fig. 3.3) (DNA circular no genòmic que conté el gen que la codifica). En tots dos casos cal purificar la proteïna. El procés realitzat per la proteïna del connector està explicat a la plana 77 [24].

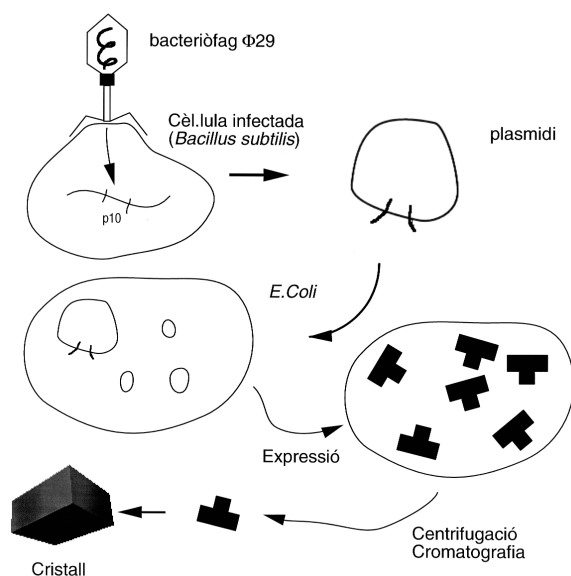


Figura 3.3: Obtenció de la proteïna. S'insereix el gen que codifica la proteïna en un plasmidi, que és DNA circular no genòmic que contenen els bacteris i se l'expressa en una cèl·lula. La proteïna clonada s'aïlla de la cèl·lula per diversos mètodes. Es purifica i se n'obtenen els cristalls posant-la en una solució sobresaturada, junt amb altres agents que ajudaran a la cristal·lització.

3.2.2 Els cristalls

Molt breument, per a obtenir cristalls cal tenir una mostra de proteïna molt pura, i deixar-la en repòs, en una solució sobresaturada, que conté també agents precipitants que ajudaran a la oligomerització de la molècula, i a la posterior formació del cristall. La solució es manté segellada per mantenir un equilibri entre les fases líquida i gasosa. Cal notar que, així com a la preparació de la proteïna es fan servir tècniques de biologia molecular més o menys establertes, amb un coneixement força bo dels mecanismes que hi participen; la cristal·lització és un fenomen encara poc conegut com per poder establir tècniques de cristal·lització amb alta fiabilitat. Per això, per a obtenir

cristalls s'han de fer servir criatges de conjunts d'agents cristal·litzadors que han funcionat en altres proteïnes de característiques similars. Hi ha criatges generals, on es varien les condicions al llarg d'un gran rang, i d'altres més precisos que tracten d'afinar condicions prometedores.

El número de criatges necessaris per obtenir cristalls pot variar molt segons la "dificultat" dels cas. Generalment passen mesos fins a obtenir cristalls analitzables. A més s'ha d'afegir el temps necessari per a tal que es desenvolupin els cristalls. Un cristall pot necessitar fins i tot anys per créixer correctament. El temps d'espera habitual és de l'ordre d'una a poques setmanes. Per suposat també es pot donar el cas que el cristall es degradi si està massa temps en la solució. En el nostre cas, els cristalls es degradaven als dos dies, després de créixer molt poc.

3.2.3 Els derivats

Una estructura pot ser resolta de moltes maneres en cristal·lografia de raigs X. Quan no es coneix cap estructura similar, cal fer servir mètodes basats en la difracció més intensa, o bé anisotròpica, d'àtoms pesants (secció 3.5.1, plana 42).

Com que la major part de les proteïnes no contenen àtoms pesants amb aquestes propietats, cal introduir-los expressament. Una de les tècniques és la de xopar el cristall en una solució que contingui l'ió pesant. D'aquesta manera l'ió pesant es podrà difondre en el cristall i fixar-se en llocs on hi trobi afinitat per fer-ho. L'avantatge d'aquest mètode és que la geometria de la cel·la unitat no sol variar, condició indispensable per a segons quin mètode de resolució es vulgui afrontar. El principal problema és que no sempre l'ió pot penetrar en el cristall i situar-se ordenadament. De vegades queda desordenat en les regions de solvent, o de vegades arriba a llocs d'unió en molt poques cel·les (ocupació baixa).

Una altra tècnica és la *co-cristal·lització*. Aquest cop, proteïna i ió pesant s'intenten cristal·litzar a la vegada. Afavorint així que les unions es formin en la solució, abans de formar-se el cristall. L'avantatge és que augmenta la probabilitat de formar compostos ió pesant-proteïna, gràcies a l'agitació tèrmica. El problema és que segurament caldrà trobar condicions de cristal·lització noves. Un altre problema és que potser cristal·litza amb un empaquetament diferent i els cristalls no es poden fer servir per comparar amb els cristalls nadius.

Una tercera tècnica consisteix a substituir certs aminoàcids per altres que han estat modificats per contenir un àtom pesant. El cas més típic és el de la seleni-metionina (SeMet), on el sofre de les metionines se substitueix per seleni, un àtom que difracta anòmalament a una longitud d'ona molt conve-

nient per a cristalls de proteïna i, a més, típica en sincrotrons. L'avantatge d'aquest mètode està en que la localització dels àtoms pesants és molt precisa, i aquests estaran, en principi, idènticament presents en cadascuna de les proteïnes (ocupació alta). La dificultat consisteix, és clar, a obtenir proteïna així modificada, i que aquesta cristal·litzi.

G.E	Ió pesant	Línia	Data	Referència
<i>C2</i>		X31	1997	[24], plana 78
<i>P4₂2₁2</i>		X31,X11,id2,5.2R	1998-1999	[24], plana 78
<i>P4₂2₁2</i>	<i>Ta₆Br₁₂²⁺</i>	X31,X11,id2	1998-1999	[24], plana 78
<i>I422</i>		bm30,id2,id14EH2 id14EH3,id14EH4,bm14	1999	3.5, plana 36
<i>I422</i>	<i>Ag⁺</i>	id2	1999	3.6, plana 37
<i>I422</i>	<i>Au⁺</i>	id14EH2	1999	3.7, plana 38
<i>I422</i>	<i>Pt²⁺</i>	id14EH2	1999	3.8, plana 39
<i>I422</i>	<i>Ir⁺</i>	id14EH2	1999	
<i>I422</i>	Tk (<i>Hg²⁺</i>)	id14EH2	1999	
<i>I422</i>	<i>Ta₆Br₁₂²⁺</i>	bm30, id14EH4, bm14	1999	3.9, plana 40
<i>I422</i>	<i>Hg²⁺</i>	id14EH2	1999	[42], plana 102
<i>I422</i>	SeMet	id14EH4, bm14	2000	

Taula 3.2: Llista dels cristalls analitzats. Tipus de cristalls del connector analitzats per difracció de raigs X, ió pesant i línia de difracció usada. G.E: grup espacial. Les línies de difracció estan llistades a la taula 3.3. Les referències corresponen a les taules on hi ha les estadístiques de cada cristall processat. Les taules dels cristalls fets servir als resultats estan referenciades cap a l'article corresponent. El cristall que es va fer servir a l'afinament és el I422 Hg, tractat com natiu.

3.3 Raigs X: generació, òptica, interacció amb els cristalls i detecció

3.3.1 Generació

Per a l'anàlisi de cristalls de proteïna per difracció de raigs X, aquests es generen de dues maneres: amb generadors d'ànode rotatori i amb la radiació de sincrotró.

Un generador d'ànode rotatori, [20] consisteix en una cambra al buit, on un filament (el càtode) emet electrons que són accelerats per una diferència de

potencial, fins que xoquen amb un ànode fet d'un material adequat per a que emeti raigs X a la longitud d'ona necessària per a l'experiment. Generalment, per a proteïnes, l'ànode és de coure, que té una K_α d'1.5418Å. Aquest procés és molt ineficient, ja que la majoria de l'energia consumida es transforma en calor. Per evitar que l'ànode es fongui, se'l manté rotant per anar canviant el punt d'incidència dels electrons i se l'envolta amb un circuit de refrigeració. L'avantatge és que l'aparell té unes dimensions raonables per instal·lar en un laboratori. És adequat per a cristalls relativament grans, amb cel·les relativament petites. El problema és l'elevada quantitat de temps necessari per realitzar un experiment (d'uns pocs dies a setmanes), a causa de la poca intensitat del feix de raigs X. Un altre problema és que no es pot canviar la longitud d'ona, per la qual cosa no són possibles, en general, experiments de difracció anòmala.

La radiació de sincrotró [58] és la que es desprèn en fer canviar de direcció partícules carregades que estan viatjant a velocitats properes a la de la llum. Les partícules carregades, en un sincrotró típicament electrons o positrons, circulen per un tub poligonal que està al buit. En els vèrtexs d'aquests polígons, els *bending magnets* corben la trajectòria de les partícules, generant alhora els raigs X. Un dispositiu que genera raigs més intensos és l'*insertion device magnet* (ondulador), que s'intercala en les seccions rectes. En un ondulador, uns camps magnètics en sèrie fan oscil·lar els electrons generant una radiació a una energia relacionada amb el període de l'oscil·lació. Un *wiggler* és un ondulador on els camps magnètics són d'intensitat molt més alta. El resultat és la generació d'harmònics del període de l'ondulador. Aquests harmònics formen un espectre similar al del *bending magnet*, però d'intensitat proporcional al nombre d'imants.

En aquest treball s'ha fet servir únicament radiació de sincrotró, ja que la mida petita dels cristalls i la gran mosaïcitat, obligava a tenir feixos intensos, i de secció petita.

3.3.2 Òptica

Els raigs X surten tangencialment de l'anell del sincrotró amb un espectre d'energies molt gran. Abans que arribin als aparells per fer l'experiment, s'han de preparar adequadament.

L'òptica consisteix en una sèrie d'aparells situats tangencialment a l'anell. Principalment, els monocromadors filtren les longituds d'ona sobrants, i els col·limadors afinen el raig [20].

El típic monocromador és un cristall de silici. Com que l'angle de difracció varia amb la longitud d'ona, es pot seleccionar la longitud d'ona orientant aquest cristall convenientment. Hi ha diverses configuracions estudiades per

Nom	Tipus	Sincrotró	Detector
BW7A	mad	Doris	MarCCD
BW7B	fixe	Doris	Mar 345 Image Plate
X31	mad	Doris	Mar 345 Image Plate
X11	mad	Doris	MarCCD
id2	fixe	ESRF	Image Plate
id14EH2	fixe	ESRF	Quantum CCD
id13	mad	ESRF	Quantum CCD
id14EH3	fixe	ESRF	MarCCD/Mar 345 Image Plate
id14EH4	mad	ESRF	Quantum CCD
bm14	mad	ESRF	MarCCD/Mar 345 Image Plate
bm30	mad	ESRF	Mar 345
5.2R	fixe	Elettra	Mar 300 Image Plate

Taula 3.3: Línies de difracció. fixe/mad: longitud d'ona fixe/ajustable. Doris: Desy, Hamburg. ESRF: Grenoble. Elettra: Trieste. id: *insertion device*. bm: *bending magnet*. X, BW i R: *wiggler*.

millorar l'enfoc, monocromacitat i intensitat del feix. Típicament el cristall monocromador es corba mecànicament per a comprimir el feix, i se'n munten dos en seqüència per a resoldre millor la longitud d'ona i orientar el feix en direcció a l'estació experimental.

Un col·limador serveix per definir un feix el més paral·lel possible. Els més senzills consisteixen en un tub allargat per on passa el raig. Aquest tub té unes obertures (*slits*) ajustables que aturen els raigs més divergents. En calen almenys tres convenientment separats per aturar també els raigs difractats per les pròpies obertures. Aquestes obertures es poden ajustar per a definir la secció del raig.

La divergència és l'angle d'obertura que té el raig en sortir de les últimes obertures. Aquesta divergència ha de ser la menor possible, ja que provocaria reflexions escampades en el detector, a causa del diferent angle d'incidència dels raigs sobre el cristall.

3.3.3 Interacció amb els cristalls

Els raigs X interaccionen amb les partícules carregades elèctricament. A les energies típiques usades (9-12 KeV), corresponents a una longitud d'ona al voltant d'1Å, l'índex de refracció està molt a prop de la unitat [20]. És per això que els raigs X no es poden enfocar, i per tant no és possible la formació

directa de la imatge. El tipus d'interacció que promou la difracció, a les energies típiques, és l'*scattering Thomson*. És tracta de la dispersió elàstica dels fotons pels electrons dels àtoms.

A l'*scattering Thomson*, si la radiació incident no està polaritzada, té una intensitat I_i , i interacciona amb una partícula lliure de càrrega e i massa m , la distribució angular de la intensitat de la radiació emesa és:

$$I_{eTh} = I_i \frac{e^4}{m^2 r^2 c^4} \frac{1 + \cos^2(2\theta)}{2},$$

essent c la velocitat de la llum, i r i 2θ les coordenades polars del punt d'observació, respecte de la direcció incident del raig.

La relació és inversament proporcional al quadrat de la massa de la partícula. Els electrons són 2000 cops més lleugers que els protons, i el valor absolut de la seva càrrega és el mateix. Per tant, la contribució dels electrons és 4 milions de vegades major que la dels protons. La difracció estarà només relacionada amb la distribució electrònica dins del cristall.

En un cristall els electrons no estan aïllats, pertanyen a àtoms, que alhora pertanyen a molècules i a cristalls. Cada raig difractat interfereix amb els difractats pels altres electrons, d'una manera característica donada per la geometria de tot el cristall. En particular, l'ordre cristal·lí afavorirà unes poques direccions en les que es produiran una interferència constructiva dels raigs (*les reflexions*) i l'estructura de la molècula modularà les intensitats de cadascuna d'aquestes reflexions.

Si f_j , \mathbf{r}_j i β_j són, respectivament, el factor de difusió atòmic, les coordenades i el tensor dels desplaçaments tèrmics de cadascun dels $j = 1..N$ àtoms de la cel·la unitat, i $\mathbf{h} = (\mathbf{k}_D - \mathbf{k}_I)/2\pi$ és l'anomenat vector recíproc que representa la diferència entre els vectors de propagació del raig difractat (\mathbf{k}_D) i l'incident (\mathbf{k}_I), la intensitat difractada és el mòdul al quadrat de [43]:

$$F_{\mathbf{h}} = \sum_{j=1}^N f_j e^{2\pi i(\mathbf{h}^T \mathbf{r}_j - \mathbf{h}^T \beta_j \mathbf{h})}$$

$$I_{\mathbf{h}} = |F_{\mathbf{h}}|^2$$

$F_{\mathbf{h}}$ s'anomena factor d'estructura o amplitud.

Si considerem un cristall com una distribució periòdica de densitat electrònica, també es pot posar $F_{\mathbf{h}}$ com:

$$F_{\mathbf{h}} = \int_V \rho(\mathbf{r}) e^{2\pi i \mathbf{h} \cdot \mathbf{r}} d\mathbf{r}$$

on la integració s'estén només al volum de la cel·la unitat.

El factor d'estructura és, doncs, la transformada de Fourier de la distribució electrònica en la cel·la unitat. Fent la transformada inversa, obtindrem el *mapa* de la densitat electrònica en la cel·la:

$$\rho(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{h}} |F_{\mathbf{h}}| e^{2\pi i \Phi_{\mathbf{h}}} e^{-2\pi i \mathbf{h} \cdot \mathbf{r}}$$

on s'ha posat $F_{\mathbf{h}} = |F_{\mathbf{h}}| e^{2\pi i \Phi_{\mathbf{h}}}$ per ressaltar que en fer el mòdul al quadrat la part de l'exponencial desapareix i les intensitats observades no contenen la informació de la fase ($\Phi_{\mathbf{h}}$), necessària per construir els mapes. Aquest és l'anomenat *problema de les fases* (plana 42).

3.3.4 Detectores

Per detectar els raigs difractats, en el nostre camp s'han imposat els detectors del tipus *image plate* i del tipus *CCD*. Altres mètodes, fotogràfics, estan en desús o es fan servir en casos molt específics.

Ambdós detectors tenen un aspecte exterior similar, on una pantalla d'entre 15 i 30 cm de radi se situa en línia amb la sortida dels raigs X i la mostra. La pantalla està muntada en un suport que permet variar la distància, i de vegades també l'angle del detector respecte el feix incident. Entre la mostra i el detector hi ha un *beam stop*, una peça molt petita de plom que serveix per no saturar el detector amb els raigs que no han difractat.

Els detectors del tipus *image plate* consisteixen en un disc de fòsfor d'uns 30 a 35 cm de diàmetre que va recollint la direcció i l'energia cedida pels raigs difractats durant el temps de recollida d'una imatge. Abans de recollir el següent diagrama, un làser escaneja la superfície del detector i excita un altre cop les partícules de fòsfor. La llum emesa es recull amb un fotodetector i es transforma en una imatge bidimensional [20].

Els del tipus *CCD* també tenen per pantalla un disc de fòsfor de mida similar, però en aquest cas, el fòsfor no amagatzema l'energia dels raigs difractats, si no que la converteix en llum visible. Aquesta llum es condueix per fibres òptiques a un chip *CCD* (*charge coupled device*) que l'amagatzema en forma de càrrega elèctrica fins que es digitalitza al final de la recollida de cada imatge [13].

En l'estat actual dels detectors, els *image plates* proporcionen un àrea de recollida més gran, i això es converteix en més facilitat per discernir les reflexions (*overlapping*), ja que, a igual resolució, es pot muntar més lluny. Altrament, el rang dinàmic de la recollida (diferència entre el valor més baix i el més alt) és molt més alt que en un *CCD*, que a més, quan se satura pertorba els *píxels* del voltant. També el soroll de fons de l'*image plate* és

menor que en un *CCD*. Per altra banda, el temps de *readout* és molt menor en un *CCD* (de l'ordre de 2 segons en front de més d'1 o 2 minuts). En el cas de recollides en un sincrotró, on els temps d'exposició per imatge són de l'ordre de pocs segons, l'ús de *CCDs* es fa imprescindible per accelerar el procés de recollida. En un laboratori, a causa de la menor intensitat dels raigs X, les recollides són d'hores per imatge, per la qual cosa l'increment de temps que representa un *image plate*, un parell de minuts per imatge, no és significatiu.

Ambdues tecnologies són força noves i encara poden evolucionar molt. Els *CCDs* tendeixen a augmentar la seva mida en aparells múltiples, i els *image plates* tendeixen a ser més ràpids.

3.4 L'experiment

3.4.1 Congelació i muntatge dels cristalls

Per difractar un cristall, se l'ha de *muntar* en un dispositiu adequat durant el temps que estarà exposat als raigs X. Un cristall es pot col·locar en un capil·lar de vidre, junt amb part de la gota on ha crescut. Aquesta tècnica es fa servir quan els cristalls poden aguantar la radiació sense degradar-se. També s'ha de fer servir quan els cristalls no aguanten la congelació. Llavors cal tenir una gran quantitat de cristalls per poder fer l'experiment.

Generalment els cristalls no aguanten la radiació, i per evitar *cremar-los* s'han de muntar congelats. La tècnica consisteix en mantenir-los en un flux de nitrogen a uns 100K durant tot l'experiment. Per al transport dels cristalls també és convenient la seva congelació. En aquest cas, s'enllacen amb un *pin* que consisteix en una barra metàl·lica acabada en un petit llaç de niló de les dimensions adequades [19]. La barra té unes dimensions estàndard de 18mm i el llaç té un diàmetre d'entre 0.2 a 0.05 mm segons la mida del cristall. El cristall es *pesca* amb el llaç de la gota on ha crescut, i segons les condicions en què hagi crescut se'l banya en una solució crioprotectora. Llavors se'l congela per immersió ràpida (*flash freezing*) en nitrogen o en propà líquids. Els cristalls es transporten en un *dewar* ple de nitrogen líquid al lloc de l'experiment.

Un cristall congelat aguanta generalment de l'ordre de 100 imatges en un sincrotró. En canvi un cristall no congelat només aguanta de l'ordre d'un parell d'imatges (fig. 3.4).

A l'hora de l'experiment, el cristall es col·loca davant d'on sortirà el feix, junt amb el *pin* o el capil·lar, muntat en un goniòmetre per poder ajustar l'alineament del cristall davant del feix. És important que a mesura que el

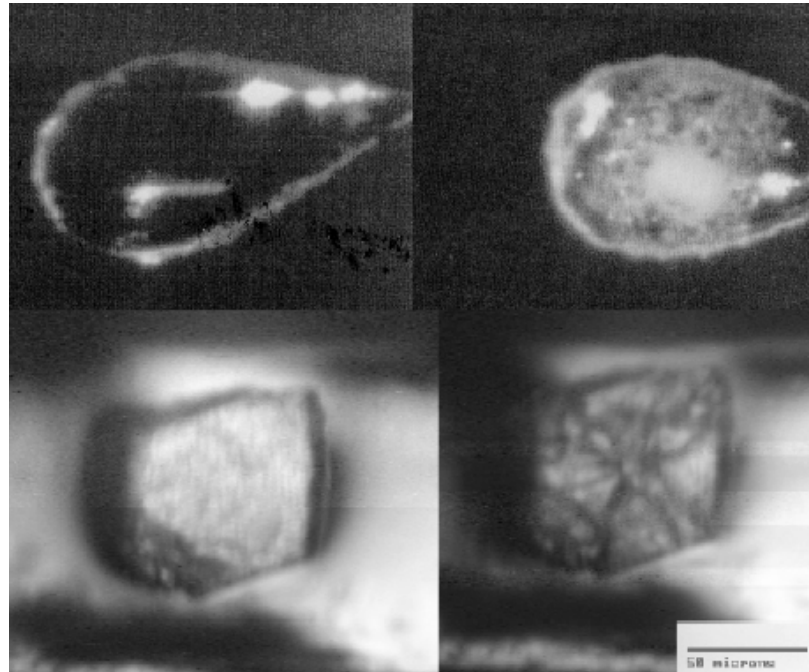


Figura 3.4: Efecte dels raigs X sobre els cristalls. Les dues imatges superiors corresponen a un dels cristalls que s'han fet servir per resoldre l'estructura. Està muntat en un loop, banyat en el criotampó que el protegeix i exposat a un flux de nitrogen a 100K. A l'esquerra, abans de començar l'experiment. El cristall és pràcticament transparent. A la dreta, després de recollir dos conjunts de dades, unes 200 imatges. El cristall està pràcticament cremat. El quadre blanc central és el punt d'incidència del feix. Les dues imatges inferiors corresponen a un cristall del virus HRV2 [56]. També abans i després de l'experiment, però el cristall no s'ha congelat, i només s'ha pogut recollir una imatge.

cristall rotí durant la difracció no es traslladi. D'aquesta manera les dades representaran un rotació completa. En el cas que hi hagués hagut translacions, s'haurien de processar les dades per separat, i probablement se'n perdrien moltes a causa dels ajustatges que caldria fer.

Quan es treu el cristall del seu contenidor de nitrogen líquid i es posa en el goniòmetre, moltes vegades hi ha un sobtat canvi de temperatura que pot formar cristalls d'aigua o de la solució que també difractaran, generalment en forma d'anells, i que segurament emmascararan la difracció de la proteïna. Si passa això, de vegades es pot intentar solucionar tallant per un segon el

flux de nitrogen (*fast annealing*). Generalment això és suficient per a què la solució es torni a recongelar o vitrificar correctament. De vegades, a més, el canvi de temperatura ajusta l'empaquetament de les proteïnes en el cristall i s'obté una difracció de millor qualitat.

Hi ha moltes variants en aquesta tècnica. La més senzilla consisteix a interrompre el feix amb una targeta de plàstic. Últimament s'estan afegint dispositius mecànics als goniòmetres que permeten graduar el temps que s'interromp el flux de nitrogen amb un pedal, i a la vegada deixar les mans lliures per treballar.

Una altra tècnica, adequada si aquesta falla és la de reescalfar el cristall a temperatura ambient. Un cop congelat el cristall se'l torna banyar en la solució crioprotectora. L'escalfament sol trencar el cristall pels punts més febles, generalment els plans de macla [46].

No és freqüent tenir gaires cristalls per fer un experiment, per la qual cosa totes aquestes tècniques tenen el risc de malmetre del tot el cristall. Si un cristall difracta mitjanament bé, és millor assegurar abans un conjunt de dades abans d'intentar aquestes millores. La poca secció transversal d'un feix de raigs X de sincrotró pot permetre tornar a difractar un cristall per una altra zona.

3.4.2 Recollida i processament de les dades

La llei de la difracció prediu raigs difractats en moltes direccions, però un detector només recull les que es troben en una direcció determinada ("endavant"). En aquestes direccions els raigs són més intensos segons la llei de Thomsom, i per altres causes com l'absorció i la correcció de Lorentz [20]. Els raigs es projecten sobre el detector donant lloc a un patró característic que permet identificar el grup espacial. (A la plana 80 es pot veure un diagrama de difracció corresponent a un dels cristalls analitzats.)

A una donada orientació del cristall respecte el raig incident, i depenent de característiques del cristall (la mida de la cel·la unitat, la mida del propi cristall, la qualitat del cristall, etc.) i de l'experiment (mida del detector, distància, longitud d'ona, etc.), el nombre de reflexions mesurades sol ser baix i les mesures de cada reflexió no són completes: En un cristall de mida finita, les reflexions no són puntuals, s'extenen sobre un angle sòlid. A més a més, els petits desordres en l'orientació de les cel·les augmenten aquest angle (*mosaicitat*). A causa d'això les reflexions no interferiran sempre constructivament i serà necessari rotar una mica el cristall per a recollir una reflexió sencera.

Per recollir un conjunt de dades complet d'un cristall (totes les reflexions necessàries per poder resoldre l'estructura) es recullen, doncs, dades en di-

I422							
Dmin	Nmes	Nref	I/sigma	Comp(%)	C.A.(%)	Mult	Rsym
6.64	12989	1955	10.3	99.3	99.3	6.6	0.058
4.70	24301	3405	10.1	100.0	99.7	7.1	0.060
3.83	31349	4353	11.0	100.0	99.9	7.2	0.055
3.32	36993	5090	6.3	100.0	99.9	7.3	0.076
2.97	42056	5724	6.6	100.0	99.9	7.3	0.112
2.71	46583	6336	3.9	99.9	99.9	7.4	0.187
2.51	50329	6840	2.5	99.9	99.9	7.4	0.289
2.35	38109	7315	1.8	99.7	99.9	5.2	0.378
2.21	26351	7744	0.5	99.2	99.8	3.4	0.521
2.10	18880	7712	1.4	94.1	98.9	2.4	0.505
Global	327940	56474	5.4	98.9	98.9	5.8	0.100

C2							
Dmin	Nmes	Nref	I/sigma	Comp(%)	C.A.(%)	Mult	Rsym
12.65	11178	3089	4.8	98.9	98.9	3.6	0.097
8.94	20880	5602	5.2	100.0	99.6	3.7	0.111
7.30	26903	7164	3.9	100.0	99.8	3.8	0.177
6.32	31833	8463	2.4	100.0	99.9	3.8	0.313
5.66	34956	9581	1.6	99.9	99.9	3.6	0.466
5.16	36951	10536	1.8	99.5	99.8	3.5	0.430
4.78	38830	11312	1.9	98.6	99.5	3.4	0.396
4.47	41067	12061	1.8	97.4	99.2	3.4	0.418
4.22	42244	12600	1.3	96.4	98.7	3.4	0.575
4.00	43043	13154	1.0	94.9	98.2	3.3	0.730
Global	327885	93562	2.3	98.2	98.2	3.5	0.316

Taula 3.4: Capes de reflexions. Estadístiques per capes del processament de les reflexions del cristall del tercer article (I422, plana 91), i d'un dels cristalls amb simetria C2. Dmin: resolució màxima de la capa. Nmes: Nombre de reflexions mesurades. Nref: Nombre de reflexions independents. I/sigma: $\langle I \rangle / \langle \sigma(I) \rangle$. Comp: Completesa. C.A.: Completesa acumulada. Mult: Multiplicitat. Rsym: Similitud entre reflexions (Veure text pla 41)

ferents orientacions del cristall. El mètode més utilitzat és de la rotació, en què es recullen les dades mentre el cristall va girant al voltant d'un eix perpendicular al raig incident. Aquesta rotació és petita, de l'ordre d'un grau, i permet recollir tota la intensitat que ha *escampat* la mosaïcitat. A continuació s'avança una quantitat similar a la de l'angle de rotació i es torna a recollir un altre cop oscil·lant al voltant d'aquesta orientació nova. Depenent

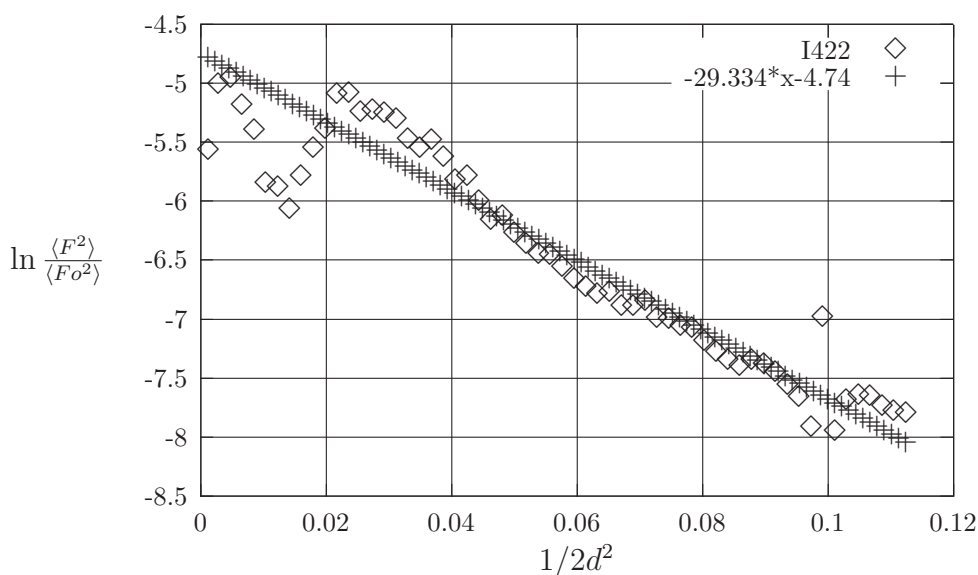


Figura 3.5: Wilson Plot. Ajustant per mínims quadrats una línia recta a la gràfica de $\ln(\frac{\langle F^2 \rangle}{\langle Fo^2 \rangle})$ enfront de $1/2d^2$ es pot calcular un B-factor mig pel cristall. Aquesta gràfica pertany al cristall del tercer article (I422, plana 91). En aquest cas $B = 29.33$.

del grup espacial i del tipus d'experiment, caldrà recollir entre 30 i 180 graus per a completar les dades.

L'experiment comença amb la recollida d'una o unes poques orientacions, les necessàries per determinar els paràmetres per la recollida completa: l'orientació del cristall, la cel·la unitat, la mosaïcitat, el rang d'angles que s'ha de cobrir i amb quina rotació. També s'ajusta la distància del detector al cristall, depenent de la resolució observada i de la qualitat de les reflexions de més alta resolució. Si són molt febles, és millor retirar el detector i recollir només les reflexions de baixa resolució amb bona qualitat i separació. També és important que les reflexions no es sobreposin, per la qual cosa s'ha de jugar amb la distància i la rotació. Aquestes reflexions no es podrien mesurar acuradament i s'haurien de descartar. Cal ajustar també el temps d'exposició de cada imatge, per tal de *resoldre* bé les reflexions i alhora no *cremar* ni el cristall ni saturar el detector.

Un cop preparades les característiques de l'experiment, es procedeix a la recollida de tot el conjunt de dades. Generalment es recullen els angles seqüencialment, per facilitar l'escalat de les dades, però es poden fer servir altres estratègies com recollir primer les zones que difracten millor, o bé, en

Estadístiques de la recollida	
Tipus	Natiu
Línia de difracció	id14EH3
Data	1999
Longitud d'ona (Å)	0.93103
Distància (mm)	600, 399.48, 179.48 (*)
Temperatura (K)	100
Detector	MAR CCD
N. d'imatges recollides	180, 74, 180
N. graus de rotació en cada imatge	1.0
Mosaïcitat (°)	0.3
Grup Espacial	I422
Cel·la unitat (Å, °)	$a = 155.751$ $b = 155.751$ $c = 160.845$ $\alpha = 90.0$ $\beta = 90.0$ $\gamma = 90.0$
Resolució (Å)	111.9 - 2.78 (2.93 - 2.78)
Completeness (%)	99.9 (99.9)
$\langle I \rangle / \langle \sigma(I) \rangle$	2.9 (5.3)
Multiplicitat	15.9 (14.6)
N. de reflexions observades	399674 (52778)
N. de reflexions <i>úniques</i>	25179 (3607)
Rmerge (%)	7.4 (27.7)
B-factor (Wilson plot, Å ²)	52.21

Taula 3.5: Estadístiques de la recollida i del processament dels cristalls I422. Els valor entre parèntesis es refereixen a la capa de més resolució. (*) tres conjunts de dades

el cas de MAD (secció 3.5.1, plana 42), recollir reflexions oposades (parells de Bijvoet, plana 43) per tal d'aconseguir dades que representin el cristall en un mateix estat.

Això també s'aplica en els canvis de longitud d'ona. Si s'ha recollit una mateixa zona (més l'oposada) en cada longitud d'ona, les dades recollides seran de major precisió. El problema és que amb aquest mètode, en fer moviments més grans del goniòmetre, s'augmenta l'error en l'angle de recollida (deriva) i que el canvi de longitud d'ona és un procés delicat fins i tot en les línies més avançades. Pot escalfar-se el monocromador, o bé desalinejar-se algun altre aparell de l'òptica, pel que cal evitar canvis innecessaris en la longitud d'ona. L'estratègia a seguir vindrà donada, en la pràctica, en les

I422Ag

Estadístiques de la recollida	
Tipus	Derivat Ag
Línia de difracció	EH2
Data	1999
Longitud d'ona (Å)	0.932
Distància (mm)	215.2
Temperatura (K)	100
Detector	MAR CCD
N. d'imatges recollides	108
N. graus de rotació en cada imatge	1.0
Mosaïcitat (°)	0.65
Grup Espacial	I4
Cella unitat (Å, °)	$a = 155.5$ $b = 155.5$ $c = 160.2$ $\alpha = 90.0$ $\beta = 90.0$ $\gamma = 90.0$
Resolució (Å)	64.7 - 2.75(*) (2.92 - 2.75)
Completeness (%)	94.6 (66.8)
$\langle I \rangle / \langle \sigma(I) \rangle$	5.7 (1.8)
Multiplicitat	4.4 (3.8)
N. de reflexions observades	206461 (20411)
N. de reflexions <i>úniques</i>	46788 (5324)
Rmerge (%)	7.5 (38.4)
B-factor (Wilson plot, Å ²)	72.2

Taula 3.6: Estadístiques de la recollida i del processament dels cristalls I422Ag. Els valor entre parèntesis es refereixen a la capa de més resolució. (*) Es van observar reflexions fins a 2.6Å

possibilitats que ofereixi la línia de difracció.

Per obtenir un bona estimació de la intensitat d'una reflexió, és necessari recollir-la diverses vegades. Generalment això s'aconsegueix augmentant el rang de rotació de cristall. En la mesura experimental, se sol considerar bona una multiplicitat mitjana de l'ordre de 4. No sempre és possible obtenir-la, sobretot en l'alta resolució, però cal tenir en compte que l'error en la mesura d'una reflexió ($\sigma(I_h)$) influenciarà totes les fases properes. En les reflexions de baixa resolució se solen obtenir valors molt alts d' $\langle I \rangle / \langle \sigma(I) \rangle$, però no és així en les que representen l'alta resolució. Generalment es fa servir aquest criteri per determinar a quina resolució màxima s'ha de recollir, que sol ser d' $\langle I \rangle / \langle \sigma(I) \rangle = 1$ o 2.

I422Au

Estadístiques de la recollida	
Tipus	Derivat Au
Línia de difracció	EH2
Data	1999
Longitud d'ona (Å)	0.932
Distància (mm)	309,5
Temperatura (K)	100
Detector	MAR CCD
N. d'imatges recollides	90
N. graus de rotació en cada imatge	1
Mosaïcitat (°)	1.1
Grup Espacial	I4
Cel·la unitat (Å,°)	$a = 155.6$ $b = 155.6$ $c = 159.2$ $\alpha = 90.0$ $\beta = 90.0$ $\gamma = 90.0$
Resolució (Å)	79.6-4.0 (4.22-4.0)
Completeness (%)	99.9 (99.7)
$\langle I \rangle / \langle \sigma(I) \rangle$	5.8 (5.1)
Multiplicitat	3.6 (3.6)
N. de reflexions observades	57464 (8402)
N. de reflexions <i>úniques</i>	16052 (2326)
Rmerge (%)	9.0 (13.3)
B-factor (Wilson plot, Å ²)	21.35

Taula 3.7: Estadístiques de la recollida i del processament dels cristalls I422Au. Els valor entre parèntesis es refereixen a la capa de més resolució.

El procés de la indexació ja s'ha fet en recollir les primeres imatges del cristall. Molt breument, consisteix a determinar el grup espacial, l'orientació del cristall i la cel·la a partir de la distribució de reflexions d'un grup petit d'imatges. Donades les característiques físiques de l'experiment (distància del detector, longitud d'ona, polarització del feix, etc) i de les de la imatge (centre, resolució, separació entre reflexions, etc.) es pot predir la mida i orientació de la cel·la unitat:

La projecció de les reflexions en el detector forma unes *llunes* que permeten ordenar les reflexions i calcular la orientació del cristall. A una longitud d'ona fixa, la distància entre les reflexions és inversament proporcional a la mida de la cel·la unitat i proporcional a la distància entre el detector i el cristall. Un cop predita una cel·la i una orientació, cal comprovar si és correcta,

I422Pt

Estadístiques de la recollida	
Tipus	Derivat Pt
Línia de difracció	EH2
Data	1999
Longitud d'ona (Å)	0.932
Distància (mm)	309.7
Temperatura (K)	100
Detector	MAR CCD
N. d'imatges recollides	100
N. graus de rotació en cada imatge	1
Mosaïcitat (°)	1.1
Grup Espacial	I4
Cel·la unitat (Å, °)	$a = 155.4$ $b = 155.4$ $c = 158.7$ $\alpha = 90.0$ $\beta = 90.0$ $\gamma = 90.0$
Resolució (Å)	79.4-4.0 (4.22-4.0)
<i>Completeness</i> (%)	96.7 (98.2)
$\langle I \rangle / \langle \sigma(I) \rangle$	3.6 (4.4)
Multiplicitat	3.0 (3.0)
N. de reflexions observades	46075 (6729)
N. de reflexions <i>úniques</i>	15436 (2279)
Rmerge (%)	8.3 (13.4)
B-factor (Wilson plot, Å ²)	73.6

Taula 3.8: Estadístiques de la recollida i del processament dels cristalls I422Pt. Els valor entre parèntesis es refereixen a la capa de més resolució.

tractant de predir el patró en altres imatges no fetes servir en la indexació. Moltes mesures de la fiabilitat dels resultats, tant en la indexació com en càlculs posteriors es basen en la predicció de reflexions no fetes servir en els càlculs.

Un cop acabada la recollida de totes les dades, a partir de la indexació, es podrà predir la situació de cada reflexió en cada una de les imatges recollides. La integració consisteix a escanejar la quantitat d'energia dipositada en cadascuna de les reflexions. Per això es limita cada reflexió dins d'una capsa i s'ajusta el perfil de la reflexió a una gaussiana. Aquesta capsa està formada per dues zones, la que pertany pròpiament a la reflexió, i la que forma el soroll de fons de cada reflexió, que es farà servir per modular l'error de la mesura. No es pot fer servir un soroll de fons general per a tota la

I422TaBr

Estadístiques de la recollida			
Tipus	MAD Ta_6Br_{12}		
Línia de difracció	bm30		
Data	1999		
Longitud d'ona (Å)	1.2651	1.2565	1.2559 1.2435
Distància (mm)	350.0		
Temperatura (K)	100		
Detector	MAR 345		
N. d'imatges recollides	30+60(*)		
N. graus de rotació en cada imatge	1.0		
Mosaïcitat (°)	0.4		
Grup Espacial	I422		
Cel·la unitat (Å,°)	$a = 156.085$	$b = 156.085$	$c = 161.080$ $\alpha = 90.0$ $\beta = 90.0$ $\gamma = 90.0$
Resolució (Å)	55.9 - 3.0		(3.16 - 3.0)
Completeness (%)	99.2		(99.8)
$\langle I \rangle / \langle \sigma(I) \rangle$	4.9		(2.8)
Multiplicitat	5.9		(5.9)
N. de reflexions observades	117274		(16973)
N. de reflexions úniques	19990		(2890)
Rmerge (%)	9.2		(26.4)
B-factor (Wilson plot, Å ²)	56.25		

Taula 3.9: Estadístiques de la recollida i del processament dels cristalls I422TaBr. Els valor entre parèntesis es refereixen a la capa de més resolució. (*) Es va recollir en dues parts a causa de que estava maclat

imatge per que aquest és molt variable, principalment amb la resolució. El soroll de fons en la zona de baixa resolució sol ser més alt que la intensitat de les reflexions d'alta resolució. Per això cal que el detector tingui un rang dinàmic molt alt. La presència de solvent desordenat en els cristalls provoca anells difusos que augmenten el soroll de fons.

El resultat de la integració és un arxiu per a cada imatge recollida on s'especifiquen les coordenades de la reflexió, la intensitat integrada, el seu σ , i altres característiques que seran necessàries per escalar-les posteriorment: angle de rotació, quina proporció de la reflexió està tota continguda en la imatge, les coordenades en el difractòmetre, etc.

L'escalat és la etapa del processament de les dades en la qual es tracta

d'ajuntar totes les reflexions repetides que s'han recollit i de reconstruir les que estan disperses en varies imatges, a causa de la mosaïcitat. Es fan servir les reflexions que estan senceres en una imatge i les equivalents en altres imatges per poder calcular un factor d'escala entre elles. Si el cristall té molta mosaïcitat, es pot intentar escalar també a partir de les reflexions parcials. Petites diferències en el temps de recollida, en la intensitat del feix, o per degradació de la mostra fan que calgui aplicar una escala entre elles, i obtenir un perfil de l'evolució de la difracció. Es per tant important definir un model d'escalat que tingui en compte si s'ha de fer fent servir el soroll de fons per determinar com evoluciona la difracció, o bé si cal un escalat independent de cada reflexió i de les seves simètriques. També cal fer separar els parells de Bijvoet si la difracció és anòmala i es volen fer servir les diferències entre aquests per resoldre l'estructura.

La qualitat de l'escalat es jutja a partir d'uns residuals que avaluen la desviació de les reflexions, ordenades per capes de resolució, i també per intensitats. Valors baixos indicaran que les reflexions relacionades per simetria, o per què s'han recollit diverses vegades, són similars i per tant es tracta de reflexions ben processades o recollides.

La similitud entre reflexions recollides diverses vegades es mesura amb un residual anomenat *Rmerge*:

$$R_{merge} = \frac{\sum_{i,h} I_{i,h} - \langle I_h \rangle}{\sum_h \langle I_h \rangle}$$

on $I_{i,h}$ és la mesura i de la reflexió h , i $\langle I_h \rangle = \frac{\sum_i I_{i,h}}{\sum_i 1}$.

Anàlogament, *Rsym* és un indicador de la similitud entre les reflexions que estan relacionades pel grup espacial. L'expressió és idèntica, només que l'índex i representa les reflexions simètriques.

Ja s'ha mencionat que una altra mesura de la qualitat és $\langle I \rangle / \langle \sigma(I) \rangle$, també tabulada en capes i per intensitats, que ajuda a discernir fins a quina resolució la difracció és prou intensa, o si hi ha àrees cremades o no aprofitables per culpa de la presència d'un anell de solvent. La última quantitat important és la completa, distribuïda per capes de resolució, que mostra quin tant per cent del total de reflexions s'han recollit en cadascuna d'elles (taula 3.4).

Una reflexió vindrà caracteritzada per 3 coordenades ($\mathbf{h} = (h, k, l)$), una intensitat $I_{\mathbf{h}}$ i una mesura de la seva qualitat $\sigma(I_{\mathbf{h}})$. En el cas d'un experiment de MAD, també s'hauran d'incloure per separat els valors dels parells de Bijvoet i les seves *sigmes*.

Generalment a l'hora d'afinar una estructura es treballa amb les amplituds, que són proporcionals a l'arrel quadrada de les intensitats. En un últim

pas del processament de les reflexions cal aplicar aquesta proporció i alhora s'aprofita per calcular el *Wilson Plot*, que és una estadística de la intensitat mitjana de les reflexions en cada capa de resolució. El pendent mitjà a la zona central indicarà el valor mitjà de la temperatura (*B-factor*) o desordre del cristall. Aquest valor mig s'ha de conservar en l'estructura afinada final. El *Wilson Plot* permet també aplicar una escala absoluta a les reflexions.

Si els àtoms estan distribuïts aleatòriament en la unitat asimètrica, llavors

$$\langle F^2 \rangle = k \langle Fo^2 \rangle e^{-2B \frac{\sin^2(\theta)}{\lambda^2}}$$

on θ és l'angle de difracció segons la llei de Bragg

$$2d \sin(\theta) = n\lambda$$

amb d la resolució i 2θ l'angle entre el raig incident i el difractat (fig. 3.5) [1].

Els àtoms no estan distribuïts aleatòriament en estructures reals. Aquest efecte s'aprecia principalment amb les reflexions de baixa resolució. Normalment es fan servir només les reflexions d'alta resolució per a calcular el *Wilson Plot* (fig. 3.5).

De cara a treballar amb mètodes de màxima versemblança, és important tenir un conjunt de dades complet, encara que no s'hagi mesurat tot. Les reflexions que falten s'afegeixen, junt amb un senyal que indiqui que no s'han observat (*missing flag*) i se n'assigna a tota la part observada un altre que indicarà si la reflexió s'haurà de separar o considerar a l'hora de afinar, per poder calcular l'*Rfree* ("reflexions *Rfree*", plana 58).

3.5 Solució de l'estructura tridimensional

3.5.1 MIR i MAD

Quan no hi ha cap model de partida (una estructura similar), el problema de les fases (plana 30) obliga a reduir l'estructura a un nombre petit d'àtoms i tractar d'afrontar el problema amb mètodes directes o de Patterson. Els àtoms pesants tenen una contribució més gran en la difracció, i presenten anisotropicitat quan la longitud d'ona dels raigs X correspon a un dels seus *edges* d'absorció.

La tècnica del MIR (*Multiple Isomorphous Replacement*) és de les més antigues que es coneix, i consisteix en recollir per separat dades de difracció de diferents cristalls isomorfs (mateixa cel·la i mateix empaquetament) que contenen àtoms pesants afegits. En principi calen un mínim de tres

cristalls (hi ha variants, plana 46). Un pot ser el natiu i els altres dos han de ser derivats diferents. Si els cristalls són prou isomorfs i es poden posar les intensitats difractades de cadascun en una mateixa escala, una síntesi de fourier de les diferències entre les intensitats del natiu i cadascun dels derivats donarà lloc a un mapa (*Patterson* plana 47) amb densitats altes en els punts corresponents a les distàncies entre els àtoms pesants. Com que és una diferència entre dos cristalls que només difereixen pels àtoms pesants, les distàncies entre els àtoms de la proteïna no apareixeran. Per això és important l'isomorfisme. Per consideracions amb el grup espacial, les diferències apareixeran formant uns motius (plans i línies de Harker) que permetran el càlcul de la posició dels àtoms pesants. La relació amb les fases de la proteïna és directa, però cal un segon cristall derivat perquè hi ha una incertesa en la fase, ja que d'aquesta relació només pot trobar el cosinus de la fase:

$$\phi_P = \phi_{H1} + \cos^{-1} \left[\frac{F_{PH1}^2 - F_P^2 - F_{H1}^2}{2F_P F_{H1}} \right]$$

$$\phi_P = \phi_{H2} + \cos^{-1} \left[\frac{F_{PH2}^2 - F_P^2 - F_{H2}^2}{2F_P F_{H2}} \right]$$

on F són els factors d'estructura i ϕ són les fases. El subíndex P representa el cristall de proteïna sol, $PH1$ i $PH2$, els cristalls de proteïna i l'ió pesant, i $H1$ i $H2$ les amplituds i fases calculades pels àtoms pesants, a partir de la seva posició en el cristall.

El principal problema del MIR és la necessitat d'obtenir dos derivats d'una proteïna nativa, i que a més hi siguin isomorfs. Si es pot aconseguir una radiació a una longitud d'ona corresponent a l'*edge* d'absorció d'un ió pesant, llavors el problema es redueix a obtenir un sol cristall derivat amb l'àtom pesant, i optativament, un natiu si cal verificar que no hi hagi hagut modificacions importants en l'estructura. Això podria passar, per exemple, si l'àtom pesant interacciona amb la proteïna en aminoàcids actius. A aquestes longituds d'ona els electrons absorbeixen l'energia dels fotons per canviar a una òrbita superior i, per després emetre fluorescència en tornar al nivell energètic original. El resultat és que la difracció es torna anisòtropa (anòmala), i les reflexions oposades (parells de Friedel) ja no són d'igual intensitat. En aquest cas els parells de reflexions, i els seus simètrics, s'anomenen parells de Bijvoet.

Un experiment de MAD (*Multiwavelength Anomalous Dispersion*) consisteix a recollir dades de difracció del cristall més d'un cop (típicament tres), cada vegada a una longitud d'ona diferent. Les longituds d'ona es trien de manera que una d'elles estigui en el màxim d'absorció, l'altra en el punt

d'inflexió i l'altra després d'aquest màxim (figs. 3.6, plana 44 i 3.7, plana 45).

Per comprovar la presència de l'àtom pesant dins del cristall, i calcular quina és l'energia d'absorció, se sotmet el cristall a un rang d'energies variable i es mesura la fluorescència produïda (XAFS, X ray Absorption Fine Structure). Aquesta fluorescència és màxima en l'*edge* d'absorció. La fluorescència és una radiació isotròpica que no depèn de l'estat cristal·lí, per la qual cosa es pot produir tant per la presència d'àtoms pesants en la gota que protegeix el cristall com per per àtoms pesants desordenats dins del cristall (fig. 3.6).

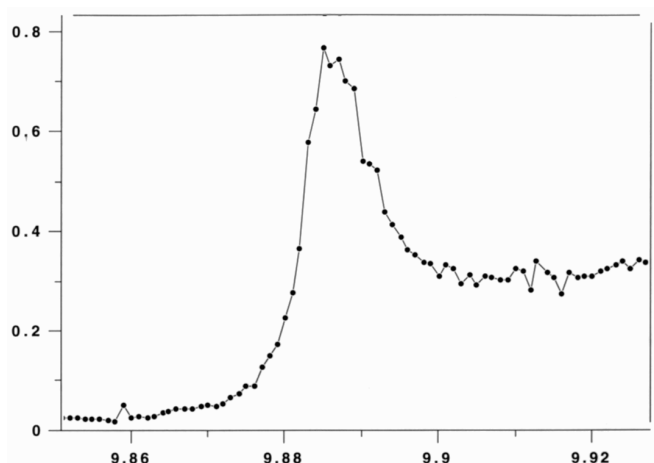


Figura 3.6: Fluorescència TaBr. Màxim d'absorció per un cristall del connector derivat amb TaBr ($1422 Ta_6Br_{12}$). A l'eix horitzontal es representa el rang d'energies en KeV i al vertical la fluorescència en unitats relatives.

A partir de les dades recollides, es poden tractar els resultats com en un experiment de MIR, substituint les diferències entre cristalls, per diferències entre longituds d'ona i entre parells de Bijvoet. Els factors de difusió atòmics s'han d'ajustar a les propietats de l'àtom pesant en cada longitud d'ona.

Un factor d'estructura en una dispersió anòmala es pot separar en dues parts,

$$F(\mathbf{h}) = F'(\mathbf{h}) + iF''(\mathbf{h})$$

(fig. 3.7) on F' representa la contribució no anòmala (proteïna i àtoms pesants) i F'' l'anòmala, a causa només dels àtoms pesants (o dispersor anòmal). A causa de l'anisotropicitat hi haurà un factor d'estructura diferent per a cada parell de reflexions oposades. Aquesta contribució es pot trobar en taules, o millor per fluorescència, ja que l'entorn molecular la modifica una mica. Si F^+ i F^- són les amplituds mesurades per a cada parell de Bijvoet, es pot calcular el sinus de la fase i trencar l'ambigüitat aprofitant una de les

expressions d'abans pel MIR:

$$\phi_{PH} = \phi_H \pm \sin^{-1} \left[\frac{F_{PH}^+ - F_{PH}^-}{2F_H''} \right]$$

$$\phi_P = \phi_H + \cos^{-1} \left[\frac{F_{PH}^2 - F_P^2 - F_H^2}{2F_P F_H} \right]$$

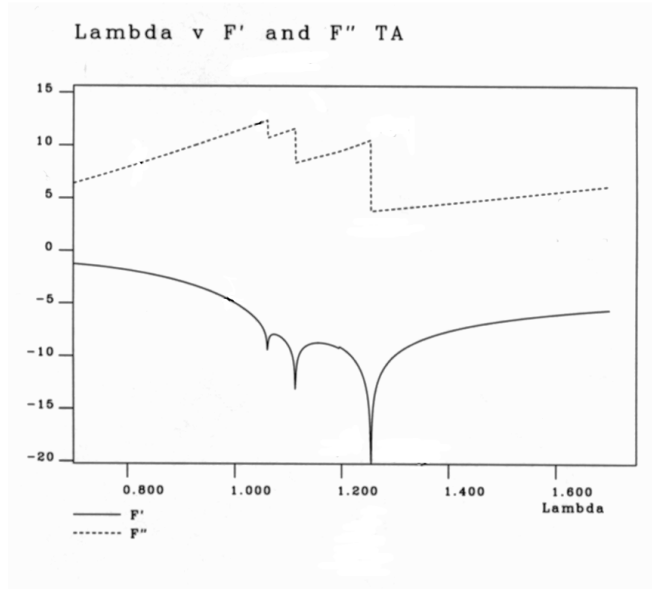


Figura 3.7: Factors de difusió atòmics. A partir de la corba 3.6 es poden calcular els factors de difusió atòmics dels àtoms pesants. Un factor de difusió al voltant d'un *edge* d'absorció té la forma $f = f_0 + \Delta f' + i f''$ on f_0 és la contribució *normal* i $\Delta f'$ i f'' són, respectivament les parts real i imaginària degudes al moment dipolar dels electrons que oscil·len entre dos nivells energètics. Els factors d'estructura F' i F'' s'obtenen agrupant les parts real i imaginària en fer la suma sobre tots els àtoms. A la figura F' i F'' representen $\Delta f'$ i f'' . Els coeficients d'absorció $\mu(E)$ de la corba 3.6, estan relacionats amb $f''(E)$ segons $\mu(E) = (4\pi N \hbar e^2 / mE) f''(E)$ (N és el nombre atòmic, m i e són la massa i la càrrega de l'electró). A partir d'una de les relacions de Kramers-Krönig s'obté l'altre factor de difusió: $f'(E) = \frac{2}{\pi} P \int_0^\infty \frac{E' f''(E')}{E^2 - E'^2} dE'$ amb P , el valor principal de la integral a la singularitat, i es suposa que no hi ha dependència entre f' i f'' respecte la direcció en l'espai recíproc [40].

Com que es tenen diferències entre longituds d'ona i entre parells de reflexions, per calcular el Patterson es poden fer més combinacions que en

el cas del MIR. Per això, de vegades un experiment de MAD es pot resoldre amb menys longituds d'ona, si les dades són prou bones. Un requisit per fer un experiment de MAD és disposar d'una línia de sincrotró que permeti ajustar la longitud d'ona. En un aparell d'ànode rotatori no es poden fer aquest tipus d'experiments ja que la longitud d'ona és fixa.

Per la similitud entre els dos mètodes, existeixen altres combinacions com el SIRAS (*Single Isomorphous Replacement and Anomalous Scattering*), on es fa servir un natiu i un derivat que tingui un àtom pesant amb contribució anòmala significativa. A la pràctica, però, el que s'intenta és preparar alhora cristalls nadius, derivats normals i derivats per MAD i combinar els que s'hagin pogut obtenir i fer un experiment de MIRAS (*Multiple Isomorphous Replacement and Anomalous Scattering*). La qualitat (o l'existència) de la difracció determinarà *a posteriori* el mètode a utilitzar per resoldre l'estructura.

3.5.2 Reemplaçament Molecular

El reemplaçament molecular és la tècnica que més es fa servir actualment per a resoldre el problema de les fases. A mesura que el nombre d'estructures conegudes augmenta, el nombre de plegaments o estructures terciàries noves augmenta a un ritme cada cop més lent. A l'any 1999 només un 9% de les molècules resoltes contienien un *folding* nou [34]. El reemplaçament molecular és capaç de determinar l'orientació i posició d'una proteïna en el cristall a partir de l'estructura tridimensional d'una altra proteïna o fragment similar, una solució de RMN (ressonància magnètica nuclear), o d'un model de baixa resolució provinent de microscòpia electrònica, que és el que s'ha intentat en aquest treball.

Consisteix en la cerca sistemàtica per correlació entre les intensitats difractades i les que produiria el model, tractat com un cos rígid, en successius empaquetaments en la cel·la cristal·logràfica. Es tracta d'un problema de 6 dimensions per molècula en la unitat asimètrica (tres rotacions i tres translacions), per la qual cosa cal dividir-lo en cerques més senzilles, o seria impracticable el temps de CPU necessari, ja que per a cada empaquetament cal calcular els factors d'estructura, per FFT, i després calcular-ne la correlació amb les intensitats observades.

El problema es divideix en dos. La primera part consisteix en una cerca de l'orientació de cada molècula model posada a l'origen de la cel·la. Molt bàsicament, es calculen les intensitats per cadascuna de les rotacions i es correlacionen amb les observades. Els valors de millor correlació correspondran a l'orientació de les molècules en el cristall. Es treballa amb intensitats per que la transformada de Fourier d'aquestes dona lloc un mapa que és l'auto-

convolució de la cel·la unitat. En aquests mapes es mantenen les orientacions de la molècula, però totes les translacions es desplacen a l'origen (mapes de Patterson):

$$P(\mathbf{u}) = \int_V \rho(\mathbf{r})\rho(\mathbf{r} + \mathbf{u})d\mathbf{r} = \frac{1}{V} \sum_{\mathbf{h}} |F_{\mathbf{h}}|^2 \cos(2\pi\mathbf{h} \cdot \mathbf{u})$$

\mathbf{u} (vector de Patterson) representa totes les distàncies interatòmiques, traslladades a l'origen. Es calcula una funció de rotació

$$\mathcal{R}(R) = \frac{1}{V} \int_V P(\mathbf{u})P_m(\mathbf{R}^{-1}\mathbf{u})d^3\mathbf{u}$$

entre el Patterson natiu de les reflexions (P) i el del model situat a l'origen de coordenades i en diverses orientacions (P_m).

De tots els valors de correlació, corresponents a orientacions de la molècula s'extreuen tantes orientacions com molècules en la unitat asimètrica. Generalment s'observen grups de solucions similars que corresponen a petites diferències en l'orientació, a causa del soroll. El nombre de molècules per unitat asimètrica es pot estimar senzillament a partir del volum de Matthews (V_m):

$$1.66 \leq V_m = \frac{V(\text{\AA}^3)}{M(Da)N_{asu}N_{mol}} \leq 4$$

on V és el volum de la cel·la en \AA^3 , M és la massa de la proteïna en Dalton, N_{asu} és el nombre d'unitats asimètriques i N_{mol} és el nombre de molècules per unitat asimètrica. Es tracta d'una llei empírica corresponent a cristalls amb un rang del 30% al 75% de solvent, pel que l'última paraula la té l'empaquetament trobat.

Un cop resolt el problema de les orientacions, s'aplica un procediment similar per les translacions. Aquest és més ràpid, per que una translació només implica un canvi en la fase de les amplituds calculades a partir de les rotacions. Es trasllada, tridimensionalment, primer una de les molècules fins arribar a la millor correlació, i després es fixa en aquesta posició per tenir-la en compte en el càlcul de les amplituds de les següents molècules. Els factors d'estructura de cada molècula i orientació s'han calculat prèviament en la rotació. Si apliquem una translació T a una molècula orientada segons R , els nous factors d'estructura seran:

$$F_{\mathbf{h}}^c(\mathbf{T}) = \sum_s^{N_{asu}} g_s(\mathbf{h})e^{2\pi i\mathbf{h}R\mathbf{T}}$$

on $g_s(\mathbf{h})$ és un factor d'estructura global calculat prèviament per a cadascuna de les N_{asu} molècules de la unitat asimètrica.

I422, un cos										
	rz	ry	rz	tx	ty	tz	CCF	RF	CCI	Dist
cos1	38.5	-0.9	-18.0	0.0023	1.0011	0.1906	54.6	44.6	59.6	0.3
cos1	57.6	-0.8	-7.1	0.0012	0.0016	0.1909	53.7	44.9	59.8	0.4

C2, tres cossos										
	rz	ry	rz	tx	ty	tz	CCF	RF	CCI	Dist
cos1	89.7	89.7	194.8	0.1934	-0.0004	0.3487	37.4	46.1	42.7	98.6
cos2	90.4	89.8	295.1	0.0277	0.5041	0.6819	53.4	40.9	62.5	59.7
cos3	89.7	90.2	65.1	0.3592	0.5034	0.0149	69.2	35.0	75.8	59.4

Taula 3.10: Reemplaçament Molecular. Solucions del reemplaçament molecular del cristall del tercer article. (I422, plana 91), i d'un dels cristalls amb simetria C2. La posició de la molècula model està definida per tres angles d'Euler (rz,ry,rz) i tres translacions (tx,ty,tz) respecte al model situat a l'origen de coordenades i amb l'eix principal d'inèrcia perpendicular al pla a-b. CCF és el coeficient de correlació entre les Fo i les Fc del model, RF és l'Rfactor entre ells i CCI el coeficient de correlació calculat amb les intensitats. Dist és la distància (en Å) entre els centres de massa de molècules simètriques. A la taula I422 es mostren les dues millors solucions fent servir un model del connector sencer [48]. Com que cada connector ocupava 4 unitats asimètriques (3 proteïnes p10 per unitat asimètrica i rotades 30°), es van trobar dues solucions equivalents rotades 30°. La tercera solució no es va trobar, segurament a causa de que aquest model no és totalment simètric. La recerca es va fer a una resolució de 4 Å i fixant el centre al voltant de l'origen. Les distàncies entre centres de massa són nul·les a causa de que les solucions es sobreposen. A la taula C2 es mostra una única solució amb tres connectors per unitat asimètrica (una en cada línia). En aquesta cel·la hi havia espai per a tres connectors per unitat asimètrica. La recerca es va fer a 4.5 Å amb el model I422 afinat. A part d'aquesta solució es van trobar moltes solucions equivalents, cadascuna corresponent a diferents rotacions de múltiples de 30° del connector. En aquesta segona taula es pot veure com la correlació i l'Rfactor milloren, augmentant i disminuint, respectivament en afegir més cossos.

El resultat de les rotacions i translacions és un conjunt de solucions possibles. Moltes es poden descartar per un mal empaquetament, però la funció de correlació és molt sensible a petites diferències en la posició, i cal fer un afinament final de les 6 dimensions, i de cada molècula, al voltant de les millors posicions i orientacions (afinament del sòlid rígid).

La correlació està calculada com:

$$C = \frac{\sum_h (|Fo_h| - \langle |Fo| \rangle) (|Fc_h| - \langle |Fc| \rangle)}{[\sum_h (|Fo_h| - \langle |Fo| \rangle)^2 \sum_h (|Fc_h| - \langle |Fc| \rangle)^2]^{\frac{1}{2}}}$$

on F_o representa el factor d'estructura observat, i F_c , el calculat pel Reemplaçament Molecular.

Un altre valor, l'*Rfactor* es farà servir també en l'afinament posicional. Representa la desviació dels valors d'amplitud calculats per la proteïna que s'està modelant respecte als observats:

$$R = \frac{\sum_h ||Fo_h| - |Fc_h||}{\sum_h |Fo_h|}$$

Per tant, és un valor que ha de disminuir amb la qualitat del model. Els valors típics finals d'*Rfactor* en estructures proteïques afinades oscil·len entre el 20% i el 30%.

A la taula 3.10 es mostren les solucions de reemplaçament molecular de dos cristalls diferents del connector. El primer (I422) correspon al presentat en aquesta tesi (tercer article, plana 91). El segon (C2), a un dels cristalls obtinguts al principi (primer article, plana 77), que s'ha pogut resoldre *a posteriori* a partir de la solució afinada. Aquest segon model no s'ha afinat, ja que la qualitat i resolució de les dades no aportaria detalls nous al model actual (taula 3.4, plana 34).

3.5.3 NCS

Quan una proteïna està composta de diverses parts idèntiques, és molt freqüent que aquestes estiguin relacionades per una operació de simetria local dins de la unitat asimètrica. Això passa quan l'operació de simetria no pertany a un grup espacial, o bé quan l'empaquetament està relaxat, i li falta poc per constituir grup espacial (veure la figura 3.1 de la plana 22).

En aquest casos es diu que hi ha *Simetria No Cristal·logràfica* (NCS). La NCS és d'una gran ajuda en totes les etapes de la resolució d'una estructura:

- A la difracció provoca relacions entre reflexions més enllà de les del grup espacial. Això permet trobar les direccions dels eixos de simetria de la molècula amb la funció d'autorotació. La funció d'autorotació és la correlació del Patterson de les reflexions amb el mateix Patterson, amb una rotació diferent. A les planes 86 i 87 es poden veure dos exemples de mapes d'aquest tipus.

- Per solucionar una estructura, parts que no s'han resolt bé poden ser completades gràcies a la informació de parts simètriques. Aquesta tècnica forma part de l'extensió de fases, que es comentarà més endavant (secció 3.5.5, plana 51).
- En l'afinament de la estructura, la NCS relaciona cadenes peptídiques, disminuint el nombre de paràmetres lliures per afinar.

Hi ha dos tipus d'NCS, l'estricta o rígid, definit per les matrius que relacionen les parts repetides, i el restringit, on només s'imposa que la “forma” de la cadena d'aminoàcids ha de ser el més igual possible. En aquest cas, enlloc de fer servir, per exemple, les coordenades (i temperatures, torsions, etc) per a cada aminoàcid, es fan servir les d'un dels relacionats, més les diferències amb els altres. Aquestes diferències es deixen créixer segons un pes, que pel cas de simetria NCS estricta, seria infinit i per no simetria seria de zero.

3.5.4 Màscares

Una màscara és equivalent a un mapa de densitat, però composta només d'uns i zeros. Multiplicant els punts equivalents dels mapes de densitat i de les màscares s'obtenen mapes on només hi ha les zones on la màscara val 1.

Una màscara és necessària quan cal aplicar certes operacions a una part d'una estructura. Per modular l'efecte del solvent en les amplituds (plana 60), es pot fer servir una màscara que separi la part que és “cristal·logràfica” de la que està composta només de solvent desordenat. Es tracta d'una operació iterativa que forma part del cicle de l'afinament, ja que aquesta màscara variarà lleugerament a cada cicle. La idea consisteix a calcular quina seria l'aportació de la part del solvent en la difracció, calculant uns factors d'estructura (Fbs) a partir d'un mapa on només hi ha un valor constant de la densitat ($0.36 e^-/\text{Å}^3$). Aquesta aportació es calcula només per a les reflexions de baixa resolució, fins uns 6 o 8 Å, ja que és fins aquest límit quan es pot observar l'efecte del solvent desordenat. Aquest límit és el que fa que el *Wilson Plot* (fig. 3.5, plana 35) perdi el comportament rectilini.

La màscara es construeix perfilant la proteïna, a una distància d'uns 3 Å, i a la màxima resolució. Per això cal que sigui iteratiu, ja que al llarg de l'afinament és la part d'alta resolució la que variarà més. A l'hora de calcular els *Rfactors*, es sumen les Fbs a les Fc per aproximar millor el valor a Fo .

En la extensió de fases amb simetria no cristal·logràfica (NCS, plana 51) és necessari promitjar les densitats en zones relacionades en una unitat asimètrica. Aplicant la màscara es pot aïllar cadascuna d'aquestes zones, normalment un monòmer, i llavors, aplicar les transformacions necessàries

per sobreposar-les i promitjar-les. Una altra aplicació en NCS és la de generar la unitat asimètrica a partir d'una zona que es consideri millor definida. Es tracta de fer el mateix, però enlloc de promitjar, es copien els valors d'un dels mapes emmascarats.

Per construir les màscares se sol partir d'un model d'àtoms, però en el nostre cas, es partia d'un model de microscòpia electrònica. Les màscares inicials es van construir d'una manera similar al model de microscòpia (plana 69), però tallant un sector cilíndric de 30° , seguint la forma d'un monòmer. Més endavant, en estendre les fases, la resolució del model de microscòpia no era suficient. Es va fer servir una màscara editada manualment a partir dels mapes a aquella resolució o bé una màscara construïda a partir d'un model de pseudoàtoms que cobria una unitat asimètrica. Aquest mètode està explicat a la secció de materials i mètodes del tercer article (plana 101).

3.5.5 L'extensió de fases

L'extensió de fases és un procediment iteratiu que intenta obtenir fases a una resolució més alta de la que es disposa amb només un model o amb un traçat fet a partir de fases experimentals a menor resolució.

El típic cas és amb cristalls de derivats. El fet d'introduir àtoms pesants en un cristall, moltes vegades provoca un empitjorament de la qualitat del cristall, i, per tant de la difracció.

Un altre cas, el fet servir per aquest treball, és el d'estendre fases a partir d'un model de molt baixa resolució. Aquest model pot ser una estructura de microscòpia electrònica, o, fins i tot un model geomètric, com un icosaèdre en el cas de càpsides de virus, si es parteix de molt baixa resolució (menys que 10 o 20 Å).

La tècnica consisteix en construir un model que contingui la informació esperada a baixa resolució. Per exemple, un model de microscòpia tindrà el perfil de la proteïna, i una certa distribució de densitat a l'interior corresponent a zones més denses de la molècula. Amb aquest model es construeix una cel·la igual que l'experimental, aprofitant totes les dades que es puguin obtenir a partir de la síntesi de Patterson[6], l'empaquetament, la funció d'autorotació, o de micrografies electròniques de la mateixa cel·la. S'ha d'intentar ajustar aquest model perquè la seva correlació amb les reflexions de baixa resolució sigui la millor possible, el que és un problema no gaire ben resolt, ja que la informació del solvent desordenat s'obté, com hem vist abans, a partir de l'alta resolució (plana 50).

Un cop s'ha obtingut un bon ajust del model i de les reflexions, s'intenten extrapolar les fases a una resolució lleugerament més alta, per transformació de Fourier del model. Si pot ser, que inclogui només una reflexió nova, o

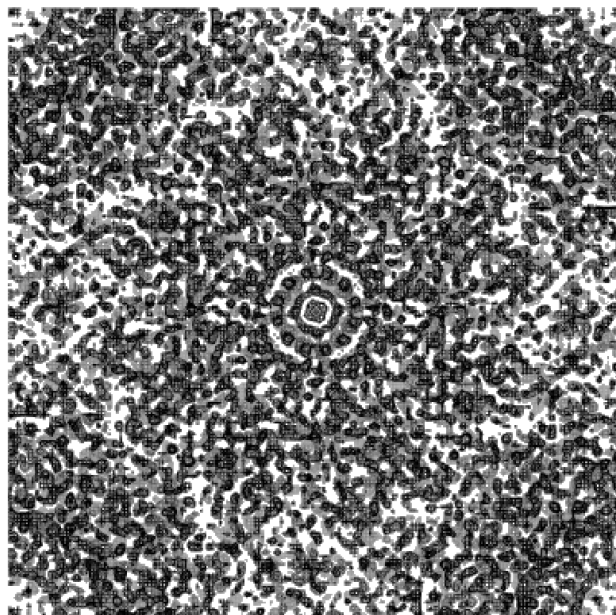


Figura 3.8: Extensió de fases I. Projectió al llarg de l'eix c del mapa resultant de l'extensió de fases explicada a la plana 91 [42].

almenys només una capa de resolució ben petita. L'extrapolació es fa per un procés iteratiu, on, en cada increment de resolució, es fa una modificació de la densitat del model per que tingui una distribució adequada a aquella resolució, comparant amb distribucions d'altres proteïnes (*histogram matching*).

Una altra modificació molt efectiva és la de l'aplanament de solvent. La densitat corresponent a la zona de solvent desordenat es posa a un valor constant per augmentar el contrast entre proteïna i solvent.

La més efectiva de les modificacions a fer en cada increment de resolució és la del promitjat de la simetria no cristal·logràfica. Per això cal, és clar, que la molècula tingui aquesta simetria, però en grans molècules sol ser el cas. En casos com aquests, la baixa resolució inicial permet usar màscares tal com abans s'ha explicat, ja que les diferències entre els monòmers no se solen apreciar fins mitja o alta resolució.

La densitat es pot modificar també per una millora de les màscares en cada iteració. També per una esqueletonització, que consisteix en connectar parts que tenen una densitat més alta i intentar seguir el traç de la proteïna. I també per mètodes de màxima versemblança a l'espai real, semblants als fets servir per l'afinament de l'estructura (plana 60), però substituint el paper de les reflexions "Rfree" per petites esferes de densitat [2].

Un cop s'han aplicat totes les modificacions de densitat adequades, es calculen les Fc del model i es comparen amb les experimentals. Si la tendència

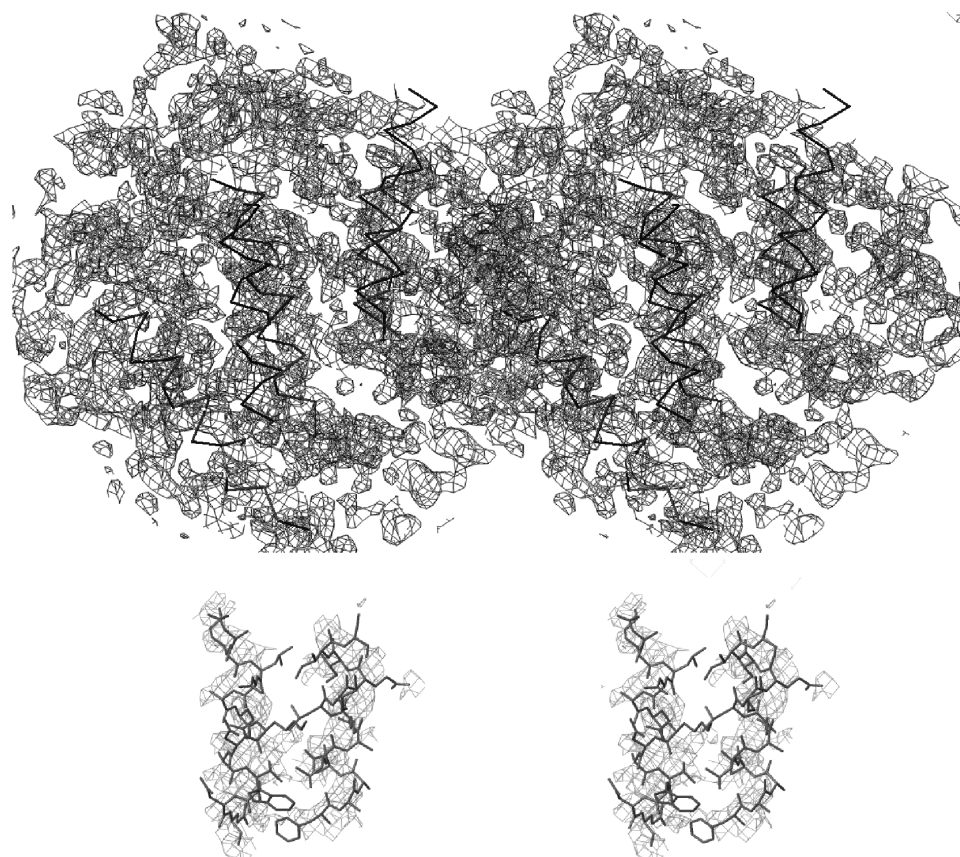


Figura 3.9: Extensió de fases II. Detall del mapa resultant de l'extensió de fases, corresponent a les hèlixs del tronc central del connector. A la imatge superior, mapa estèreo d'aquesta zona amb la cadena principal de les hèlixs de l'estructura final traçada amb barres. A la inferior, detall d'una d'aquestes hèlixs amb cadenes laterals.

no és bona, es poden afegir més ajustos, si és bona, es calculen les amplituds i les fases a una resolució una mica més alta, i es tornen a comparar amb les experimentals. En el cas que siguin bones, es prendran aquestes fases noves per a un proper cicle.

La mesura de la bondat de les reflexions noves es fa a partir de la versemblança tal com s'explica a l'afinament (secció 3.6.2, plana 60). En aquest cas, però, en lloc de predir les reflexions “*R_{free}*”, es tracten de predir les amplituds de les reflexions a la propera capa de resolució. La precisió d'aquesta predicció queda reflectida en la *figura de mèrit* (FOM), que és una constant que multiplicarà a cada reflexió a l'hora de construir els mapes. Una FOM

baixa, fa que la reflexió tingui poc pes en la síntesi de Fourier, i a més informa al programa que és una reflexió (i una fase) que cal millorar en el proper cicle.

L'aplicació pràctica d'aquesta extensió de fases està explicada a materials i mètodes del tercer article (plana 91).

3.5.6 El traçat de l'estructura

Es considera que s'ha trobat la *solució* d'una estructura quan s'obté un mapa interpretable en termes de cadena d'aminoàcids (o dels àtoms que la formen). A partir d'aquest moment comença l'etapa de traçat i d'afinament.

El traçat consisteix a inspeccionar el mapa i anar situant la cadena peptídica al llarg d'ell. Depenent de la qualitat de les dades s'ha de fer gràficament o es pot fer automàticament si la resolució és molt alta. Generalment es comença introduint elements d'estructura secundària en els mapes inicials. Quan es té un traçat inicial, encara que només sigui aproximat o de només certes parts de les cadenes, es calculen mapes de diferències $2Fo - Fc, \phi c$ i $Fo - Fc, \phi c$. Aquests mapes s'obtenen fent la síntesi de Fourier amb les diferències entre les amplituds observades (Fo) i les calculades amb el material traçat (Fc), i com a fases (ϕ) es fan servir les disponibles, és a dir, les provinents del càlcul de Fc o bé una combinació amb altres fonts que també donin informació de fases, com la provinent d'àtoms pesants, o de l'extensió de fases.

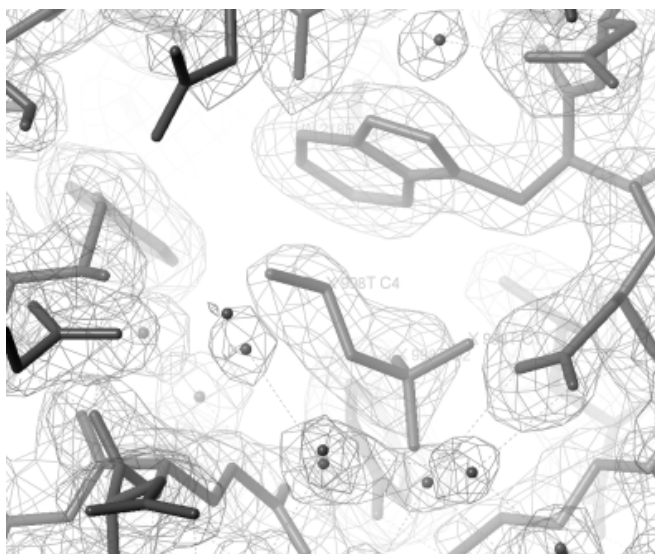
Els mapes $2Fo - Fc, \phi c$, tenen un pes doble en la amplitud observada respecte de la calculada. El motiu és aconseguir mapes que estaran més desviats cap a l'estructura correcta, que no pas síntesis del tipus $Fo, \phi c$, ja que el pes negatiu d' Fc ajudarà a eliminar el biaix dels mapes. En una estructura afinada Fo s'aproxima molt a Fc i el resultant és un mapa que en la pràctica representa un mapa $Fo, \phi c$. És a dir, un mapa on les amplituds són les experimentals i les fases són les del model afinat.

Els mapes $Fo - Fc, \phi c$ ajuden a la interpretació de parts conflictives dels mapes. Generalment són més intensos en les zones on hi ha errors en l'estructura, o aquesta no ha estat traçada encara. Les zones a on falten àtoms tindran una densitat positiva ($Fo > Fc$), i les zones on n'hi ha de mal traçats tindran densitat negativa ($Fc > Fo$). En l'estructura final, aquests mapes han de desaparèixer ja que s'igualen els dos components. De totes maneres, com que s'està treballant amb quoficients propers a zero (un cop feta la resta), els mapes $Fo - Fc, \phi c$ són molt més sorollosos que els $2Fo - Fc, \phi c$, per la qual cosa és necessari inspeccionar-los a un nivell o σ més elevat. Generalment es treballa amb densitats d' 1σ pel cim del valor mitjà dels mapes $2Fo - Fc, \phi c$ i de ± 2 o 2.5σ en els mapes $Fo - Fc, \phi c$.

També són possibles altres mapes. Els programes automàtics, per exemple, treballen amb mapes $3Fo - 2Fc$, ϕc , la *suma* dels dos mapes anteriors, encara que per al treball manual és més pràctic separar els mapes.

Tot aquest treball és iteratiu. Entre cada cicle de traçat, manual o automàtic, s'intercalen cicles d'afinament fins que totes les cadenes de la proteïna resultin traçades el millor possible.

Figura 3.10: Molècula del criotampó. De vegades el criotampó es difon dins del cristall. En el centre, una de les 4 molècules d'MPD (2-methyl-2,4-pentanediol) que es va localitzar per unitat asimètrica (16 per connector). Els carbonis de la part superior de la molècula d'MPD estan encarats cap a un triptòfan, mentre que l'oxigen inferior està lligat per ponts d'hidrogen a molècules d'aigua. El mapa correspon a l'estructura afinada del cristall I422 del tercer article (plana 91).



3.6 Estructura final de la proteïna

3.6.1 L'afinament de l'estructura

Quan una estructura està traçada en la seva major part, els mapes resultants ja haurien de ser de prou qualitat com per poder afinar les posicions dels àtoms individualment. Tant a l'afinament com en el traçat, s'han de combinar cicles automatitzats amb cicles manuals, fent servir un programa gràfic que permeti modelar manualment l'estructura en les zones on un programa automàtic no sigui capaç d'obtenir bons resultats.

L'afinament automàtic intenta alhora situar els àtoms del model en el lloc correcte i mantenir l'estereoquímica de la proteïna. Per això construeix una funció d'energia formada per termes cristal·logràfics (S_x) i estructurals (S_g):

$$S_x = \sum_i w_i (Fo_i - Fc_i)^2$$

$$S_g = \sum_i w_i (d_i - dc_i)^2$$

L'afinament consistirà a minimitzar aquesta funció, generalment pel mètode del gradient conjugat. Tots aquestes termes tenen forma d'oscil·lador, de manera que l'energia augmenta si els valors s'allunyen dels valors promig que s'han descrit per als aminoàcids. Els termes cristal·logràfics representen la desviació dels F_c respecte els F_o d'una manera similar als mapes. Els estructurals representen distàncies d'enllaç, angles de torsió, angles dièdrics, temperatura o agitació dels àtoms, fases experimentals, etc. Depenent de la qualitat i la resolució de la difracció i del tipus d'afinament es dona més pes (w_i) a uns termes o a altres. Els termes estructurals s'afinen contra valors (d_i) calculats o tabulats.

Cada estructura és diferent i no hi ha una recepta exacta del que cal fer per afinar-la. Se sol començar amb un rescalfament simulat per eliminar de l'estructura les rigideses imposades en el traçat inicial o en la solució de reemplaç molecular.

El rescalfament simulat consisteix a simular els efectes d'escalfar la proteïna al voltant d'uns 5000K, per deixar-la refredar a poc a poc després. La intenció és que els enllaços i torsions de la proteïna guanyaran mobilitat amb el calor, i podran ajustar-se més fàcilment als mapes, saltant mínims locals de potencial que no es podrien saltar a baixa temperatura.

A continuació s'intercalen cicles automàtics i manuals d'afinament individual de les posicions dels àtoms.

Altres paràmetres que s'afinen són la *temperatura* i l'ocupació de cada àtom. La *temperatura* és un paràmetre relacionat amb la fluctuació al voltant de la posició mitjana de cada àtom. Marca el grau de precisió amb què s'ha vist l'àtom en aquella coordenada del cristall, ja que d'una cel·la a una altra aquesta posició pot estar lleugerament modificada. Els valors més alts de la temperatura es donen en zones exposades al solvent o en zones *actives* de la proteïna, que presenten diferents conformacions per poder exercir la seva funció. L'ocupació és un paràmetre que representa la proporció en què aquell àtom es troba a totes les cel·les d'un cristall. A les resolucions típiques de la cristal·lografia de proteïnes l'ocupació no sol ser afinable, ja que es podrien emascarar errors de l'afinament general a base de modificar les ocupacions. De vegades, en llocs mòbils de l'estructura que presenten unes poques conformacions estables, es possible traçar una cadena diferent per a cada cas, i llavors assignar a cadascuna una ocupació proporcional.

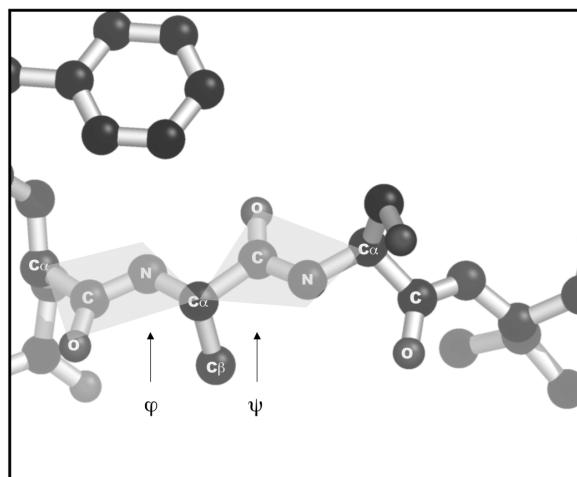
L'ocupació és més útil quan hi ha ions pesants dintre d'un cristall. Com que generalment no s'hauran difós igual per tot el cristall, llavors afinar l'ocupació té sentit, ja que les seves coordenades espacials es poden fixar.

L'objectiu de l'afinament és aconseguir que el model construït predigui el

millor possible les reflexions observades, alhora que conservi una geometria correcta en termes estructurals. S'ha observat les conformacions dels aminoàcids, dels enllaços entre ells i les interaccions entre cadenes properes estan restringits a una sèrie de valors que estan tabulats. Una molècula ha de complir aquest requisits o ha de tenir algun motiu especial per no complir-los. Per exemple es poden trobar desviacions en centres actius o per la influència d'algun metall.

En un enllaç peptídic, es poden definir uns rangs d'angles i unes interaccions mitjançant *ponts d'hidrogen* amb altres aminoàcids que el classifiquen en una determinada estructura secundària.

Figura 3.11: Angles peptídics. Els angles Ψ i Φ de la cadena principal estan limitats per l'estereoquímica de l'enllaç peptídic.



Totes aquestes restriccions geomètriques s'han de complir en una proteïna i es poden anar comprovant al llarg de l'afinament. Són molt útils els diagrames de Ramachandran (fig. 3.13).

Per mesurar la qualitat de la predicció de la difracció amb el model, es fa servir un concepte molt similar a l'escalat, l'*Rfactor*. Es tracta simplement de la desviació entre les amplituds observades i de les calculades, amb un pes per compensar la variació de la intensitat de la difracció a diferent resolució:

$$R = \frac{\sum_h |F_o h - k F_c h|}{\sum_h F_o h}$$

on k és el factor d'escala entre els valors mitjans d' F_o i d' F_c en cada capa de resolució.

Encara que és difícil per les restriccions geomètriques, és possible obtenir una estructura almenys parcialment incorrecta, deguda a sobreestimacions en l'afinament. Per evitar-ho, es defineix un altre residual similar a l'*Rfactor*,

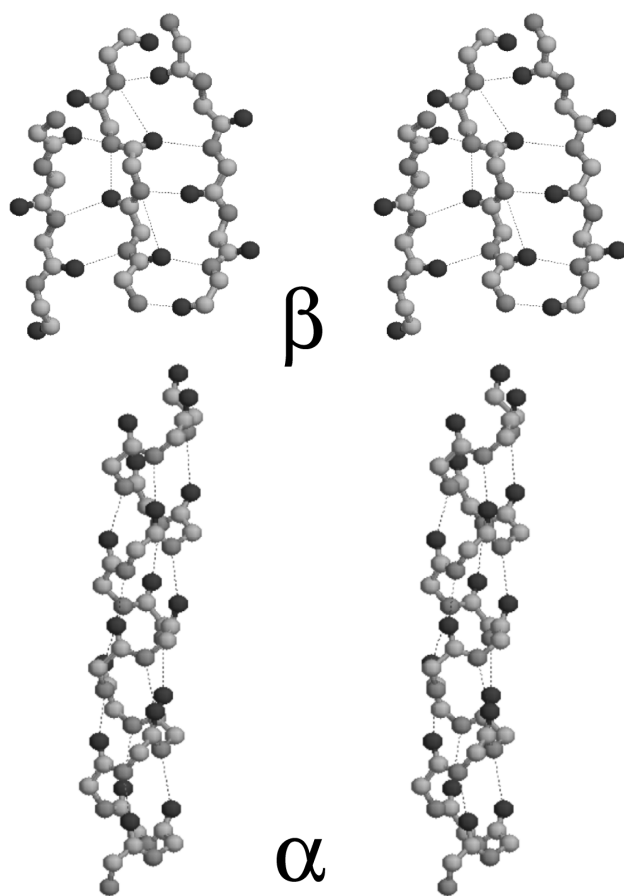


Figura 3.12: Hèlix α i fulla β . Cadena principal de la fulla β d'una part del domini SH3 del connector (plana 91) i de la hèlix $\alpha 3$. Els ponts d'hidrogen (línies de punts) segueixen un ordre característic en cada cas. En les fulles β aquests s'estableixen entre els carbonis α i els oxígens d'aminoàcids paral·lels, formant una làmina. En les hèlixs α , entre els mateixos àtoms, però d'aminoàcids separats per altres dos en la seqüència. Els angles formats s'especifiquen en la figura 3.13.

Rfree. Es construeix amb un grup de reflexions que des d'un principi se separa de les que es fan servir per resoldre, traçar i afinar l'estructura. Generalment entre un 5 i un 10% de les reflexions [5, 35]. Es pot monitoritzar la qualitat de l'afinament calculant aquest paràmetre, ja que en no ser fet servir pels càlculs, la seva correcta predicció indicarà un bon funcionament de l'afinament.

En una estructura l'aigua hi és present de dues maneres: localitzada, al voltant de la proteïna, i formant part d'ella, ja que l'estructura tridimensional i el funcionament correcte d'una proteïna depenen també de la presència d'aigua (excepte en proteïnes liposolubles, com les de la membrana cel·lular) i també deslocalitzada, formant part del gruix de solvent que forma el cristall, que normalment és de l'ordre d'un 50% de tota la matèria que conté.

L'aigua localitzada (*aigua cristal·logràfica*) es construeix i afina amb la resta de l'estructura. A les resolucions típiques, no són més que una sèrie

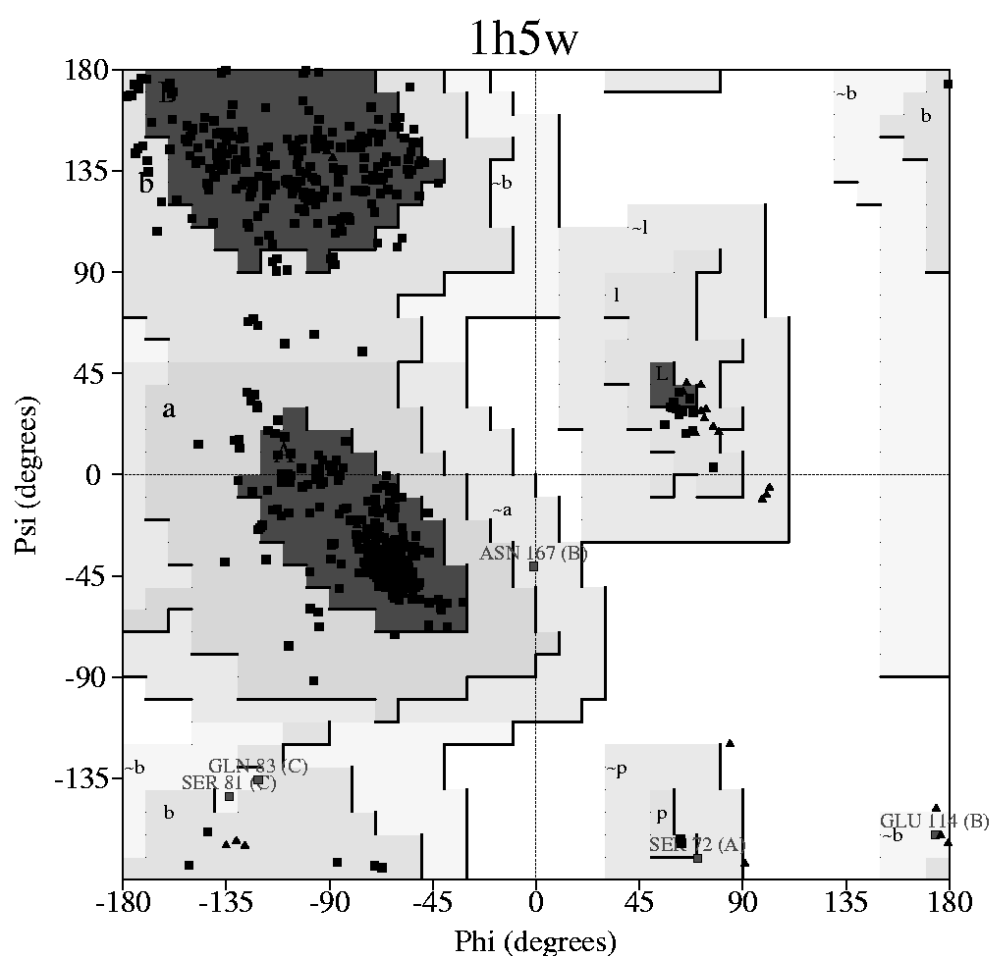


Figura 3.13: Ramachandran del connector. Diagrama de Ramachandran: Les zones fosques representen els valors més típics dels angles peptídics. La zona superior representa conformacions en fulla β . Les centrals representen hèlixs α dextrògires (esquerra) i levògires (dreta). Les zones grises representen valors menys habituals d'aquests angles. Aquest valors s'han calculat a partir d'estructures conegudes de proteïnes. Els quadres negres corresponen als angles en l'estructura afinada del connector, i els triangles a glicines, que en no tenir cadena lateral, poden abraçar més angles.

d'esferes, ja que els hidrògens no s'arriben a veure. Normalment s'espera a les últimes etapes de l'afinament per afegir-les, ja que podrien ser confoses amb altres àtoms, o bé sobreafinar l'estructura. És força fàcil aconseguir bons valors de l'*Rfactor* (però no de l'*Rfree*) posant aigua en tots els llocs dels mapes on hi aparegui densitat.

L'aigua deslocalitzada (solvent desordenat) no es pot afinar, ja que no representa àrees repetides al llarg del cristall. Només es repeteix el seu perfil a baixa resolució. Les reflexions de baixa resolució contenen aquesta informació i s'han de modular per tenir-la en compte. Una manera consisteix en construir una màscara al voltant de la proteïna, i calcular la contribució del que queda fora d'aquesta màscara en les reflexions (plana 50).

Les estadístiques de l'afinament estan resumides a la taula 1 del tercer article (plana 101).

3.6.2 Màxim de versemblança i Mínims quadrats

L'objectiu de l'afinament és obtenir una sèrie de coordenades que representin la situació i els seus errors dels àtoms de l'estructura en la cel·la unitat del cristall, a partir d'un model ja prèviament traçat. Aquestes coordenades han de complir una sèrie de restriccions geomètriques donats per les conformacions dels aminoàcids i les relacions que defineixen la seva estructura secundària, terciària i quaternària (hèlixs α , fulles β , girs, ponts d'hidrògen, altres interaccions amb molècules veïnes, etc). Alhora han de representar l'estructura que ha donat lloc a la difracció, per la qual cosa han de complir altres restriccions amb les amplituds obtingudes en la difracció.

La quantitat de variables és molt gran i les restriccions geomètriques no permeten (avui dia) generar una estructura mitjana *de novo* a partir de només càlculs. La posició avantatjosa de la cristal·lografia recau en el fet que que es poden fer servir resultats experimentals per resoldre aquest problema.

De totes maneres, la quantitat de variables (3 coordenades + 1 *b-factor* per a cada àtom) obliga a obtenir una quantitat mínima de reflexions igual. Per exemple, per resoldre l'estructura d'aquest treball, cal conèixer la posició de 12 monòmers amb 290 aminoàcids aproximadament cadascun, i unes 3500 molècules d'aigua, considerant una de promig per a cada aminoàcid. Això obliga a treballar amb unes 120000 reflexions i per tant amb una equació com:

$$F_{o_{\vec{h}}} = \sum_j f_j e^{2\pi i \vec{h} \cdot \vec{x}_j} e^{2\pi i \vec{h} \cdot \vec{b}}$$

per a cadascuna de les reflexions $F_{o_{\vec{h}}}$. La simetria del cristall i la NCS (plana 3.5.3) permeten reduir la quantitat d'equacions necessàries.

El sistema conté les incògnites en exponencials i per tant no es pot resoldre directament. En realitat, la situació és pitjor, ja que cal resoldre els mòduls d'aquesta expressió.

L'única manera de resoldre'l és disposar d'un model de partida prou bo (el traçat) i minimitzar l'equació de l'energia (plana 55) per mantenir les

restriccions. Aquesta equació es minimitza pel mètode del gradient conjugat.

L'aplicació d'aquest mètode es fa generalment minimitzant l'equació per *mínims quadrats*, però la minimització pot conduir molt fàcilment a un mínim local que no permeti millorar més enllà l'estructura. Per evitar això, s'ha d'avançar a poc a poc, començant en la resolució en què s'ha resolt l'estructura i incrementant-la només quan està correcte en aquella resolució.

Una manera molt efectiva d'evitar el perill dels mínims locals és fer servir tota la resolució de cop. D'aquesta manera es treballa amb el conjunt sencer de les dades, i estadísticament és més difícil caure en un mínim local. Cal, però un mètode per discernir què hi ha de correcte en el model, ja que ara no hi ha el criteri d'avançar cap a l'alta resolució.

Un model correcte suposa unes fases correctes, i per tant es pot tractar de jutjar *com* de correctes són unes fases. Com que no es té cap evidència experimental d'aquestes fases, cal jutjar-les a partir de l'únic que es té experimentalment: les amplituds [44]. Un model correcte predirà unes amplituds correctes, però l'afinament tendeix a situar els àtoms de manera que es prediguin amplituds correctes. Per evitar aquest cercle tancat es fan servir les reflexions deixades de banda (les de l'*Rfree*). Si aquestes es prediuen bé, se suposarà que les fases també estan ben predites, en la seva corresponent capa de resolució, i s'assignarà un pes (figura de mèrit, FOM) alt a les reflexions corresponents. En cas contrari, el pes serà baix. En treballar amb les reflexions *lliures*, s'elimina en gran part el problema del biaix [2].

La funció a minimitzar és (funció de màxima versemblança):

$$\ln L \sim \sum_i \ln[p(x_i)]$$

on $p(x_i)$ és la probabilitat (anomenada versemblança) de que el paràmetre x_i del model s'ajusti a les observacions [44].

Quan tots els errors tenen forma gaussiana, el màxim de versemblança és equivalent al mètode dels mínims quadrats:

$$p(x) \sim \frac{1}{\sqrt{2\pi}\sigma(x)} e^{-\frac{(x-\langle x \rangle)^2}{2\sigma^2(x)}}$$

$$-\ln L \sim \sum_i \ln(\sigma_i) + \frac{e_i^2}{\sigma_i^2}$$

on e_i i σ_i són els errors i les seves desviacions associats a cadascun dels paràmetres que es minimitzen en la funció d'energia.

3.7 Macles

Un cristall no sempre creix ben ordenat. Apart de petites diferències a causa de la mosaïcitat (plana 33), el cristall pot contenir dominis macroscòpics orientats de maneres diferents. Cada domini és, en la pràctica, un cristall diferent, i difracta independentment dels altres. En un lupa de laboratori es poden veure els dominis amb un polaritzador. Si són prou grans es pot veure directament per la forma del cristall: branques, estrelles, barres. . .

De vegades és possible separar els cristalls abans de muntar-los, trencant-ne una part. Pot ser amb una agulla molt fina prement al punt d'unió, pot ser canviant-lo a una altra solució a diferent temperatura (plana 33). Les interfícies entre els cristalls són menys fortes que entre cel·les d'un mateix cristall i qualsevol acció sobre elles pot ajudar a separar-les.

Quan no és possible separar els components d'una macla, però es poden diferenciar visualment les parts, es pot ajustar el raig posant les obertures de la sortida dels raigs X a una mida molt petita i només difractar una part del cristall (fig. 3.14).

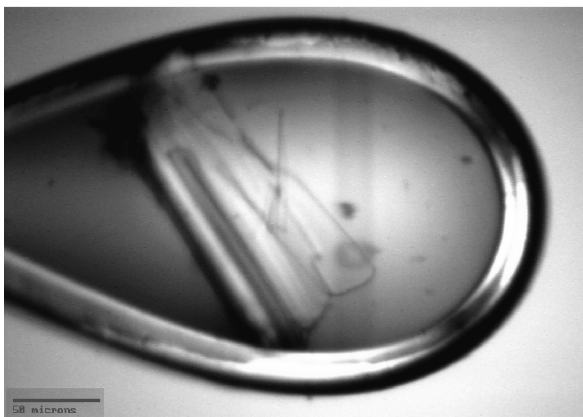
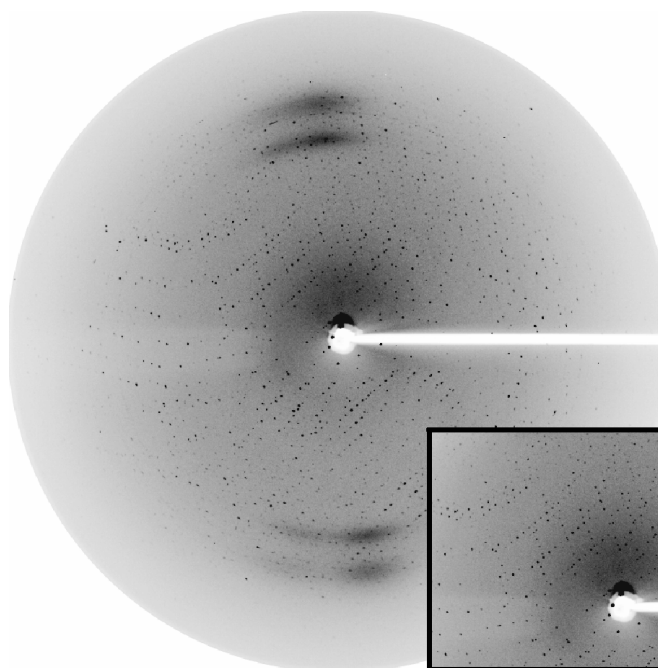


Figura 3.14: Difracció amb microfocus. Si un cristall maclat no es pot separar en parts, és convenient utilitzar un feix de secció molt petita i enfocar-lo només a una zona del cristall. Aquesta imatge correspon a la proteïna PhoB [4]. El feix està centrat en una làmina lateral del cristall. Es pot veure el punt d'incidència del feix en el cristall (la taca fosca) i el senyal que ha fet el feix de raigs X en el criotampó (la banda vertical). La secció del feix és d'uns $20 \mu\text{m}$.

Un problema que generalment passa en aquest casos es que en anar girant el cristall durant la recollida, és molt possible que s'arribi a una orientació en la qual es difracten altres parts de la macla. El resultat són patrons de difracció que contenen diferents conjunts de reflexions (fig. 3.15). Si les orientacions dels dominis són prou diferents, aquestes imatges es poden processar tenint cura de no associar les reflexions al cristall equivocat. Algunes de les reflexions se sobreposaran i es molt difícil en aquests casos determinar a quina part del cristall pertanyen o bé quina proporció de cada cristall conté cada reflexió.

Figura 3.15: Difracció d'una macla. Difracció d'un cristall maclat. La difracció correspon a uns cristalls del connector amb TaBr difractats a la línia BM30 de l'ESRF. Les reflexions es van separar processant-les per separat. A la part inferior, detall de la imatge, o es veuen els patrons de punts en direccions casi perpendiculars.



Molt sovint els cristalls que formen la macla tenen orientacions casi similars. Aquest és el pitjor dels casos, ja que les reflexions estan gairebé allà mateix i no es poden separar. La forma de la *taca* és allargada, amb varis màxims al llarg d'ella. No solen ser útils per a resoldre una estructura. Com a molt per determinar els paràmetres de la cel·la i tractar d'hipotetitzar què es pot fer per millorar aquests cristalls.

3.7.1 Macles hemièdriques

Determinats grups espacials, per exemple els tetragonals o els hexagonals contenen una simetria que permet que les cel·les creixin en diferents orientacions. Això és a causa de, per exemple, que certs cantons de la cel·la són iguals i el grup espacial deixa que es puguin anar formant cel·les alternades en diferents orientacions.

Si aquesta alternança és estricta (sempre es repeteix), la cel·la resultant serà una de més gran que contindrà les dues. En la difracció, però, la cel·la aparent serà més petita i contindrà una barreja de les dues. La solució a aquest problema, si es detecta a posteriori, és la de reprocessar les dades forçant una cel·la més gran.

Si aquesta alternança és per petits dominis en dues orientacions diferents, la macla s'en diu hemièdrica [21]. La difracció proporcionarà una cel·la apa-

rent que és la superposició dels dos dominis i no serà possible trobar una altra indexació que les separi [18]. A la figura 2 del primer article (plana 80) hi ha una imatge de difracció d'un cristall d'aquest tipus.

En una macla d'aquest tipus les reflexions es sobreposen segons la simetria d'aquesta macla. La intensitat de les reflexions serà la suma ponderada de les dues parts:

$$I(h, k, l) = \alpha I_1(h, k, l) + (1 - \alpha) I_2(h, k, l)$$

$$\alpha = V_1 / (V_2 - V_1); V_2 > V_1$$

on I_1 i I_2 són les intensitats individuals de cadascun dels components de la macla i α és la proporció en volum de cadascun dels dos components.

Per determinar si unes dades de difracció corresponen a una macla d'aquest tipus es poden comprovar diverses estadístiques. En primer lloc, la distribució acumulada de les intensitats de les reflexions hauria de presentar un perfil en forma de corba de saturació [50, 45, 21] (fig. 3.16). Un cristall maclat d'aquesta manera té un perfil sigmoide.

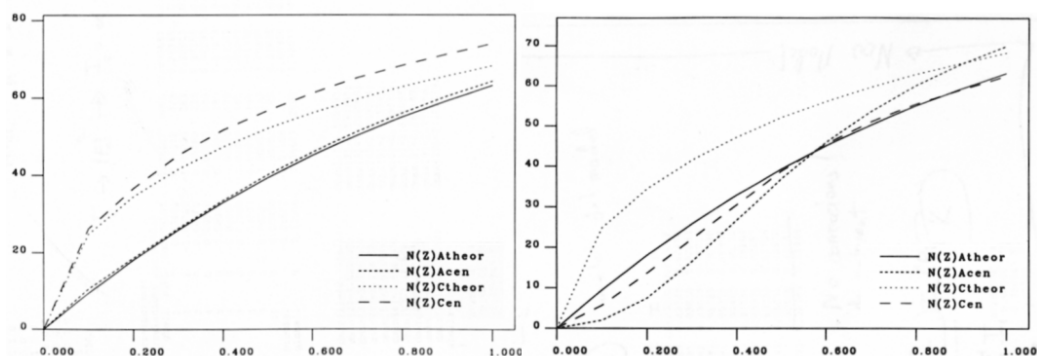


Figura 3.16: Distribució acumulada. Distribucions acumulades d'un cristall no maclat i un de maclat. La corba amb perfil sigmoide indica que el cristall està maclat. Esquerra: proteïna TrwB [22]. Dreta: un dels cristalls del connector amb simetria $P4_22_12$. A l'eix horitzontal es representa el valor de $I/\langle I \rangle$ amb el promig fet per capes de resolució i a l'eix vertical el tant per cent de reflexions.

Altres característiques que es poden comprovar són les distribucions de Wilson [59] i d'Stanley [50]. S'ha comprovat que aquestes distribucions presenten uns valors mitjans determinats que diferencien si la difracció prové d'una cel·la maclada o no:

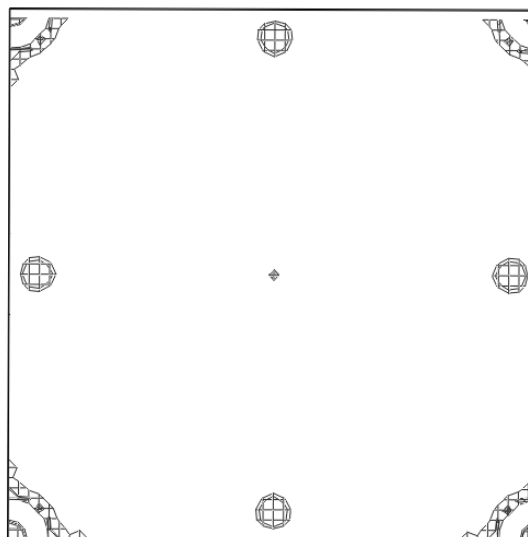
$$\text{Stanley : } \frac{\langle I^2 \rangle}{\langle I \rangle^2} = \begin{cases} > 2 & \text{desordre} \\ 2 & \text{normal} \\ 1.5 & \text{macla} \end{cases}$$

$$\text{Wilson : } \frac{\langle |F| \rangle^2}{\langle I \rangle} = \begin{cases} > 0.785 & \text{macla} \\ = 0.785 & \text{normal} \\ < 0.785 & \text{desordre} \end{cases}$$

Tots aquest mètodes estan basats en l'estadística de les reflexions per una distribució aleatòria d'àtoms dins d'una cel·la unitat. Pot passar que per motius geomètrics la distribució real dels àtoms no sigui prou aleatòria. Per exemple, si hi ha simetria no cristal·logràfica o bé si hi ha grans arees de solvent. Llavors les reflexions no segueixen tant fidelment aquestes estadístiques, i és d'esperar desviacions al comportament estàndard.

Per això és necessari assegurar-se també amb altres mètodes *a posteriori* si la cel·la pot estar maclada. El més evident, és un aturament de la qualitat de l'afinament, a valors alts d'R i d'Rfree, però a l'hora mostrant mapes millors que els esperats a aquests valors. Una altra possibilitat és l'obtenció de conjunts de solucions sobreposades en el Reemplaçament Molecular, o bé solucions molt clares a partir de la síntesi de Patterson, però que impliquen empaquetaments impossibles (fig. 3.17).

Figura 3.17: Mapa de Patterson. Projectió al llarg de l'eix c del mapa de Patterson d'un dels cristalls amb simetria $P4_22_12$. L'empaquetament que implica un connector aproximadament centrat en cada unitat asimètrica no es pot donar, ja que les dimensions del connector ho impedeixen. El mapa està representat a 4σ del valor mig i els quatre pics estan situats a $(0.5, 0.05, 0.5)$ i simètrics.



Gràcies a un mapa de Patterson d'aquest tipus, es pot preveure quines són les cel·les individuals que han format la macla. Han de ser cel·les idèntiques, relacionades per una rotació simple, per exemple de 90 graus.

Resoldre aquest tipus de macla comporta dificultats, que degraden la qualitat del resultat i actualment no s'han resolt gaires estructures d'aquest tipus. Sol ser molt més efectiu tractar d'aconseguir cristalls no maclats.

Si no es troba la manera d'aconseguir cristalls simples, la difracció és de prou bona qualitat, (i l'esforç val la pena), és possible separar la macla (*desmaclar*) a partir de certes suposicions sobre l'estructura:

Si prenem dues reflexions (h i h') relacionades per la simetria d'una macla hemièdrica [21]

$$Im(h) = \alpha I(h) + (1 - \alpha)I(h')$$

$$Im(h') = (1 - \alpha)I(h) + \alpha I(h')$$

on Im representa les reflexions observades (maclades) i I les del grup espacial no maclat, l'equació només té solució quan α és diferent de 0.5:

$$\begin{cases} I(h) = \frac{(1-\alpha)Im(h') - \alpha Im(h)}{1-2\alpha} \\ I(h') = \frac{(1-\alpha)Im(h) - \alpha Im(h')}{1-2\alpha} \end{cases}$$

Per resoldre macles amb $\alpha=0.5$ es pot construir un model de partida similar al de reemplaçament molecular, bé traçant un mapa obtingut per derivats i separant les cadenes, o bé a partir d'un model de criomicroscòpia.

A partir d'aquest model, i amb ajuda del mapa de Patterson, o alguna altra manera que ens permeti situar la molècula al seu lloc en la unitat asimètrica, es construeix la macla afegint les contribucions de cada orientació. D'aquesta manera es construeix una "cel·la maclada" tal com s'ha observat. L'afinament es farà només per una molècula individual, però en cada pas en el que calgui comprovar la fiabilitat del afinat caldrà reconstruir la macla i calcular les estadístiques a partir de les reflexions calculades d'aquesta manera.

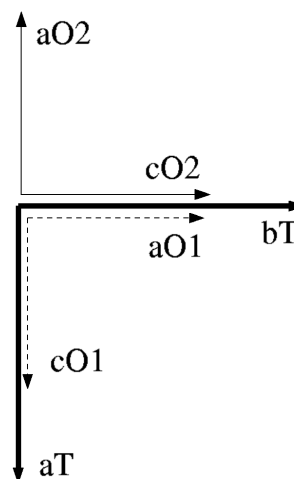
3.7.2 Procediment pràctic

Per a desmaclar les reflexions dels cristalls amb grup espacial aparent $P4_22_12$ es va haver d'afrontar un altre problema: com que el model era de molt baixa resolució, el nombre de reflexions a ajustar contra un model era proporcionalment molt baix, pel que calia alhora tractar d'extendre les fases a més resolució per comprovar la fiabilitat del procediment.

El grup espacial aparent era el $P4_22_12$, amb $a=b=170.2 \text{ \AA}$ i $c=156.9 \text{ \AA}$ (tetragonal o T, d'ara endavant). Per consideracions d'empaquetament i de la síntesi de Patterson (fig. 3.17) es va veure que es podia descomposar en dues cel·les amb simetria $P2_12_12$ $a=c=170.2 \text{ \AA}$ i $b=156.9 \text{ \AA}$ (ortoròmbic o

O, d'ara endavant). Una rotada 90° al voltant de l'eix b respecte l'altra. La cel·la tetragonal s'aconsegueix sumant aquestes dues i rotant 90° al voltant del nou eix eix c (fig. 3.18).

Figura 3.18: Eixos de la macla. Relació entre els eixos de les cel·les que formen la macla i la cel·la aparent. O1 i O2 són les dues cel·les ortoròmbiques i T és la tetragonal. O1 i O2 ($a=c$) estan entre elles rotades 90° al voltant de l'eix b. La cel·la T ($a=b$) es construeix permutant els eixos de manera que l'eix b de les ortoròmbiques coincideixi amb l'eix c de la tetragonal.



Per tant una reflexió (h, k, l) en el grup espacial maclat seria la suma

$$I_T(h, k, l) = \alpha I_O(k, l, h) + (1 - \alpha) I_O(-h, l, k)$$

de les reflexions del grup no maclat.

Aquesta idea la vàrem implementar escrivint un programa que combinava la búsqueda de la posició del connector i la formació de la macla. Les coordenades del centre del connector estaven ben definides en el pla a-b gràcies al mapa de Patterson, però calia trobar també la posició al llarg de l'eix c i també la orientació dels lòbuls.

El programa feia una cerca de diferents posicions i orientacions del connector en una cel·la $P2_12_12$. Primer es construïa un mapa de densitat electrònica a baixa resolució a partir del model de microscòpia (de 20 a 12Å). Llavors es calculaven les reflexions exteses al grup espacial $P1$, per tenir totes les simètriques. Amb elles es calculaven les reflexions maclades en $P4_22_12$ a partir de la fórmula anterior. La correlació entre les reflexions es calculava amb una fórmula idèntica a la de la correlació en el Reemplaçament Molecular (plana 3.5.2), però fent servir intensitats. Les correlacions es calculaven globalment per totes les reflexions, per capes de resolució i també per les reflexions $(h,k,0)$ i $(h,0,l)$.

Les reflexions $(h,k,0)$ representen, en el grup espacial $P4_22_12$, la projecció sobre el pla a-b, que era convenient ja que en aquesta direcció es coneixia bé el centre de la molècula i a més coincidia amb la màxima resolució del model

de microscòpia. Les reflexions $(h,0,l)$ representen la projecció en el pla a - c i es feien servir per buscar la màxima correlació en l'eix c . No es varen poder fer servir reflexions $(0,k,l)$ a causa de que se'n van mesurar molt poques. A més es podia aplicar un decaïment a cada reflexió proporcional a la resolució i dependent de la direcció per a simular l'anisotropicitat de les reflexions.

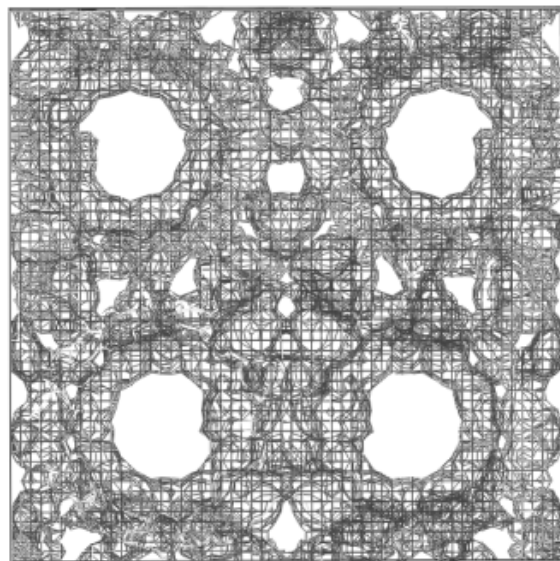


Figura 3.19: Macla reconstruïda. Cel·la tetragonal maclada reconstruïda a partir d'una de les possibles orientacions. El connector està aproximadament centrat en la unitat asimètrica, però la mida d'aquest no permet que hi càpiga. En les cel·les ortoròmbiques, al tenir menor simetria hi ha menys connectors en la cel·la i és possible l'empaquetament.

Per a poder comprovar la fiabilitat de la reconstrucció, les orientacions de major correlació se'ls aplicaven cicles d'extensió de fases amb $NCS=12$, alternats amb desmaclat de les reflexions a partir de les noves fases.

Per assegurar la fiabilitat del mètode es va construir una macla fictícia a partir d'un model fet modificant l'estructura d'una proteïna de dimensions similars i donant-li una forma semblant a la del connector. Amb aquest model es va simular un de microscòpia, calculant un mapa de baixa resolució (a 10\AA), i també les reflexions d'una cel·la maclada. Els tests realitzats localitzaven correctament la proteïna test, fins i tot si la forma es modificava una mica, o treballant a menys resolució. Les correlacions eren de l'ordre del 99% tant en les búsquedes en dues dimensions com en les de tres. En canvi, quan es van fer servir les dades reals, les correlacions van caure a menys del 78% en dues dimensions i a un 46% en tres, no mostrant cap solució prominent. Les búsquedes al llarg de l'eix c no van donar cap posició especial. A la secció 3.8.2 de la plana 70 hi ha una argumentació de per què aquestes búsquedes tenien tant poca precisió.

3.8 Microscòpia electrònica

Els electrons poden ser conduïts per mitjà de lents electromagnètiques. Així, en la microscòpia electrònica, al contrari que a la cristal·lografia de raigs X, és possible enfocar els electrons, i obtenir imatges directes de la mostra en estudi. Aquestes imatges representen projeccions de l'estructura perpendiculars al pla focal de microscopi. Cal, per tant, un procés de reconstrucció de la imatge a partir de diverses projeccions.

Si hi ha cristalls bidimensionals disponibles, aquests són adients per experiments de microscòpia electrònica (difracció d'electrons) (figs. 1.4, plana 6 i 1.9, plana 9). Els electrons tenen poca penetració en les mostres, i cal evitar alhora interaccions extres dels electrons amb el cristall. Es poden recollir també dades de difracció d'electrons d'aquests cristalls. D'aquesta manera, s'obtenen també factors d'estructura que es poden combinar amb les imatges directes per a obtenir major resolució.

3.8.1 Reconstrucció de la M.E.

Els models de microscòpia electrònica ens eren presentats en forma de seccions perpendiculars al llarg de l'eix del del connector. Calia, doncs, construir un programa per obtenir un mapa tridimensional compatible amb els mapes usats en cristal·lografia.

Les seccions venien ordenades en forma de fitxers numerats seqüencialment. Cadascun contenia les dades de densitat electrònica en forma seqüencial en valors enters d'1 bit o flotants de 4 bits, representant una matriu de 128×128 punts. No contenien cap informació sobre la mida real. El volum tridimensional s'aconseguia sobreposant aquestes seccions en una matriu tridimensional de $128 \times 128 \times (\text{nombre de seccions})$. El nombre de seccions era variable segons la resolució del model. En la criomicroscòpia la resolució obtinguda en el pla perpendicular del connector era més alta que al llarg de la seva alçada, pel que el nombre de seccions era només de 30 a 60.

Depenent de quin model, calia primer fer una *neteja* per eliminar soroll de fons i imatges residuals d'altres connectors veïns. Aquests filtres es van incorporar dins dels programes: Es podia determinar un valor de tall i eliminar tota la densitat menor a aquest, o bé fer servir una màscara per a cada imatge. La màscara era un fitxer binari del mateix format que les seccions. Es construïa la secció modificada multiplicant prèviament punt a punt per la màscara.

Es va decidir de construir mapes en format "newEZD" ja que resultaven més fàcils d'editar a posteriori al ser composts només de caràcters ASCII. El pas a altres formats s'aconseguia fàcilment amb el programa `mapman`[36]. La

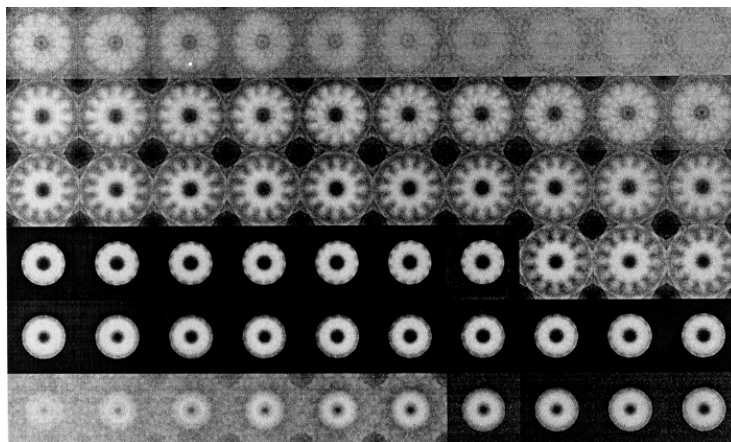


Figura 3.20: Seccions M.E.. Seccions del model de microscòpia electrònica, que es van fer servir per a construir els models inicials per l'extensió de fases a alta resolució.

mida del model s'obtenia escalant la cel·la unitat de la reconstrucció.

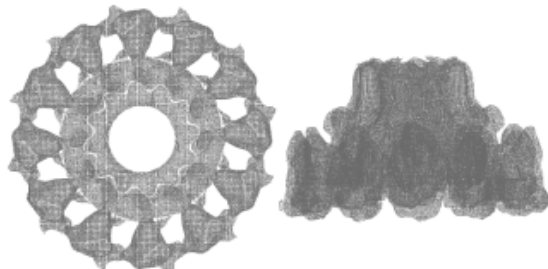


Figura 3.21: Reconstrucció de la M.E.. Vista frontal i lateral de la reconstrucció del model de microscòpia electrònica, construït a partir de les seccions de la figura 3.20

Per conveniència amb els programes estàndard, a partir de la reconstrucció es van construir altres programes que transformaven la densitat a pseudoàtoms, a pseudoaminoàcids o a màscares per aplicar NCS.

3.8.2 Ajust de la microscòpia i del model afinat

Amb les coordenades del model final afinat es va fer un ajust amb el model de microscòpia per a veure les diferències entre aquests dos mètodes. A la figura 3.22 es mostra aquest ajust, junt amb un mapa del connector afinat a una resolució equivalent. Com es pot veure, la forma general d'ambdós mapes és similar, tal com era d'esperar, hi ha diferències prou importants

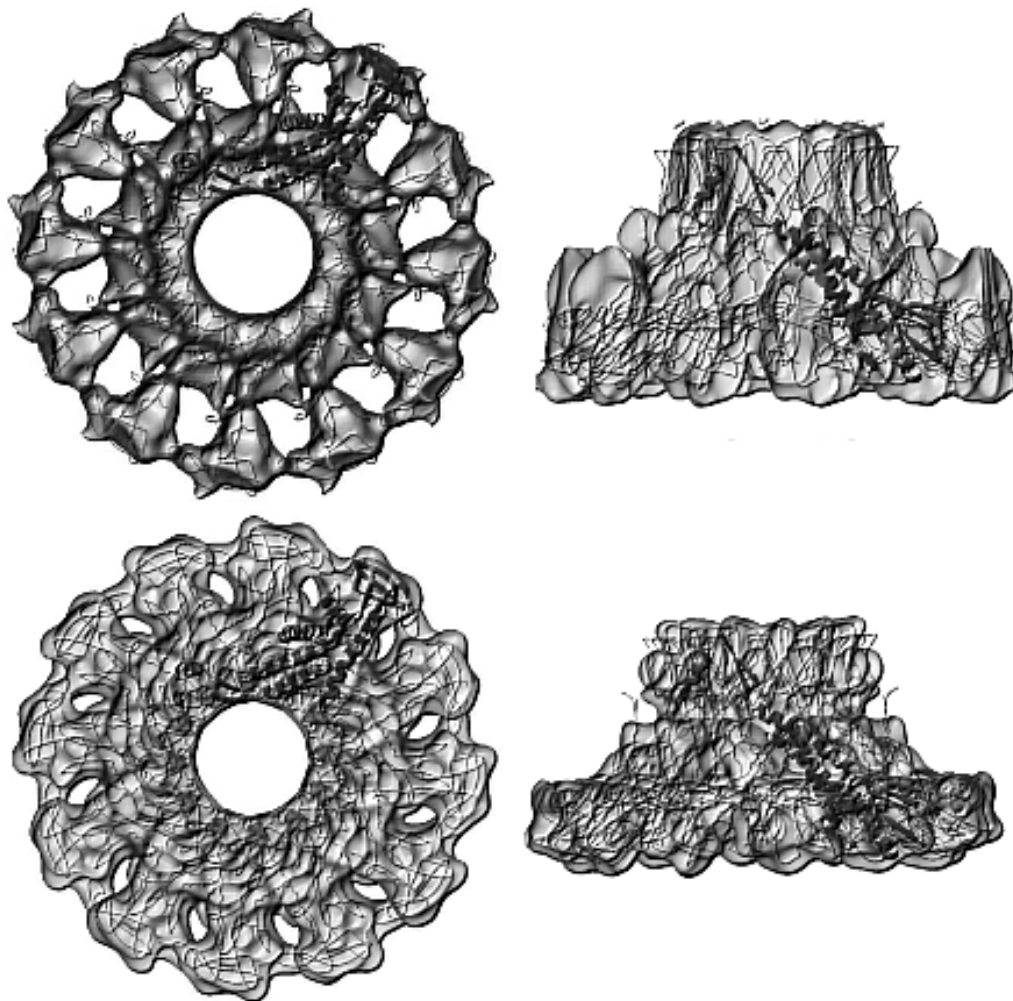


Figura 3.22: Ajust de la microscòpia i del model afinat. Vistes axials i laterals del mapa de microscòpia (a dalt) i del mapa final (a baix), a baixa resolució (10 Å), amb la cadena principal superposada. La forma general és similar, però els mapes difereixen força en la zona dels lòbuls, essent menys alts i més amples en el model final.

com per a impedir que la cerca per correlacions fos prou acurada a baixa resolució.

El model final és més cilíndric en la zona dels lòbuls, impedit un bon contrast entre solvent i proteïna, pel que la cerca per correlació a baixa resolució no podia discernir clarament l'orientació exacta del connector en la projecció al llarg de l'eix *c*. També és força diferent en l'alçada dels lòbuls.

Els del model afinat són la meitat d'alts. Aquest detall impedia també una bona localització al llarg de l'eix *c* fent servir el model de microscòpia. Per això les búsquedes per correlació donaven poc contrast entre les orientacions possibles.

Altres detalls que difereixen estan en el tronc central. En aquest cas, els lòbuls del model final no són tant prominents ja que corresponen a zones poc ordenades de la densitat. En el dibuix inferior es veuen les cadenes traçades fora del mapa, ja que aquest no era prou dens pel contrast de la figura.

3.9 El PDB

El Protein Data Bank (PDB, www.pdb.org) és la base de dades centralitzada on es publiquen les estructures de proteïnes i d'àcids nucleics fetes per difracció de raigs X, per ressonància magnètica nuclear, difracció de neutrons, models teòrics, etc.

Un "pdb" és un arxiu ASCII on cada línia conté una entrada amb les coordenades dels àtoms de la molècula, de quin àtom es tracta, a quin aminoàcid, nucleòtid o compost pertany, quina ocupació i quin *B factor* tenen. Es tracta d'un format ordenat per columnes per permetre búsquedes i modificacions ràpides, fàcils d'automatitzar i d'incorporar en programes. Altres entrades del pdb contenen informació sobre l'estructura secundària, el cristall, com s'ha aconseguit la molècula, com s'ha fet l'experiment, on s'ha publicat, i quines estadístiques ha donat. Es tracta de que amb un sol fitxer es trobi tota la informació per regenerar l'estructura i per calibrar la seva qualitat, així com la informació genètica necessària, generalment posada en forma de referències a bases de dades de gens. A l'hora d'escriure aquest text (30 de novembre de 2001) el nombre de molècules dipositades és de 16651.

El connector s'ha dipositat en el PDB amb el codi **1h5w**.

Part III
Resultats

Capítol 4

Resultats

- plana 77 Purification, crystallization and preliminary X-ray diffraction studies of the bacteriophage $\phi 29$ connector particle.
A. Guasch, A. Párraga, J. Pous, J.M. Valpuesta, J.L. Carrascosa, and M. Coll.
FEBS Lett, 430:283–287, 1998.
- plana 83 Crystallographic analysis reveals the 12-fold symmetry of the bacteriophage $\phi 29$ connector particle.
A. Guasch, J. Pous, A. Párraga, J.M. Valpuesta, J.L. Carrascosa, and M. Coll.
J Mol Biol, 281(2):219–225, 1998.
- plana 91 Detailed architecture of a DNA translocating machine: The high-resolution structure of the bacteriophage $\Phi 29$ connector particle.
J. Pous, A. Guasch, B. Ibarra, F.X. Gomis-Rüth, J.M. Valpuesta, N. Sousa, J.L. Carrascosa and M. Coll.
J Mol Biol 315, 663–676, 2002

Purification, crystallization and preliminary X-ray diffraction studies of the bacteriophage ϕ 29 connector particle

Alicia Guasch^a, Antonio Parrága^a, Joan Pous^a, José M. Valpuesta^b, José L. Carrascosa^b,
Miquel Coll^{a,*}

^aDepartament de Biologia Molecular i Cel·lular, Centre d'Investigació i Desenvolupament, CSIC, Jordi Girona 18–26, 08034 Barcelona, Spain

^bCentro Nacional de Biotecnología, CSIC, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain

Received 13 April 1998; revised version received 26 May 1998

Abstract The connector or portal particle from double-stranded DNA bacteriophage ϕ 29 has been crystallized. This structure, which connects the head of the virus with the tail and plays a central role in prohead assembly and DNA packaging and translocation, is formed by 12 subunits of the p10 protein and has a molecular weight of 430 kDa. The connector structure was proteolysed with endoproteinase Glu-C from *Staphylococcus aureus* V8, which removes 13 and 18 amino acids from the amino- and carboxy-terminal regions of the p10 protein, respectively. Two crystal forms were grown from drops containing an alcohol solution and paraffin oil. Crystals of form I are monoclinic, space group C2 with cell dimensions $a = 416.86$ Å, $b = 227.62$ Å, $c = 236.68$ Å and $\beta = 96.3^\circ$ and contain four connector particles per asymmetric unit. Crystals of form II are tetragonal, space group P4₂2₁2 with cell dimensions $a = b = 170.2$ Å, $c = 156.9$ Å and contain half a particle per asymmetric unit. X-ray diffraction data from both native crystal forms have been collected to 6.0 and 3.2 Å respectively, using synchrotron radiation. Crystals of form II are likely to have the same packing arrangement as the two-dimensional crystals analyzed previously by electron microscopy.

© 1998 Federation of European Biochemical Societies.

Key words: Bacteriophage ϕ 29; Connector; Portal particle; Crystallization; X-ray diffraction

1. Introduction

ϕ 29 is a complex double-stranded DNA bacteriophage that infects *Bacillus subtilis* cells. The viral particles are formed by an elongated icosahedral capsid, or head, and a tail structure. Between these two structures there is a connecting region called the portal structure or connector. This head-to-tail connector plays an important role in the first steps of head assembly and the packaging of DNA. The connector is also involved in the process of DNA transfer into the host cell [1,2].

Electron microscopy studies based on two-dimensional projections show that the ϕ 29 connector is an oligomeric structure built up by multiple copies of the 36 kDa gen 10 product. This structure is composed of two morphological domains: a wider disk-like domain of about 14.5 nm in diameter and 3.2 nm in height with protruding lobules and a narrower domain with a cylindrical shape of 8 nm in diameter and 5 nm in height, which presents a longitudinal channel of

40 nm in diameter along its axis [3]. The symmetry of the connector has been controversial, although there is strong evidence for a 12-fold local symmetry of the particle [1,4].

The connector contains two independent domains for interaction with DNA [5] and with a small RNA of viral origin (prohead RNA or pRNA) which is necessary for DNA packaging into the ϕ 29 prohead [6]. Treatment of ϕ 29 connectors with endoprotease Glu-C from *Staphylococcus aureus* suppresses the connector DNA binding and abolishes the DNA packaging ability [5].

Although the pRNA seems to play a role in the selection of the DNA to be packaged [7], its role in the DNA translocation process remains unclear. Also, the molecular details of the interaction of both the DNA and the RNA with the connector have not been analyzed in detail due to the lack of a high resolution model of the packaging structure. To understand these complex interactions as well as the molecular basis underlying the DNA translocation during DNA packaging, it is critical to determine the atomic structure of the connector. To this end, we have crystallized proteolysed connector particles; the crystals are suitable for X-ray diffraction and a preliminary crystallographic analysis is described. This is the first crystallographic report of a bacteriophage connector particle.

2. Materials and methods and results

2.1. Preparation of inactive and proteolysed connectors

The connectors were purified as described [8] with some modifications. Briefly, the protein p10 that builds the connector was overproduced in *Escherichia coli* carrying a recombinant plasmid containing the gene encoding this protein under the control of the PL promoter of phage λ . The overproduced protein was purified in native form, yielding p10 dodecamers that closely resembled the necks extracted from ϕ 29 viral particles. The cells were lysed with lysozyme treatment and the connectors isolated from the cell debris by centrifuging the cell extract. The connector solution was loaded onto a DE-32 cellulose (Whatman) column equilibrated with 0.3 M KCl 50 mM Tris-HCl, pH 7.7 and washed with the same buffer to remove the nucleic acids. The pass-through was precipitated with 0.26% ammonium sulfate and redissolved in 5% glycerol, 50 mM Tris-HCl, pH 7.7. The solution was loaded onto a P11 phosphocellulose (Whatman) column equilibrated with 5% glycerol, 50 mM Tris-HCl, pH 7.7 and washed with a 0–1 M NaCl gradient. Pure connectors were obtained at 0.8 M NaCl. The connector solution was then dialyzed against 5% glycerol, 50 mM Tris-HCl, pH 7.7 and subsequently treated with V8 (endoproteinase Glu-C from *S. aureus* V8) (100:1, p10/V8 (w/w)) for 60 min at 25°C. Proteolysis was terminated by addition of 1 mM phenylmethylsulfonyl fluoride. After digestion, the protein solution was loaded onto a p11 phosphocellulose column equilibrated with 5% glycerol, 1 mM DTT, 50 mM Tris-HCl, pH 7.7 and washed with a 0–1 M NaCl gradient. Pure proteolysed connectors, devoid of the proteolytic fragments, were obtained at 0.4 M NaCl.

*Corresponding author. Fax: (34) (3) 2045904.
E-mail: mceeri@cid.csic.es

2.2. Sample preparation

The initial protein concentration was 1 mg/ml and the final concentration of 7 mg/ml, used for crystallization experiments, was reached by two different ways, either lowering the volume with Centricon 30 or by precipitation with ammonium sulfate. With the first method, sometimes depending the sample preparation, significant protein loss was observed due to its binding to the Centricon membrane. The second method consisted in an ammonium sulfate precipitation (0.26 g/ml) at 4°C, followed by centrifugation at 4000×g, and dialysis against 0.4 M NaCl, 20 mM Tris pH 7.7, 1 mM DTT. In both cases the same type of crystals were obtained.

The molecular weight of each monomer, determined by mass spectral analysis, was 32208 Da, which coincides with the theoretical molecular weight of the proteolysed p10 protein. However, in some preparations which resulted in lower quality crystals the molecular mass showed an increase of 30 Da. To prevent it, 1 mM DTT was added to the protein solution.

The initial protein solution for crystallization contained 7 mg/ml of protein in 0.4 M NaCl, 5 mM Tris-HCl, pH 7.7. Higher protein concentrations led to the formation of two-dimensional crystals. High salt concentration was needed to maintain the connectors in solution.

2.3. Crystallization

Initial conditions for crystallization were obtained by the hanging drop vapor diffusion method upon a broad screening following the conditions of Jancarik and Kim [9] and Cudney et al. [10] and included in the Crystal Screening I and II kits (Hampton Research), respectively. The native connector crystallized in 21% *tert*-butyl alcohol, 50 mM CaCl₂, 0.1 M Tris-HCl pH 8.0 at 20°C. The proteolysed connector crystallized in 20–30% isopropyl alcohol, 0.2 M MgCl₂, 0.1 M HEPES pH 7.5, at 20°C.

A modified batch method [11] using paraffin oil made it possible to increase the crystal size of the proteolysed connector, keeping constant the drop volume. The mixture of protein for crystallization was prepared as follows: 64 µl protein (7 mg/ml) in 0.4 M NaCl, 5 mM Tris-HCl, pH 7.7, 20 µl 0.2 M MgCl₂, 0.1 M HEPES, pH 7.5, and 24 µl pRNA 2 mg/ml. The protein solution was filtered through 0.22 µm Ultrafree-MC at 4°C. Crystallization boxes with plastic bridges were used for seeding drop experiments. On each bridge depression 10 µl of protein mixture and 50 µl of paraffin oil were applied. The bottom of the box was filled with reservoir solution containing initially 25% isopropyl alcohol. Then, it was progressively increased up to 30% isopropyl alcohol over 2 days. Following this procedure, chunky crystals (form I) appeared after 4 days and reached a maximum size of approx. 0.4×0.3×0.3 mm (Fig. 1). In some drops, plate-like crystals measuring 0.4×0.4×0.1 mm (crystal form II) appeared simultaneously.

2.4. Flash-freezing and heavy atom derivative preparation

Both crystals were extremely unstable under the X-ray at 20 or 10°C and diffracted at a very low resolution (i.e. 10 Å) using a rotating anode source. Under these conditions only a single diffraction image could be taken before complete diffraction decay. However, flash-freezing of the crystals lead to their stabilization and improvement of the diffracting resolution. The flash-frozen procedure was as follows. For crystal form I each crystal was taken directly from the crystallization drop with an appropriated loop attached to a metal tip. Immediately it was introduced in a tube with liquid propane and the tube kept in liquid nitrogen until use. The metal tip fits on a magnetic piece attached to the goniometer head for rapid mounting under the nitrogen stream, were the solid propane evaporates [12]. For crystal form II the crystals were taken from the drop with a loop, soaked on a solution containing 50% glycerol and flash-frozen directly under the nitrogen stream.

Two putative Ta₆Br₁₂²⁺ derivatives of crystal form II were prepared with different soaking times. A few crystals of Ta₆Br₁₂²⁺ were intro-

duced in the crystallization drop containing the connector crystals. After 2 h a pale-green crystal was taken and flash-frozen directly on the nitrogen stream. The same procedure was repeated with a second crystal of intense dark green color after 24 h of soaking.

2.5. X-ray data collection and processing of crystal form I

X-ray intensity data from a single frozen crystal were collected with a MAR Research imaging plate detector, using 0.869 Å synchrotron radiation at the BW7B beamline at the EMBL outstation in the Deutsches Elektronensynchrotron (Hamburg). Many crystals had to be tested before higher resolution than 10.0 Å could be observed. A 6.0 Å data set was collected at 120 K with a total of 180 frames of 1° rotation. Data indexing and reduction was performed with DENZO [13]. Crystal data and diffraction statistics are shown in Tables 1 and 2.

2.6. Crystal density determination

For the crystal form I the crystal volume per unit of protein molecular mass V_m [14] varies from 4.84 Å³/Da considering three connectors to 2.44 Å³/Da for six connectors in the asymmetric unit. Because of the size of the particles and their large central channel it was difficult to estimate a reasonable water content in the crystals. For this reason an experimental determination of density was necessary. The density was therefore measured experimentally using a ficoll gradient according to the method of Bode and Schirmer [15]. The protein crystals showed a continuous increase of apparent density during the first 3 min and after this time the crystals descended slowly. The density was determined at the inflection point of the slope as 1.09 g/cm³ which implies four connector particles in the asymmetric unit of the crystal. With 0.5% glutaraldehyde cross-linked crystals the same value was obtained. The V_m value, assuming four connectors in the asymmetric unit, is 3.61 Å³/Da.

2.7. Data collection and processing of crystal form II and derivatives

X-ray diffraction data for native and Ta₆Br₁₂²⁺ derivatives of soaking crystals of form II were collected with a MAR Research imaging plate detector, using 1.105 Å synchrotron radiation at the BW7B beamline at the EMBL outstation in the Deutsches Elektronensynchrotron (Hamburg). A 3.2 Å native data set was collected at 120 K from two crystals with a total of 90° rotation. The detector was set at 385 and 440 mm distance respectively. Data indexing and reduction was performed with DENZO [13]. Crystal data and diffraction statistics are shown in Tables 1 and 2. For crystal form II the calculated crystal volume per unit of protein molecular mass, V_m , is 2.94 Å³/dalton considering half connector in the asymmetric unit of the crystal.

3. Discussion

Three-dimensional crystals of the φ29 connector particle can be obtained from alcohol-containing solutions. The strong tendency of the connectors to aggregate in two-dimensional layers has to be prevented by adjusting a precise protein and salt concentration. The removal of micro-aggregates or crystal nuclei by filtering the protein solution is also necessary in order to avoid the formation of large two-dimensional crystal layers in the crystallization drops (Fig. 1).

Three-dimensional crystals of both native and proteolysed connectors are obtained this way, but the later grow to larger size and can be stabilized for X-ray data collection. The proteolytic cleavage removes 13 residues of the N-terminus and 18 residues of the C-terminus. The N-terminal region of p10 contains a stretch of basic charged residues that could hinder

Table 1
Crystal data

Crystal form	Space group	Cell dimensions (Å)	V_m (Å ³ /Da)	Molecules/connectors in the asymmetric unit
I	C2	$a = 416.8$ $b = 227.6$ $c = 236.7$ $\beta = 96.3$	3.61	48/4
II	P4 ₂ 2 ₁ 2	$a = b = 170.2$ $c = 156.9$	2.94	6/0.5

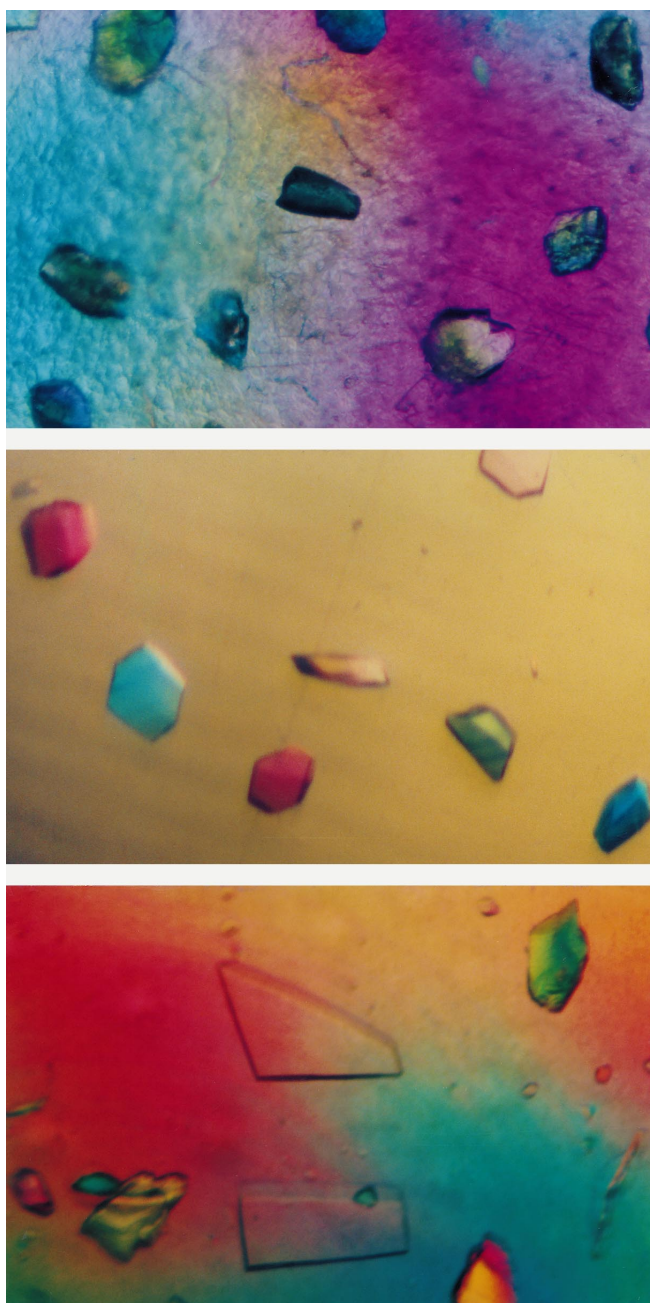


Fig. 1. Crystals of the bacteriophage ϕ 29 connector obtained with the proteolyzed p10 protein. Top: Crystals of form I obtained without filtering the protein solution. A background of two-dimensional layers as the ones used for EM studies covers the bottom of the drop. Middle: After filtering the initial p10 protein solution crystals of form I grow from clear drops and show faces and sharp edges. Bottom: Plate-like tetragonal crystals of form II coexist in some drops with crystals of form I.

the crystallization process. It is also possible that because both N- and C-terminal ends are presumably located at the narrow edge of the connector cylinder [1], the elimination of the two

tails in this region must favor the packing of the particles in the c axis direction, that is, the piling of the connectors along their 12-fold axis. Thus, three-dimensional crystals are favored

286

A. Guasch et al./FEBS Letters 430 (1998) 283–287

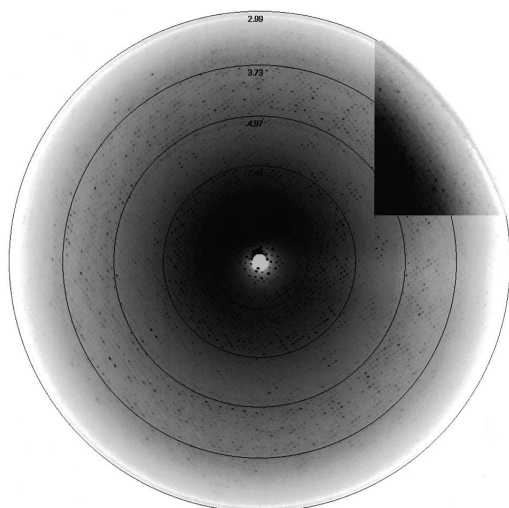


Fig. 2. Diffraction image obtained from a frozen crystal of form II, using synchrotron radiation. The crystal is oriented with the *c* axis approximately parallel to the X-ray beam. Resolution rings are labeled in Å.

over two-dimensional layers, which appear easily in the crystallization drops containing the native protein.

The use of paraffin oil covering the crystallization drops is also critical in order to enlarge the crystals for X-ray diffraction analysis. Without oil, the volume of the drop increases during the crystallization process because a volatile precipitant is used and therefore the protein is diluted during the vapor diffusion experiment. Being the protein concentration already low in the initial solution, the small crystals obtained in drops without oil are unable to grow further. The use of paraffin oil is an efficient alternative crystallization method to dialysis buttons, which were also tried unsuccessfully in order to prevent the increase in the drop volume. The paraffin oil also retards nucleation by slowing the alcohol diffusion into the crystallization drop.

The p10 monomer contains two cysteine residues that form a disulfide bridge (J.J. Calvete, J.M. Valpuesta and J.L. Carrascosa, unpublished results), and seven methionines. Preventing oxidation – most probably at a methionine residue – is also important in order to reproduce crystallization. The pro-

tein sample was treated with *N*-methylmercaptoacetamide, an effective reducing agent that converts methionine sulfoxide to methionine under relatively mild conditions [16].

Two crystal forms have been obtained, appearing sometimes in the same crystallization drops. Although both crystals diffract to high resolution, they display very different diffraction patterns. Crystals of form I diffract anisotropically and weakly beyond 6 Å. Even though reflections can be observed up to 3.4 Å resolution, no useful data can be collected in the 6–3.4 Å resolution range. An attempt to do so resulted in low completeness, high R-merge values and unacceptable signal to noise ratio. On the other hand, for crystal form II diffraction intensities can be observed up to 3.2 Å, with good completeness in the last resolution shell and reasonable $\langle I/\sigma I \rangle$ values (Table 2). The diffraction pattern (Fig. 2) does not show the strong anisotropy and dramatic intensity drop after 6 Å seen in crystal form I. In crystal form II a packing arrangement as that described for the two-dimensional sheets of $\phi 29$ connectors is expected to occur. Electron diffraction patterns obtained from frozen-hydrated thin crystals show unit cell parameters of $a = b = 165$ Å and 422 symmetry [3], similar to those found in three-dimensional crystals of form II. An analysis of the internal symmetry of the particle will be published elsewhere [17]. Crystal of form II will be used for further crystallographic analysis, including the search for new heavy atom derivatives suitable for phasing.

Acknowledgements: This work was supported by the Ministerio de Educación y Ciencia (Grants PB95-0224 to M.C. and PB96-0818 to J.L.C.), by the CSIC (Acciones especiales 1993–95) and the Generalitat de Catalunya (Centre de Referència en Biotecnologia and Grant 1997SGR-275 to M.C.). Data collection at EMBL-DESY was supported by the EU Large Installations Project CHGECT93-0040. We thank R. Huber for providing the $\text{Ta}_6\text{Br}_{12}^{2+}$ compound and for useful discussion.

References

- [1] Valpuesta, J.M. and Carrascosa, J.L. (1994) *Q. Rev. Biophys.* 27, 107–155.
- [2] Casjens, S. and Hendrix, R. (1988) in: *The Bacteriophages* (Calendar, R.A., Ed.), Vol. 1, pp. 15–91, Plenum, New York.
- [3] Valpuesta, J.M., Carrascosa, J.L. and Henderson, R. (1994) *J. Mol. Biol.* 240, 281–287.
- [4] Müller, D.J., Engel, A., Carrascosa, J.L. and Vélez, M. (1997) *EMBO J.* 16, 2547–2553.
- [5] Donate, L.E., Valpuesta, J.M., Rocher, A., Méndez, E., Rojo, F., Salas, J.L. and Carrascosa, J.L. (1992) *J. Biol. Chem.* 267, 10919–10924.
- [6] Donate, L.E., Valpuesta, J.M., Mier, C., Rojo, F. and Carrascosa, J.L. (1993) *J. Biol. Chem.* 268, 20198–20204.

Table 2
Data collection statistics

Crystal form	I	II	TaBr derivative
Resolution range (Å)	40–6.0	40–3.2	40–7.0
Number of unique reflections	41 193	35 898	3 437
Completeness (%)			
All data	75.7	89.0	88.1
Highest resolution shell (range in Å)	64.9 (6.21–6.0)	80.8 (3.31–3.2)	87.6 (7.25–7.0)
Average redundancy	3.3	3.5	2.7
$\langle I/\sigma I \rangle$			
All data	3.9	6.1	8.0
Highest resolution shell (range in Å)	1.2 (6.21–6.0)	2.3 (3.31–3.2)	4.1 (7.25–7.0)
R_{merge}^a	11.3	15.4	11.5
Highest resolution shell (range in Å)	15.7 (6.21–6.0)	20.3 (3.31–3.2)	19.8 (7.25–7.0)

^a $R_{\text{merge}} = \sum I_h - \langle I_h \rangle / \sum \langle I_h \rangle$ where $\langle I_h \rangle$ is the average *I* of reflection *h* and their symmetry related reflections.

- [7] Valpuesta, J.M., Donate, L.E., Mier, C., Herranz, L. and Carrascosa, J.L. (1993) *EMBO J.* 12, 4453–4459.
- [8] Ibáñez, C., García, J.A., Carrascosa, J.L. and Salas, M. (1984) *Nucleic Acids Res.* 12, 2351–2356.
- [9] Jancarik, J. and Kim, S.H. (1991) *J. Appl. Crystallogr.* 24, 409–411.
- [10] Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. and McPherson, A. (1994) *Acta Crystallogr. D* 50, 414–423.
- [11] Blow, D.M., Chayen, N.E., Lloyd, L.F. and Saridakis, E. (1994) *Protein Sci.* 3, 1638–1644.
- [12] Rodgers, D.W. (1994) *Structure* 2, 1135–1140.
- [13] Otwinowski, Z. (1993) DENZO: an Oscillation Data Processing Program for Macromolecular Crystallography, Yale University, New Haven, CT.
- [14] Matthews, B.W. (1968) *J. Mol. Biol.* 33, 491–497.
- [15] Bode, W. and Schirmer, T. (1985) *Biol. Chem. Hoppe-Seyler* 366, 287–295.
- [16] Houghten, R.A. and Li, C.H. (1979) *Anal. Biochem.* 98, 36–46.
- [17] Guasch, A., Pous, J., Párraga, A., Valpuesta, J.M., Carrascosa, J.L. and Coll, M. (1998) *J. Mol. Biol.* (in press).

JMB

COMMUNICATION

Crystallographic Analysis Reveals the 12-fold Symmetry of the Bacteriophage ϕ 29 Connector Particle**Alicia Guasch¹, Joan Pous¹, Antonio Párraga¹, José M. Valpuesta²
José L. Carrascosa² and Miquel Coll^{1*}**

¹*Departament de Biologia Molecular i Cel·lular, Centre d'Investigació i Desenvolupament, CSIC Jordi Girona, 18-26 08034 Barcelona, Spain*

²*Centro Nacional de Biotecnología, CSIC Universidad Autónoma de Madrid, Cantoblanco 28049 Madrid, Spain*

The internal symmetry of the connector or portal particle from the double-stranded DNA bacteriophage ϕ 29 has been examined by X-ray crystallography. This large multimeric structure (420 kDa) is built up by a number of identical subunits of the p10 protein. It connects the head of the virus with the tail and plays a central role in the prohead assembly and DNA packaging. For the first time a bacteriophage connector has been crystallized and X-ray data have been collected up to a resolution of 3.2 Å. A self-rotation function has been calculated, unambiguously revealing the 12-fold symmetry of the particle and its orientation in the crystal lattice. The orientation has been confirmed by calculating a cross-rotation function using a low resolution model based on electron microscopy reconstructions.

© 1998 Academic Press

Keywords: bacteriophage ϕ 29; connector structure; internal symmetry; X-ray crystallography

*Corresponding author

ϕ 29 is a complex double-stranded DNA bacteriophage that infects *Bacillus subtilis* cells. The viral particles are formed by a prolate icosahedral capsid, or head, and a tail. Between these two structures there is a connecting region called portal vertex or connector. This head-to-tail connector plays an important role in the first steps of head assembly and the packaging of DNA. The connector is also involved in the process of DNA transfer into the host cell (for reviews see Casjens & Hendrix, 1988; Valpuesta & Carrascosa, 1994).

The ϕ 29 connector not only interacts with the prohead and the tail, but also with the DNA that is going to be packaged. The DNA binding domain has been located in the N-terminal region of the connector protein p10, and its removal by treatment with endoprotease Glu-C from *Staphylococcus aureus* V8 (V8 from now on) generates a connector structure that is unable to bind and package DNA but maintains its quaternary structure (Donate *et al.*, 1992). For the packaging to occur, an RNA of viral origin (prohead RNA or pRNA) is absolutely required (Guo *et al.*, 1987). The RNA binds to the connector and its binding site has been located in the N-terminal region of the protein p10, not far

from the DNA-binding domain (Donate *et al.*, 1993). Finally, an ATPase of viral origin is also required to provide energy to the translocating machinery. This ATPase interacts with the connector although its binding site is not known yet (Valpuesta & Carrascosa, 1994).

The symmetry of the particle

The low resolution three-dimensional structure of the ϕ 29 connector, built up by several copies of the 35.8 kDa gene 10 product, has been obtained by electron microscopy and image processing of tilted specimens (Carazo *et al.*, 1986). It is composed of two morphological domains: a wider domain of about 14.5 nm in diameter with protruding lobules and a narrower cylindrical domain of 8 nm in diameter. Running through these two domains, there is a channel of about 4 nm diameter through which the DNA has been hypothesized to enter the viral capsid. The three-dimensional structures of the connector from bacteriophages T3 (Valpuesta *et al.*, 1992) and SPP1 (Tavares *et al.*, 1995) have shown similar general features to that of the ϕ 29 connector. The projection image of several phage connectors (ϕ 29, T3, λ , p22, T4, SPP1) has been obtained from electron microscopy and image processing of untilted specimens. The projection structure of all these connectors, albeit their

E-mail address of the corresponding author: mcccri@cid.csic.es

different dimensions, look very similar: a central mass surrounding the channel and a set of protruding lobules around the central mass. The difference at the level of resolution obtained for these connectors (20 to 25 Å) resides in the number of lobules surrounding the central mass and therefore the number of copies upon which the connector is made. In the case of the connectors from phages λ (Kochan *et al.*, 1984), T4 (Driedonks *et al.*, 1981) and P22 (Bazin *et al.*, 1988), the results obtained show a clear 12-fold symmetry. In the case of phage connector SPP1 (Dube *et al.*, 1993), the projection structure reveals a clear 13-fold symmetry. The symmetry of other phage connectors is more controversial and reports postulating 12-fold or 13-fold symmetry have been published for phage connectors $\phi 29$ (Carazo *et al.*, 1986; Dube *et al.*, 1993; Tsuprun *et al.*, 1994), T3 (Valpuesta *et al.*, 1994; Kocsis *et al.*, 1993) and T7 (Cerritelli & Studier, 1996). In the case of the $\phi 29$ connector, a projection obtained at medium resolution (10 Å) using cryoelectron microscopy of unstained two-dimensional crystals (Valpuesta *et al.*, 1994) indicates that the connector has 12 mass units protruding from a central mass. Recently, the same crystals have been subject to atomic force microscopy studies (Müller *et al.*, 1997) and 12 lobules around the central mass have been imaged. The discrepancy in the results obtained for the $\phi 29$ connector have been tried to explain as a result of the non-controlled overexpression of the recombinant connector protein in the bacteria, that could lead to the production of 12-fold or 13-fold symmetry oligomers. Other reports have tried to explain the discrepancy as a result of some of the image processing methods used, which could prevent the identification of the 13-fold symmetric components (Taveres *et al.*, 1995). Underlying these discrepancies, there is an important question, as the bacteriophage connector is placed in one of the vertices of the prohead, which has 5-fold symmetry. The mismatch in symmetry between the capsid and the connector has been hypothesized to be an important part of the mechanism of DNA packaging (Hendrix, 1978). To understand the DNA translocating process, it may be important to determine the correct symmetry of the connector.

Three-dimensional crystals diffract up to 3.2 Å

A detailed description of the protein overexpression, connector particle preparation and crystallization will be published elsewhere (Guasch *et al.*, 1998). Briefly, the difficulty in obtaining crystals of the $\phi 29$ connector suitable for X-ray diffraction arises from the tendency of the connector particle to aggregate in two-dimensional crystal layers at high protein concentration. Only at very precise protein and salt concentrations can the connectors be maintained in solution. Removing aggregates from the protein solution by filtration through 0.22 μm pore filters was also critical in order to obtain high-resolution diffracting three-dimen-

Table 1. Data collection statistics

Resolution (Å)	40-3.2
No of measured reflections	120,145
No of unique reflections	35,898
Average multiplicity	3.3 (2.2)
R-merge ^a	12.2 (31.2)
Completeness (%)	80.8 (65.8)
$\langle I/\sigma(I) \rangle$	8.1 (2.4)

X-ray data statistics of crystal form II. Figures in parenthesis are for the outermost 3.31 to 3.2 Å resolution shell. 90° rotation data were collected from two frozen crystals at 120 K with a MAR Research imaging plate detector, using 1.105 Å synchrotron radiation at beamline BW7B at the EMBL outstation at DESY (Hamburg). Data indexing and reduction was performed with DENZO-SCALEPACK (Otwinowski & Minor, 1997).

^a R-merge = $\sum_h \sum_i |I_i(h) - \langle I(h) \rangle| / \sum_h \sum_i I_i(h)$, where h are unique reflections indices, $I_i(h)$ the i th measurement of reflection h and $\langle I(h) \rangle$ the mean intensity of all measurements of $I(h)$.

sional crystals. Crystallization was achieved using the vapour diffusion method at 20°C and sitting drops. Initial crystals of both native and proteolysed connector grown from alcoholic precipitants were too small for X-ray diffraction, likely due to the increase of the drop volume during the crystallization process and the subsequent dilution of the protein concentration in the drop. A modified batch method using paraffin oil (Blow *et al.*, 1994) allowed the increase of crystal size, keeping constant the drop volume. Larger crystals could be obtained with the proteolysed connector. Two crystal forms appeared, sometimes in the same drop, depending on the alcohol concentration. Chunky crystals of form I appeared first, after four days, reaching a maximum size of aprox. 0.4 mm × 0.3 mm × 0.3 mm. Plate-like crystals of form II have a slower growth habit and grow up to 0.4 mm × 0.4 mm × 0.1 mm. For the structural analysis described here only crystals of this second form were used.

Crystals are extremely unstable under the X-ray beam at 20 or 10°C and diffract to very low resolution (i.e. 10 Å) using a rotating anode source. Under these conditions only a single diffraction image could be taken before complete diffraction decay. However, flash-freezing of the crystals led to their stabilization and improvement of the diffracting power. The crystals were taken from the drop with a loop, soaked rapidly on a crystallization solution containing 30% (v/v) glycerol and flash-frozen directly under the nitrogen stream. X-ray diffraction data were collected using synchrotron radiation at the EMBL-BW7B beam-line of the Deutsches Elektronensynchrotron in Hamburg, Germany. Many crystals had to be tested before resolution higher than 6.0 to 7.0 Å could be observed. Crystals of form II diffract to 3.2 Å resolution (Table 1). With these crystals, the intensity of the reflections does not decay sharply beyond 7 Å as in form I (A diffraction pattern of crystal form II is shown in Figure 2 of Guasch *et al.*, 1998). Form II crystals are tetragonal, space group $P4_22_12$ and

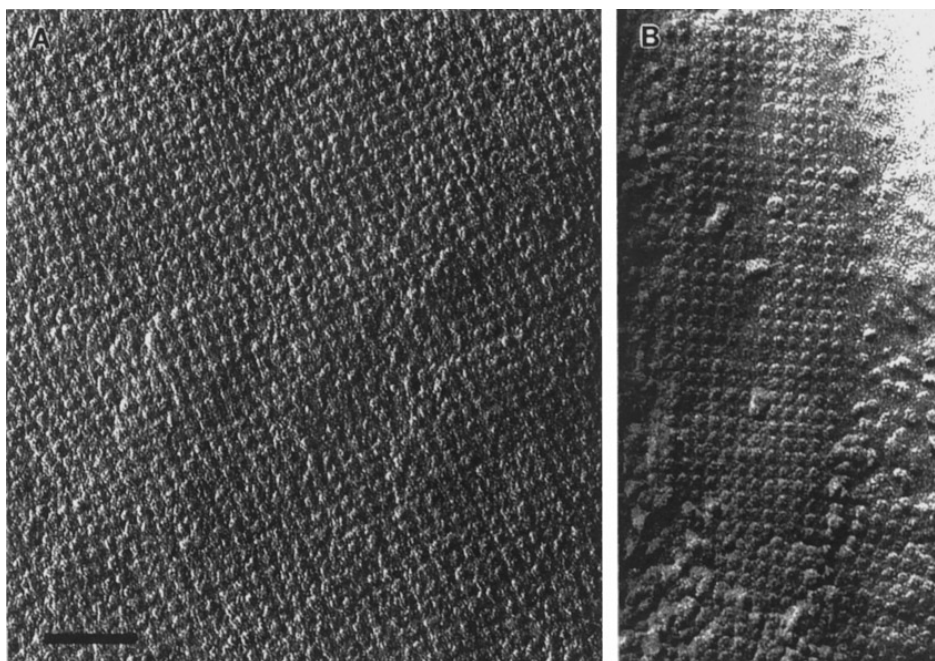


Figure 1. Surface replicas of crystal form II. Samples containing the crystals (30 μl) were spread over the surface of freshly cleaved mica (5 mm \times 5 mm) and left to adsorb at room temperature for five minutes. After removing the liquid to leave a very thin sample layer, the mica was plunged on liquid nitrogen and loaded into a Balzers 400 unit cooled at -170°C . The samples were dried at -80°C for two hours and then shadowed with a layer of around 2 nm of Pt/C at an inclination angle of 45 degrees, followed by a carbon backing of around 25 nm. The shadowed samples were subsequently floated on distilled water, mounted onto nickel grids and visualized in a JEOL 1200 EXII electron microscope, using nominal magnifications of $50,000\times$. Measurements of lattice spacing were performed using catalase crystals as standard (Wrigley, 1968). Surface replicas showed under the electron microscope a clear tetragonally ordered pattern that extended over several microns (A). Occasionally, very clean areas revealed the characteristic substructure of the connectors (B), allowing the unambiguous definition of the lattice ($a = b = 15.1(\pm 0.6)$ nm). Bar represents 100 nm.

have unit cell dimensions $a = b = 170.2$ \AA , $c = 156.9$ \AA . The calculated V_m value (Matthews, 1968) is 2.94 $\text{\AA}^3/\text{Da}$, assuming half of a connector particle per asymmetric unit.

Comparison of the crystal packing in form II and two-dimensional crystals

In an attempt to define the packing of the connectors in form II crystals, surface replicas of the plate-like crystals were obtained by Pt/C shadowing of frozen and dried samples (Figure 1). Extended areas showing tetragonal arrangement of connectors could be identified. The approximate unit cell dimensions were found to be $a = b = 15.1$ nm. This fact, together with the resemblance of these surface relieves with those already published for multi-layered crystals of $\phi 29$ connectors (Figures 1 and 2 of Carazo *et al.*, 1986), led us to suggest that crystal form II has a similar unit packaging to that described for the two-dimensional sheets of $\phi 29$ connectors. In fact, the unit cell dimensions obtained by X-ray diffraction of the crystals (170.2 \AA) are very close to those obtained either by cryoelectron microscopy of

frozen-hydrated two-dimensional sheets (165 \AA , Valpuesta *et al.*, 1994), or by atomic force microscopy of these sheets in solution (165 \AA , Müller *et al.*, 1997). The smaller unit cell dimensions found in the surface replicas could be due to a shrinkage of around 10% due to extensive drying (Kellenberger, 1987). In fact, similar amounts of shrinkage were observed in ordered arrays of viral capsomeres (Carrascosa *et al.*, 1984). The most conspicuous characteristic of the packaging of the connectors in those crystals is that the unit connector is oriented with the longitudinal axis perpendicular to the ab plane. Also, the unit cell accommodates, in one layer, two connectors with alternate orientations, one facing up, while the other is facing down relative to the plane of the crystal surface. Crystal form II has two such layers, i.e. four connectors per unit cell.

Self-rotation function

The local symmetry in the connector was investigated. Self-rotation functions were calculated in Patterson space (Huber, 1965) as implemented in XPLOR (Brünger, 1992) and in reciprocal space

(Rossmann & Blow, 1962) as implemented in GLRF (Tong & Rossmann, 1997), using polar angles. With XPLOR, the self-rotation function was calculated using data from 40 Å to 3.5 Å and Patterson vectors up to 45 Å. A very strong peak of value 14.6 (second peak = 4.5, $\sigma = 2.97$) appears at $\psi = 90^\circ$, $\phi = 90^\circ$, $\kappa = 30^\circ$ (Figure 2A). This peak corresponds to the local 12-fold axis of the particle and is parallel to the crystallographic c axis. Perpendicular to it, at section $\kappa = 180^\circ$ and $\phi = 90^\circ$, peaks of the same height (14.6) appear at 15° and 30° on ψ and $\phi = 0^\circ$ (Figure 2B). The crystallographic 2-fold axes (21.9) are at $\psi = 0^\circ$, 45° , 90° , $\phi = 0^\circ$ and $\psi = 90^\circ$, $\phi = 90^\circ$ as expected for the 422 symmetry. The peaks at 15° and 30° are equivalent to the maximum at section $\kappa = 30^\circ$ (12-fold) and are a consequence of the crystallographic 2-fold axes being exactly perpendicular to the 12-fold local axis of the particle. All these peaks are con-

sistently maintained when varying the resolution range (i.e. using 40 to 6 Å resolution), the integration radius or the self-rotation function evaluation (GLRF). Figure 3 shows a plot of the self-rotation function at $\psi = 90^\circ$, $\phi = 90^\circ$ calculated in either space group $P4_22_12$ or $P1$. The crystallographic 4-fold axis is seen at $\kappa = 90^\circ$. The local symmetry of the particle is unambiguously 12-fold as indicated by the peaks at $\kappa = 30^\circ$, 60° , 120° and 150° . This rotation function was calculated with data in the resolution range 20 to 6 Å. Using higher resolution data the rotation function results in sharper peaks at the same κ values. A 13-fold local symmetry would show a peak at $\kappa = 27.69$, which is not the case.

In order to confirm the orientation of the connector particle in the crystals of form II, a cross-rotation function was calculated with AMoRe (Navaza, 1994) using as a starting model the 20 Å

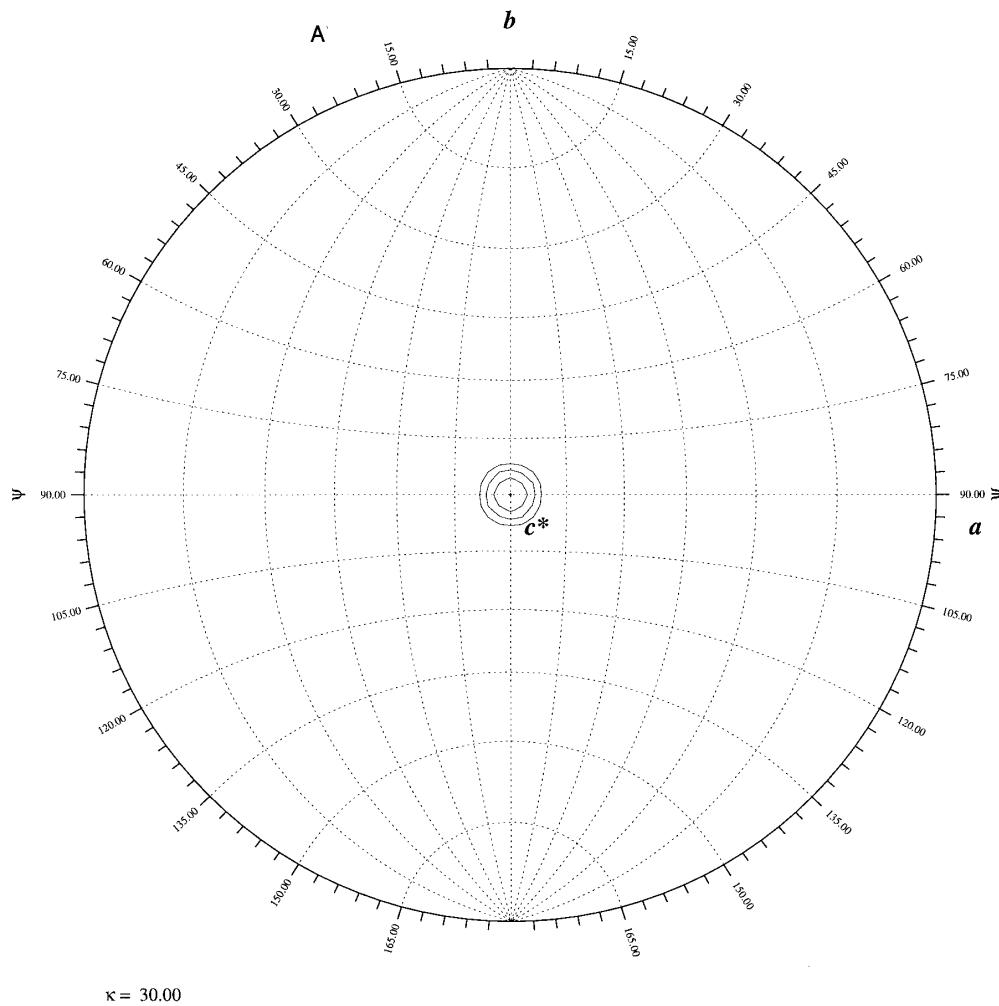


Figure 2 (legend opposite)

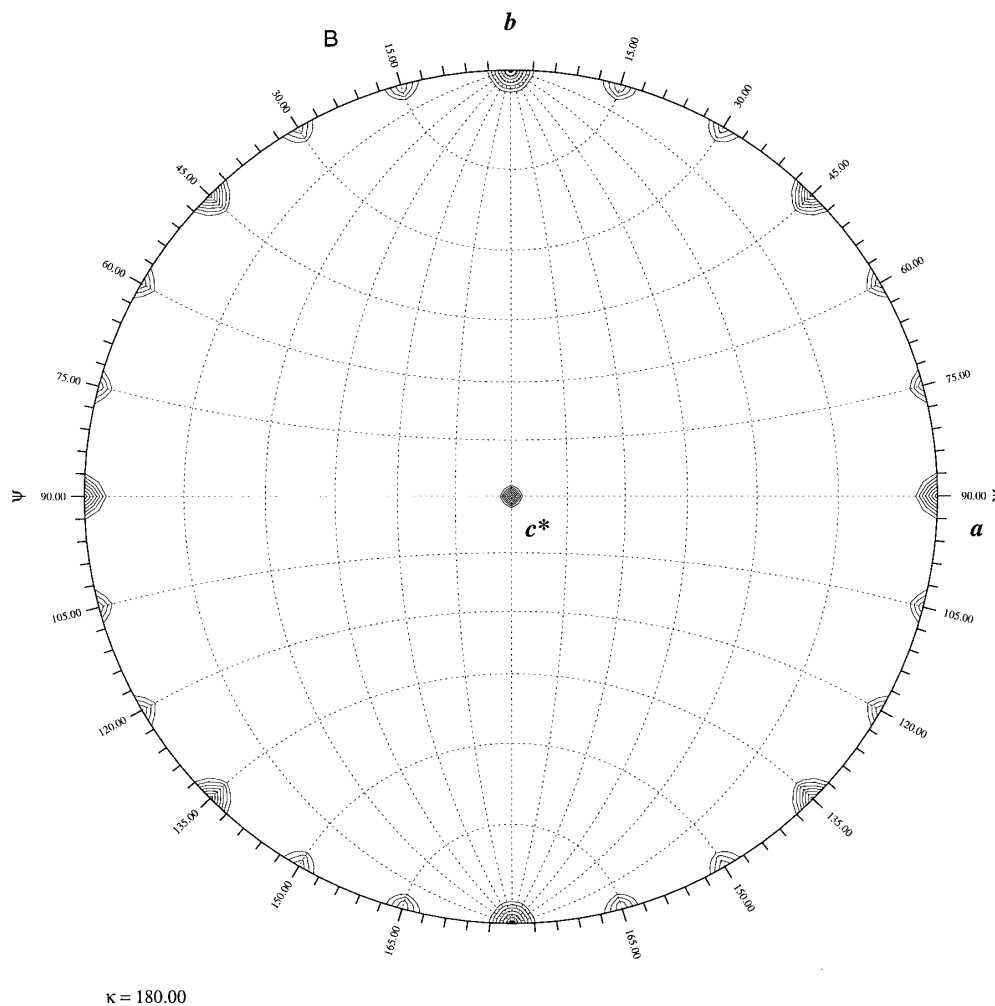


Figure 2. Stereographic projections of the self-rotation function calculated with GLRF (Tong & Rossmann, 1997) and viewed along the crystallographic c axis. Contour levels are drawn every 0.74σ . A, Section at $\kappa = 30^\circ$. The peak at $\psi = 90^\circ$, $\phi = 90^\circ$ has a value of 9.65σ . B, Section at $\kappa = 180^\circ$. The peaks at $\psi = 15^\circ$, $\phi = 0^\circ$ and $\psi = 30^\circ$, $\phi = 0^\circ$ have also a value of 9.65σ .

electron microscopy structure (Carazo *et al.*, 1986). An electron density map was built based on the EM model, oriented with its 12-fold axis parallel to the crystallographic c axis and structure factors were calculated with the option MAP of the TABLING routine in AMoRe. The cross-rotation function between the calculated and the observed amplitudes, computed in the resolution range 30 to 7 Å, gave a set of five similar maxima of value 4.7 to 4.9 (correlation coefficient = 6.1 to 7.0) at Eulerian angles $\beta = 0.0$, $\gamma = 0.0$ and $\alpha = 1.4$, 14.5, 35.3, 61.8 and 79.2. The sixth peak at $\alpha = 80.3$, $\beta = 89.7$, $\gamma = 358.0$ has a value of 3.1 (correlation coefficient = 3.7). Similar results are obtained when using a resolution range from 30 to 10 Å. The same orientation for the particle resulted from the cross-rotation function when using an atomic model,

built with atoms randomly filling a 12 Å 3D-model obtained by cryoelectron microscopy (J. M. V., Fernandez & J.L.C., unpublished). These results are consistent with the self-rotation function and confirm that the particle is oriented with its 12-fold along the crystallographic c axis although the rotation of the particle around its 12-fold is not resolved using this low resolution EM model.

Conclusion

This is the first crystallographic report of a bacteriophage connector or portal particle. The difficulties in obtaining suitable crystals for X-ray diffraction analysis have been overcome by using a proteolyzed p10 protein which makes a slightly modified connector cylinder. This probably facili-

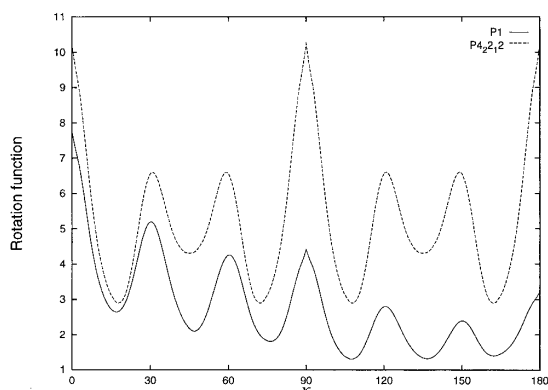


Figure 3. Plot of the self-rotation function at $\psi = 90^\circ$, $\phi = 90^\circ$, calculated in space groups $P4_2,2,2$ (dotted line) and $P1$ (continuous line). The function was computed with XPLOR (Brünger, 1992), for the κ interval 0° to 180° , using data from 20 to 6 Å and Patterson vectors up to 45 Å. The grid value in κ is 0.5° .

tates the interlayer interactions, avoiding the formation of bi-dimensional crystal aggregates. The use of oil in the crystallization drops was also critical in order to enlarge the size of the crystals. High resolution diffraction data, up to 3.2 Å, were obtained from crystal form II, a form that is related to the two-dimensional crystal layers used for previous electron microscopy and atomic force microscopy studies.

The self-rotation and cross-rotation functions allow us to determine the orientation of the particle in the crystal, with its local axis parallel to the crystallographic c axis. The self-rotation function also clarifies a controversial issue showing unambiguously that the internal symmetry of the particle is 12 rather than 13.

Acknowledgements

This work was supported by the Ministerio de Educación y Ciencia (grants PB95-0224 to M.C. and PB96-0818 to J.L.C.), by the CSIC (Acciones especiales 1993-95) and the Generalitat de Catalunya (Centre de Referència en Biotecnologia and grant 1997SGR-275). Data collection at EMBL-DESY was supported by the EU Large Installations Project CHGECT93-0040. We thank Jorge Navaza for help with AMoRe and Robert Huber for useful discussion.

References

Bazinot, C., Benbasat, J., King, J., Carazo, J. M. & Carrascosa, J. L. (1988). Purification and organization of the gene 1 portal protein required for phage P22 DNA packaging. *Biochemistry*, **27**, 1849–1856.

- Blow, D. M., Chayen, N. E., Lloyd, L. F. & Saridakis, E. (1994). Control of nucleation of protein crystals. *Protein Sci.* **3**, 1638–1634.
- Brünger, A. T. (1992). *XPLOR, Version 3.1, A System for X-ray Crystallography and NMR*, Yale University Press, New Haven and London.
- Carazo, J. M., Donate, L. E., Herranz, L., Secilla, J. P. & Carrascosa, J. L. (1986). Three-dimensional reconstruction of the connector of bacteriophage $\phi 29$ at 1.8 nm resolution. *J. Mol. Biol.* **192**, 853–867.
- Carrascosa, J. L., Carazo, J. M., Carrascosa, A. L., García, N., Santisteban, A. & Viñuela, E. (1984). General morphology and capsid fine structure of African Swine Fever Virus particles. *Virology*, **132**, 160–172.
- Casjens, S. & Hendrix, R. (1988). Control mechanisms in dsDNA bacteriophages. In *The Bacteriophages* (Calendar, R. A., ed.), vol. 1, pp. 15–91, Plenum Publishing Corp., New York.
- Cerritelli, M. E. & Studier, F. W. (1996). Purification and characterization of T7 head-tail connectors expressed from the cloned gene. *J. Mol. Biol.* **258**, 299–307.
- Donate, L. E., Valpuesta, J. M., Rocher, A., Méndez, E., Rojo, F., Salas, M. & Carrascosa, J. L. (1992). The role of the amino-terminal domain of bacteriophage $\phi 29$ connector in DNA binding and packaging. *J. Biol. Chem.* **267**, 10919–10924.
- Donate, L. E., Valpuesta, J. M., Mier, C., Rojo, F. & Carrascosa, J. L. (1993). Characterization of a RNA-binding domain in the bacteriophage $\phi 29$ connector. *J. Biol. Chem.* **268**, 20198–20204.
- Driedonks, R. A., Engel, A., ten Heggeler, B. & van Driel, R. (1981). Gene 20 product of bacteriophage T4. Its purification and structure. *J. Mol. Biol.* **152**, 641–662.
- Dube, P., Tavares, P., Lurz, R. & van Heel, M. (1993). The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry. *EMBO J.* **12**, 1303–1309.
- Guasch, A., Pous, J., Párraga, A., Valpuesta, J. M., Carrascosa, J. L. & Coll, M. (1998). *FEBS Letters*, in the press.
- Guo, P., Erickson, S. & Anderson, D. (1987). A small RNA is required for in vitro packaging of bacteriophage $\phi 29$ DNA. *Science*, **236**, 690–694.
- Hendrix, R. W. (1978). Symmetry mismatch and DNA packaging in large bacteriophages. *Biochemistry*, **75**, 4779–4783.
- Huber, R. (1965). Die automatisierte Faltmolekül-methode. *Acta Crystallog.* **19**, 353–356.
- Kellenberger, E. (1987). The response of biological macromolecules and supramolecular structures to the physics of specimen cryopreparation. In *Cryotechniques in Biological Electron Microscopy* (Steinbrecht, R. A. & Zierold, Z., eds), pp. 35–63, Springer Verlag, Berlin, Heidelberg.
- Kochan, J., Carrascosa, J. L. & Murialdo, H. (1984). Bacteriophage lambda preconnectors. Purification and structure. *J. Mol. Biol.* **174**, 433–447.
- Kocsis, E., Cerritelli, M. E., Cheng, N., Trus, B. & Steven, A. C. (1993). Polymorphism of the head to tail connector complex of bacteriophage T7: 13-fold and 12-fold symmetric variants. Proceedings of the XIII International Conference on Virus and Phage Assembly, Syria, Virginia.
- Mathews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.

- Müller, D. J., Engel, A., Carrascosa, J. L. & Vélez, M. (1997). The bacteriophage ϕ 29 head-tail connector imaged at high resolution with the atomic force microscope in buffer solution. *EMBO J.* **16**, 2547–2553.
- Navaza, J. (1994). AMoRe: an automated package for molecular replacement for molecular replacement. *Acta Crystallog. sect. A*, **50**, 157–163.
- Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
- Rossmann, M. G. & Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallog.* **15**, 24–31.
- Tavares, P., Dröge, A., Lurz, R., Graeber, I., Orlova, E., Dube, P. & van Heel, M. (1995). The SPP1 connection. *FEMS Microbiol. Rev.* **17**, 47–56.
- Tong, L. & Rossmann, M. G. (1997). Rotation function calculations with GLRF program. *Methods Enzymol.* **276**, 594–611.
- Tsuprun, V., Anderson, D. & Egelman, E. H. (1994). The bacteriophage ϕ 29 head-tail connector shows 13-fold symmetry in both hexagonally-packed arrays and as single particles. *Biophys. J.* **66**, 2139–2150.
- Valpuesta, J. M. & Carrascosa, J. L. (1994). Structure of viral connectors and their function in bacteriophage assembly and DNA packaging. *Quart. Rev. Biophys.* **27**, 107–155.
- Valpuesta, J. M., Fujisawa, H., Marco, S., Carazo, J. M. & Carrascosa, J. L. (1992). Three-dimensional structure of T3 connector purified from overexpressing bacteria. *J. Mol. Biol.* **224**, 103–112.
- Valpuesta, J. M., Carrascosa, J. L. & Henderson, R. (1994). Analysis of electron microscope images and electron diffraction patterns of thin crystals of ϕ 29 connectors in ice. *J. Mol. Biol.* **240**, 281–287.
- Wrigley, N. G. (1968). The lattice spacing of crystalline catalase as an internal standard in electron microscopy. *J. Ultrastruct. Res.* **24**, 454–464.

Edited by R. Huber

(Received 5 March 1998; received in revised form 11 May 1998; accepted 12 May 1998)

JMB

Detailed Architecture of a DNA Translocating Machine: The High-resolution Structure of the Bacteriophage ϕ 29 Connector Particle

Alicia Guasch^{1†}, Joan Pous^{1†}, Borja Ibarra², F. Xavier Gomis-Rüth¹, José María Valpuesta², Natalia Sousa², José L. Carrascosa² and Miquel Coll^{1*}

¹*Institut de Biologia Molecular de Barcelona, CSIC, Jordi Girona 18-26, E-08034 Barcelona, Spain*

²*Centro Nacional de Biotecnología, CSIC Universidad Autónoma de Madrid, Canto Blanco E-28049, Madrid, Spain*

The three-dimensional crystal structure of the bacteriophage ϕ 29 connector has been solved and refined to 2.1 Å resolution. This 422 kDa oligomeric protein connects the head of the phage to its tail and translocates the DNA into the prohead during packaging. Each monomer has an elongated shape and is composed of a central, mainly α -helical domain that includes a three-helix bundle, a distal α/β domain and a proximal six-stranded SH3-like domain. The protomers assemble into a 12-mer, propeller-like, super-structure with a 35 Å wide central channel. The surface of the channel is mainly electronegative, but it includes two lysine rings 20 Å apart. On the external surface of the particle a hydrophobic belt extends to the concave area below the SH3-like domain, which forms a crown that retains the particle in the head. The lipophilic belt contacts the non-matching symmetry vertex of the capsid and forms a bearing for the connector rotation. The structure suggests a translocation mechanism in which the longitudinal displacement of the DNA along its axis is coupled to connector spinning.

© 2002 Academic Press

Keywords: bacteriophage ϕ 29; X-ray diffraction; connector particle; DNA packaging machinery

*Corresponding author

Introduction

DNA translocation is a central biological issue since it is essential to the processing and transfer of genetic information, including cell infection by viruses. It involves a number of basic mechanisms that are probably common to many systems (nucleic acid replication, transcription, translation, conjugation, etc.). Many proteins involved in DNA translocation, or processive movements along the DNA, function as oligomers with a general toroidal morphology.¹ Among these toroidal particles are the bacteriophage head-to-tail connectors, present in one particular vertex of viral heads. The head-to-tail connecting region of bacteriophage ϕ 29 is one of the best known. ϕ 29 is a small double-stranded DNA bacteriophage that infects *Bacillus subtilis* cells. The viral particles are formed by a

prolate icosahedral capsid or head, which contains a 19 kb long DNA, and a tail of complex structure. Between these two structures there is a connecting region called portal vertex or connector. This head-to-tail connector plays an important role in the first steps of head assembly and is the central piece of the DNA packaging machinery. In addition, during viral infection, the DNA is ejected through the tail and hence traverses the connector in the opposite direction (for reviews see references^{2,3}).

The packaging of DNA into the head involves, besides the connector, two other essential components that interact in its distal part: an RNA called pRNA,⁴ and the ATPase p16, required to provide energy to the translocation machinery. Both RNA and ATPase are of viral origin and, once the DNA has been packaged, they are released from the connector while the collar and tail components assemble from its distal part.

Electron microscopy studies, based on two-dimensional projections and three-dimensional reconstructions, showed that the ϕ 29 connector is an oligomeric protein built up from 12 copies of

[†]Both authors contributed equally to this study and share first authorship.

E-mail address of the corresponding author: mcccri@ibmb.csic.es

the 36 kDa gene 10 product.⁵ This structure is shaped like a propeller, in two parts: a wide disk, with protruding lobes that interacts with the capsid proteins, and a narrower cylinder that has a longitudinal channel along its axis and contacts the collar and tail proteins.^{6,7}

In spite of the absence of sequence homology, viral connectors have similar structural features. The three-dimensional reconstructions of the connector from bacteriophages T3⁸ and SPP1⁹ have revealed striking similarities to the ϕ 29 connector. This common architecture is also seen in other viruses: the connectors from bacteriophages, λ ,¹⁰ p22¹¹ and T4¹² show projection images that are fully consistent with a ring of appendages enclosing a narrower domain with a channel, similar to those obtained from ϕ 29,^{13,14} T3,¹⁵ or SPP1.¹⁶ These connectors differ only in dimensions (which correlate to the molar mass of the connector proteins), and all of them present 12 outer appendages, except for SPP1, which has 13.

We have previously reported the crystallization of the ϕ 29 connector particle in two crystal forms (C2 and $P4_2$).¹⁷ The crystals diffracted to 3.6 and 3.0 Å, respectively, the second form showing merohedral twinning. These initial crystallographic studies, together with the cryo electron microscopy data¹⁸ and the atomic force microscopy images obtained from connector aggregates in solution,¹⁹ solved a controversial issue about the symmetry of the ϕ 29 connector (reviewed in reference³). A self-rotation function calculation demonstrated unambiguously the 12-fold local symmetry of the particle.²⁰ Recently, Simpson *et al.*²¹ reported an unrefined low-resolution crystal structure of the connector, determined from C2 crystals. Although our present data fit the general outline of that structure, the previous model differs markedly in both the chain tracing and the side-chain positioning at the proximal part, where five β -strands in each of the 12 protomers are missing, and the main chain is out of register from residue 271 onward. Therefore, the crucial SH3-like domain is not described in the Simpson model.

Here we report the high-resolution 2.1 Å structure of the bacteriophage ϕ 29 connector particle, obtained from a new crystal form. The quality of these data allowed us to correctly trace the polypeptide chain and refine the structure. The resulting structural details reveal an amazing DNA translocation device and suggest a simple rigid-body rotary mechanism of the connector coupled to the linear movement of the double helix through its central channel.

Results

Quality of the data

All previous attempts to obtain high-quality diffracting data on the connector particle were unsuccessful. The size of the particle and, most probably, the presence of flexible regions at the N and C

terminus of the protein have so far prevented the collection of high-resolution data. This problem was solved by the controlled proteolysis of p10 by the protease Glu-C from *Staphylococcus aureus* V8, which removes the first 13 residues from the N terminus and the last 18 residues from the C terminus.²² The C2 crystals obtained previously by us¹⁷ and others²¹ contain one or several particles in the asymmetric unit. This feature facilitates the phase extension procedure, because 12-fold averaging can be applied, but the monoclinic crystals hardly diffract to 3.5 Å. On the other hand, the orthorhombic $P4_2$ crystals¹⁷ diffracted to higher resolution (about 3 Å) but showed a merohedral twinning problem, where a 50% mixture of up and down connector orientations was present.

The crystals here reported have a new form, space group $I422$. One connector particle occupies a special position with its 12-fold local axis coincident with the crystallographic 4-fold axis. Therefore, the asymmetric unit includes only three protomers. The particles are tightly packed with many inter-particle contacts and the crystals diffract to 2.1 Å, a remarkably high resolution for a particle of this size.

Structure determination

A 10 Å resolution electron microscopy model⁸ was used for initial phasing in the averaging/phase extension procedure. The coincidence of the 12-fold axis of the connector with the crystallographic 4-fold axis and the fact that only 3-fold averaging could be applied resulted in an uninterpretable map. However, because high resolution was available, this map was used for an automatic structure determination with the program Arp/Warp. By alternating Arp/Warp cycles with 3-fold averaging cycles, the map improved substantially and some secondary structure elements showed up, in particular the long helices of the central domain. At this stage, however, the coordinates of the Simpson X-ray model became available (pdb code 1fou) and they were also used for molecular replacement. The model was not good enough to provide a solution using a protein monomer, but a search with the whole connector (see Materials and Methods), gave an identical solution to that obtained with the electron microscopy model. From the resulting maps, given the high resolution of the data, the chain tracing could be corrected. Subsequently, the structure was fully refined, with the addition of solvent molecules (Figure 1(b)).

Structure of the monomer

Each monomer consists of three domains (Figure 2(a)), a central domain that extends to the internal face of the wide proximal end, an SH3-like domain at the exterior of the wide proximal end, and a distal domain at the tip of the narrow end. The central domain contains five α -helices (α 1, α 2, α 3, α 5 and α 6) and two β -strands (β 1 and β 11).

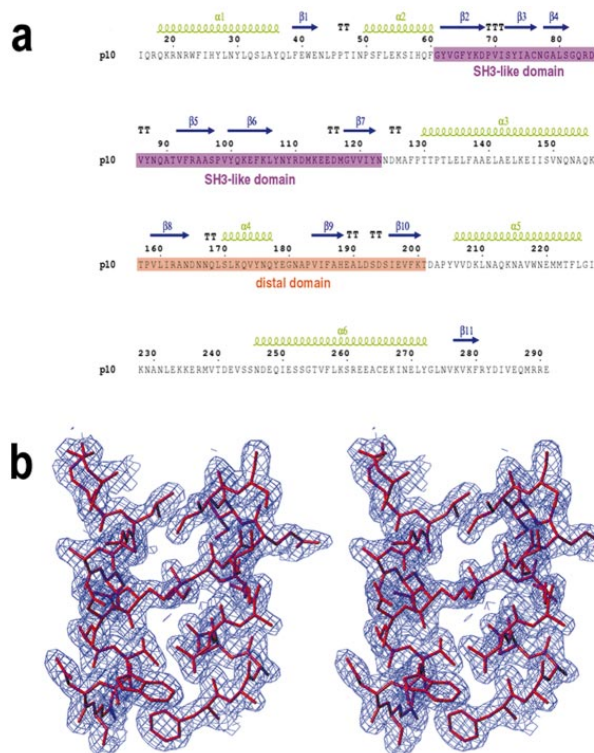


Figure 1. (a) Primary and secondary structure of bacteriophage $\phi 29$ connector protein. Helices ($\alpha 1$ - $\alpha 6$) are displayed in green and strands ($\beta 1$ - $\beta 11$) as blue arrows, as determined with DSSP.⁵⁷ Distinct subdomains of the structure are shadowed in pink and magenta. (b) Stereo pair of the final sigmaA-weighted $2F_o - F_c$ electron density map contoured at the 1σ level, superimposed over the final refined model shown as red sticks, around the region of interaction between helices $\alpha 5$ and $\alpha 6$.

Figure 1(a) shows the residues included in each secondary structure element. Helices $\alpha 1$, $\alpha 3$ and $\alpha 5$ form a bundle in which $\alpha 3$ and $\alpha 5$ are anti-parallel and $\alpha 1$ forms an angle of 35° with the other two, running anti-parallel to $\alpha 3$. Helix $\alpha 2$ is located at the C terminus of $\alpha 1$ and the N terminus of $\alpha 3$, almost perpendicular to both, while helix $\alpha 6$ is placed at the C terminus of $\alpha 5$. Helix $\alpha 6$ has a kink between residue Val256 and Phe257. This causes a deviation of the helical axis of about 30° . Interestingly, the axes of helices $\alpha 5$ and the second part of $\alpha 6$ (after the kink) are almost coincident and thus, this part of $\alpha 6$ appears as a continuation of $\alpha 5$. Finally strands $\beta 1$ and $\beta 11$ form an anti-parallel β -ribbon, at the proximal part of the central domain.

A total of 14 residues, from 231 to 244, in the loop between $\alpha 5$ and $\alpha 6$, are not seen in the electron density maps. We checked the molecular weight of the crystallized protein by mass spectroscopy and ruled out proteolytic cleavage at this site. Therefore, this loop is a flexible region, facing the interior of the channel, and it might play a role in closing the gate after DNA packaging. The rest of the loops of this central domain are well defined and include two β -turns, between $\beta 1$ and $\alpha 2$ and between $\beta 7$ (of the proximal SH3-like domain) and $\alpha 3$.

The SH3-like domain is inserted between $\alpha 2$ and $\alpha 3$ of the central domain, and forms the protruding lobules of the proximal wide part of the particle. It consists of six β -strands ($\beta 2$ to $\beta 7$) with an SH3-like β -barrel topological arrangement, where $\beta 2$, $\beta 3$ and $\beta 7$ form an anti-parallel three-stranded β -sheet which lies perpendicular to the second anti-parallel three-stranded β -sheet formed by $\beta 4$, $\beta 5$ and $\beta 6$ (Figure 2(b)). The two sheets enclose a compact hydrophobic core that includes six aromatic residues, three of them (Tyr66, Phe104 and Phe93) perfectly stacked. The connections of $\beta 2$ with $\beta 3$, $\beta 3$ with $\beta 4$ and $\beta 5$ with $\beta 6$ are rather short, with a β -turn between $\beta 2$ and $\beta 3$. On the other hand, the connections from $\beta 4$ to $\beta 5$ and $\beta 6$ to $\beta 7$ are ten-residue long loops or hairpins, both including β -turns. A structural homology search with the program DALI²³ did not reveal any similarity with other known structures. Nevertheless, it is clear from Figure 2(b) and superposition studies that the topology is similar to the SH3 protein-protein recognition domains. However, it does not form a cleft for polyproline binding, because loop $\beta 6$ - $\beta 7$ would occupy part of the cleft space and loop $\beta 2$ - $\beta 3$ is too short to form one of the walls of the cleft, in contrast to the long equivalent RT-*Src* loop of typical SH3 domains.

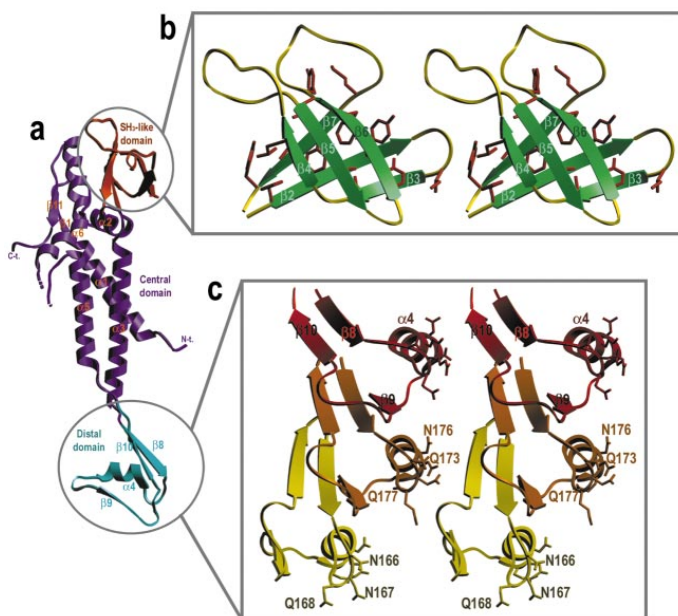


Figure 2. (a) Ribbon cartoon of a bacteriophage $\phi 29$ connector protein monomer displaying the labelled secondary structure elements. The three domains of the connector monomer are shown in different colours. (b) and (c) Close up stereo views of the two distinct apical domains of the bacteriophage $\phi 29$ connector protein monomer, the SH₃-like domain (b) and the distal domain (c). The regular secondary structure elements are labelled. In (b) the hydrophobic residue side-chains making up the filling of the β -sandwich are displayed as orange sticks. In (c) three vicinal distal domains are displayed in red, orange and yellow, respectively. In the latter two, the side-chains of the residues making up the Gln/Asn belt are displayed as sticks. Figure created with SETOR.⁵⁸

The distal domain, at the narrow tip of the connector, consists of three β -strands ($\beta 8$, $\beta 9$ and $\beta 10$), one α -helix ($\alpha 4$) and a β -turn located between the last two β -strands (Figure 2(c)). Two strands ($\beta 8$ and $\beta 10$) of one monomer and one strand ($\beta 9$) of an adjacent monomer form a mixed β -sheet where $\beta 8$ and $\beta 9$ are parallel and $\beta 10$ anti-parallel. Helix $\alpha 4$ runs anti-parallel to $\beta 8$ and also belongs to the adjacent monomer (Figure 2(c)). In this way the monomers are strongly interconnected.

The architecture of the p10 protein monomer does not share any similarity with other known protein structures. Only the disposition of the three-helix bundle at the central domain is reminiscent of other three-helix bundles, like that found in the N-terminal domain of the low-density lipoprotein receptor related protein RAP (pdb code 1lre).

Structure of the dodecamer

Twelve protomers assemble in a propeller-like super-structure that has an external diameter of 146 Å at its proximal part and 77 Å at its distal end (Figure 3(a) and (b)). The height of the connector is 75 Å. The three helices forming the bundle of the central domain are not parallel with the 12-fold axis of the particle, but laterally inclined, about 45°. This implies that the proximal part of monomer n , including the SH3-like protuberance, projects over the distal domain of monomer $n + 2$, as seen when looking down the 12-fold axis of the particle.

The interactions between monomers are different in the three parts of the structure. For example (as

indicated above), in the distal domain one of the β -strands of the three-stranded β -sheet belongs to the adjacent monomer and, therefore, multiple hydrogen bonds form between monomers (Figures 2(c) and 4(c)). In addition, there are several hydrophobic contacts, involving eight non-polar residues: Tyr175A-Pro183B, Leu171A-Ile185B, Asp165A-Ala187B, Thr199A-Ile196B and Leu160A-Ile196B (A, B and C denote adjacent monomers). Furthermore, there is an H-bond between Asp194A and Ser100B. Interestingly, residues of three consecutive monomers (Thr157A, Tyr175B and Pro183C) converge at one point establishing hydrophobic contacts between each other.

In the central domain most of the contacts are formed by residues belonging to the helices (Figure 4(b)). Helices $\alpha 5$ of two consecutive monomers run parallel along almost their entire length, establishing multiple contacts. Helix $\alpha 3$ contacts an adjacent $\alpha 3$ with its C-terminal half, while its N-terminal half contacts $\alpha 1$. The N-terminal part of the kinked $\alpha 6$ helix interacts with the last turns of $\alpha 1$ and $\alpha 5$ of a neighbouring protomer.

In the proximal part, the C-terminal half of helix $\alpha 6$ contacts the C terminus of helix $\alpha 2$ and residues of strand $\beta 1$ (Figure 4(a)). More important are the interactions at the SH3-like domain. Loop $\beta 5$ - $\beta 6$, together with the loop $\beta 1$ - $\alpha 2$ (which does not belong to the SH3-like domain, but which is also in the proximal part of the connector) define a hydrophobic cleft with strand $\beta 3$ at the bottom. Loop $\beta 4$ - $\beta 5$ of a neighbouring molecule intrudes into this cleft establishing hydrophobic interactions. Tyr87A is at the tip of the loop and forms several hydro-

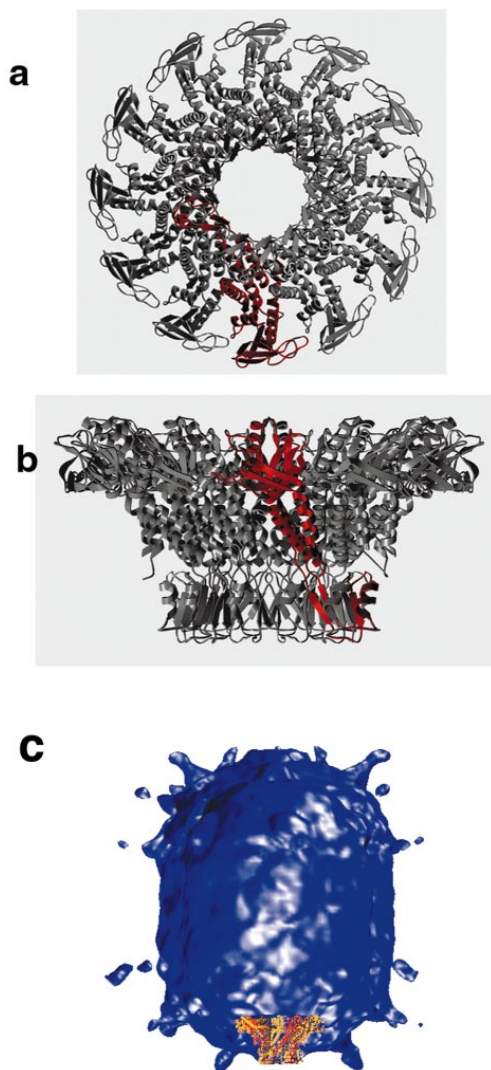


Figure 3. (a) and (b) Ribbon representations of a whole dodecamer of the bacteriophage ϕ 29 connector protein. One of the 12 monomers is displayed in distinct red. The structure is shown in its axial (a) and lateral (b) views. Figure made with VOLUMES/BOBSCRIPT^{59,60} and rendered with Raster-3D.⁶¹ (c) Location of the connector structure in the viral proheads. The connector X-ray structure was fitted into the three-dimensional reconstruction of the ϕ 29 prohead obtained by cryo-electron microscopy at 3 nm resolution³⁴. The fits of both volumes and their graphic representation were generated using the VMD program.⁶²

phobic contacts with Ala75B and Ile74B, both from β 3. It is edge-to-face stacked on Phe52B. Furthermore, its OH group forms H-bonds with the carbonyl oxygen of Thr47B and the amide nitrogen of

Ala75B. Val86A of the same β 4- β 5 loop contacts Ile74B and Val99B, while Gln89A forms an H-bond with Asn49B. Further contacts are observed between loop β 6- β 7 and loop β 1- α 2 of a neighbouring protomer, including Arg110A-Pro46B and Tyr109A-Thr47B.

The buried surface at a monomer-monomer interface is 5519 Å², which means that, in the dodecamer, only one third of the total area of each monomer is accessible. Thus, a large proportion of the surface is involved in the inter-protomer contacts, resulting in a strongly bound particle. This is in accordance with previous biochemical data indicating that single monomers are never observed, even at high salt concentrations. The connector structure is disassembled²⁴ only when incubated at high chaotropic agent concentrations (2.5 M guanidinium hydrochloride).

Internal channel and surfaces

A channel traverses the particle longitudinally. Its width is about 35 Å at the distal end and widens up at the proximal end. The width at the proximal end is difficult to estimate accurately, however, since the disordered and hence not modeled residues 231-244 and the excised C terminus are present in that area. The proximal end of the channel may thus be much narrower than is shown in this model.

In Figure 5(a), a space filling representation highlights different features of the exterior surface of the particle. Going from the proximal wide end to the distal narrow end, we find first the protuberance that corresponds to the SH3-like domain with positively and negatively charged polar residues. At the subsequent concave surface, there is a belt of hydrophobic residues, including Phe24, Ile25, Leu28, Phe60, Phe128, Pro129, and Pro132. Then, another area of mixed positively and negatively charged residues follows. This part corresponds to the N terminus of the protein, where 14 residues were excised. Further down, a ring with exposed Asn and Gln residues forms the external surface of the distal domain. It encompasses a large concentration of residues with side-chains having amide moieties, including Asn166, Asn167, Gln168, Gln173, Asn176 and Gln177. All these residues are found at the exposed face of helix α 4 (Figure 2(c)). Finally, at the apical end of the connector, surrounding the entrance to the channel, there is an acidic ring formed by the side-chains of Glu189, Asp192 and Asp194.

Figure 5(b) shows the electrostatic potential interior surface of the connector, which appears to be highly electronegative. However, closer inspections reveals two electropositive rings, due to the presence of residues Lys200 and Lys209, the first in the middle of the narrow channel and the second at the beginning of the wide channel. The distance between the two rings (i.e. from Lys200B C $^{\alpha}$ to Lys209A C $^{\alpha}$) is 20 Å, while the distance between adjacent lysine residues of the same ring is 9-10 Å.

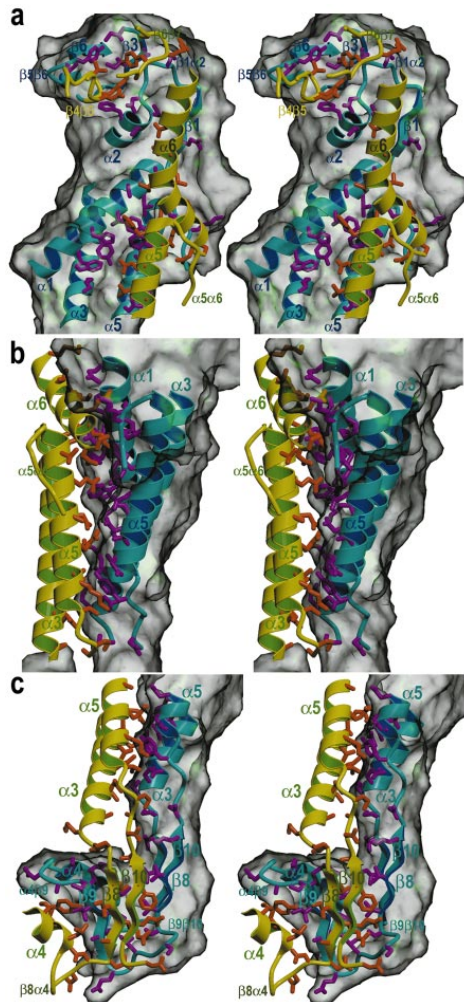


Figure 4. Close up stereo view of a protein monomer in light blue (ribbons) and magenta (selected side-chains), superimposed over its semi-transparent Connolly surface, and the segments of a vicinal monomer (yellow ribbons, orange side-chains) taking part in inter-monomer contacts. The three subfigures display the SH₃-like domain and the upper part of the central domain (a), the central helical domain (b) and the distal domain with the lower part of the central domain (c). Figure prepared with VOLUMES/BOBSRIPT^{59,60} and rendered with Raster-3D.⁶¹

The electron density is well defined for most of these lysine side-chains. The diameters of the rings are 36 and 41 Å, respectively. In addition, the disordered 231-244 area contains two lysine residues, Lys234 and Lys235. Both residues 231 and 244 are at 20 Å from the second lysine ring towards the proximal end. Therefore, there are two (and prob-

ably three) positively charged rings along the channel, at 20 Å intervals. Both Lys200 and Lys209 have their electronegative counterparts in close proximity, Asp202 and Asp208. It is interesting that the position of the aspartate side-chains is quite similar in both rings, with an alternation of Asp and Lys residues around a channel section. In the Lys200 ring, however, we observe a clear salt bridge, at least in two out of the three monomers of the crystal asymmetric unit, between the lysine and one of the flanking aspartate residues, while the other is at a distance of 6.3 Å. This is because the lysine side-chains are not oriented radially towards the centre of the channel. In contrast, in the Lys209-ring they point to the centre of the channel and are equidistant, 5 to 6 Å away, from the two flanking aspartate residues, without establishing any interaction with them.

Discussion

Requirements for DNA packaging

The packaging of DNA into preformed proheads is a mechanism exhibited by complex double-stranded DNA bacteriophages, as well as other animal viruses (adenovirus, herpesvirus).^{25,26} The packaging process²⁷ has two important aspects: First, there is a processive movement of the DNA through the packaging machinery.⁷ The other characteristic feature of the viral DNA packaging process is the condensation of the DNA inside the viral head to crystal-like concentrations. The translocation of a long, rather stiff DNA polymer into a preformed container imposes severe limitations on the function of the packaging machinery and the movement of the DNA, as the acquisition of the solenoid-like ordered parallel arrangement of the DNA inside the viral prohead generates pressure and a high concentration of mass and charges. Thus, the viral DNA packaging machinery has to deal with the general problems of translocation of nucleic acids by protein pumps, and at the same time has to find suitable solutions to release the tensions accumulated by the forced concentration of a highly charged polymer in a limited container.

The connector shows a unique architecture

It was reasonable to believe that the movement of the DNA would share some common mechanistic features with other nucleic acid processive systems,¹ such as the sliding clamps,²⁸ certain helicases^{29,30} or conjugation systems.³¹ However, the $\phi 29$ connector analysed here has a completely novel architecture, which suggests differences in the way the system functions. For example, ring helicases and the conjugation protein TrwB have a basic RecA/F1-ATPase architecture, including a core formed by a parallel doubly wound β -sheet with helices on both sides that encompasses a nucleotide binding domain. These proteins may contain additional domains, but neither these nor

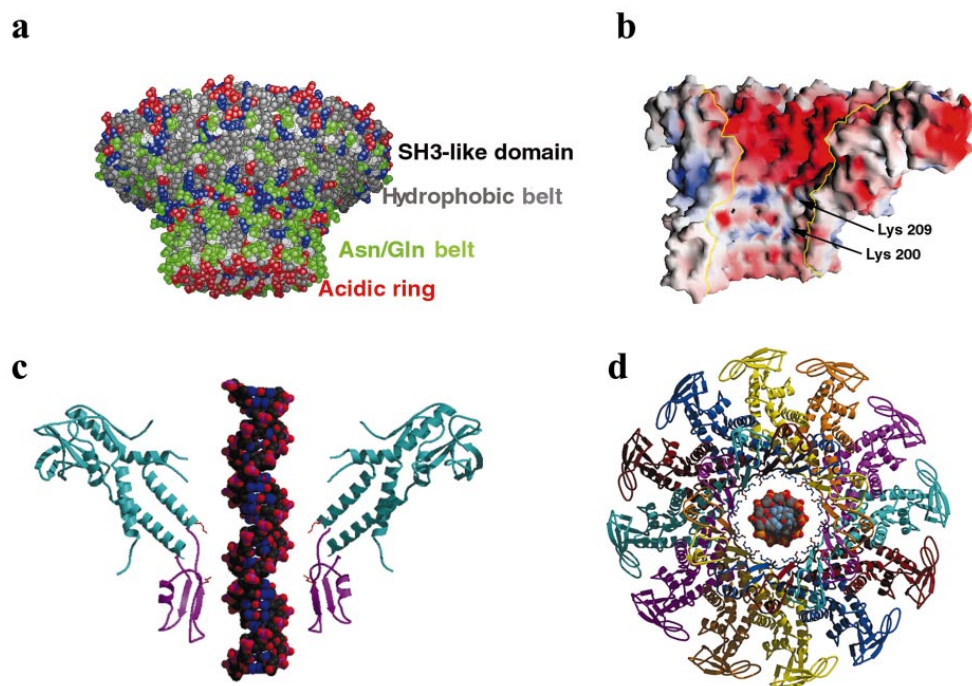


Figure 5. (a) Atom-sphere model of the p10 connector particle. Colour coding of the side-chains: green for Asn/Gln, red for Asp/Glu, grey for hydrophobic residues, and dark blue for Lys/Arg. The model was generated with GRASP.⁶³ (b) Molecular surface coloured according to electrostatic potential (red for negative, blue for positive) belonging to six vicinal protein monomers generated with GRASP.⁶³ The position of the two interior lysine rings (made up by Lys209 and Lys200 of each monomer, respectively) is indicated. (c) Ribbon plot showing a longitudinal cut of the connector particle. Note that the colours of the distal domains are different from the central and SH3-like domains since they belong to different monomers. A dsB-DNA molecule is shown traversing the central channel. The side-chains of Lys200 and Lys209 are indicated. (d) View along the connector axis with the modelled DNA. The side-chains of Lys200 and Lys209 are displayed as a balls and sticks.

the central nucleotide binding domain share any resemblance with the connector p10 protein monomer. Regarding the size of the whole particle, the connector is also unique, larger than all the others. Ring helicases are normally hexamers, while all the functional connectors are formed by 12 subunits.⁷ The shape of the connector is unique too, with its distal narrow cylinder and a wide proximal disc of a diameter twice as large as the first, quite different from the more ellipsoidal or cylindrical shape of helicases and TrwB. A notable functional difference with ring helicases is that the connectors do not contain ATPase activity, and thus need ancillary ATP hydrolases to function, which is similar to what happens in sliding clamps. The protomer of a sliding clamp has a completely different fold from a connector protomer, but their oligomeric structure shows striking coincidences with the ϕ 29 connector particle: the central channel is similarly defined by 12 α -helices that are inclined with respect to the particle axis, giving a propeller-like appearance.²⁸ The inclination is such that the

helices would lie perpendicular to the sugar-phosphate backbone of a putative DNA traversing the channel. Kong *et al.*²⁸ suggest that this feature prevents the entry of the protein into the DNA groove, facilitating its rapid motion. A further coincidence between sliding clamps and the connector is the diameter of the channel, which is obviously determined by the diameter of the DNA double helix.

B-DNA fits smoothly into the channel

The connector defines a central channel of appropriate size to accommodate a double-stranded DNA molecule. The diameter of the channel is 35 Å at its narrowest part, while the DNA B-form has an average diameter of 23 Å. This leaves about 6 Å between the DNA and the channel wall. This wall is mostly electronegative (Figure 5(b)), as is the tip of the connector where the DNA enters the channel. The electronegative nature of the channel surface is not incompatible

with its function since it does not have to bind tightly to DNA, but rather forces it to pass smoothly through at 140 bp s⁻¹.^{32,33} The DNA would traverse the channel floating away from the repulsive walls. There is room for a layer of water molecules, between the DNA and the protein. It is notable that the recently solved TrwB protein, involved in single-stranded DNA translocation, also displays a rather electronegative channel.³¹

The exceptions to the extended electronegative nature of the channel surface are the two lysine rings (200 and 209) (and putatively two more at Lys234 and Lys235). When fitting the DNA into the channel, if a particular backbone phosphate is aligned with one of the Lys200, another lysine of the same ring would be aligned with another phosphate from the other DNA strand (Figure 5(c) and (d)). This second lysine is not opposite the first one, across the channel, but that belonging to the fifth monomer. Keeping these two residues of ring 200 aligned with two phosphates, we again observe that at the second lysine ring (209) two lysine side-chains, five monomers apart, lie in front of two phosphate groups. This is because the 20 Å that separates the two lysine rings is about the same distance that separates two phosphates of the DNA across the major groove, in the helix axis direction. In consequence, at any particular moment during its translocation, the DNA molecule seems to be able to interact electrostatically at four points of the channel (and presumably two more at the next disordered lysine ring).

Sliding groove and SH3-like ridge

The X-ray structure of the ϕ 29 connector fits very well in the three-dimensional reconstructions from cryo-electron microscopy of both proheads³⁴ (Figure 3(c)) and purified connectors⁵ (not shown). Although the docking does not reveal details due to the limited resolution and the averaging imposed in the microscopy reconstruction, it is clear that the wider part of the connector, comprising some elements of the central domain and the whole SH3-like domain, is buried inside the cavity of the head, while the α -helical bundle of the central domain is inserted into the five-folded prohead vertex. In particular, the hydrophobic belt at the external concave area or groove of the connector is facing the corresponding areas of the head protein building the vertex.

As indicated in Results, the SH3-like domain does not have the canonical RT-*Src* loop for polyproline recognition. Its topology, however, is similar to SH3 domains, with two sets of three-stranded β -sheets arranged at 90 degrees to each other, forming a barrel, which encloses a tight hydrophobic core. A comparison with other structures with the same topology, but not strictly SH3, reveals that it closely resembles the DNA-binding domain of the HIV1 integrase.³⁵ This protein, also a viral protein like ϕ 29 p10, forms dimers using one of the β -sheets as a dimerization surface. In the

ϕ 29 connector the SH3-like domains from the various monomers also contact each other but in a different way. They assemble in a circling head-to-tail fashion, using the protruding loops β 4- β 5 and β 6- β 7, which fit into the groove defined by β 5- β 6 and β 1- α 2 of the next monomer (Figures 4(a) and 5(c)). Thus, the distinct arrangement of the loops in the SH3-like domain facilitates protein-protein contacts, which finally hold together the protruding lobules of the wide proximal end. The SH3-like domain is therefore a sort of protein recognition motif, since it serves to assemble the compact protuberances that form the outer crown of the connector.

Spinning mechanism

Hendrix^{36,37} suggested a way to correlate translocation of DNA and its packaging into the viral prohead based on the symmetry mismatch between the 5-fold vertex of the prohead and the 6*n*-folded symmetry of the tail components. The hypothesis that the rotation of the connector is actively involved in the actual DNA translocation was strongly supported by the description of 12-fold symmetry in several head-to-tail connectors either extracted from viral particles or assembled from overexpressed protein.^{14,15,38-40} A longitudinal movement of the DNA, coupled to the connector rotation, has been proposed for both SPP1¹⁶ and ϕ 29²¹ phage connectors.

The surface features as defined now in the high-resolution connector structure are consistent with the spinning hypothesis. As previously discussed,⁸ the rotation of the connector with respect to the non-matching 5-fold symmetry head vertex could take advantage of the hydrophobic groove used as a sliding furrow. The adjacent crown of SH3-like domain would retain the connector inside the capsid while spinning.

Is the rotation of the connector enough to drive the DNA into the head? The finding of the two (or three, if we also assume the disordered loop to interact) lysine rings at regular 20 Å spacings in an otherwise electronegatively coated channel, together with the helical arrangement of the negatively charged DNA phosphate groups, strongly suggests that both movements are coupled. A one base-pair longitudinal displacement of the DNA would result in the weakening of four (or six) lysine-phosphate interactions. The restoring of the four interactions, but this time involving four new lysine residues of adjacent monomers, would occur only if the connector rotated. In other words, the energy barrier for jumping from a particular set of lysine-phosphate interactions to the next set would be overcome by the rotary force of the connector. For this model to work, the rotation of the connector has to be clockwise when looking from the wide domain, i.e. in the opposite direction of the DNA displacement. This is determined by the right-handed nature of the double helix, as the right-handed groove of a bolt determines the clock-

wise rotation of the nut in order to drive the bolt in the direction of the phage head.

A characteristic feature of this model is that the rotation of the connector has to match the symmetry of the DNA rather than the symmetry of the connector itself. In that sense, a rotation of 6° would be appropriate. Such a rotation would misalign the position of the lysine residues with respect to their phosphate counterparts but place the next monomer lysine residues at exactly the right place for interacting with the next set of phosphates, providing a DNA translocation of one base-pair into the prohead (Figure 6).

A spinning model has the beauty of simplicity. It does not need molecular rearrangements within the connector and thus the speed does not depend on time-limited inter-molecular movements. Other more complicated models, like the one put forward by Simpson *et al.*,²¹ assume, besides the rotation, a domain displacement within the p10 protomers and a spring-like elongation and contraction of the connector length, which would require the relative slipping of the central α -helices between monomers. The ϕ 29 connector quaternary structure analysed here appears to be very compact, with a high percentage of protomer surface involved in inter-subunit contacts. This arrangement gives little ground for extensive structural rearrangements, leaving only as a possible hinge candidate the region that connects the central helix bundle domain with the narrow β -structured RNA-binding tip (Figure 3(b)). However, even here, we do not observe a flexible connecting region, i.e. there are no Gly residues. A superposition of the three monomers of the asymmetric unit does not reveal any substantial differences. Although we cannot rule out a movement between domains or between subunits, we do not find any indication for such a movement in the solved high-resolution structure.

Localization and possible role of the pRNA

The packaging reaction at the ϕ 29 system requires a competent prohead, which contains, beside the capsid, the pRNA and the ATPase p16.⁴ These are indispensable components of the packaging machinery since they act, in our model, as a stator and ancillary motor, respectively. The pRNA required for the packaging of ϕ 29 DNA has been located in the narrow tip of the connector.^{21,34} The docking of the connector structure in the portal region reconstructed from proheads containing pRNA clearly shows the pRNA moiety forming a protruding ring that embraces the distal domain of interlaced β -strands at the apical region of the p10 dodecamer (Figure not shown). A peculiar aspect of this region is the presence of a ring built by helices α 4 that presents a high content of exposed glutamine and asparagine residues. This unusual arrangement offers an uncharged polar surface rich in hydrogen bond donors and acceptors that could play a role in the interaction with the pRNA.

The pRNA is a unique component of the ϕ 29 packaging machinery that interacts with the outer end of the connector in a multimeric form.^{21,34} Extensive analysis of this pRNA based on functional assays of directed mutant pRNA molecules indicated the importance of the interaction between complementary loops in adjacent pRNA molecules (inter-subunit complementation).⁴¹ The fact that two or three different serially complementary mutant pRNAs are required for restoring packaging activity in DNA packaging assays *in vitro* led to the conclusion that the pRNA should be arranged in a multimer containing $2n$ and $3n$ subunits. This conclusion is supported by the finding that a set of six serially complementary mutants are fully active in packaging, strongly suggesting the hexameric arrangement of the pRNA as the functional oligomer that is assembled in proheads competent for DNA packaging.^{41,42} A different pRNA arrangement has recently been proposed,²¹ where a pentamer has been found associated with the presence of pRNA in proheads reconstructed from electron microscopy.

Assuming the rotation of the connector, there are two possible scenarios for the pRNA. In one of them, it rotates with the connector. In the other, it is fixed to the head and does not follow the connector movement. The latter is an attractive hypothesis since the pRNA could act as a stator fixed to the head vertex. A 5-fold symmetry would suit this case better. The alternative scenario is also acceptable: the pRNA would be fixed to the connector, constituting a washer to hold it in place, and would be the equivalent piece of the SH3-like crown, but on the outside of the capsid wall. A 6-fold symmetry would suit this role better, to be compatible with the 12-fold symmetry of the connector. In this scenario, the head vertex would be the sole stator.

Localization and role of ATPase p16

The other basic component of the packaging reaction is protein p16.⁴³ In our model, the energy necessary for the connector spinning is provided by this ancillary 39 kDa ATP hydrolase. This protein interacts transiently with the prohead during packaging, but it is not present in the final viral particle.⁴⁴ The pRNA is required to couple the ATPase activity to packaging.

There are no direct data on the location of p16 in the packaging machinery, although Simpson *et al.*²¹ provide evidence for its presence in the pRNA domain of proheads competent for packaging. We have found that p16 cosediments with proheads in gradients (B.I., J.M.V. and J.L.C., unpublished results). Our results also indicate that p16 interacts with the connector. The interaction with the isolated connector is dependent on the presence of the amino-terminal region, as connectors lacking the 13 N-terminal amino acid residues cannot interact with p16. Taken together, these results support the hypothesis that p16 contacts the narrow part of the

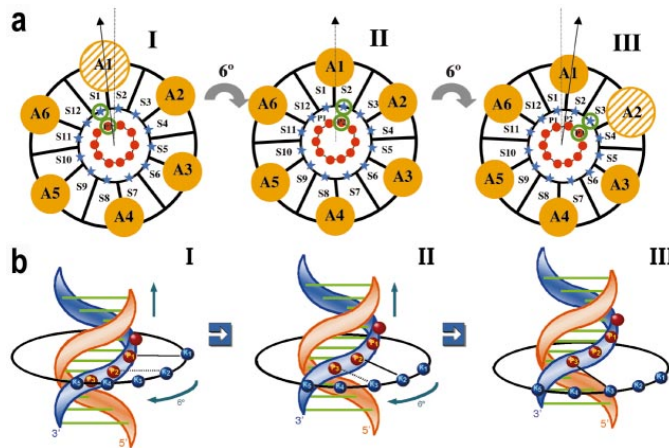


Figure 6. The spinning connector model. (a) Schematic arrangement of a portal vertex in projection view from the viral prohead. A1 to A6 represent the ATPase (p16), S1 to S12 represent the connector subunits, and P1, 2, 3, etc. represent the DNA phosphates of a DNA chain. For clarity, the model is restricted to the description of one single DNA chain and one lysine ring. The starting position (I) aligns a Lys (blue star) of subunit S1 with the phosphate P1. Firing of the p16 ATPase (A1) rotates the connector 12° clockwise while the DNA moves linearly along the connector axis by 2 bp. A transient interaction among the corresponding Lys from S2 and the phosphates P2 (II) occurs at 6° rotation. The final

stage (III) is characterized by the interaction of Lys S3 and the phosphate P3, which is accomplished after a further clockwise rotation of the connector by 6° . This movement positions the next p16 ATPase (A2) in an equivalent position with respect to the DNA, to that of A1 in step I. (b) Schematic representation of a side view of the model, corresponding to the same steps as in (a). The lysine residues are depicted as blue spheres labelled K1, K2, ... In step I there is a strong interaction between lysine K1 and phosphate P1 (continuous line), while the interaction of lysine K2 with phosphate P2 is still weak because P2 is on a lower plane than the ring of lysine residues (broken line). In step II the 6° rotation of the connector places lysine K2 in an optimal position to interact with phosphate P2, provided there is a longitudinal displacement of the DNA by 1 bp. In step III there has been a further 6° rotation of the connector and another 1-bp displacement of the DNA, implying that lysine K3 is now aligned with phosphate P3. The movement of two base-pairs of DNA from I to III is, thus, correlated with a rotation of the connector by 12° .

connector, which remains accessible from outside the prohead. As the pRNA is arranged in a ring around the outer end of the dodecameric connector, this region comprises a complex set of concentric structures with different symmetries: the DNA inside the channel (10_1 -fold), the connector walls (12-fold), and the outer ring of pRNA coupled to p16 (most probably, 6-fold). At the same time, the other side of the connector (12-fold) is embedded in the 5-fold vertex of the prohead.

At present there are insufficient data to confirm a direct interaction of p16 with the DNA. Therefore, we cannot discern whether the ATPase stroke pushes the DNA directly while the connector rotates passively, or whether the stroke only drives the connector rotation. In any case, as explained above, the connector would have to rotate at least 6° until the next base-pair establishes an electrostatic interaction with the following connector subunit. If the pRNA-p16 complex were a hexamer, the connector would have to rotate twice 6° (i.e. 12°) to place the p16/pRNA into an equivalent position with respect to the DNA. This would correlate the three main factors in the model: the linear translocation of two DNA base-pairs, a clockwise rotation of 12° of the connector, and the activation of one p16 ATPase. Each packaging step would then involve two base-pairs of DNA, two p10 subunits and one ATPase/pRNA subunit. Every ATPase would then be sequentially fired, followed by the clockwise rotation of the connec-

tor. According to our model, the activation of one ATPase would induce the translocation of two base-pairs into the prohead, the rotation of the connector by 12° , and the placing of the next ATPase in an equivalent position with respect to the DNA. Even if the connector rotates passively to the DNA translocation, its rotation would control the sequential firing of the ATPase.

The hydrolysis of one ATP molecule to package two base-pairs fits with the experimental data obtained for packaging $\phi 29$ DNA and other viruses *in vitro*.^{41,45,46} In fact, this ATP consumption is slightly lower than that associated with the walking of helicases along the DNA (one to two ATP molecules per base-pair); but in the latter case these proteins separate DNA strands besides moving it. However, despite the structural and mechanical differences mentioned above, the translocation of DNA by connectors and helicases or sliding clamps could have in common the hydrolysis of ATP as a power stroke to processively move the DNA in a confined channel created by the toroid topology of these protein assemblies.

Other alternatives for the packaging of DNA into the viral heads could involve the rotation of the DNA and the connector (reviewed in reference³⁶). Yet, any model involving rotation of the DNA would be penalized by the additional tension produced during its packaging into the prohead container. The sum of the strain imposed

by the coiling of the DNA within the head, following the inner wall of the prohead, together with the twisting forces induced by the rotation of DNA during packaging would lead to intolerable strain in the DNA inserted into the viral head (supercoiling and partial strand separation) and, eventually, to the arrest of the packaging process.

Further studies on the structure of the complexes of the various components of the packaging machinery and the connector will be required to provide deeper insight into the molecular events involved in the processive translocation of DNA, but the present high-resolution structure already suggests a simple and elegant mechanism that can now be validated with other complementary techniques.

Materials and Methods

A detailed description of the protein expression, proteolyzed connector particle preparation, crystallization and preliminary crystallographic analysis has been published elsewhere.^{17,20} A new crystal form was obtained using the vapour diffusion method, at 20°C, from sitting drops containing 70% MPD (2-methyl-2,4-pentanediol), 0.1 M Tris-HCl buffer (pH 8.0). Typically, crystals appear after two days and reach a maximum size of approx. 0.5 mm × 0.1 mm × 0.1 mm. These crystals were picked up with a cryo-loop and directly flash-cryocooled by immersion in liquid propane. X-ray diffraction data were collected using synchrotron radiation at the ESRF ID14-EH2 beam-line in Grenoble, France. The crystals diffract to 1.9 Å resolution, although the very high-resolution shell decayed rapidly after collecting one frame. Complete data were collected up to 2.1 Å resolution (Table 1). Crystals are tetragonal, with space group *I*422, and have unit cell dimensions of $a = b = 155.3$ Å and $c = 160.4$ Å. The calculated V_m value⁴⁷ is 2.8 Å³/Da, assuming one quarter of a connector particle per asymmetric unit. Data were evaluated with MOSFLM⁴⁸ and SCALA, of the CCP4 suite.⁴⁹ All attempts to obtain suitable heavy-atom derivatives were unsuccessful. Crystals obtained from Se-Met preparations did not diffract appropriately either.

A self-rotation function, calculated with GRLF⁵⁰ showed a 12-fold axis parallel to the crystal *c*-axis. This and packing considerations suggested that the particle was located so that its centre of mass was approximately at $x = 0$, $y = 0$, $z = 0.25$. A starting model for phasing was obtained from the electron microscopy reconstructions.³ The model was positioned as mentioned above and adjusted by two-dimensional correlation fitting, using the *hk0* reflections between 40 and 12 Å (our programs). A phase extension protocol was carried out using DM,⁴⁹ as follows. The phase extension started from a map calculated at 8 Å and the increment steps were of 0.1 Å. The DM-generated values for the figure-of-merit (FOM) were not used. Instead, FOM-values were set, at each extension cycle, to a lower value in the high-resolution range, to avoid an overestimation of the likelihood. The typical values for the manually set FOM ranged between 0.6 and 0.9 for stabilized resolution shells and 0.1 to 0.3 for the new shell increments. The mask used for the phase extension was made initially from a 30° cylindrical sector of the EM model. This mask was manually edited (our programs) after three or four cycles of phase extension according to the maps. The extension was continued in this way until

Table 1. Data collection and refinement statistics

Data collection	
Wavelength (Å)	0.932
Temperature (K)	100
Resolution range (Å)	42-2.1
No. of observed reflections	327,940
No. of unique reflections	56,472
Completeness (%)	98.9 (94.1)
R_{merge}^a	0.10 (0.40)
$\langle I \rangle / \langle \sigma I \rangle$	5.4 (2.1)
Multiplicity	5.8 (2.4)
Refinement	
Resolution range used (Å)	42.0-2.1
No. of reflections in working set	53,611
No. of reflections in test set	2861
No. of protein atoms in the asymmetric unit (a.u.)	6517
No. of solvent molecules in the a.u.	523
No. of MPD molecules in the a.u.	4
R^b , R_{free}^c	0.205, 0.247
R.m.s deviations from target values	
Bond lengths	0.006 Å
Bond angles	1.13 degrees
Average B factors (Å ²)	
Protein atoms	35.4
Water molecules	51.9
All atoms	37.0
Wilson plot	29.3

The outermost resolution shell (2.21-2.10 Å) values are shown in parentheses.

^a $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where I_i is the *i*th measurement of reflection *hkl* and their symmetry related reflections.

^b $R = \sum_{hkl} |F_o(hkl) - |F_c(hkl)|| / \sum_{hkl} F_o(hkl)$, where $F_o(hkl)$ and $F_c(hkl)$ are the observed and calculated structure factors, respectively.

^c $R_{\text{free}} = R$ -value calculated for 5% of reflections not used for refinement

2.3 Å. Although the global shape of the molecule connector was identifiable in the resulting maps, no clear secondary structure elements could be located. This map, however, was used to build a pseudo-atom model with MAPMAN,⁵¹ filling the asymmetric unit. This model was then divided in three 30° cylindrical sectors, which were then superimposed, resulting in an averaged pseudo-atom model (our programs) that was used both to generate a new mask and to further improve the phases with ARP/WARP,⁵² using the whole resolution of the data to 2.1 Å. Each cycle consisted of an unrestrained water search and refinement using REFMAC. After each ARP/WARP cycle the model was reconstructed from an averaged pseudo-atom model as before.

As in this protocol the FOM-values and the crystallographic *R*-factors are unreliable, checks were made between cycles by trying different combinations of FOM-values, looking for features in maps, and mainly checking the correlation between the cylindrical sectors of the map. This protocol was successful, and after a few runs zones of well connected electron density appeared, with several alpha helices identified.

At this stage, the coordinates of the Simpson *et al.* model (PDB code 1fou) became available in the PDB. They were used as an initial model for molecular replacement with AMoRe.⁵³ No solution, however, could be obtained using a monomer as a starting model. A

different strategy was then adopted: a whole connector model, with 12 monomers, was fitted into the density of the phase extension/WARP map. A one-dimensional translation search was then carried out along the *c*-axis, followed by rigid-body refinement. This gave a solution with an *R*-factor of 44.6 and a correlation of 54.6 (next peak 37.0 and 50.6, respectively). SigmaA-weighted $2F_o - F_c$ and $F_o - F_c$ electron density maps calculated with this solution showed that it was correct, although in many areas, and in particular in the proximal disc-like part, the density did not coincide with the model. These areas were manually rebuilt with TURBO,⁵⁴ including the tracing of five new β -strands and the correct placing of the amino acid sequence, which in the search model was out of register from residue 271 onwards. The structure was refined with CNS⁵⁵ and further manual interventions. Initially, 3-fold NCS restraints were applied. These were relaxed and finally suppressed, once the *R*-free fell below 26% at full resolution. At this stage, water molecules were added to the model. Atoms not seen in the $2F_o - F_c$ and $F_o - F_c$ sigmaA-weighted maps were set to zero occupancy. This mainly affected the N and C-terminal residues, and some residues before $\beta 10$. The sequence of 15 amino acid residues between residues 230 and 245 of each monomer was not seen on the maps and was not included, as explained in Results. The final model is of excellent geometry as confirmed by PROCHECK.⁵⁶ The final refinement statistics are summarized in Table 1.

Atomic coordinates

The atomic coordinates have been deposited with the PDB under accession number 1h5w.

Acknowledgements

This study was supported by grants from the Ministerio de Educación y Cultura of Spain (PB98-1631 and 2FD97-0518 to M.C., PB97-1225-C02 to J.M.V., and PB96-0818 to J.L.C.) and the Generalitat de Catalunya (1999SGR-188 and Centre de Referència en Biotecnologia) to M.C. Synchrotron data collection was supported by the EU grants ERBFMGECT980134 and HPRI-CT-1999-00017 to the EMBL-DESY and by the ESRF. J.P. acknowledges a fellowship from the Ministerio de Educación y Cultura. We are indebted to Rosa Pérez-Luque for assistance in crystallization experiments and Ignacio Fita and Jaime Martín-Benito for useful suggestions.

References

- Hingorani, M. M. & O'Donnell, M. (1998). Toroidal proteins: running rings around DNA. *Curr. Biol.* **8**, R83-R86.
- Casjens, S. & Hendrix, R. W. (1988). Control mechanisms in dsDNA bacteriophage assembly. In *The Bacteriophages* (Calendar, R., ed.), vol. 1, pp. 15-90, Plenum Press, New York.
- Valpuesta, J. M. & Carrascosa, J. L. (1994). Structure of viral connectors and their function in bacteriophage assembly and DNA packaging. *Quart. Rev. Biophys.* **27**, 107-155.
- Guo, P. X., Erickson, S. & Anderson, D. (1987). A small viral RNA is required for *in vitro* packaging of bacteriophage phi 29 DNA. *Science*, **236**, 609-614.
- Valpuesta, J. M., Fernandez, J. J., Carazo, J. M. & Carrascosa, J. L. (1999). The three-dimensional structure of a DNA translocating machine at 10 Å resolution. *Struct. Fold. Des.* **7**, 289-296.
- Tao, Y., Olson, N. H., Xu, W., Anderson, D. L., Rossmann, M. G. & Baker, T. S. (1998). Assembly of a tailed bacterial virus and its genome release studies in three dimensions. *Cell*, **95**, 431-437.
- Carrascosa, J. L. & Valpuesta, J. M. (1999). Bacteriophage connectors: structural features of a DNA translocating motor. *Recent Res. Dev. Virol.* **1**, 449-465.
- Valpuesta, J. M., Sousa, N., Barthelemy, I., Fernandez, J., Fujisawa, H., Ibarra, B. & Carrascosa, J. L. (2000). Structural analysis of the bacteriophage T3 head-to-tail connector. *J. Struct. Biol.* **131**, 146-155.
- Orlova, E. V., Duve, P., Beckmann, E., Zemlin, F., Lurz, R., Trautner, T. A. *et al.* (1999). Structure of the 13-fold symmetric portal protein of bacteriophage SPP1. *Nature Struct. Biol.* **6**, 842-846.
- Kochan, J., Carrascosa, J. L. & Murialdo, H. (1984). Bacteriophage lambda preconnectors. Purification and structure. *J. Mol. Biol.* **174**, 433-447.
- Bazinet, C., Benbasat, J., King, J., Carazo, J. M. & Carrascosa, J. L. (1988). Purification and organization of the gene 1 portal protein required for phage P22 DNA. *Biochemistry*, **27**, 1849-1856.
- Driedonks, R. A., Engel, A., ten Heggeler, B. & van Driel, R. (1981). Gene 20 product of bacteriophage T4. Its purification and structure. *J. Mol. Biol.* **152**, 641-662.
- Carrascosa, J. L., Carazo, J. M., Ibañez, C. & Santisteban, A. (1985). Structure of phage phi 29 connector protein assembled *in vivo*. *Virology*, **141**, 190-200.
- Carrascosa, J. L., Viñuela, E., Garcia, N. & Santisteban, A. (1982). Structure of the head-tail connector of bacteriophage phi 29. *J. Mol. Biol.* **154**, 311-324.
- Carazo, J. M., Donate, L. E., Herranz, L., Secilla, J. P. & Carrascosa, J. L. (1986). Three-dimensional reconstruction of the connector of bacteriophage phi 29 at 1.8 nm resolution. *J. Mol. Biol.* **192**, 853-867.
- Dube, P., Tavares, P., Lurz, R. & van Heel, M. (1993). The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry. *EMBO J.* **12**, 1303-1309.
- Guasch, A., Parraga, A., Pous, J., Valpuesta, J. M., Carrascosa, J. L. & Coll, M. (1998). Purification, crystallization and preliminary X-ray diffraction studies of the bacteriophage phi29 connector particle. *FEBS Letters*, **430**, 283-287.
- Valpuesta, J. M., Carrascosa, J. L. & Henderson, R. (1994). Analysis of electron microscope images and electron diffraction patterns of thin crystals of $\phi 29$ connectors in ice. *J. Mol. Biol.* **240**, 281-287.
- Muller, D. J., Engel, A., Carrascosa, J. L. & Velez, M. (1997). The bacteriophage phi29 head-tail connector imaged at high resolution with the atomic force microscope in buffer solution. *EMBO J.* **16**, 2547-2553.
- Guasch, A., Pous, J., Parraga, A., Valpuesta, J. M., Carrascosa, J. L. & Coll, M. (1998). Crystallographic analysis reveals the 12-fold symmetry of the bacteriophage phi29 connector particle. *J. Mol. Biol.* **281**, 219-225.
- Simpson, A. A., Tao, Y., Leiman, P. G., Badasso, M. O., He, Y., Jardine, P. J. *et al.* (2000). Structure of

- the bacteriophage ϕ 29 DNA packaging motor. *Nature*, **408**, 745-750.
22. Donate, L. E., Valpuesta, J. M., Rocher, A., Méndez, E., Rojo, F., Salas, M. & Carrascosa, J. L. (1992). Role of the amino-terminal domain of bacteriophage ϕ -29 connector in DNA binding and packaging. *J. Biol. Chem.* **267**, 10919-10924.
 23. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138.
 24. Urbaneja, M. A., Rivas, S., Carrascosa, J. L. & Valpuesta, J. M. (1994). An intrinsic-tryptophan-fluorescence study of phage ϕ 29 connector/nucleic acid interactions. *Eur. J. Biochem.* **225**, 747-753.
 25. Casjens, S., van Vugt, R., Tilly, K., Rosa, P. A. & Stevenson, B. (1997). Homology throughout the multiple 32-kilobase circular plasmids present in Lyme disease spirochetes. *J. Bacteriol.* **179**, 217-227.
 26. Wikoff, W. R. & Johnson, J. E. (1999). Virus assembly: imaging a molecular machine. *Curr. Biol.* **9**, 296-300.
 27. Black, L. W. (1989). DNA packaging in dsDNA bacteriophages. *Annu. Rev. Microbiol.* **43**, 267-292.
 28. Kong, X. P., Onrust, R., O'Donnell, M. & Kuriyan, J. (1992). Three-dimensional structure of the beta subunit of *E. coli* DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell*, **69**, 425-437.
 29. Sawaya, M. R., Guo, S., Tabor, S., Richardson, C. C. & Ellenberger, T. (1999). Crystal structure of the helicase domain from the replicative helicase-primase of bacteriophage T7. *Cell*, **99**, 167-177.
 30. Fass, D., Bogden, C. E. & Berger, J. M. (1999). Crystal structure of the N-terminal domain of the DnaB hexameric helicase. *Struct. Fold. Des.* **7**, 691-698.
 31. Gomis-Ruth, F. X., Moncalian, G., Perez-Luque, R., Gonzalez, A., Cabezon, E., de la Cruz, F. & Coll, M. (2001). The bacterial conjugation protein TrwB resembles ring helicases and F1-ATPase. *Nature*, **409**, 637-641.
 32. Shibata, H., Fujisawa, H. & Minagawa, T. (1987). Characterization of the bacteriophage T3 DNA packaging reaction *in vitro* in a defined system. *J. Mol. Biol.* **196**, 845-851.
 33. Catalano, C. E., Cue, D. & Feiss, M. (1995). Virus DNA packaging: the strategy used by phage lambda. *Mol. Microbiol.* **16**, 1075-1086.
 34. Ibarra, B., Caston, J. R., Llorca, O., Valle, M., Valpuesta, J. & Carrascosa, J. L. (2000). Topology of the components of the DNA packaging machinery in the phage ϕ 29 prohead. *J. Mol. Biol.* **298**, 807-815.
 35. Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R. & Davies, D. R. (1994). Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science*, **266**, 1981-1986.
 36. Hendrix, R. W. (1998). Bacteriophage DNA packaging: RNA gears in a DNA transport machine. *Cell*, **94**, 147-150.
 37. Hendrix, R. W. (1978). Symmetry mismatch and DNA packaging in large DNA bacteriophages. *Proc. Natl Acad. Sci. USA*, **75**, 4779-4783.
 38. Carazo, J. M., Santisteban, A. & Carrascosa, J. L. (1985). Three-dimensional reconstruction of bacteriophage ϕ 29 neck particles at 2×2 nm resolution. *J. Mol. Biol.* **183**, 79-88.
 39. Donate, L. E., Herranz, L., Secilla, J. P., Carazo, J. M., Fujisawa, H. & Carrascosa, J. L. (1988). Bacteriophage T3 connector: three-dimensional structure and comparison with other viral head-tail connecting regions. *J. Mol. Biol.* **201**, 91-100.
 40. Valpuesta, J. M., Serrano, M., Donate, L. E., Herranz, J. & Carrascosa, J. L. (1992). DNA conformational change induced by the bacteriophage ϕ 29 connector. *Nucl. Acids Res.* **20**, 5549-5554.
 41. Guo, P., Zhang, C., Chen, C., Garver, K. & Trottier, C. (1998). Inter-RNA interaction of phage ϕ 29 pRNA to form a hexameric complex for viral DNA transportation. *Mol. Cell*, **2**, 149-155.
 42. Zhang, F., Lemieux, S., Wu, X., St-Arnaud, D., McMurray, C. T. & Anderson, D. (1998). Function of hexameric RNA in packaging of bacteriophage ϕ -29 DNA *in vitro*. *Mol. Cell*, **2**, 141-147.
 43. Bjornsti, M. A., Reilly, B. E. & Anderson, D. (1982). Morphogenesis of bacteriophage ϕ 29 of *Bacillus subtilis*: oriented and quantized *in vitro* packaging of DNA protein gp3. *J. Virol.* **45**, 383-396.
 44. Viñuela, E., Camacho, A., Jimenez, F., Carrascosa, J. L., Ramirez, G. & Salas, M. (1976). Structure and assembly of phage ϕ 29. *Phil. Trans. Roy. Soc. Lond. B, Biol. Sci.* **276**, 29-35.
 45. Guo, P., Peterson, C. & Anderson, D. (1987). Prohead and DNA-gp3-dependent ATPase activity of the DNA packaging protein gp16 of bacteriophage ϕ 29. *J. Mol. Biol.* **197**, 229-236.
 46. Morita, M., Tasaka, M. & Fujisawa, H. (1993). DNA packaging ATPase of bacteriophage T3. *Virology*, **193**, 748-752.
 47. Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.
 48. Leslie, A. G. W. (1991). Macromolecular data processing. In *Crystallographic Computing* (Moras, D., Podjarny, A. D. & Thierry, J. C., eds), vol. 5, pp. 27-38, Oxford University Press, Oxford.
 49. CCP4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760-763.
 50. Tong, L. (1993). REPLACE, a suite of computer programs for molecular-replacement calculations. *J. Appl. Crystallog.* **26**, 748-751.
 51. Kleywegt, G. J. & Jones, T. A. (1999). Software for handling macromolecular envelopes. *Acta Crystallog. sect. D*, **55**, 941-944.
 52. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458-463.
 53. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A*, **50**, 157-163.
 54. Roussel, A. & Cambilleau, C. (1989). Turbo-Frodo. In *Silicon Graphics Geometry Partners Directory*, pp. 77-79, Silicon Graphics, Mountain View, CA.
 55. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905-921.
 56. Laskowsky, R. A., McArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of a protein structure. *J. Appl. Crystallog.* **24**, 946-950.
 57. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577-2637.

58. Evans, S. V. (1993). SETOR: hardware lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graph.* **11**, 134-138.
59. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.
60. Esnouf, R. M. (1997). An extensively modified version of Molscript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 133-138.
61. Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D Version 2.0. A program for photorealistic molecular graphics. *Acta Crystallog. sect. D*, **50**, 869-873.
62. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD-visual molecular dynamics. *J. Mol. Graph.* **14.1**, 33-38.
63. Nicholls, A., Bharadwaj, R. & Honig, B. (1993). GRASP: graphical representation and analysis of surface properties. *Biophys. J.* **64**, A166-.

Edited by R. Huber

(Received 17 September 2001; received in revised form 13 November 2001; accepted 16 November 2001)

Part IV

Discussió i Conclusions

Capítol 5

Discussió

5.1 Estructura del connector

El connector és una estructura troncocònica de 75 Å d'alçada rodejada de dotze lòbuls en la part més ampla amb un diàmetre total de 146 Å i de 77 Å en la part més estreta. Està format per 12 còpies de la proteïna p10 posades en forma de vòrtex al voltant d'un canal central de manera que l'extrem contrari al lòbul de cada proteïna està situat radialment sobre el lòbul de tres proteïnes p10 més enllà (fig. 3 de l'article 3, plana 95).

El domini lobular té un plegament tipus SH3, que consta de 6 làmines β en forma gairebé esfèrica i està emmarcat per dues hèlixs. La zona central del connector està formada per 3 hèlixs α per a cada proteïna i la zona distal (respecte els lòbuls) per unes nanses, unes fulles β curtes i una petita hèlix. Les hèlixs del domini central estan inclinades uns 45° respecte del radi del connector i aquesta inclinació augmenta fins a ser perpendicular al radi en el domini distal, donant l'aparença de turbina a l'estructura (fig 5.1).

El canal central té un diàmetre variable des de 35 Å en la part estreta del connector i s'eixampla en la part ampla, tot i que no podem assegurar quin és el seu diàmetre ja que la densitat no és clara en aquesta zona. Les reconstruccions de microscòpia electrònica veuen aquesta zona com uns lòbuls interns que donen al canal una forma cilíndrica de diàmetre aproximadament constant. L'interior del canal és majoritàriament negatiu excepte per uns anells paral·lels formats per lisines separats aproximadament 20 Å. Hi ha 3 d'aquests anells tot i que el tercer estaria en la zona de baixa densitat esmentada abans.

La unitat asimètrica del cristall de grup espacial I422 conté tres proteïnes p10 (cadena A, B i C). A la cadena A s'observen 253 residus, des del 17 al 284 (de 291, fig 1 de l'article 3, plana 93), exceptuant la seqüència que va

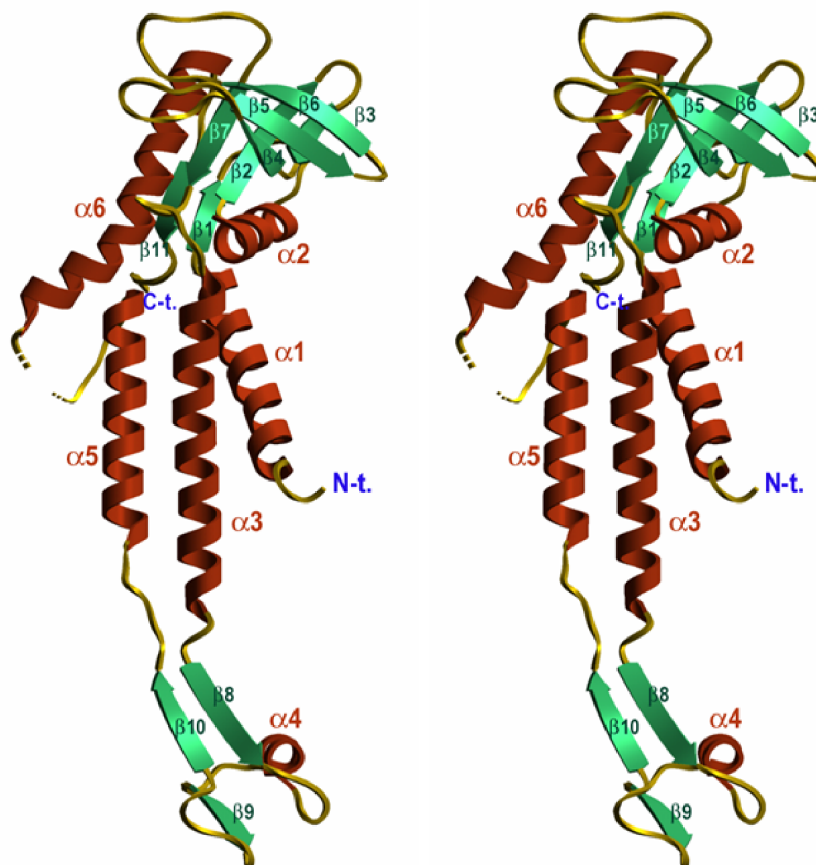


Figura 5.1: Monòmer. Representació estéreo en cintes d'un monòmer, amb la classificació dels elements d'estructura secundària.

des del 230 al 244, que correspon als lòbuls interns abans mencionats. A la cadena B se n'observen 255, des del 16 al 284, exceptuant del 231 al 244 i a la cadena C, se n'observen 251, des del 18 al 283 i exceptuant del 231 al 245 (fig. 5.2).

La unitat funcional (el connector sencer) es construeix a partir de la simetria de l'eix d'ordre 4 que el travessa longitudinalment pel centre del canal. El nombre de residus observats a la unitat funcional és de 3036, de 3492. La majoria dels residus que falten corresponen als extrems N i C-terminal, per on es va tallar la seqüència per propiciar la cristallització. El nombre total d'àtoms observats en la unitat funcional és de 24696, a més de 1976 molècules d'aigua directament lligades a l'estructura per ponts

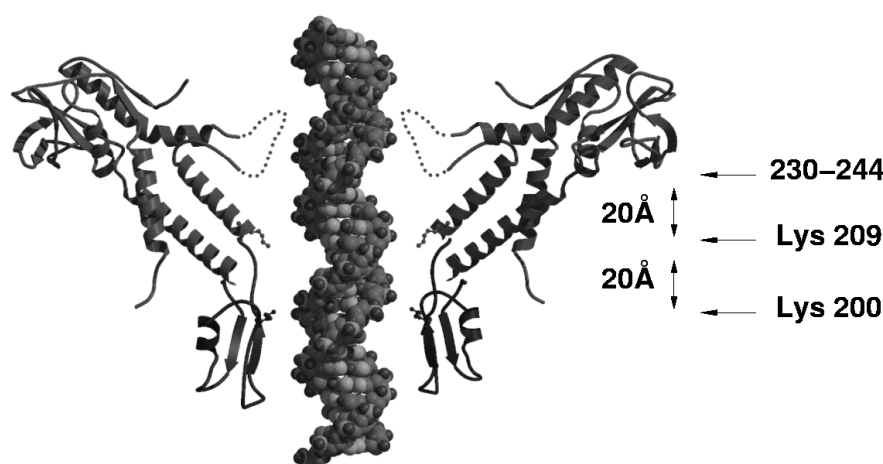


Figura 5.2: Secció lateral. Representació en cintes de dos monòmers oposats amb un DNA en esferes. Les dues lisines de cada monòmer estan representades independentment i també, amb línia discontinua l'espai que ocuparien els residus mòbils que no s'han pogut resoldre. La forma d'aquesta línia discontinua està basada en el model de microscòpia electrònica, on es veu que ha de formar un lòbul interior. La distància entre cada grup és d'uns 20 Å, que és el pas de rosca del DNA.

d'hidrogen i 16 molècules d'MPD situades en la part central de l'estructura (fig. 3.10, plana 55).

Comparant amb l'altra estructura del connector ja publicada [48, 47], no s'observen grans diferències en la forma global del connector. També coincideixen les zones desordenades dels lòbuls interns. De totes maneres, segurament a causa de la menor resolució d'aquesta estructura (3.2 Å), no han observat dues característiques molt importants pel model del mecanisme del connector.

La primera característica correspon a la forma de domini SH3 dels lòbuls externs. Tal com manifesten en la seva discussió, creuen que aquesta zona és flexible, ja que observen desviacions a la simetria 12 en aquesta zona. En el nostre cas la simetria és perfecta, i la zona està a més molt ben definida. Per això vàrem poder-la traçar correctament i determinar que tenia estructura tipus SH3, altament hidrofòbica, que és la propietat ideal per a que el connector roti lliurement dins de la càpsida (fig. 5.2).

L'altra característica es troba en el canal central. Probablement pels mateixos motius que abans, no han observat els anells de càrrega positiva

situats cada 20 Å. Si bé la situació de les lisines que formen els anells és la mateixa en la seva estructura, no tenien bona densitat en les cadenes laterals, per la qual cosa no es devien tenir en compte en calcular la càrrega de la molècula. Aquesta característica provoca una diferència radical en el seu model del mecanisme del connector.

5.2 Mecanisme del connector

Hi ha dos aspectes principals en el procés de l'empaquetament del DNA dins de la càpsida dels bacteriòfags. Un és el pas del DNA a través de la maquinària empaquetadora i l'altre l'alta concentració que ha d'assolir el DNA dins d'aquesta càpsida. Aquesta alta concentració genera una pressió molt alta i també una gran densitat de càrrega. Per tant, la maquinària empaquetadora ha d'ajuntar característiques d'altres proteïnes processives del DNA [29] amb un sistema capaç alhora d'alliberar les tensions que es poguessin acumular en un DNA que s'ha d'empaquetar en un espai molt limitat.

El diàmetre del canal del connector té la mida justa per tal que passi una cadena doble de DNA. En algunes zones una mica més amples del canal hi ha l'espai just per què hi càpiga també una capa de molècules d'aigua.

La paret interior és majoritàriament negativa, a l'igual que ho és la superfície exterior d'una cadena de DNA. La repulsió de càrregues evita que el DNA quedi lligat a l'interior del connector, ajudant-lo també que travessi suaument el connector i el manté centrat al llarg de l'eix de simetria.

L'excepció a aquesta paret negativa són unes lisines que formen dos anells de densitat positiva (i potser un tercer) separats 20Å, que és el pas de rosca del DNA (figura 5b del tercer article, plana 97). Si orientem un dels fosfats del DNA davant d'una de les lisines, es pot formar un pont d'hidrogen entre aquest parell. A més es pot formar, en el mateix anell, un altre pont entre àtoms equivalents de la segona cadena del DNA. No diametralment, sinó a 5 monòmers de distància. A la vegada s'observa una situació idèntica en el segon anell.

Si el connector està girant respecte al DNA, en qualsevol moment hi haurà aquestes quatre interaccions entre connector i DNA. És possible que també n'hi hagi dues més al tercer anell de lisines no ordenades que també es troba a uns 20 Å del segon anell.

L'interior del canal conté dotze hèlixs inclinades de tal manera que quedarien perpendiculars a l'esquelet de sucres i fosfats del DNA que travessaria el canal. Es creu que aquesta qualitat evita que la proteïna penetri dins del solc major del DNA i facilita així que pugui lliscar al llarg de la proteïna. A

part d'aquesta coincidència, el connector té una estructura molt diferent que altres proteïnes processives, per la qual cosa el seu mecanisme ha de ser d'un tipus nou.

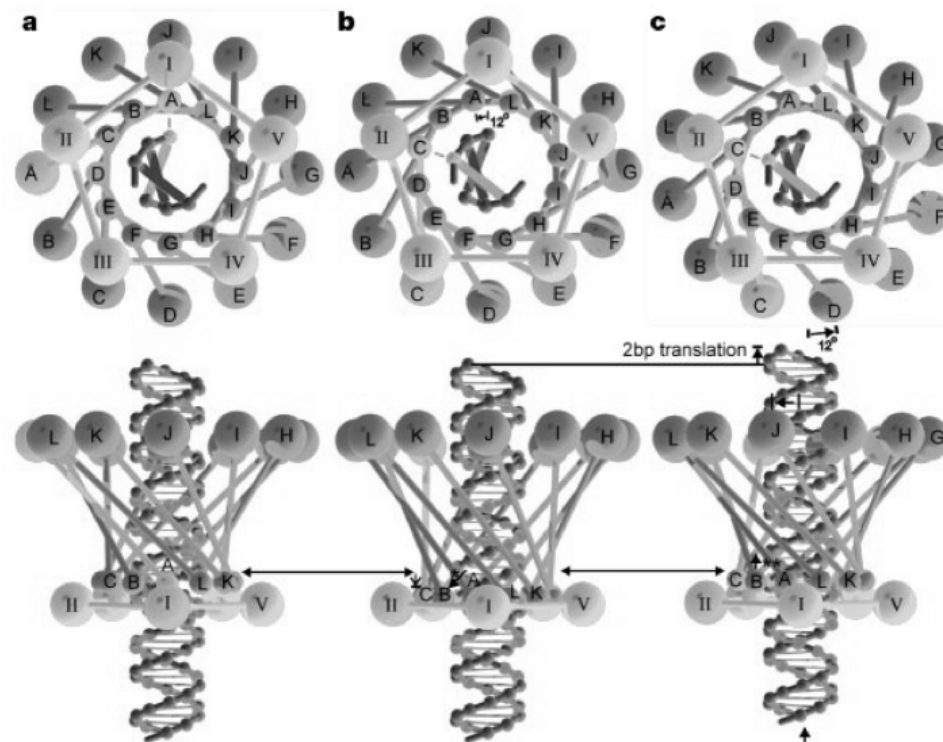


Figura 5.3: Un model per a l'empaquetament. Model proposat a [48] per l'acció del connector. El pRNA i l'ATPasa estan formats per 5 subunitats (pentàgon). Les 12 esferes grans exteriors i les interiors petites representen, respectivament, els dominis amples i estret del connector. (Veure text, plana 111)

A partir de l'estructura cristal·logràfica ja s'ha proposat un model de l'empaquetament del DNA [48]. La falta de simetria 12 perfecta de l'estructura del connector que presenten fa creure que el processament del DNA es basa en la flexibilitat del connector.

En contrast al que han proposat altres autors (plana 14), han observat per microscòpia electrònica que el pRNA té simetria 5, enlloc de 6. L'ATPasa p16 que estaria unida al pRNA també estaria composta de 5 unitats. Aquest conjunt podria estar unit al vèrtex de la càpsida del $\Phi 29$, formant un tap

inferior que evitaria que el connector s'escapés de la càpsida. L'element giratori seria únicament el connector. L'asimetria entre el conjunt pRNA-p16 i el connector permetria que aquest girés amb poca energia aportada per la hidròlisi de l'ATP. Aquesta hidròlisi actuaria seqüencialment en cada subunitat de la p16, cada cop que un dels 12 monòmers del connector estigués enfront. La part interactuant entre el connector i la p16 està en el domini estret del connector. L'energia empenyeria el connector fent-lo rodar en dues etapes. En la primera només giraria 12° el domini estret del connector. En la segona, el seguiria el domini ample i acompanyaria alhora en DNA, que avançaria dos parells de bases (fig. 5.3).

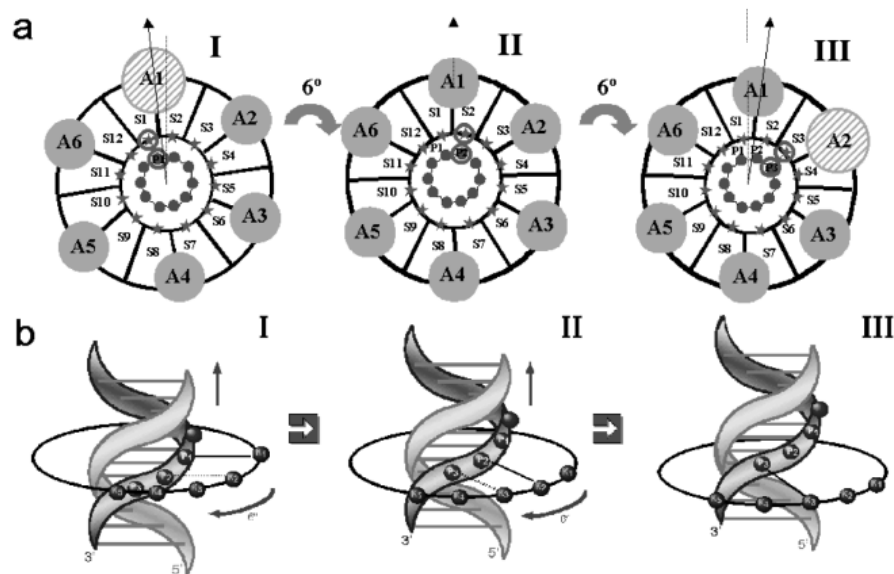


Figura 5.4: El nostre model per a l'empaquetament. Model proposat en aquest treball per a l'empaquetament del DNA. El pRNA i l'ATPasa estan formats per 6 subunitats (cercles grans). El connector està representat pels 12 sectors circulars i les lisines són les 12 estrelles. Els 10 cercles interns representen els fosfats del DNA que entra pel canal del connector. Per a una explicació, veure el text (plana 112) i la figura 6 de l'article 3 (plana 100)

En el nostre model, considerem que el pRNA i la p16 tenen ambdós

simetria 6, tal com prediuen estudis de microscòpia electrònica, bioquímics, funcionals i la longitud de la seva seqüència (plana 14). Com que el pRNA i la p16 han de desaparèixer en finalitzar l'empaquetament del DNA (plana 15), seria més senzill si el complex no estigués unit a la càpsida. Podria ser que la p11 els desplaçés en el moment d'unió de la cua del fag. El connector, el pRNA i la p16 formarien el conjunt rotatori (fig. 5.4).

El DNA té els fosfats repartits circularment cada 36° (simetria 10_1), i el connector té les lisines cada 30° (simetria 12). La rotació necessària del connector per a que passi d'un grup d'interaccions a les següents és de només 6° . Després d'aquesta rotació es tornarà a repetir la situació inicial, però ara els enllaços seran amb fosfats del següent nivell, 3.6 \AA per sota. L'atracció per ponts d'hidrogen és prou propera com per a que durant la rotació de 6° es puguin començar a formar enllaços amb els fosfats que estan pujant. A la vegada, els fosfats que s'allunyen aniran perdent aquesta interacció.

Una volta de 360° del connector representaria la translació de 60 parells de bases. S'ha mesurat que el DNA travessa el connector a una velocitat d'uns 140 parells de bases per segon [49], enlentint-se a mesura que queda més comprimit dins de la càpsida. El connector rotaria així a una mica més de dues revolucions per segon, encapsulant els 19000 parells de bases del genoma del $\Phi 29$ en menys de 3 minuts.

El connector encaixa bé en el vèrtex de la càpsida (fig 3c del tercer article, plana 95). El tronc central del connector té una estructura hidrofòbica que permetria que rotés lliurement, aprofitant la diferència de simetries entre aquest i el vèrtex de la càpsida. El disc ample, amb els dotze lòbuls formant estructures tipus SH3, seria qui mantindria el connector en el seu lloc, evitant que pogués desplaçar-se longitudinalment.

L'avantatge d'aquest model és que no cal cap variació en l'estructura del connector. No són necessaris ni canvis en cadenes laterals, ni moviments dels dominis. Un connector rígid pot complir la seva funció molt més ràpid, com una turbina. L'estructura observada és molt compacte i no sembla quedar lloc per canvis relatius entre monòmers. Aquesta afirmació està reforçada pel fet que el connector ha empaquetat en un grup espacial amb alta simetria, amb l'eix 12 situat al llarg d'un de quaternari. Si el connector fos flexible, les diferències entre monòmers no haguessin produït un grup espacial d'aquesta mena. Els monòmers estan separats per un angle de 30 graus exactes.

Reconstruccions de microscòpia electrònica prediuen que el pRNA podria estar localitzat a l'extrem estret del connector. En l'estructura obtinguda, observem que aquest extrem està coronat per una curta hèlix α rica en glutamines i asparagines exposades. Aquesta configuració li dota de potencials ponts d'hidrogen que podrien lligar-se amb el pRNA. El pRNA té simetria 6 i, lligat a l'extrem distal, formaria amb el connector una estructura rígida

on el pRNA seria qui reconeix i condueix el DNA cap a l'entrada del connector. Alhora, el pRNA faria el paper equivalent dels lòbuls del connector i n'evitaria moviments longitudinals d'aquest. L'encaixaria, però no el fixaria en el vèrtex de simetria cinc de la càpsida.

La proteïna ATPasa p16 és l'últim component necessari en el mecanisme d'empaquetament del DNA. Aquesta proteïna intervé en l'empaquetament del DNA dins de la càpsida, subministrant energia, però desapareix en el virus format, ja que per a la infecció és suficient l'energia adquirida pel DNA en comprimir-se en la càpsida.

Es coneix que connectors sense els primers 13 residus de l'extrem N-terminal no es lliguen amb p16. Aquest extrem està situat en la zona central del connector, per la qual cosa aquesta proteïna estaria formant un anell al voltant del tronc del connector i establiria alhora contacte amb el pRNA, situat immediatament a sota. El bloc giratori estaria llavors format pel connector, el pRNA i la p16. L'energia subministrada per la p16 faria girar tot el bloc gràcies a l'impuls proveït per la p16 interaccionant amb la càpsida.

Considerant que la p16 és un hexàmer, són necessàries dues rotacions de 6° per deixar el sistema en un estat igual a l'inicial, respecte de la cadena de DNA que està travessant el connector. Llavors, per a cada hidròlisi d'ATP el DNA avançaria dos parells de bases i coincidiria amb les dades experimentals.

Ambdós models prediuen un avanç de dos parells de bases de DNA cada 12° de rotació del connector, per la qual cosa l'eficiència i el consum d'ATP dels dos models és similar.

El fet que el conjunt rotatori estigui format només pel connector, o bé pel conjunt de connector, pRNA i p16, dependrà de si futurs estudis confirmen la simetria d'aquests dos últims. De totes maneres ja està establert que es tracta d'un sistema motor a base de peces concèntriques i asimètriques que li permeten superar la gran pressió de l'empaquetament del DNA amb un consum reduït d'energia.

El paper dels anells de lisines com a guies del DNA també s'haurà de demostrar en futurs experiments. L'avanç del DNA en aquest model proposat per nosaltres resultaria també producte d'un conjunt d'asimetries que s'anirien ajustant seqüencialment i permetrien l'aparició i desaparició de ponts d'hidrogen donant una aparença que s'estan movent circularment.

Capítol 6

Conclusions

6.1 Conclusions globals

Sobre l'estructura del connector:

- L'estructura del connector s'ha resolt i afinat a 2.1Å a partir d'uns cristalls amb simetria I422 i mostra una estructura amb una forma global similar a les reconstruccions de microscòpia electrònica, però amb detalls a nivell atòmic: Un grau més alt de vorticitat de les 12 còpies de la proteïna p10, un canal interior majoritàriament de càrrega negativa, però amb uns anells de càrrega positiva separats una distància igual al pas de rosca del DNA, i una zona externa amb afinitat al pRNA i a l'ATPasa.

Sobre l'extensió de fases:

- L'extensió de fases a partir d'un model de microscòpia té el problema de la baixa resolució d'aquest model. A aquesta resolució, els efectes del solvent podrien ser prou grans com per impedir un bon ajust de les reflexions. Caldria estudiar una manera d'afegir aquesta contribució, sense haver de fer servir un model d'alta resolució.

Sobre la derivació dels cristalls:

- L'estructura del connector ha mostrat un canal de càrrega negativa i molt uniforme de grans dimensions. Si bé els ions pesants podien entrar per aquest canal, no tenien un lloc únic on fixar-se, per la qual cosa quedaven desordenats dins del cristall.

Sobre el mecanisme d'empaquetament del DNA:

- L'estructura del canal interior del connector forçaria a que el DNA avanci centrat al llarg del seu eix, i sigui impulsat en la direcció correcta gràcies a uns anells de càrrega positiva que establirien en seqüència ponts d'hidrogen entre aquest i el DNA.

6.2 Conclusions dels articles

- Article 1
 - Els cristalls de la proteïna p10 difractaven a molt baixa resolució, eren molt inestables i aguantaven molt poc temps el feix de raigs X a temperatura ambient. La congelació ràpida en propà líquid i la difracció sota un flux de nitrogen va permetre superar aquests problemes, incrementant la resolució fins al voltant dels 3 Å.
 - Els cristalls eren de dos tipus. Uns de grans dimensions, amb 4 connectors per unitat asimètrica, i uns altres amb mig connector per unitat asimètrica. Aquesta segona forma va difractar a major resolució, i en ella s’hi van introduir àtoms pesants per intentar difracció anòmala.
- Article 2
 - Els cristalls de la segona forma van permetre determinar, amb la funció d’autorotació, que la simetria del connector és 12, és a dir, el connector està format per 12 còpies de la proteïna p10. Es tracta del primer anàlisi estructural d’un connector de bacteriòfag, per cristal·lografia de raigs X. Tots els estudis previs s’havien fet per microscòpia electrònica.
 - La funció d’autorotació va determinar, a més, que el connector estava orientat amb el seu canal paral·lel a l’eix c dels cristalls del tipus II, del grup espacial $P4_22_12$.
- Article 3
 - Els cristalls del tipus II van resultar ser macles formades per dos conjunts de cel·les rotades 90°. Amb un model de microscòpia electrònica es va intentar el desmaclat de les dades, obtenint només resultats de l’estructura a baixa resolució.
 - Una nova forma cristal·lina, en el grup espacial $I422$ va difractar a major resolució (2.1 Å), i amb ella es van intentar diverses derivacions.
 - No es va poder localitzar cap del àtoms pesants introduïts en els cristalls, probablement a causa de que no tenen un lloc específic d’unió. Per això es va continuar intentant resoldre l’estructura per extensió de fases a partir d’un model de microscòpia electrònica a baixa resolució.

- A causa de l'empaquetament del connector en el cristall, l'extensió de fases només es podia fer aprofitant un eix de simetria no cristal·logràfica d'ordre 3, resultant en un avanç molt lent de l'extensió de fases. Amb aquest sistema es van poder començar a definir les zones helicoidals del tronc central del connector.
- Paral·lelament, un model sense afinar, a menor resolució va ser publicat i, amb ell, es va poder acabar de resoldre l'estructura a alta resolució.
- L'estructura del connector del fag $\Phi 29$, té una forma general troncocònica, rodejada per uns lòbuls en la part més ampla, i travessada per un canal longitudinal d'un diàmetre suficient per al pas d'una cadena doble de DNA.
- L'interior del canal està molt carregat negativament, a l'igual que la superfície exterior del DNA. Únicament uns anells de lisines, separats igual que una volta de DNA, donen càrrega positiva al canal.
- Aquests anells de lisina probablement guien el DNA al llarg del seu pas, interaccionant amb els àtoms d'oxigen que estan units als de fòsfor del DNA. En rotar el connector, aquestes interaccions profereixen un moviment lineal al DNA al llarg del canal, sense necessitat de tòrcer-lo. Molt semblant a la manera com avança linealment un cargol sota l'acció d'una rosca giratòria.

Referències

- [1] The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D*, 50:760–763, 1994.
- [2] J.P. Abrahams. Bias reduction in phase refinement by modified interference functions: introducing the γ correction. *Acta Cryst D.*, 53:371–376, 1997.
- [3] M. O. Badasso, P. G. Leiman, Y. Tao, Y. He, D. H. Ohlendorf, M. G. Rossmann, and D. Anderson. Purification, crystallization and initial x-ray analysis of the head-tail connector of bacteriophage ϕ 29. *Acta Crystallogr D*, 56:1187–90, 2000.
- [4] R.A. Bosacoma, A.G. Blanco, M. Solà, B.L. Wanner, F.X. Gomis-Rüth, and M. Coll. Structural and functional analysis of a constitutively-active PhoB receiver domain provides further insights into the activation mechanism. *J. Mol. Biol.*, *in press*.
- [5] A.T. Brunger. Free r value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, 355:472–475, 1992.
- [6] Buerger. The patterson function and image theory. pages 276–316.
- [7] A.M. Campbell. *Bacteriophages*, chapter 15. Lippincott-Raven Publishers, Philadelphia, 1996.
- [8] J. M. Carazo, N. Garcia, A. Santisteban, and J. L. Carrascosa. Structural study of tetragonal-ordered aggregates of phage ϕ 29 necks. *J Ultrastruct Res*, 89:79–88, 1984.
- [9] J. M. Carazo, A. Santisteban, and J. L. Carrascosa. Three-dimensional reconstruction of bacteriophage ϕ 29 neck particles at 2·2 nm resolution. *J Mol Biol*, 183:79–88, 1985.

- [10] J. L. Carrascosa and J.M. Valpuesta. Bacteriophage connectors: Structural features of a DNA translocating motor. *Recent Res. Devel. Virol.*, 1:449–465, 1999.
- [11] L. Carrascosa, J. E. Viñuela, N. García, and A. Santiesteban. Structure of the head to tail connector of bacteriophage $\phi 29$. *J. Mol. Biol.*, 154:311–324, 1982.
- [12] C. Chen and P. Guo. Sequential action of six virus-encoded DNA-packaging RNAs during phage $\phi 29$ genomic DNA translocation. *Journal of virology*, 71:3864–3871, 1997.
- [13] Area Detector Systems Corporation. Quantum CCD X-ray detectors. *Manual*.
- [14] L. E. Donate, H. Murialdo, and J. L. Carrascosa. Production of $\lambda - \phi 29$ phage chimeras. *Virology*, 179:936–40, 1990.
- [15] L.E. Donate, J.M. Valpuesta, C. Mier, F. Rojo, and J.L. Carrascosa. Characterization of a RNA-binding domain in the bacteriophage $\phi 29$ connector. *J.Biol.Chem.*, 268:20198–20204, 1993.
- [16] L.E. Donate, J.M. Valpuesta, A. Rocher, E. Méndez, F. Rojo, M. Salas, and J.L. Carrascosa. Role of the amino-terminal domain of bacteriophage $\phi 29$ connector in DNA binding and packaging. *J.Biol.Chem.*, 267:10919–10924, 1992.
- [17] P. Dube, P. Tavares, R. Lurz, and M. van Heel. The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry. *EMBO J.*, 12:1303–1309, 1993.
- [18] R.G. Fisher and R.M. Sweet. Treatment of diffraction data from protein crystals twinned by merohedry. *Acta Cryst A*, 36:755–760, 1980.
- [19] E.F. Garman and T.R. Schneider. Macromolecular cryocrystallography. *J. Appl. Cryst.*, 30:211–237, 1997.
- [20] C. Giacovazzo, H.L. Monaco, D. Viterbo, F. Scordari, G. Gilli, G. Zanotti, and M. Catti. *Fundamentals of Crystallography*. International Union of Cristallography, Oxford University press, 1992.
- [21] F.X. Gomis-Rüth, I. Fita, R. Kiefersauser, R. Huber, F.X. Avilés, and J. Navaza. Determination of hemihedral twinning and initial structural analysis of crystals of the procarboxipeptidase a ternary complex. *Acta Cryst D*, 51:819–823, 1995.

- [22] F.X. Gomis-Rüth, G. Moncalián, R. Pérez-Luque, A. González, E. Cabezón, F. de la Cruz, , and M. Coll. The bacterial conjugation protein TrwB resembles ring helicases and F1-ATPase. *Nature*, 409:637–41, 2001.
- [23] S. Grimes and D. Anderson. RNA dependence of the bacteriophage $\phi 29$ DNA packaging ATPase. *J. Mol. Biol.*, 215:559–566, 1990.
- [24] A. Guasch, A. Párraga, J. Pous, J.M. Valpuesta, J.L. Carrascosa, and M. Coll. Purification, crystallization and preliminary X-ray diffraction studies of the bacteriophage $\phi 29$ connector particle. *FEBS Lett.*, 430:283–287, 1998.
- [25] A. Guasch, J. Pous, A. Párraga, J.M. Valpuesta, J.L. Carrascosa, and M. Coll. Crystallographic analysis reveals the 12-fold symmetry of the bacteriophage $\phi 29$ connector particle. *J Mol Biol.*, 281(2):219–225, 1998.
- [26] P. Guo, C. Zhang, C. Chen, K. Garver, and C. Trottier. Inter-RNA interaction of phage $\phi 29$ pRNA to form a hexameric complex for virial DNA transportation. *Molecular Cell*, 2:149–155, 1998.
- [27] R. W. Hendrix. Bacteriophage DNA packaging: RNA gears in a DNA transport machine. *Cell*, 94:147–150, 1998.
- [28] R.W. Hendrix. Symmetry mismatch and DNA packaging in large bacteriophages. *PNAS*, 75:4779–4783, 1978.
- [29] M.M. Hingorani and M. O’Donnell. Toroidal proteins: running rings around DNA. *Curr. Biol.*, 8(3):R83–86, 1998.
- [30] C Ibáñez, J.A. García, J.L. Carrascosa, and M. Salas. Overproduction and purification of the connector protein *Bacillus subtilis* phage $\phi 29$. *Nucleic Acids Res.*, 12:2351–2356, 1984.
- [31] P. Jardine, M. C. McCormick, C. Lutze-Wallace, and D.H. Coombs. The bacteriophage T4 DNA packaging apparatus targets the unexpanded prohead. *J. Mol. Biol.*, 284:647–659, 1998.
- [32] P. J. Jardine and D.H. Coombs. Capside expansion follows the initiation of DNA packaging in bacteriophage T4. *J. Mol. Biol.*, 284:661–672, 1998.
- [33] P. Jekow, S. Schaper, D. Günter, P. Tavares, and W. Hinrichs. Crystallization and preliminary X-ray crystallographic studies of the 13-fold

- symmetric portal protein of bacteriophage SPP1. *Acta Cryst D*, 54:1008–1011, 1998.
- [34] L.N. Johnson and D. Turk. Erice 2000: Methods for macromolecular crystallography. *IUCR Newsletter*, 9(2):11–12, 2001.
- [35] G.J. Kleywegt and A.T. Brünger. Checking your imagination: applications of the free R value. *Structure*, 4:897–904, 1996.
- [36] G.J. Kleywegt and T.A. Jones. xdlMAPMAN and xdlDATAMAN, programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta Cryst D*, 52:826–828, 1996.
- [37] A.L. Lehninger. *Bioquímica*. Ed. Omega, 1990.
- [38] M. F. Moody. Geometry of phage head construction. *J. Mol. Biol.*, 293:401–433, 1999.
- [39] D.J. Müller, A. Engel, J.L. Carrascosa, and M. Vélez. The bacteriophage ϕ 29 head-tail connector imaged at high resolution with the atomic force microscope in buffer solution. *EMBO J.*, 16:2547–2553, 1997.
- [40] M. Newville and J.O. Cross. Diffkk manual.
- [41] M. Perutz. *Early days of protein crystallography*, volume 114, chapter 1. Academic Press, 1985.
- [42] J. Pous, A. Guasch, B. Ibarra, F.X. Gomis-Rüth, J.M. Valpuesta, N. Sousa, J.L. Carrascosa, and M. Coll. Detailed architecture of a DNA translocating machine: The high-resolution structure of the bacteriophage ϕ 29 connector particle. *J.Mol.Biol*, 315:663–676, 2002. (AG and JP share first authorship).
- [43] E. Prince. *Mathematical Techniques in Crystallography and Materials Science*. Springer-Verlag, 1994.
- [44] R.J. Read. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Cryst D*, 57:1373–1382, 2001.
- [45] D.C. Rees. The influence of twinning by merohedry on intensity statistics. *Acta Cryst A*, 36:578–581, 1980.

- [46] V.R. Samygina, S.V. Antonyuk, V.S. Lamzin, and A.N. Popov. Improving the x-ray resolution by reversible flash cooling combined with concentration screening, as exemplified with PPase. *Acta Crystallographica D.*, 56:595–603, 2000.
- [47] A. A. Simpson, P. G. Leiman, Y. Tao, Y. He, M. O. Badasso, P. J. Jardine, D. L. Anderson, and M. G. Rossmann. Structure determination of the head-tail connector of bacteriophage $\phi 29$. *Acta Cryst. D*, 57:1260–1269, 2001.
- [48] A. A. Simpson, Y. Tao, P. G. Leiman, M. O. Badasso, Y. He, P. J. Jardine, N. H. Olson, M. C. Morais, S. Grimes, D. L. Anderson, T. S. Baker, and M. G. Rossmann. Structure of the bacteriophage $\phi 29$ DNA packaging motor. *Nature*, 408:745–50, 2000.
- [49] D.E. Smith, S.J. Tans, S.B. Smith, S. Grimes, D.L. Anderson, and C. Bustamante. The bacteriophage $\Phi 29$ portal motor can package DNA against a large internal force. *Nature*, 413:748–752, 2001.
- [50] E. Stanley. The identification of twins from intensity statistics. *J. Appl Cryst.*, 5:191–194, 1972.
- [51] S. Turnquist, M. Simon, E. Egelman, and D. Anderson. Supercoiled dna wraps around the bacteriophage $\phi - 29$ head-tail connector. *Proc. Natl. Acad. Sci. USA*, 89:10479–10483, 1992.
- [52] M. Valle, L. Kremer, F. Martinez, A. C. and Roncal, J. M. Valpuesta, J. P. Albar, and J. L. Carrascosa. Domain architecture of the bacteriophage $\phi 29$ connector protein. *J Mol Biol*, 288:899–909, 1999.
- [53] J. M. Valpuesta, J. J. Fernandez, J. M. Carazo, and J. L. Carrascosa. The three-dimensional structure of a DNA translocating machine at 10 Å resolution. *Structure*, 7:289–96, 1999.
- [54] J.M. Valpuesta, L. Carrascosa, J, and R. Henderson. Analysis of electron microscope images and electron diffraction patterns of thin crystals of $\phi 29$ connectors in ice. *J. Mol. Biol*, 240:281–287, 1994.
- [55] J.M. Valpuesta, H. Fujisawa, S. Marco, J.M. Carazo, and L. Carrascosa, J. Three-dimensional structure of T3 connector purified from overexpressing bacteria. *J.Mol.Biol.*, 224:103–112, 1992.
- [56] N. Verdager, B. Blaas, and I. Fita. Structure of Human Rhinovirus Serotype 2 (HRV2). *J. Mol. Biol.*, 300:1179–1194, 2000.

-
- [57] J.D. Watson. *The Double Helix, a personal account of the discovery of the structure of DNA*. W.W.Norton & Company, 1968.
- [58] H. Wiedemann. Synchrotron radiation primer. Technical report, Stanford Synchrotron Radiation Laboratory, 1998.
- [59] A.J.C. Wilson. The probability distribution of X-ray intensities. *Acta Cryst*, 2:318–321, 1949.
- [60] W.B. Wood. Assembly of a complex bacteriophage *in vitro*. *BioEssays*, 14:635–639, 1992.
- [61] M. Wurtz. Bacteriophage structure. *Electron Microsc. Rev.*, 5:283–390, 1992.