Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure determination of the head–tail connector of bacteriophage φ29

Alan A. Simpson,[a] Petr G. Leiman,[a] Yizhi Tao,[a]† Yongning He,[a] Mohammed O. Badasso,[b] Paul J. Jardine,[b] Dwight L. Anderson[b] and Michael G. Rossmann[a]*

[a]Department of Biological Sciences, Purdue University, West Lafayette, IN 47907-1392, USA, and [b]Departments of Microbiology and Oral Science, 18–246 Moos Tower, University of Minnesota, Minneapolis, MN 55455, USA

† Present address: Department of Molecular and Cellular Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138, USA.

Correspondence e-mail: mgr@indiana.bio.purdue.edu

Received 30 March 2001
Accepted 22 June 2001

PDB References: crystal form 3 of head–tail connector of bacteriophage φ29, 1jnb; crystal form 4 of head–tail connector of bacteriophage φ29, 1ijg.

The head–tail connector of bacteriophage φ29 is composed of 12 36 kDa subunits with 12-fold symmetry. It is the central component of a rotary motor that packages the genomic dsDNA into preformed proheads. This motor consists of the head–tail connector, surrounded by a φ29-encoded, 174-base, RNA and a viral ATPase protein, both of which have fivefold symmetry in three-dimensional cryo-electron microscopy reconstructions. DNA is translocated into the prohead through a 36 Å diameter pore in the center of the connector, where the DNA takes the role of a motor spindle. The helical nature of the DNA allows the rotational action of the connector to be transformed into a linear translation of the DNA. The crystal structure determination of connector crystals in space group $C2$ was initiated by molecular replacement, using an approximately 20 Å resolution model derived from cryo-electron microscopy. The model phases were extended to 3.5 Å resolution using 12-fold non-crystallographic symmetry averaging and solvent flattening. Although this electron density was not interpretable, the phases were adequate to locate the position of 24 mercury sites of a thimerosal heavy-atom derivative. The resultant 3.2 Å single isomorphous replacement phases were improved using density modification, producing an interpretable electron-density map. The crystallographically refined structure was used as a molecular-replacement model to solve the structures of two other crystal forms of the connector molecule. One of these was in the same space group and almost isomorphous, whereas the other was in space group $P2_12_12$. The structural differences between the oligomeric connector molecules in the three crystal forms and between different monomers within each crystal show that the structure is relatively flexible, particularly in the protruding domain at the wide end of the connector. This domain probably acts as a bearing, allowing the connector to rotate within the pentagonal portal of the prohead during DNA packaging.

1. Introduction

The bacteriophage φ29 infects *Bacillus subtilis* and consists of a fivefold-symmetric prolate head, about 540 Å long and 420 Å wide, onto which is attached a 500 Å long tail (Tao *et al.*, 1998). Assembly of a prohead is nucleated by the head–tail connector, which is positioned at a fivefold vertex (Tao *et al.*, 1998). The connector itself is a 12-fold symmetric dodecamer of gene product 10 (gp10), which forms a hollow cone about 75 Å long and with its diameter tapering from 66 Å at the narrow end to 138 Å at the wide end. The central pore is approximately 36 Å in diameter at the narrow end (Simpson *et*

Table 1
φ29 head–tail connector crystal data used in the present analysis.

All data were collected at the Advanced Photon Source (APS) BioCARS beamline 14BM-C using the ADSC Quantum 4 2K × 2K CCD detector with a wavelength of 1.000 Å. The data were collected using a 1.0° oscillation angle with an exposure time of 20 s for each image and a crystal-to-detector distance of 250 mm.

	Native data	Native data	Thimerosal derivative	Native data
Type†	1	3	3	4
Space group	$P2_12_12$‡	$C2$	$C2$	$C2$
Unit-cell parameters (Å, °)	a = 157.6 b = 245.2 c = 143.0	a = 177.2 b = 169.2 c = 185.4 β = 114.1	a = 177.9 b = 171.1 c = 184.6 β = 114.1	a = 176.8 b = 171.6 c = 184.7 β = 112.2
V_M (Å^3 Da^{-1})	3.2	2.9	2.9	2.9
No. of connectors per crystallographic asymmetric unit	1	1	1	1
Resolution limit (Å)	3.5	3.2	3.5	2.9
R_{merge} (%)	11.2 (31.3)	6.5 (24.0)	11.8 (23.7)	8.8 (33.3)
Completeness (%)	90.6 (71.7)	99.0 (99.0)	98.2 (98.2)	97.8 (99.0)
Redundancy	2.53	3.77	3.22	3.54
Crystallization conditions	30–35% MPD, 0.2 M $MgCl_2$, 0.1 M HEPES pH 7.5	36–40% MPD, 0.05 M $CaCl_2$, 0.1 M Tris–HCl pH 8.0	36–40% MPD, 0.05 M $CaCl_2$, 0.1 M Tris–HCl pH 8.0	36–40% MPD, 0.05 M $CaCl_2$, 0.1 M Tris–HCl pH 8.0
Maximum crystal size (mm)	0.3	0.3	0.3	0.3
Direction of 12-fold axis	Parallel to a	Perpendicular to b, at an angle of 108.4° with a	Perpendicular to b, at an angle of 108.4° with a	Perpendicular to b, at an angle of 108.4° with a

† See Badasso *et al.* (2000). ‡ The unit-cell directions have been permuted relative to those given by Badasso *et al.* (2000) in order to represent the space group in its standard form.

al., 2000). The connector is surrounded by an oligomeric φ29-encoded, 174-base, RNA molecule (prohead RNA or pRNA) attached to the head. The 19.3 kbp dsDNA genome, covalently attached to the viral protein gp3, is packaged into the assembled prohead by being translated through the center of the connector, requiring the hydrolysis of one ATP molecule for the packaging of two base pairs (Guo *et al.*, 1987). ATPase molecules (gp16) are associated with the pRNA. We have proposed a mechanism for DNA translation by a rotary motor mechanism containing elements of earlier suggestions by Hendrix (1978) and Dube *et al.* (1993), but also incorporating results from both our near-atomic resolution crystal structure of the connector and our cryo-electron microscopy (cryoEM) images of proheads at several stages during packaging (Simpson *et al.*, 2000; Morais *et al.*, 2001).

All known dsDNA phages have head–tail connectors with functions similar to those found in phage φ29 (Valpuesta & Carrascosa, 1994). These connectors have been found to have 12- or 13-fold symmetry (Orlova *et al.*, 1999) with similar dimensions. The head–tail connector of φ29 has been studied by a variety of physical methods, including atomic force microscopy (Müller *et al.*, 1997), cryoEM of two-dimensional arrays (Valpuesta *et al.*, 1999), immune-electron microscopy (Valle *et al.*, 1999) and X-ray crystallography (Guasch *et al.*, 1998; Badasso *et al.*, 2000). Here, we present details of the first near-atomic resolution structure determination of the φ29 connector. This was accomplished by using a low-resolution cryoEM structure (Valpuesta *et al.*, 1999) to initiate phasing to higher resolution sufficient to determine the position of heavy-atom markers. We examine the reasons why non-crystallographic symmetry (NCS) averaging and solvent flattening had been insufficient to solve the structure and we analyze the structures from three different crystal forms in terms of the biological functions of the connector.

2. Crystallization and data collection

Bacteriophage φ29 connectors were produced in *Escherichia coli* and crystallized in several forms (Badasso *et al.*, 2000) (Table 1). Crystal form 3 (Table 1) was used for the initial structure determination. The resultant structure was used to solve crystal forms 1 and 4 (Table 1).

The X-ray diffraction data were processed using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997) as well as *DPS* (Steller *et al.*, 1997; Rossmann & van Beek, 1999). One native and one heavy-atom-derivatized type 3 crystal were used for the initial structure determination. Excellent rotation functions (Badasso *et al.*, 2000) demonstrated the presence of 12-fold symmetry and determined the direction of the 12-fold molecular axes with respect to the unit-cell axes.

3. Molecular replacement using a low-resolution phasing model

A rotation function (Badasso *et al.*, 2000) had shown that the 12-fold NCS connector axis was perpendicular to the crystallographic b axis and at an angle of 108.4° to the a axis (Fig. 1) in the better diffracting type 3 crystals. A self-translation function (Argos & Rossmann, 1980) showed that the NCS axis intersected the crystallographic dyad [see Badasso *et al.* (2000) for the $T(\mathbf{S})$ function for type 2 crystals]. The 12-fold NCS combined with the crystallographic dyad results in overall D_{12} symmetry for two connector oligomers, aligned across a crystallographic twofold axis. This results in identical lattice contacts between opposing connectors (Fig. 1), constraining the position of the NCS axis. A similar situation exists for the packing of southern bean mosaic virus particles into a rhombohedral crystal lattice, where the viral fivefold axes of neighboring particles are aligned (Abad-Zapatero *et al.*, 1981). Given the exact position of the NCS axis in the cell, a search with a phasing model requires the exploration of two parameters, namely rotation about and translation along the NCS axis.

An initial low-resolution phasing model was obtained from two orthogonal views of the connector determined from a cryoEM study of two-dimensional crystals (Valpuesta *et al.*, 1999). The structure of the two-dimensional array had been determined to a resolution of 10 Å in the plane of the lattice, decreasing to 17 Å perpendicular to the lattice. The two perpendicular views published by Valpuesta *et al.* (1999), together with the 12-fold symmetry of the connector, were sufficient information for us to derive a moderately accurate three-dimensional envelope plus a rough estimate of the electron distribution within the envelope of the connector. The model was represented by atoms placed on a 1 Å cubic lattice. The weight of each atom was derived in part from the density in the projected image down the 12-fold axis.

The model was used to perform two types of searches, one based on packing considerations and the other on structure-amplitude agreement. The searches explored rotations about the NCS axis in two intervals over an angle of 30° (there being a repeat every 30° on account of the 12-fold symmetry axis of the connector) and in steps of 3.7 Å along the axis. The packing searches counted the number of steric clashes, defined as the number of atoms in the reference connector within 4 Å of an atom in a neighboring connector molecule. (For this purpose, the number of atoms were reduced by randomly retaining only one out of every 300 atoms, giving a mean atomic separation of 6.8 Å.) Structure-factor correlation searches for data between 60 and 20 Å resolution were calculated using the program *AMoRe* (Navaza, 1994). Two types of structure factors were calculated, the normalized set (E values), in which the mean of all amplitudes in each resolution shell was adjusted to be constant, and the sharpened set, in which the calculated amplitudes were sharpened with a Gaussian to have the same fall-off of intensity with resolution as the corresponding observed data set. All searches were performed twice, once with the narrow end closest to the crystallographic twofold that intersects the NCS axis (Fig. 1) and once with the wide end closest to the same dyad.

Table 2
Refined Hg parameters.

Site number†	Fractional coordinates			Occupancy	B factor (Å^2)	Mean deviation‡ (Å)
	x	y	z			
A1	0.295	−0.284	0.919	0.407	42	0.9
A2	0.113	−0.373	0.858	0.278	38	1.3
A3	−0.095	−0.360	0.803	0.207	51	1.2
A4	−0.266	−0.252	0.742	0.141	63	1.6
A5	−0.351	−0.077	0.726	0.182	51	1.1
A6	−0.338	0.118	0.732	0.197	46	1.3
A7	−0.238	0.274	0.755	0.201	50	2.1
A8	−0.048	0.369	0.813	0.266	49	1.1
A9	0.158	0.355	0.881	0.380	52	1.0
A10	0.327	0.250	0.928	0.238	43	1.0
A11	0.422	0.070	0.961	0.218	37	1.2
A12	0.409	−0.121	0.955	0.352	59	1.0
B1	0.204	−0.249	0.958	0.237	25	1.0
B2	0.060	−0.291	0.917	0.242	20	1.3
B3	−0.108	−0.256	0.864	0.192	26	2.1
B4	−0.207	−0.162	0.835	0.155	24	1.5
B5	−0.265	−0.013	0.819	0.148	63	1.3
B6	−0.230	0.130	0.827	0.157	62	1.0
B7	−0.125	0.236	0.859	0.172	45	1.4
B8	0.027	0.284	0.908	0.192	54	1.2
B9	0.181	0.257	0.952	0.248	42	1.1
B10	0.296	0.155	0.988	0.206	30	1.0
B11	0.343	0.018	1.007	0.233	38	1.7
B12	0.315	−0.144	0.994	0.171	49	1.9

† Two sets of NCS-related heavy atoms were found (A and B), which were later found to be associated with Cys76 and Cys265, respectively. ‡ The mean deviation is defined as $\sum_{j=1}^{11} d_{i,i+j}$, where $d_{i,i+j}$ is the distance between atom i and atom $i + j$ after the monomer i is superimposed onto monomer $i + j$.

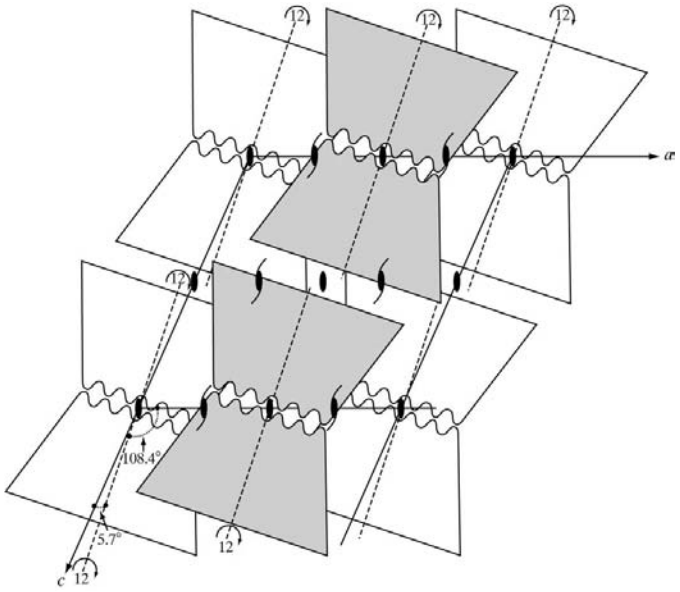


Figure 1
Packing arrangement of connectors, where pairs of connectors face each other across crystallographic dyads, within the type 3 $C2$ crystal unit cell. Connector pairs at ($\frac{1}{2}$, $\frac{1}{2}$, 0) are in gray.

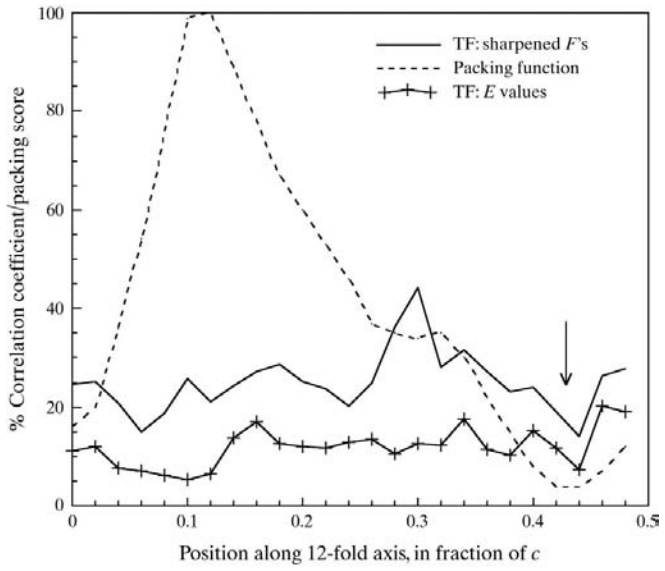


Figure 2
Translation searches showing a packing function and two different structure-factor correlation searches. The 12-fold axis selected for the searches passes through the origin. The orientation of the model was placed as in Fig. 1. The abscissas measure the distance of an arbitrary origin on the model relative to the origin of the unit cell. The arrow shows the chosen position of the model.

Table 3
Phasing statistics.

Resolution range	No. of acentric reflections	Acentric FOM†	Phasing power‡ acentric	Cullis *R* factor§ acentric	No. of centric reflections	Centric FOM†	Phasing power‡ centric	Cullis *R* factor§ centric	Total No. of reflections	Overall FOM
20.34	161	0.26	1.03	0.84	27	0.44	0.66	0.81	188	0.29
11.51	910	0.33	1.20	0.76	77	0.52	0.77	0.79	987	0.34
8.02	2252	0.31	1.30	0.79	129	0.52	0.95	0.72	2381	0.32
6.16	4947	0.31	1.38	0.78	207	0.47	1.04	0.80	5154	0.32
5.00	8148	0.33	1.28	0.79	265	0.52	0.85	0.78	8413	0.33
4.20	11938	0.28	0.87	0.86	319	0.47	0.59	0.80	12257	0.28
3.63	16725	0.16	0.73	0.95	378	0.31	0.49	0.92	17103	0.16
3.19	16398	0.12	0.69	0.97	332	0.22	0.48	0.96	16730	0.12
Total	61479	0.21	0.90	0.89	1734	0.40	0.67	0.84	63213	0.22

† Figure of merit. ‡ Phasing power = $\langle|F_H|\rangle/\langle|\text{lack of closure}|\rangle$, where $\langle|F_H|\rangle$, $\langle|F_{PH}|\rangle$ and $\langle|F_P|\rangle$ are the mean of the heavy-atom, heavy-atom derivative and native structure amplitudes, respectively, and lack of closure = $|F_{PH} - |F_P + F_H||$. § Cullis *R* factor = $\langle|\text{lack of closure}|\rangle/\langle||F_{PH}| - |F_P||\rangle$.

The packing-function search (Fig. 2) had a clear minimum with only 250 steric clashes, rising to a maximum of 6000 clashes away from the best position. However, the packing search could not readily discriminate between different rotational angles around the 12-fold connector axis because the connector envelope, but not its density distribution, was roughly circularly symmetric. The structure-amplitude search using the normalized structure factors gave a maximum correlation coefficient of 0.2 (Fig. 2) within 5.5 Å from the best position found with the packing function. In contrast, the search with the sharpened structure factors gave a well defined maximum (correlation coefficient of 0.45), but in disagreement with the packing and normalized structure-factor searches. Therefore, the normalized search and packing search were used for positioning the model for subsequent phasing with the narrow end of the connector closest to the crystallographic dyad.

With the translation and orientation of the model in the crystal cell tentatively determined, a set of phases were calculated and applied to the observed amplitudes in the 60–20 Å resolution range. The phases were extended in small incremental steps to 10 Å resolution, assuming the envelope of the original model (Arnold & Rossmann, 1988) for solvent flattening and averaging, using the program *DM* (Cowtan, 1994). Once this resolution had been reached, it was possible to refine the solvent-flattening envelope (Wang, 1985; Leslie, 1989). At this stage, the monomer was also divided into three domains (the narrow end, the central portion and the wide end) in order to refine the NCS operators for each domain separately. Eventually, the phases were extended to 3.5 Å resolution (Fig. 3). Although these phases were of insufficient accuracy to produce an interpretable electron-density map, they were adequate to produce an interpretable isomorphous heavy-atom difference map.

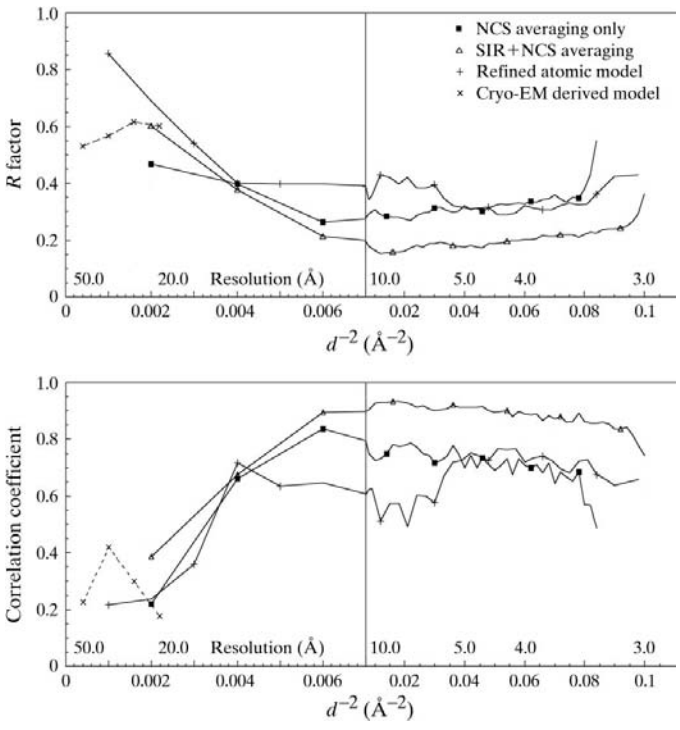


Figure 3
R factor and correlation-coefficient distribution at different stages in the structure-factor determination.

4. Isomorphous replacement

Fortunately, an isomorphous thimerosal heavy-atom derivative diffraction data set (Table 1) had been collected earlier. The native and thimerosal data sets were put onto the same relative scale with the program *SCALEIT* (Collaborative Computational Project, Number 4, 1994). The overall $R_{\rm diff}$ value was 24.4%, rising to 27.2% at 3.2 Å resolution. The difference Fourier map computed with *SIGMAA* weights (Read, 1986) showed two sets of 12-fold related sites at radii of 64 and 50 Å from the common NCS axis, both in the wide-end domain. It was later verified that these sites corresponded to the only two cysteine residues, Cys76 and Cys265, respectively. The parameters for each of the heavy atoms in the crystallographic asymmetric unit (Table 2) were refined using the program *MLPHARE* (Otwinowski, 1991). The resultant 3.5 Å resolution single isomorphous replacement (SIR) phases (Table 3) were improved and extended to 3.2 Å resolution by NCS averaging and solvent flattening using the program *DM*. The SIR phases were combined with the phases derived by Fourier transformation of the modified map in each cycle.

5. Structure refinement of the type 3 crystal data

The averaged 3.2 Å resolution electron-density map from the improved SIR phases was of reasonable quality and allowed rapid tracing of the polypeptide chain with the programs *O*

(Jones *et al.*, 1991) and *XtalView* (McRee, 1999). The structure was then refined with the program *CNS* (version 1.0; Brunger *et al.*, 1998), using rigid-body refinement of each of the three domains followed by simulated annealing and restrained positional and individual temperature-factor refinement with a maximum-likelihood-weighted target function against *F*s (Adams *et al.*, 1997). The three sets of 12 NCS-related domains were restrained to be similar (Fig. 4). The wide-end domain Cα atoms deviated from overall 12-fold symmetry by as much as 2.5 Å. However, in the other two domains deviations from the overall 12-fold symmetry were insignificant. Although bulk water corrections were applied, the agreement of model-calculated and observed structure factors was very poor at low resolution (Fig. 3). The final $R_{\rm working}$ and $R_{\rm free}$ factors were 28.4 and 30.1%, respectively, for data in the resolution range 9.0–3.2 Å. All but 5.6% of the main-chain dihedral angles were within the generous limits of the Ramachandran plot (Laskowski *et al.*, 1993). The bond lengths and bond angles deviated from idealized values by 0.009 Å and 1.6°, respectively.

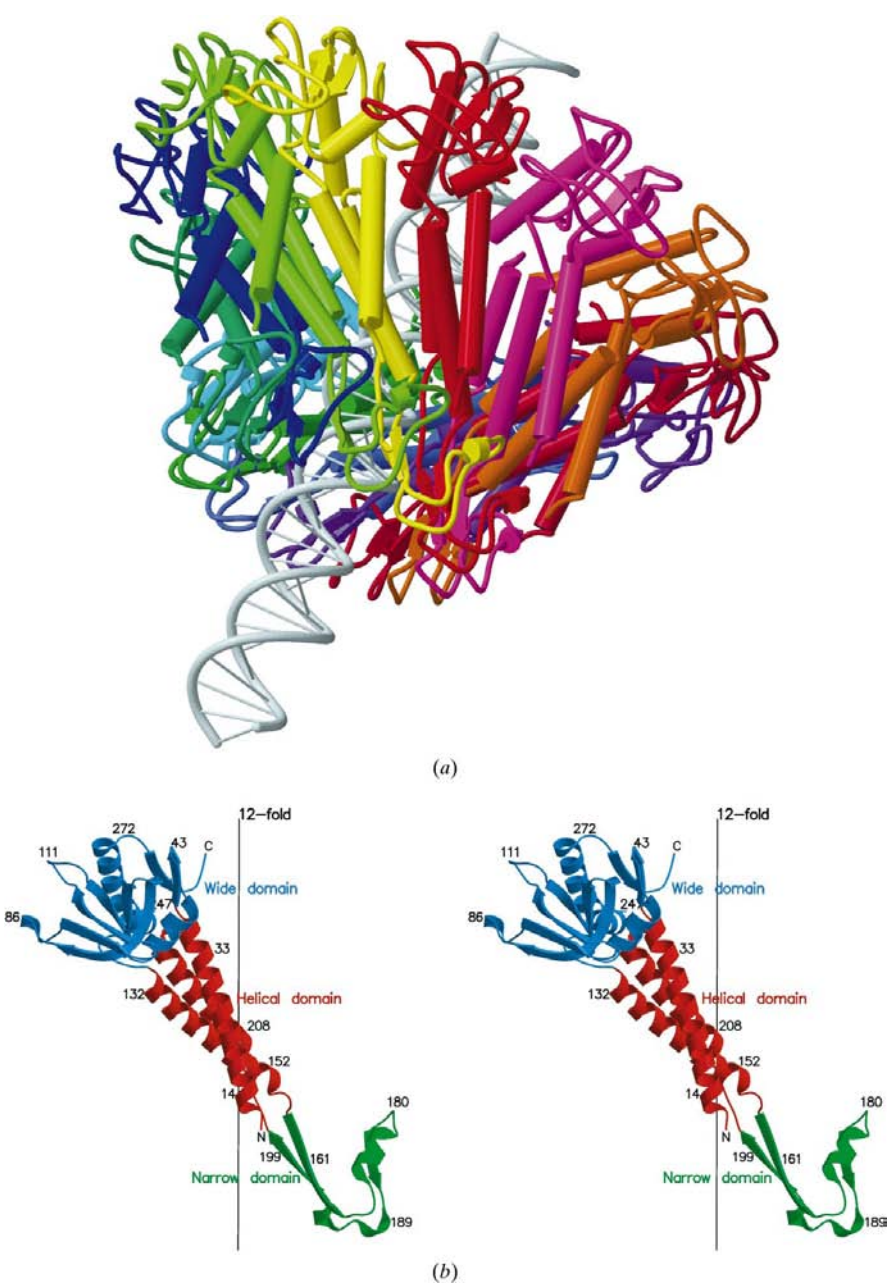


Figure 4
(*a*) The connector structure, including a model of DNA passing through its center. Each monomer is represented by a different color. (*b*) Stereo diagram of a single monomer seen with respect to the central 12-fold axis. Selected amino-acid residue numbers are shown. The three domains are colored differently. [These figures were prepared using the programs *MOLSCRIPT* (Kraulis, 1991) and *Raster3D* (Merritt & Bacon, 1997).]

6. Molecular-replacement solution of other crystal forms

Although the unit-cell parameters of type 3 and 4 crystals were almost the same (Table 1), the $R_{\rm diff}$ between the native data sets was 46%. Attempts at refining the partially refined type 3 atomic structure against the type 4 observed data failed. Nevertheless, as the type 4 data extended to 2.9 Å resolution, it was decided to solve the structure using the program *AMoRe* (Navaza, 1994). The outstanding top solution of the rotation function, calculated in the resolution range 50–5 Å, was used to compute a translation function for the same data. The top peak gave an *R* factor and correlation coefficient of 46% and 0.41, respectively. Rigid-body refinement of the whole monomer constrained by NCS decreased the *R* factor to 45% and increased the correlation coefficient to 0.48. A comparison of the directions of the NCS axes relative to the crystal axes in the two crystal forms are given in Table 1. Initial refinement with the program *CNS* (version 1.0; Brunger *et al.*, 1998), treating three domains in each monomer as rigid bodies and using a bulk-solvent correction, gave $R_{\rm working}$ = 41.1% and $R_{\rm free}$ = 40.5% in the 9.0–2.9 Å resolution range. Subsequent simulated annealing with restrained NCS using torsion-angle dynamics with 9.0–2.9 Å resolution data lowered $R_{\rm working}$ to 37.0% and $R_{\rm free}$ to 38.4%. Analysis of the electron density showed that residues 108–113 in the wide-end domain were significantly different to each other. Hence, the NCS restraints were removed in this region for further refinement. Model rebuilding, individual restrained temperature-factor refinement and positional refinement with NCS

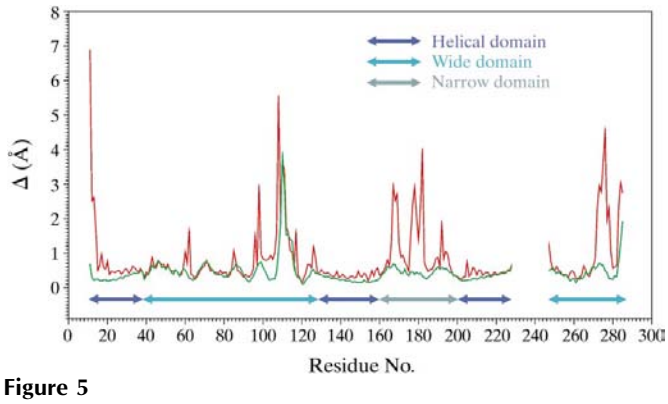


Figure 5
Mean deviations of C^{α} positions between the monomers of the type 4 crystal form in green and between the connectors in the type 3 and type 4 crystal forms in red.

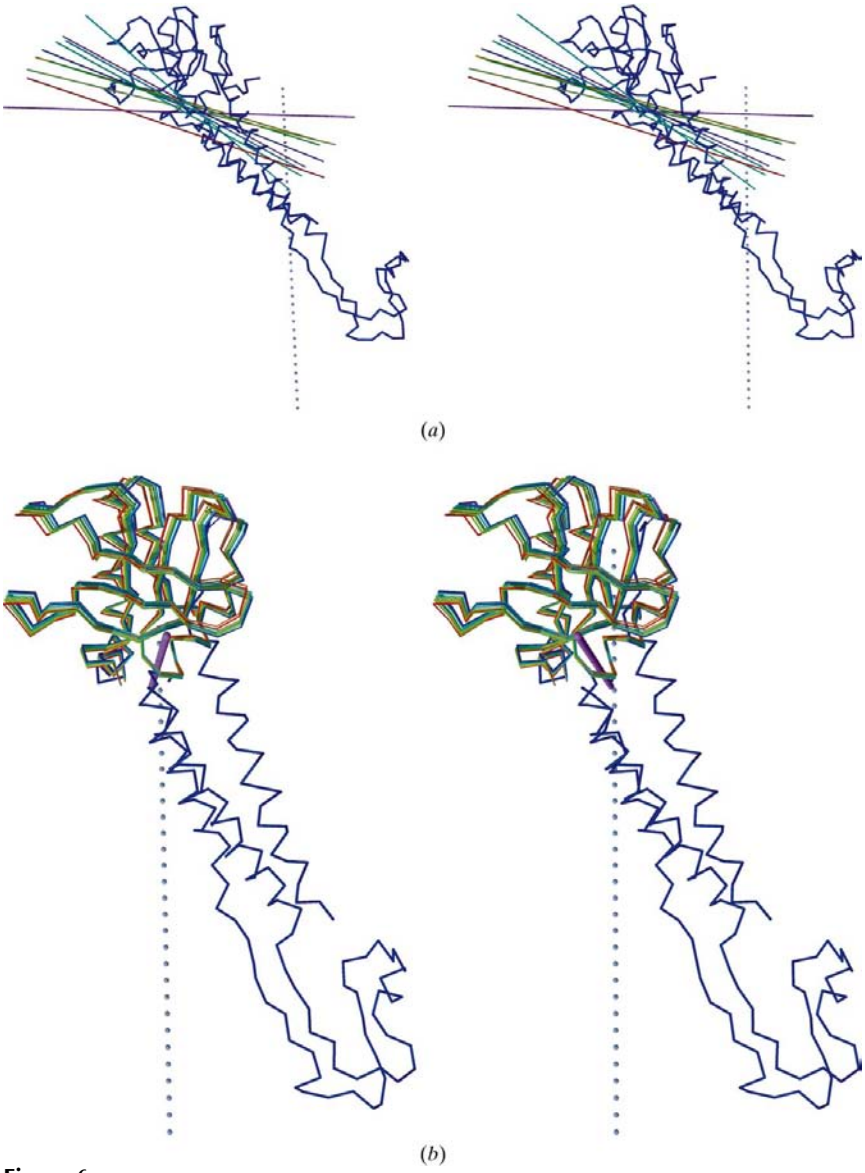


Figure 6
Stereo diagram based on the superposition of the helical domain in crystal form 3. Sequence of colors follows the order of displacement shown in Table 4. (*a*) Axes of rotation relating any one wide-end domain to the reference monomer *F*. (*b*) The variation in position of the wide-end domain in the type 3 crystal form. [These figures were prepared using the programs *XtalView* (McRee, 1999) and *Raster3D* (Merritt & Bacon, 1997).]

restraints applied to main-chain atoms eventually resulted in an $R_{\rm working}$ factor of 22.4% and an $R_{\rm free}$ of 26.4%. The r.m.s. difference from idealized bond lengths was 0.008 Å and from idealized bond angles was 1.3°. All residues except for Asn108 of monomer *C* were within the generously allowed regions of the Ramachandran plot.

The programs *AMoRe* and *CNS* were also used to solve and refine the type 1 crystal form (Table 1). Independent rigid-body refinement of the three domains, as described for the type 3 crystal form above, showed a variation of 2.7 Å in the Cα atoms of the wide-end domain from what their positions would be if there had been proper 12-fold NCS symmetry. Individual atomic refinement did not improve $R_{\rm free}$. The final $R_{\rm working}$ was 42.2% and $R_{\rm free}$ was 42.9%. Phases from this model were used for NCS density modification, resulting in a final density-modification correlation coefficient of 0.85. The resultant electron-density map of the type 1 crystal form was readily interpretable in the central helical domains, although it was of slightly lower quality than the map of the type 3 crystal form. However, the electron density corresponding to the narrow and wide domains was poorly defined, particularly in the surface loops of the wide domain. Since the electron density of the central domain was well defined, the poor density in the other two domains is probably caused by structural variation in different crystal unit cells.

7. The structure

The connector monomer consists of a central well defined helical domain with a loop forming the narrow-end domain and a somewhat complex less well ordered wide-end domain. Residues 158–202 comprise the narrow-end domain, forming mostly an antiparallel β-sheet that binds the pentameric pRNA (Simpson *et al.*, 2000). The central domain consists of a bundle of three α-helices (residues 11–14, 130–157 and 203–226) twisted in a left-handed superhelix around the central 12-fold axis at an angle of approximately 40° (Fig. 4). The wide-end domain (residues 42–129 and 248–285) is flexibly attached to the central domain, creating the wide end of the connector, and forms a contact surface with the inside of the head in the assembled phage. Of the 309 residues in the connector monomer, three regions (residues 1–10, 229–246 and 286–309) were sufficiently disordered to make it impossible to build a reasonable model to the averaged or model-phased electron

Table 4
Differences in the position of the wide-end domains after superposition of helical domains, using monomer *F* as a standard.

Monomers are labelled sequentially *A*, *B*, *C*, ..., *L*.

Crystal form 4				Crystal form 3			
Domain	δC_α† (Å)	κ‡ (°)	t§ (Å)	Domain	δC_α† (Å)	κ‡ (°)	t§ (Å)
F	0	0	0	*F*	0	0	−0.36
C	0.19	0.6	0.02	*C*	0.0	0.0	0.00
A	0.73	1.4	0.04	*I*	0.70	2.0	0.01
H	0.80	1.8	0.04	*A*	1.29	2.1	0.01
J	1.10	3.1	0.05	*L*	1.40	2.6	0.01
K	1.69	4.0	0.02	*H*	1.53	3.1	0.03
E	1.72	4.25	0.05	*E*	1.49	3.1	0.04
I	2.70	6.41	0.13	*J*	1.58	3.2	0.06
L	2.75	6.46	0.14	*K*	1.78	3.5	0.01
G	2.95	7.18	0.10	*G*	2.16	4.7	0.10
B	3.03	7.11	0.13	*B*	2.51	5.4	0.02
D	3.26	7.54	0.43	*D*	2.34	5.1	0.23

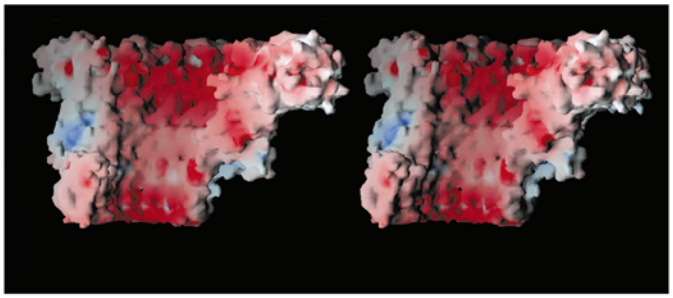
† δC_α (Å) is the distance of residue 273 C_α in *F* to C_α in any other monomer after superposition of the helical domain. ‡ κ (°) is the magnitude of rotation about the axis of rotation. § t (Å) is the translation parallel to the instantaneous axis of rotation.

density in any of the three determined crystal structures. However, the second two disordered regions were evident in the 22 Å resolution cryoEM reconstructed image (Simpson *et al.*, 2000).

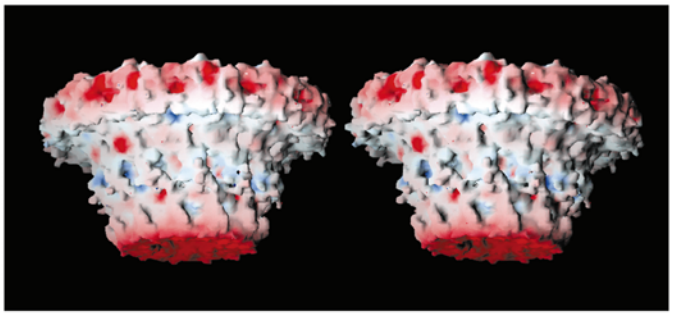
In both type 3 and 4 crystal forms, two connectors are packed with narrow-end-to-narrow-end in a D_{12} arrangement (Fig. 1). This probably prevented the narrow-end domains from deviating significantly from exact 12-fold symmetry. However, it is possible that there would be larger deviations when the 12-fold symmetric narrow end is turning inside the fivefold-symmetric pRNA (Simpson *et al.*, 2000) during DNA packaging. The type 1 crystals are packed differently, with the narrow end of the connector fitting between narrow ends of four connectors in the adjacent layer. A similarly packed two-dimensional crystal form has been examined previously using atomic force microscopy (Müller *et al.*, 1997). The considerable disorder in the narrow-end domain of the type 1 crystal form supports the concept that the structure of this domain is normally variable except when stabilized by opposing connector structures as in crystal types 3 and 4.

The better quality type 4 crystals showed significant conformational differences in the electron densities representing residues 108–113 in the wide-end domains (Fig. 5). Comparison of the monomer of the type 3 and 4 crystal forms showed significant differences, not only in the same region but also in the narrow-end domain (Fig. 5). The narrow-end domain differences corresponded to the D_{12} connector-interface regions at residues 165–172 and 190–196, as well as the central section of the domain, residues 177–184. The change in the interface between the D_{12}-related connectors would account for the slight difference in the type 3 and type 4 crystal packing. If the atoms in the connector wide-end domains are excluded, then the difference between the type 3 and 4 crystals is a 0.6° rotation about the 12-fold NCS axis. The difference in the central region of the narrow-end domain may be important for the adaptation of this domain in binding to the fivefold pRNA during the transmission of energy from the ATPase to the connector.

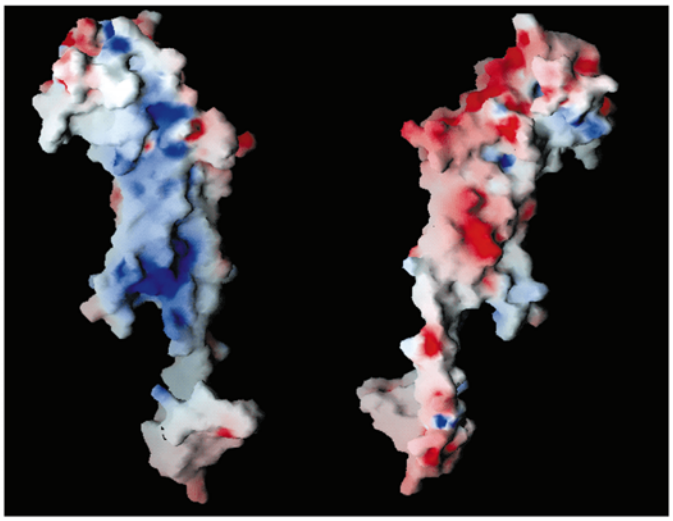
Differences in the structure of the connector between the different crystal forms are mainly localized in the wide-end domain. Superposition of NCS-related subunits (Table 4) shows that the narrow-end and middle domains obey almost perfect 12-fold symmetry, but the wide-end domains show



(*a*)



(*b*)



(*c*)

Figure 7
Charge distribution. (*a*) Internal view of the connector showing the negatively charged surface forming the DNA channel. (*b*) External view of the whole connector showing the highly charged negative regions at either end of the connector and the relatively hydrophobic region in the center that would make contact with the phage head. (*c*) Opposing views of a monomer showing the oppositely charged surfaces that are in contact between the subunits. (These figures were prepared using the program *GRASP*; Nicholls *et al.*, 1993.)

rigid rotations relative to their middle domains of up to 5.4°. These rotations are about local axes that are at an angle of 75 ± 5° relative to the common 12-fold axis, corresponding to a 'rolling' motion of the wide-end domain on the inside surface of the head (Fig. 6). The ability of specific domains to be slightly altered in their position is consistent with the mechanism proposed for connector rotation and DNA packaging (Simpson *et al.*, 2000), in which one monomer of the 12 is required to have a temporarily altered conformation.

8. Charge distribution

The charge distribution over a monomer is shown for the type 3 crystal structure refined coordinates (Fig. 7*a*). The monomer has a highly negatively charged buried surface on one side and highly positively charged surface on the other. The total buried surface area per monomer is 6040 Å^2. Therefore, the connector is held together by the charge complementarity of neighboring monomers within the dodecamer. A surface-charge distribution was also computed for

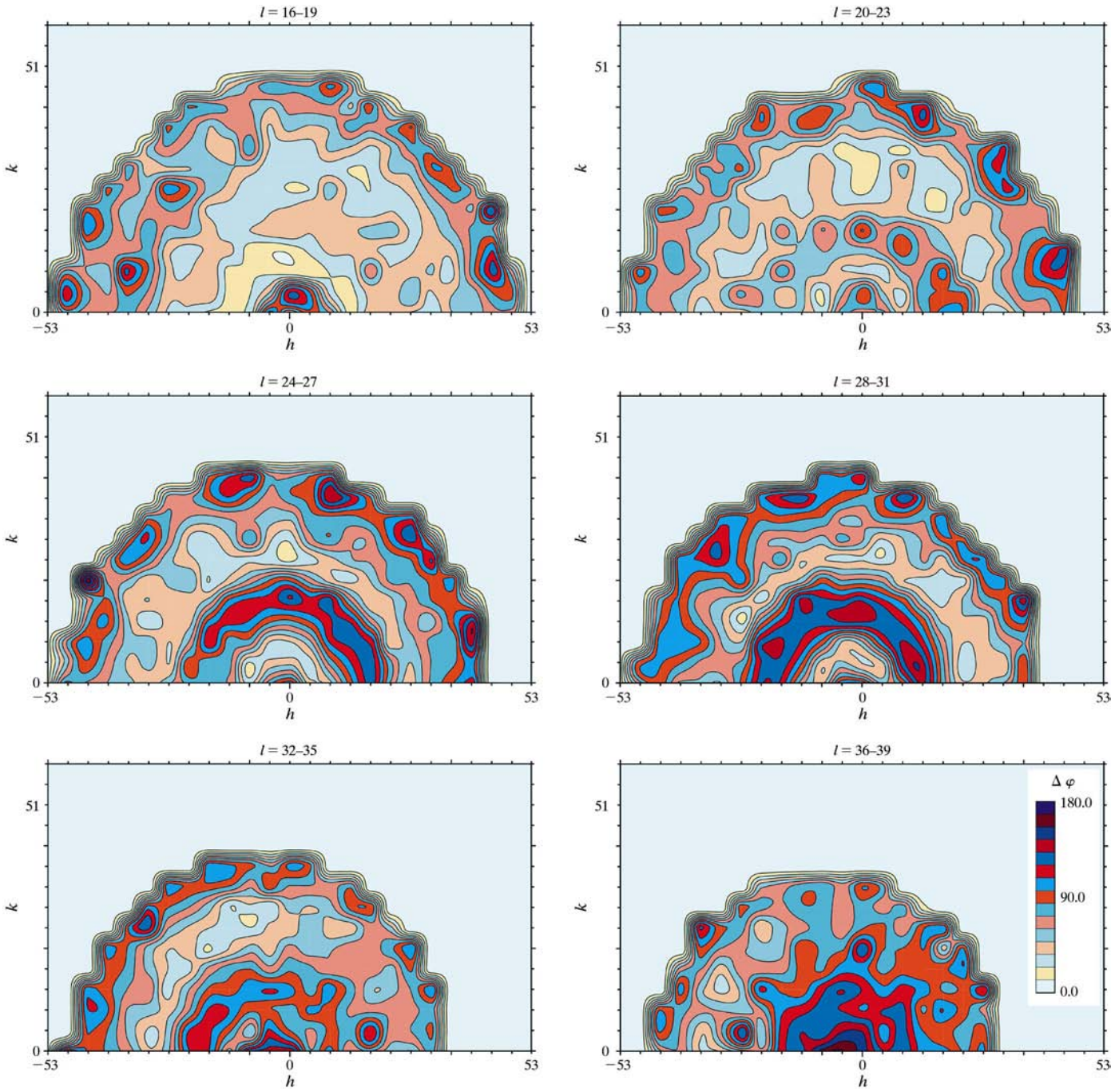


Figure 8
Plots of the mean phase differences for different regions of reciprocal space between the final NCS density-modified *SIR* maps for the type 3 crystals and the original phases derived from the cryoEM model after phase extension to 3.5 Å. Systematic differences vary with respect to both resolution and the *l* index. Asymmetry of the distribution is caused by incompleteness of the data.

the whole connector. The region identified as making contact with the phage head, consisting of parts of the central helical domain and the wide-end domain, is largely hydrophobic (Fig. 7*b*). As this is the part of the connector that must rotate within a pentameric opening of the phage head, the hydrophobic external surface provides an oily interface suitable for rotation, assuming that the phage capsid has a similar hydrophobic opposing surface. A similar analysis of the internal cylindrical hole through the connectors (Fig. 7*c*) shows this surface to be negatively charged and therefore acting as a focus to center the opposing negatively charged exterior of the DNA that is translated through the connector (Simpson *et al.*, 2000).

9. Analysis of the phases derived from the initial model and NCS averaging

A plot of the phase differences between the final NCS density-modified SIR maps for the type 3 crystals α_1 and the original phases α_2 derived from the cryoEM model after phase extension to 3.5 Å shows systematic differences varying with respect to both resolution and the l index (Fig. 8). This figure shows the coexistence of two types of solutions, one where the difference is approximately zero (the correct solution, $\alpha_1 \rightarrow \alpha_2$) and the other where the difference is approximately π (the Babinet inverted solution, $\alpha_1 \rightarrow \alpha_2 + \pi$). Once the final structure had been determined, it was found that both these structures had a hand inversion. This was because the original cryoEM model had approximate symmetry 12*m* and therefore the initial phasing was unable to differentiate easily between opposite enantiomorphic solutions. In the case of icosahedral NCS, initiating phases with a poor model frequently causes a Babinet inversion between sets of reflections in different resolution shells during phase extension (Valegård *et al.*, 1991; McKenna *et al.*, 1992; Tsao *et al.*, 1992). However, in the case of the connector crystals, the 12-fold NCS axis is almost parallel to the crystallographic *c* axis to within 5.7°. Because of the slight tilt of the NCS 12-fold axis with respect to the *a*b** planes, there will be some transfer of phase information at the lines of intersection between NCS-related reciprocal lattice points in adjacent *a*b** reciprocal-lattice planes (Fig. 8). Thus, if the tilt were zero, it would be expected that the phases in any one layer of l would be either all α or $\alpha + \pi$. The tilt will introduce opposing Babinet solutions in different resolution shells. Initial errors were probably introduced in the φ29 connector structure determination because the very low resolution data were in very poor agreement with both the calculated structure factors from density-modified maps and from the refined atomic model (Fig. 3). In the present case, the problem was corrected by using the isomorphous heavy-atom data set, which initiated phasing for the same model on each reciprocal-lattice plane perpendicular to the NCS axis.

We are grateful for the assistance of the personnel at the Cornell High Energy Synchrotron Source, the Advanced Photon Source BioCARS and the Advanced Photon Source Structural Biology Center for help with data collection. We are thankful also for the help of Hiro Tsuruta and the use of beamline 4 at Stanford Synchrotron Radiation Laboratory (SSRL) for very low resolution data collection and to April Burch for her assistance in this collection. We thank Sharon Wilder for help in the preparation of the manuscript. The work was supported by a National Science Foundation grant to MGR and a National Institutes of Health grant to DLA.

References

Abad-Zapatero, C., Abdel-Meguid, S. S., Johnson, J. E., Leslie, A. G. W., Rayment, I., Rossmann, M. G., Suck, D. & Tsukihara, T. (1981). *Acta Cryst.* B**37**, 2002–2018.
Adams, P. D., Pannu, N. S., Read, R. J. & Brünger, A. T. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 5018–5023.
Argos, P. & Rossmann, M. G. (1980). *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. Ladd & R. A. Palmer, pp. 361–417. New York: Plenum.
Arnold, E. & Rossmann, M. G. (1988). *Acta Cryst.* A**44**, 270–282.
Badasso, M. O., Leiman, P. G., Tao, Y., He, Y., Ohlendorf, D. H., Rossmann, M. G. & Anderson, D. (2000). *Acta Cryst.* D**56**, 1187–1190.
Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. D. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Dube, P., Tavares, P., Lurz, R. & van Heel, M. (1993). *EMBO J.* **12**, 1303–1309.
Guasch, A., Pous, J., Párraga, A., Valpuesta, J. M., Carrascosa, J. L. & Coll, M. (1998). *J. Mol. Biol.* **281**, 219–225.
Guo, P., Peterson, C. & Anderson, D. (1987). *J. Mol. Biol.* **197**, 229–236.
Hendrix, R. W. (1978). *Proc. Natl Acad. Sci. USA*, **75**, 4779–4783.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kraulis, P. (1991). *J. Appl. Cryst.* **24**, 946–950.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Leslie, A. G. W. (1989). *Proceedings of the CCP4 Study Weekend. Improving Protein Phases*, edited by S. Bailey, E. Dodson & S. Phillips, pp. 13–24. Warrington: Daresbury Laboratory.
McKenna, R., Xia, D., Willingmann, P., Ilag, L. L. & Rossmann, M. G. (1992). *Acta Cryst.* B**48**, 499–511.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Morais, M. C., Tao, Y., Olson, N. H., Grimes, S., Jardine, P. J., Anderson, D. L., Baker, T. S. & Rossmann, M. G. (2001). In the press.
Müller, D. J., Engel, A., Carrascosa, J. L. & Vélez, M. (1997). *EMBO J.* **16**, 2547–2553.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Nicholls, A., Bharadwaj, R. & Honig, B. (1993). *Biophys. J.* **64**, 166–170.
Orlova, E. V., Dube, P., Beckmann, E., Zemlin, F., Lurz, R., Trautner, T. A., Tavares, P. & van Heel, M. (1999). *Nature Struct. Biol.* **6**, 842–846.
Otwinowski, Z. (1991). *Proceedings of the CCP4 Study Weekend. Isomorphous Replacement and Anomalous Scattering*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 80–86. Warrington: Daresbury Laboratory.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Rossmann, M. G. & van Beek, C. G. (1999). *Acta Cryst.* D**55**, 1631–1640.
Simpson, A. A., Tao, Y., Leiman, P. G., Badasso, M. O., He, Y., Jardine, P. J., Olson, N. H., Morais, M. C., Grimes, S., Anderson, D. L., Baker, T. S. & Rossmann, M. G. (2000). *Nature (London)*, **408**, 745–750.
Steller, I., Bolotovsky, R. & Rossmann, M. G. (1997). *J. Appl. Cryst.* **30**, 1036–1040.
Tao, Y., Olson, N. H., Xu, W., Anderson, D. L., Rossmann, M. G. & Baker, T. S. (1998). *Cell*, **95**, 431–437.
Tsao, J., Chapman, M. S. & Rossmann, M. G. (1992). *Acta Cryst.* A**48**, 293–301.
Valegård, K., Liljas, L., Fridborg, K. & Unge, T. (1991). *Acta Cryst.* B**47**, 949–960.
Valle, M., Kremer, L., Martínez-A., C., Roncal, F., Valpuesta, J. M., Albar, J. P. & Carrascosa, J. L. (1999). *J. Mol. Biol.* **288**, 899–909.
Valpuesta, J. M. & Carrascosa, J. L. (1994). *Quart. Rev. Biophys.* **27**, 107–155.
Valpuesta, J. M., Fernández, J. J., Carazo, J. M. & Carrascosa, J. L. (1999). *Structure*, **7**, 289–296.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.