Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Exploiting large non-isomorphous differences for phase determination of a G-segment invertase–DNA complex

Christopher J. Ritacco,[a] Thomas A. Steitz[a,b,c] and Jimin Wang[a]*

[a]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA, [b]Department of Chemistry, Yale University, New Haven, CT 06520, USA, and [c]Howard Hughes Medical Institute, Yale University, New Haven, CT 06520, USA

Correspondence e-mail: jimin.wang@yale.edu

Received 3 September 2013
Accepted 27 November 2013

PDB reference: G-segment invertase–DNA complex, 4m6f

Crystals of the G-segment invertase in complex with a 37-base-pair asymmetric DNA duplex substrate had an unusually high solvent content of 88% and diffracted to a maximal resolution of about 5.0 Å. These crystals exhibited a high degree of non-isomorphism and anisotropy, which presented a serious challenge for structure determination by isomorphous replacement. Here, a procedure of cross-crystal averaging is described that uses large non-isomorphous crystallographic data with *a priori* information of an approximate molecular boundary as determined from a minimal amount of experimental phase information. Using this procedure, high-quality experimental phases were obtained that have enabled it to be shown that the conformation of the bound substrate DNA duplex significantly differs from those of substrates bound in other serine recombinase–DNA complexes.

1. Symbols and abbreviations

F_P: structure factors of parent native protein or macromolecular crystals.
F_{PH}: structure factors of derivative crystals containing heavy atoms.
F_H: structure factors of heavy-atom substructures.
ρ_{Mol}: electron density of molecules within the boundary of one macromolecular assembly in one crystallographic asymmetric unit.
R_i: rotation and translation matrices of molecules for the ith non-isomorphous native data set, with $i = 1, 2, 3, \ldots N$.
R_{iso}: amplitude-differences R factor between isomorphous pairs of data sets or between non-isomorphous pairs.
FT: Fourier transformation function.
$\sum_{symm}$: summation over all crystallographic symmetry operations.
Q: minimization targets.
Gin: G-segment invertase.
HA: heavy atom(s).
IR, SIR, MIR: isomorphous replacement, single and multiple isomorphous replacement methods.
SAD, MAD: single anomalous dispersive, multi-wavelength anomalous dispersive diffraction methods.
NCS: noncrystallographic symmetry.

2. Background and theoretical considerations of cross-crystal averaging

In the early days of macromolecular crystallography, two mathematical relationships were proposed to potentially solve

the phase problem: shrinking-and-swelling and isomorphous replacement (IR) (Law, 1973; Sayre, 2002). In the swelling-and-shrinking method (1*a* and 1*b*), the constant in the equation is the structure of the macromolecule, which can occupy slightly different positions varying from one crystal to the next. This method relies on non-isomorphous changes to the transform of the crystal that may be induced by chemical or physical means. This change allows alternative samplings in reciprocal space of the same structure. In the IR method (2*a* and 2*b*), the constant of the equation is again the structure of the macromolecule, but changes are recorded to the transform of the crystal by the addition of heavy atoms (HA) and the macromolecule does not change position in the crystals. This addition directly induces measurable changes to the structure factors, allowing us to use difference Patterson methods or other direct methods to determine the positions of the HA.

$$\mathrm{FT}\left[\sum\nolimits_{\mathrm{symm}} R_i(\rho_{\mathrm{Mol}})\right] = F_{\mathrm{P}i},\ i = 1, 2, 3, \ldots, N, \qquad (1a)$$

$$Q = \sum_{i=1}^{N}\left\{\left|FT\left[\sum\nolimits_{\mathrm{symm}} R_i(\rho_{\mathrm{Mol}})\right] - F_{\mathrm{P}i}\right|\right\}^2, \qquad (1b)$$

$$F_{\mathrm{PH}} = F_{\mathrm{P}} + F_{\mathrm{H}}, \qquad (2a)$$

$$Q = (|F_{\mathrm{PH}}| - |F_{\mathrm{P}} + F_{H}|)^2. \qquad (2b)$$

The first step in utilizing the shrinking-and-swelling method as defined in (1) is the determination of the position and the orientation of a reference molecular envelope or solvent boundary in one asymmetric unit as well as in one unit cell of the reference crystal. Once the reference envelope has been properly positioned, the next step is to calculate and refine its re-orientation matrices (R_i) in other non-isomorphous crystals using diffraction data. Although some progress has been made over the years in the direct determination of such a molecular envelope (Carter *et al.*, 1990; Lunin *et al.*, 2000), an explicit method for solving this equation has not yet been developed. It is also unclear how much initial phase information is needed for this equation to have a unique solution besides the information of the initial envelope.

The initial step in the single isomorphous replacement (SIR) method (2) is the determination of the HA positions that have been introduced into the target crystal (Blundell & Johnson, 1976). With the known HA structure, there are two possible solutions to (2) for each non-centric reflection, one corresponding to the correct phase (α_{P}) and the other to the incorrect phase angle ($\alpha_{\mathrm{H}} - \alpha_{\mathrm{P}}$), which is mirror-symmetry-related to the correct solution through the HA structure factors, α_{H}. In some cases, multiple HA derivatives may be required to resolve the phase ambiguity (Blundell & Johnson, 1976). More often, this ambiguity can be resolved through density-modification procedures, in which an approximate molecular envelope is determined from a set of partially correct phase angles (Wang, 1985). An isomorphous requirement for IR methods to succeed is that any amplitude changes owing to the repositioning of the macromolecule in the crystals must be relatively small compared with the changes introduced by the added HA. Otherwise, the IR methods are not effective, and the single anomalous dispersion (SAD) or multi-wavelength anomalous dispersion (MAD) methods are preferred (Hendrickson & Ogata, 1997). There are also two possible solutions of the SAD phase equation for each non-centric reflection, whose ambiguity can be resolved by MAD methods. An advantage of anomalous scattering approaches is the elimination of the requirement for isomorphism. A disadvantage is the requirement for high-resolution diffracting crystals so that relatively small anomalous/dispersive signals can be accurately measured. Because expected isomorphous differences are much larger than dispersive differences, IR methods remain preferred over SAD/MAD methods for structure determination of crystals that diffract poorly to low and very low resolutions.

As we have encountered in this study, the phase determination of a low-resolution structure presents many general challenges. Firstly, structure refinement of atomic models is not effective for phase determination, because these structures have more parameters to be fitted than the number of measurable reflections (a so-called under-determined system). Secondly, these crystals often exhibit extreme fragility during chemical and mechanical manipulation. The addition of HA compounds to crystal-stabilization solutions can introduce severe non-isomorphous changes (large R_{iso} with the parent native data) even when the HA does not bind to the macromolecules in the crystals as intended. Thirdly, the intensities of reflections are relatively weak and it is difficult to accurately measure the small changes in the structure factors induced by the added HA, resulting in inaccurate experimental phases and consequently uninterpretable maps from most experimental methods such as SIR, MIR, SAD or MAD. Moreover, when the macromolecule is reoriented from one crystal to the next, the experimental phases from independent SIR or SAD sources from different crystals cannot be simply combined using conventional phase-combination procedures. In this case, there is clearly an advantage in using the relationship in (1) for phase combination from a group of otherwise seemingly not-so-useful non-isomorphous native data sets as described in this study. Only when the model-independent experimental phases are accurately determined can they be effectively included as additional observations in structure refinement. Finally, these crystals typically have large solvent contents and may suffer from diffraction anisotropy problems owing to uneven stability in crystal packing in different directions.

In this study, we report the crystal structure of a G-segment invertase (Gin)–DNA complex in space group $P6_422$ along with unique methods for structure determination. Initial experimental phases were determined at 7.5 Å resolution using SIR/SAD, followed by phase extension to the maximal resolution limit of 5.0 Å using multi-crystal domain averaging as defined by the relationship in (1). The success of this procedure is a major achievement of this study and provides evidence that additional phase information can be extracted from nontraditional noncrystallographic symmetry (NCS) among non-isomorphous crystals even with only one copy of the complex in the asymmetric unit. Fortunately, this crystal

Table 1
Data-processing statistics as a function of resolution.

Res., resolution. % Com, completeness as a percentage. Redund., redundancy.

Native 1					Native 2					Ammonium osmate				
Res. (Å)	R_m† (%)	$I/\sigma(I)$	% Com.	Redund.	Res. (Å)	R_m† (%)	$I/\sigma(I)$	% Com.	Redund.	Res. (Å)	R_m† (%)	$I/\sigma(I)$	% Com.	Redund.
12.0	3.9	49.09	85.3	12.2	22.0	5.1	14.47	47.2	3.8	22.0	5.3	22.03	60.1	4.3
10.0	4.7	42.80	95.1	13.2	15.0	6.9	16.60	96.6	4.6	15.0	5.0	20.01	100	4.4
8.0	7.7	29.36	94.9	13.8	12.0	7.9	13.89	97.8	4.7	12.0	7.2	15.56	100	4.4
7.5	13.6	18.09	95.6	13.8	10.0	11.0	11.55	97.9	5.1	10.0	9.7	11.23	99.8	4.4
7.0	28.5	9.62	96.1	13.2	9.00	19.7	7.66	96.7	5.5	9.00	20.9	6.22	100	4.5
6.5	68.4	4.30	96.8	13.0	8.00	21.1	6.91	99.1	5.4	8.50	22.9	5.46	100	4.5
6.0	138.7	2.21	95.4	12.3	8.25	30.9	5.33	99.3	5.6	8.25	29.2	4.41	100	4.5
5.8	157.0	2.01	96.0	11.8	8.00	47.5	3.50	99.3	5.8	8.00	52.6	2.56	100	4.6
5.5	268.5	1.08	96.2	11.3	7.50	53.1	3.15	98.6	5.7	7.75	59.0	2.32	100	4.6
5.0	527.1	0.55	96.8	10.6	7.00	112.0	1.49	97.9	5.5	7.50	56.1	2.36	99.4	4.4
Total	11.0	12.35	95.1	12.2	Total	11.7	7.56	95.6	5.3	Total	8.5	8.68	98.3	4.4

† $R_m = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, merging statistics for all symmetry mates.

form exhibits internal 8.3-fold oversampling[1] (Miao *et al.*, 1998). It was possible to further increase the oversampling factor through rigid-domain averaging upon the inclusion of four non-isomorphous native data sets as well as a second Gin (apo Gin) crystal structure determined recently using SIR/SAD methods (Ritacco *et al.*, 2013). The rigid domains of Gin are defined as follows: the catalytic core domain (Core), helix E (αE) and the DNA-binding domain (DBD) with bound DNA (DBD+D). In this report, we describe the essential steps in the structure determination of the Gin–DNA complex (see Supplementary Fig. S1[2] for a computational flowchart) as well as the architecture of the complex with potential biological implications.

3. Preparation of the Gin–DNA complex, crystallization, X-ray diffraction and data processing

The wild-type construct of Gin was expressed and purified as described previously for a mutant of Gin (Ritacco *et al.*, 2013). The final purification step was modified such that the Gin protein was exchanged into a buffer consisting of 150 m*M* NaCl, 2 m*M* $MgCl_2$, 10 m*M* Tris–HCl pH 7.5. Oligonucleotides for the complementary DNA duplex of 37 base pairs corresponding to IR-R (Mertens *et al.*, 1988; Fig. 1) were synthesized by Integrated DNA Technologies (Skokie, Illinois, USA). The oligonucleotides were resuspended and annealed using a slow-cooling procedure in a buffer consisting of 50 m*M* NaCl, 10 m*M* Tris–HCl pH 7.5.

The oligonucleotides were mixed in a 2:1 molar ratio of DNA to dimers of Gin and incubated at room temperature for 5 min prior to crystallization using the hanging-drop vapor-diffusion method. Hexagonal thin plate-like crystals grew in less than 12 h after mixing the complex with an equal volume of reservoir solution consisting of 3–5% saturated ammonium sulfate solution, 10 m*M* $MgSO_4$, 20% ethylene glycol (EG), 100 m*M* MES buffer pH 6.5 at 16°C. Crystals of the Gin–DNA complex were transferred into a Gin crystal stabilization solution (GCSS) consisting of 20% saturated ammonium sulfate solution, 10 m*M* $MgSO_4$, 25% EG, 100 m*M* MES pH 6.5 before flash-cooling in liquid nitrogen. For preparation of heavy-atom (HA) derivatives, HA compounds were added to the GCSS solution at the following concentrations: 50 µ*M* to 1 m*M* for mercury or platinum, 1 m*M* to 5 m*M* for osmium, strontium, tungsten or iridium, and 50 µ*M* to 10 m*M* for tantalum bromide clusters.

Native crystals of the Gin–DNA complex diffracted anisotropically to a maximal resolution of 5.0 Å in space group $P6_422$ with unit-cell parameters $a = b = 119.8$, $c = 343.4$ Å, $\alpha = \beta = 90.0$, $\gamma = 120.0°$ (Table 1). The asymmetric unit of this crystal is composed of half of the asymmetric Gin–DNA complex dimer, which forms an averaged dimer by crystallographic symmetry. This crystal has a solvent content of 88%. The longest unit-cell axis (c = 343.4 Å) was perpendicular to

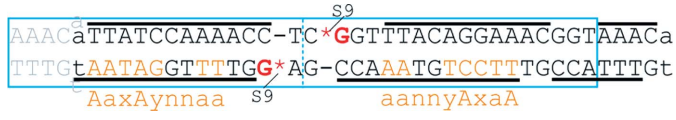

Figure 1
Asymmetric design of the substrate DNA duplex and its packing in the unit cell. The 37 bp DNA duplex corresponding to the *gix*-IR-R sequence used in the crystallization of the Gin–DNA complex is shown in dark colors and starts with the lowercase a, which is proposed to be splayed out from the duplex, and ends with AAAC. The first four gray-shaded nucleotides AAAC are at the end of the preceding end-to-end stacked duplex. A cyan box marks the sequence covered by one length of the unit-cell edge and the midpoint is indicated by a vertical dotted line. The phosphor-serine (S9) linkage after cleavage is indicated by an asterisk in the second guanine nucleotide (G) away from the pseudo-dyad (dotted line). The consensus Gin-binding sequence (see text) is in orange and the Gin-protected sequence is indicated by black bars.

[1] 88% of the unit-cell volume in this crystal is occupied by solvent, which requires only a single parameter to describe a constant density value once the molecular envelope is known. This leaves only 12% of the unit-cell volume within the molecular envelope to be determined. This property permits a sufficiently large number of observations to describe the features of the structure in the relatively small volume at any given resolution, known as oversampling. The oversampling factor of 8.3 is the ratio of the volume of a hypothetical crystal that has no solvent to the actual volume occupied by the macromolecules in this crystal. In other words, the Bragg reflections in reciprocal space of this crystal are about 8.3 times more densely distributed than the corresponding hypothetical crystal with no solvent

[2] Supporting information has been deposited in the IUCr electronic archive (Reference: DZ5308).

the plate axis of the crystals, resulting in severe reflection-overlapping problems when using conventional methods of data collection. To alleviate these problems, special angled loops were constructed in collaboration with MiTeGen (Ithaca, New York, USA). These loops aligned the long axis parallel to the oscillation axis during data collection. Native and heavy-atom data were collected on the Advanced Photon Source beamline 24ID-C, Chicago, Illinois, USA and on the National Synchrotron Light Source beamlines X-25 and X-29 at Brookhaven National Laboratory, Brookhaven, New York, USA. Data were processed using both the *HKL*-2000 and the *XDS* suites (Otwinowski & Minor, 1997; Kabsch, 2010). Experimental maps using data processed with the *XDS* suite in this study were clearly better than those from the data processed using *HKL*-2000. Figures were rendered using *RIBBONS* (Carson, 1997) and *PyMOL* (DeLano, 2002).

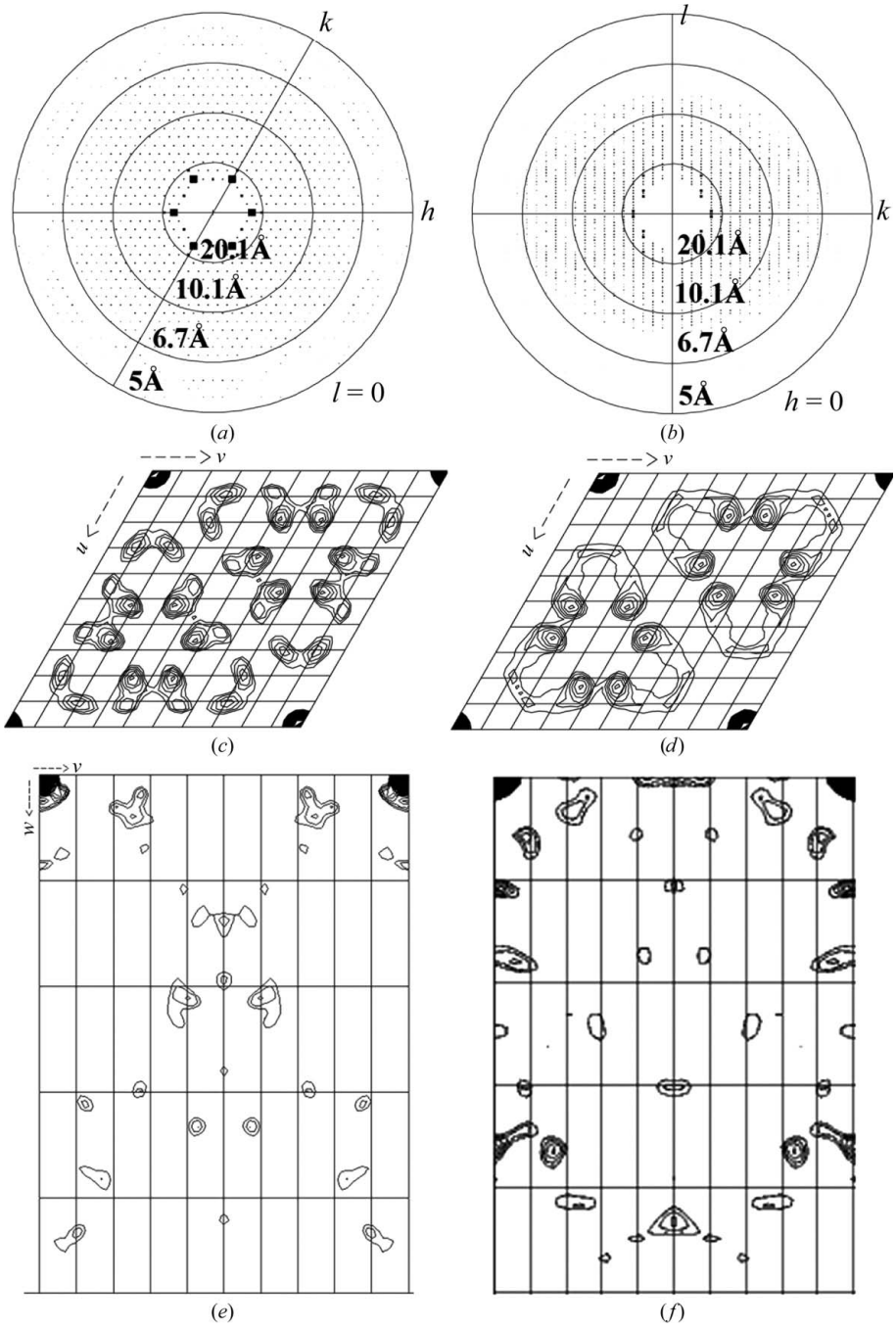

Figure 2
Data anisotropy and its effects on anomalous difference Patterson maps. (*a*, *b*) Pseudo-precession photos of the (*hk*0) and the (0*kl*) zones from a native data set (Native 1). The displayed resolution circles are at 5.0, 6.7, 10.1 and 20.1 Å, respectively. (*c*) Anomalous difference Patterson map at Harker section at 9.0 Å resolution of the ammonium osmate derivative, which provided an interpretable electron-density map from SIR/SAD phasing. (*d*) Anomalous difference Patterson map at 9.0 Å resolution of a di-platinum PIP derivative, which failed to provide useful experimental maps. (*e*) Anomalous difference Patterson maps of the Os derivative in the orthogonal Harker section. (*f*) Isomorphous difference Patterson maps between the Os-derivative and Native 2 data sets. Each section is contoured with 0.5σ increments starting at 1.5σ.

4. Severe non-isomorphism, diffraction anisotropy and effective resolution

The crystals of the Gin–DNA complex exhibited strong diffraction anisotropy, with diffraction in the *l* direction being poorer than in the other two directions (Figs. 2*a* and 2*b*). The highest resolution in the *h* and *k* directions for the best native data set (Native 1) was about 5 Å, but in the *l* direction the resolution was around 7 Å by an $I/\sigma(I)$ cutoff criterion or visual inspection (Fig. 2*a*). The anisotropic *B*-factor parameters had a spread of 85.5 Å^2, with values of −28.5, −28.5 and 57.0 Å^2 in the *h*, *k* and *l* directions, respectively. We included all measurable reflections processed at the nominal 5 Å resolution for both cross-crystal averaging and structure refinement in this study (Table 1).

Data were processed in two passes for different stages of structure determination: one with the highest-quality data using the $I/\sigma(I) = 2$ cutoff criterion in the highest resolution shell and the other with the highest-resolution data using $I/\sigma(I) = 0.5$. The highest-quality data were used in initial SIR/SAD phasing for both the HA derivative and the native data (Native 2; Table 1). The highest-resolution data were used for cross-crystal averaging and structure refinement, which increased the amount of measurable data by 60% when the resolution was extended from 7 to 5 Å, as recommended previously (Wang & Boisvert, 2003; Wang, 2010). Strong diffraction anisotropy severely skewed the data-processing statistics as well

as the structure-refinement statistics, especially when the anisotropy was not corrected or was inadequately corrected using an incomplete, unrefined atomic model. Nevertheless, the effective resolution of this structure is much higher than 7 Å using the methods defined by DeLaBarre & Brunger (2006). A measure for the success of an overall structural interpretation at such low resolution is whether model-phased maps with reasonably refinement statistics can reproduce all of the major features of the experimental maps (Fig. 3).

Soaking of most of the HA compounds into the crystals greatly reduced the diffraction (Pt, Hg, Sr, Ir and Os compounds) or completely destroyed it (Ta_6Br_{14} or W_{12} cluster compounds). Co-crystallization of a pre-derivatized Gin–DNA complex with a minimal amount of HA compound also did not improve the diffraction quality. For most derivatives, the diffraction resolution was near 8 Å in the h and k directions, but only 10–11 Å in the l direction, which resulted in an averaged resolution of about 9 Å. Even at such reduced resolution, it was possible to determine the partial HA locations in the derivatives based on strong Harker peaks in the w = 1/3 section using the anomalous difference Patterson method (Fig. 2 and Supplementary Fig. S2). Similar strong peaks were also observed for ethylmercury phosphate (EMP), di-μ-iodobis-(ethylenediamine)-diplatinum(II) nitrate (PIP) or ammonium osmate (Os). From these peaks, we obtained partial HA solutions in the x and y coordinates. However, diffraction anisotropy in the crystals did not result in any discernible features in the z direction, so the z coordinates remain undetermined.

Table 2
SHELX phasing statistics for Os *versus* Native 2 SIR/SAD.

Values are as defined and provided by *SHELXE* (Sheldrick, 2008).

	Right-hand *xyz*	Left-hand *xyz*
SHELXE pseudo-free CC (%)	51	24
SHELXE contrast	1.2	0.3

5. Experimental phasing using SIRAS and density modification

It was possible to determine the missing z coordinates of the Os positions to generate a complete HA structure solution and

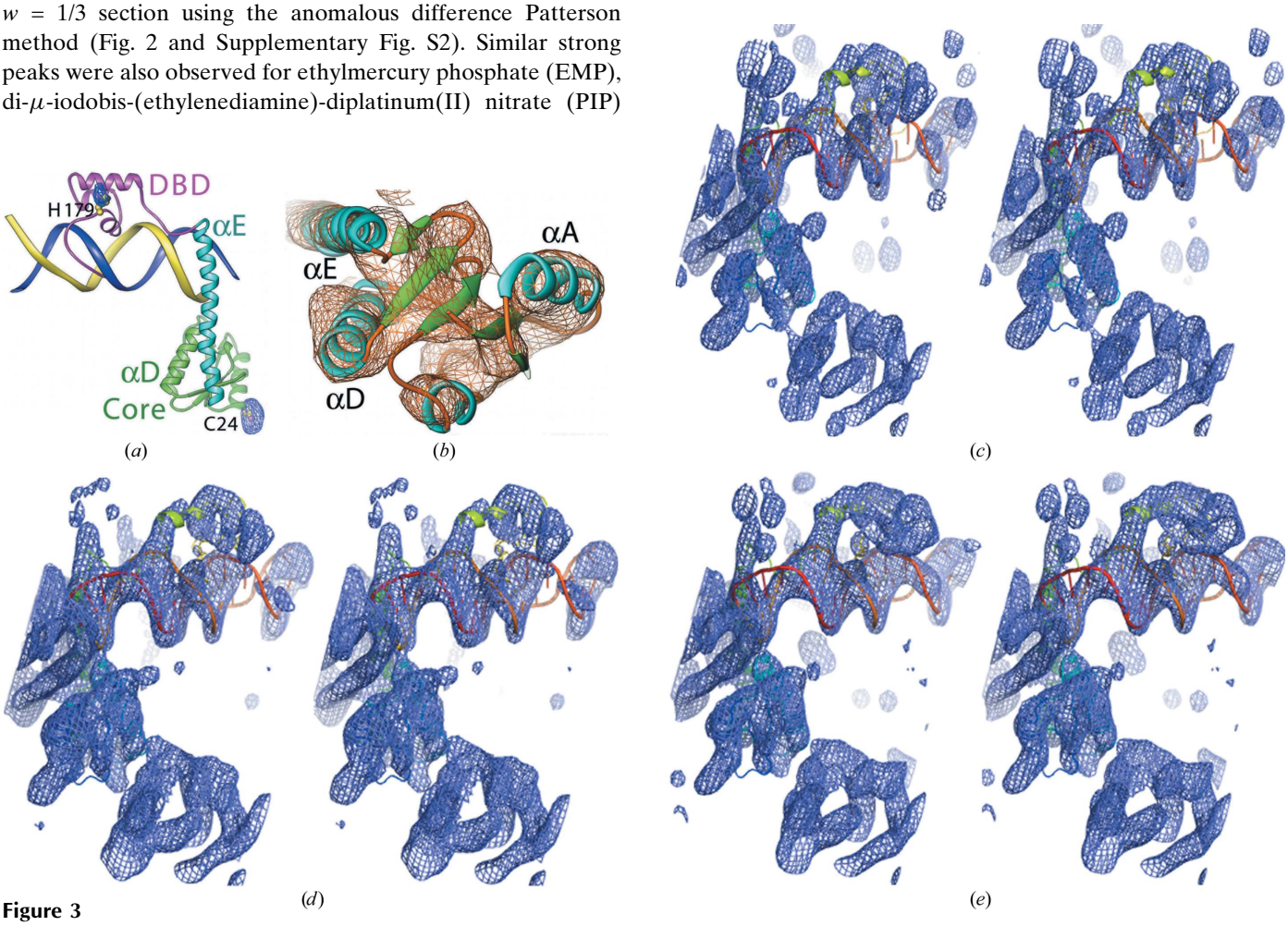

Figure 3
Experimental electron-density maps of the Gin–DNA complex. (*a*) An anomalous difference Fourier map of the ammonium osmate derivative with experimental phases contoured at 5σ is superimposed onto the refined model of Gin. The Gin model is colored by domain: the catalytic core domain (Core; Leu2–Ser97) is shown in green, the E helix (αE; Ser9–Arg135) in cyan and the DNA-binding domain (DBD; Ile136–Asp189) in magenta. (*b*) An experimental map contoured at 0.9σ after two-crystal two-domain averaging is superimposed onto the NTD of the final model. (*c*) The initial SIR/SAD maps after density modification superimposed onto the refined DBD+D model shown as a stereo diagram. (*d*) Experimental maps after two-crystal two-domain averaging (step one) using conventional cross-crystal averaging. (*e*). Experimental maps after five-crystal three-domain averaging (step two) using non-isomorphous cross-crystal averaging. The electron-density maps in (*c*–*e*) show the progression of the improvement and are contoured at 1.0σ.

initial experimental maps using the SIR/SAD method as implemented in the *SHELX* suite (Table 2; Schneider & Sheldrick, 2002; Sheldrick, 2008, 2010; Pape & Schneider, 2004). Although there was no information about the z coordinates of the Os positions in the anomalous signal (Fig. 2), there was sufficient information in the most isomorphous derivative–native pair, with a smallest isomorphous difference $R_{\rm iso}$ of 26.6% (Supporting Information §S1). Prior to this solution, extensive efforts were unsuccessfully made to solve the structure using molecular replacement or to solve the HA structures using the SAD method (Supporting Information §S2). Without molecular-replacement solutions, it is impossible to carry out conventional cross-crystal averaging (Saper *et al.*, 1991; Freymann *et al.*, 1990; Li *et al.*, 2005). We started this HA solution using the highest-quality data possible to ensure that the poor quality of the data and the presence of severe anisotropy and non-isomorphism did not adversely affect the scoring and selection of the HA solutions. Among the hundreds of possible native–derivative pairs, only one native–derivative pair between Native 2 (at 7 Å resolution) and a crystal of the Gin–DNA complex soaked with 1 m*M* ammonium osmate provided the complete coordinates of the Os atoms (Table 2).

The initial Os-derived experimental maps revealed partially interpretable features for both the Core and αE of Gin, but barely recognizable features for the DBD+D. However, we were surprised to find in this map that (i) Gin is in a dimeric complex, not in a tetrameric complex, in this crystal as in the structure of apo Gin (Ritacco *et al.*, 2013), explaining why molecular replacement failed when searching for half or full tetramers (Supplementary Table S1), (ii) the DNA duplex is straight, which has never been seen before in any previous DNA–serine recombinase complex (Li *et al.*, 2005), and (iii) there is only one half of the asymmetric Gin–DNA complex in one asymmetric unit. It was critically important to improve the experimental maps to ensure the validity of these unusual findings at such relatively low resolution. We attempted the following conventional methods for phase improvement: (i) cross-crystal averaging using Native 2 (Saper *et al.*, 1991; Freymann *et al.*, 1990; Li *et al.*, 2005; Li & Li, 2011) between the Core domains of this crystal and the previous unliganded Gin structure (Ritacco *et al.*, 2013), (ii) phase combination using partial model phases from the Core domain and (iii) phase extension using the solvent-flattening method (Wang, 1985) after the initial experimental phases were transferred to Native 1. However, none of these approaches sufficiently improved the experimental maps for unbiased interpretation of the DNA duplex structure.

6. Non-isomorphous cross-crystal three-domain averaging

We then explored the possibility of whether the phase relationship defined in (1) could help us to extract additional phase information from the large number of non-isomorphous native data sets from many different crystals accumulated during the course of this study. Some of these data sets were owing to reorientation of the complex in the crystal and others were owing to the presence of HA compounds in the solvent area so that the solvent contrast was altered (Supporting Information §S1, Supplementary Fig. S3). This non-isomorphous cross-crystal averaging differs from conventional NCS averaging (Wang *et al.*, 1998; Abrescia *et al.*, 2011) or multi-data-set averaging for the improvement of the data quality for multiple isomorphous data sets (Su *et al.*, 2010; Liu *et al.*, 2011). An advantage of this averaging is that we knew the approximate locations of the individual domains to be averaged in each crystal relative to the reference crystal. These locations were then continuously refined during averaging.

We implemented non-isomorphous cross-crystal averaging in two steps (Supplementary Fig. S1) using *dmmulti* within the *CCP*4 package (Cowtan, 1994; Winn *et al.*, 2011). In the first step we carried out conventional cross-crystal averaging as mentioned above between the experimental maps of the Gin–DNA complex and the apo Gin structure determined previously (Ritacco *et al.*, 2013). Between the two crystals, we averaged the Core and αE domains independently with initial matrices derived from a least-squares superposition of approximately placed domains in this crystal. An approximate envelope for the DBD+D was also included in this step to prevent incorrect flattening of this region as solvent, which marginally improved the experimental maps relative to an attempt without it. This averaging resulted in qualitatively improved experimental maps for both the apo Gin and Gin–DNA complex structures (Figs. 3*b* and 3*c* and Supplementary Fig. S4). The most noticeable improvement occurred in the β-sheets of the core. However, this step of averaging did not substantially improve the electron densities in the DBD+D region with or without an approximate envelope for it and we could not place a model of the DBD+D with confidence at this stage.

In the second step of averaging, we included four different native data sets for this complex (Native 1, Native 2, Native 3 and Native 4) as well as the previous apo Gin structure (Ritacco *et al.*, 2013). These native data sets differ from Native 1 used in the first step of averaging by as much as 32.7% (Supplementary Table S3) and from each other by as much as 37.8% (for example, between Native 3 and Native 4). During the averaging, we provided an envelope for each of the three domains (Core, αE and DBD+D) and initial matrices obtained in the first step for the first two domains. For the newly added native data sets without initial experimental phases (Native 1, Native 3 and Native 4), we iteratively refined domain matrices starting with an identity matrix during the averaging. As a result, the electron densities improved greatly for the DBD+D region. At the stage, we placed the straight DNA duplex along with the DBD in the complex with confidence and for structure refinement (Fig. 3, Supplementary Fig. S4).

7. Interpretation of experimental maps and structure refinement

After manually fitting individual domains into the best cross-crystal averaged maps, the structure was refined against the

Table 3
Refinement statistics.

Refinement	*REFMAC*	*CNS* with DEN
Resolution (Å)	114–5.0	50.0–5.0
No. of reflections in working set	5975	5848
No. of atoms	2194	2197
$R_{working}$† (%)	26.1	22.9
R_{free}‡ (%)	33.6	28.8
O2P ratio§	0.68	0.66
R.m.s.d., bond lengths¶ (Å)	0.011	0.003
R.m.s.d., bond angles¶ (°)	1.70	0.94
Average *B* factors (Å^2)		
Overall	316	335
B_{11}	8.8	16.3
B_{22}	8.8	16.3
B_{33}	−13.2	−32.6
NTD	215	315
DNA	346	355
DBD	450	405
Ramachandran		
Most favored and allowed (%)	93.9	97.6
Disallowed (%)	6.1	2.4
PDB code	4m6f	

† $R_{working} = \sum_{hkl} \big||F_{obs}| - |F_{calc}|\big| / \sum_{hkl} |F_{obs}|$. ‡ R_{free} is the cross-validation *R* factor for a randomly selected ∼5% of the data. § The observation-to-parameter ratio, defined as the ratio between the unique number of observations for refinement and the number of variables, which is four times of the number of atoms. ¶ Root-mean-square deviations from ideal values.

anisotropy-corrected Native 1 data at 5.0 Å resolution using *REFMAC*5 (McCoy *et al.*, 2007; Murshudov *et al.*, 2011; Winn *et al.*, 2011), resulting in crystallographic and cross-validation $R_{working}$ and R_{free} of 26.2 and 33.7%, respectively (Table 3, Supplementary Table S2 and Supporting Information §S3). We further refined the structure using deformable elastic network (DEN) restraints in *CNS* (Brünger *et al.*, 1998; Schröder *et al.*, 2007), resulting in a further improved $R_{working}$ and R_{free} of 22.9 and 28.8%, respectively (Supporting Information §S3, Supplementary Fig. S5).

In addition to the refinement statistics (Table 3), two other lines of evidence support the validity of our interpretation of the experimental maps for an overall architecture of the Gin–DNA complex structure. Firstly, despite the relatively low resolution of 5 Å, the oversampling resulted in a reasonably high observation-to-parameter ratio of 0.66 for structure refinement (5848 independent reflections for 2197 atoms). Secondly, a straightforward assessment of electron-density interpretations is that our model-phased maps such as $2F_o - F_c$ maps indeed reproduced all of the major features of the experimental maps (Supplementary Fig. S6). Given the relatively low resolution, this structure does not provide independent information on detailed interactions beyond the overall architecture (Table 3, Supplementary Fig. S6). Nonetheless, the conformation of the DNA duplex bound to this complex is clearly different from that of the dimeric $\gamma\delta$ resolvase–DNA complex (Yang & Steitz, 1995; Supplementary Fig. S4). In this Gin–DNA complex, the DNA is straight as opposed to bent by 60° as in the $\gamma\delta$ resolvase complex (Fig. 4, Supplementary Fig. S7).

In the Gin–DNA complex, the 36 bp DNA duplexes in three complete helical turns appear to stack end-to-end across the cell edges (*a*, *b* and the diagonal *a* + *b*; Fig. 5, Supplementary Fig, S6). The *gix*-site substrate DNA used in this study is 37 bp long so that one base pair at one end has to splay out (Fig. 1). In our design, there are 14 bp to the left and 23 bp to the right of the pseudo-dyad relating to two cleavage sites. This pseudo-dyad lies on the midpoint of each cell edge. In this arrangement, the duplex is 4 bp too short on the left side of the dyad and 5 bp too long on the right side of the dyad, with the extra bases crossing into an adjacent unit cell. In the experimental maps (Fig. 3, Supplementary Fig. S6), there is a partial gap in the phosphate backbone of the DNA duplex about 4 bp from one cell edge. This gap is likely to correspond to the stacked ends of two duplexes where the last base pair at one end is splayed. Even so, this end-to-end stacking should remain asymmetric with a pseudo-NCS dyad. In this case, the space group would be $P3_221$, $P3_212$ or $P6_4$, which are lower symmetry space groups than the observed $P6_422$.

To explain the discrepancy in symmetry, we ran twinning-statistics tests but found no evidence of perfect merohedral twinning (Table 1, Supplementary Fig. S8). Thus, the observed $P6_422$ symmetry is truly crystallographic. We speculate that

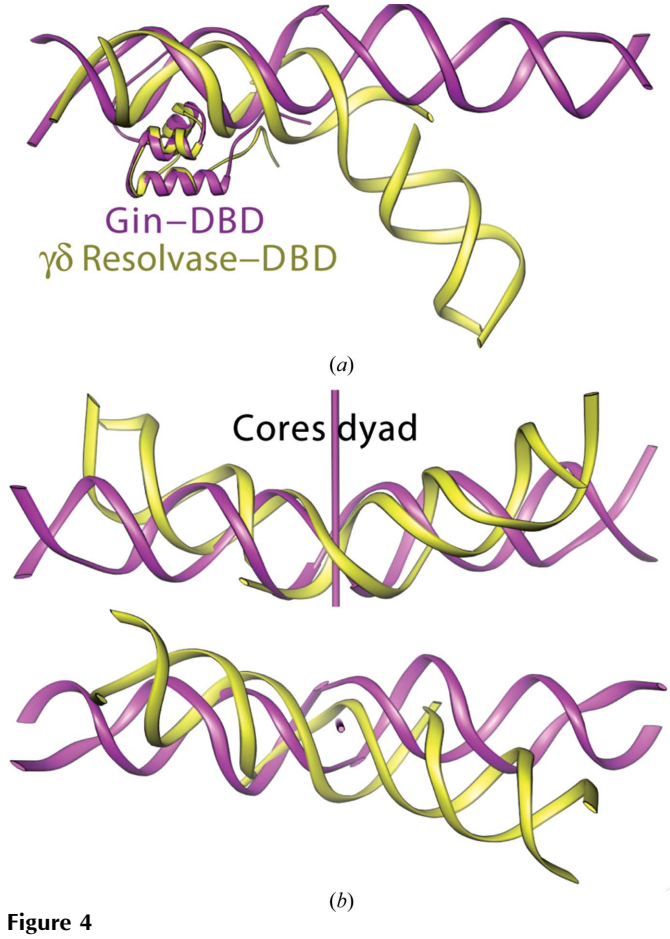

Figure 4
Conformation of the DNA duplex in the Gin–DNA complex. (*a*) Comparison of the DNA duplex of the Gin–DNA structure (magenta) with the DNA duplex of the $\gamma\delta$ resolvase–DNA complex (yellow) after alignment of their DNA-binding domains (not shown). (*b*) Two orthogonal views of the DNA duplexes from (*a*) as a comparison using the alignment of the dyad relating to the catalytic Core domains of each structure.

this higher symmetry in the observed data may result from static disorder or multiple conformations. For example, in one line of unit cells along the helical axis of the DNA duplex the infinite DNA duplex is oriented from right to left in the 5′-to-3′ direction for the top strand, but in the next line of unit cells it is oriented in the opposite orientation. Coherent interference between these two lines of unit cells averages them and places the DNA duplexes in two opposite orientations in the averaged unit cell. This averaging generates a new dyad every 18 bp. Similar unit-cell averaging could also occur between any pairs of 6_4 screw symmetry-related DNA duplexes in the lattice. Interestingly, Gin–DNA forms an unusual tetramer between two $P6_4$-symmetry related mates (Fig. 5). Every other layer of Gin–DNA dimers within the tetramer is rotated by 60° according to the $P6_4$ screw axis (Fig. 5*b*). In all of these possible unit-cell averagings the DBD does not perfectly align, so that this domain appears to have a much large apparent *B* factor relative to the Core domain of Gin (Supporting Information §S4) because the lattice contacts in the *a*, *b* and *a* + *b* directions are made mainly by the NTDs of Gin.

With our interpretation of the experimental map to give an overall architecture of the Gin–DNA complex, the DNA recognition by Gin aligns well with the consensus motif derived from Hin, γδ resolvase and all enterobacterial invertases (Feng *et al.*, 1994; Yang & Steitz, 1995). According to the HinDNA complex model, the recognition motif is 5′-*aannyA*x*a*A-3′ from residues 5 to 13 (numbering starting from the dyad), where *a* is either A or T for forming A–T base pairs, *y* is purine, *n* is any base and *x* is restricted not to be C. Whether the obtained DNA conformation in this complex is more catalytically relevant than previous complexes remains to be determined (Supporting Information §S5, Supplementary Fig. S9). Clearly, a straight DNA duplex with twofold symmetry is more suitable for committed catalysis simultaneously on both strands than any nonsymmetric conformation. However, we cannot rule out the possibility that crystal packing has some influence on the observed conformation of the DNA duplex (Supporting Information §S4, Supplementary Fig. S10).

8. Concluding remarks and perspective

Of the many challenges during the structure determination of the Gin–DNA complex at low resolution, we found that non-isomorphism is the most prominent, resulting in many extraneous native data sets. Here, we have demonstrated that we can extract phase information from them through non-isomorphous multiple-crystal averaging. Despite their potential for phase retrieval, how to define a high-resolution mask (or molecular envelope) remains a focus for further methodological development. The fact that bulk solvent contributes strongly to both the amplitudes and the phases of very low resolution reflections provides a possibility that one day we may directly extract the shape information of bulk solvent and macromolecules in the lattice from the amplitudes of these

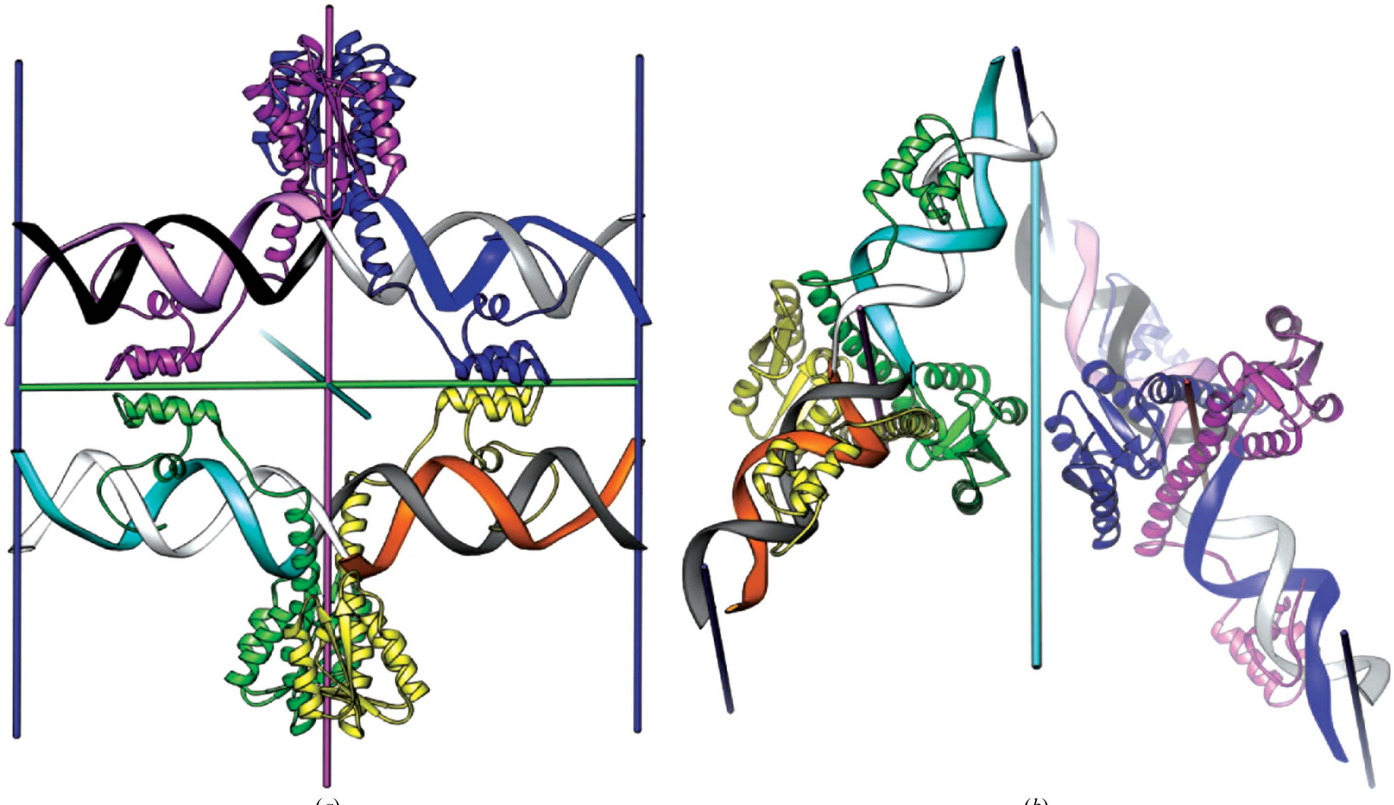

Figure 5
Packing of the tetrameric Gin–DNA complex in the crystals. (*a*) A tetrameric model is created by the crystallographic 222 dyad through contacts of DBDs. (*b*) The Core domain contacts between two dimers are related by a dyad, which is also related by 6_4 screw symmetry. Overall, the DNA duplexes are stacked to form an infinite lattice every 60° along the sixfold axis and provide a scaffold for the packing of the dimeric complex of Gin.

reflections alone. In fact, full phase retrieval has already been demonstrated to be possible for a three-dimensional non-crystalline binary image reconstruction when the amplitudes are theoretically sampled in Bragg spacing with an over-sampling ratio of greater than 2.6 (Miao *et al.*, 1998). The binary image corresponds to the shape information of the bulk solvent and macromolecules in the crystal, and the over-sampling ratio of 2.6 corresponds to macromolecular crystals with a solvent content of greater than 62%. Thus, the results of Miao and coworkers as well as our results presented here may have opened up a new unexploited avenue for phase retrieval starting with very low resolution reflections for the determination of macromolecular structures, and we may soon see a possible revival of the long-forgotten swelling-and-shrinking method. Because of our success in phase improvement using multiple non-isomorphous native data sets, we also predict that the simultaneous inclusion of severely non-isomorphous native data sets in structure refinement at low resolution could also have great potential for further improvement of the quality of refined models when properly implemented.

We acknowledge the support of this work by the National Institutes of Health (T32 GM007223, CJR; GM057510, TAS) and by the Steitz Center for Structural Biology, Gwangju Institute of Science and Technology, Republic of Korea. We thank Professor Peter Moore and Dr Richard Wing for critical reading of and comments on this manuscript.

References

Abrescia, N. G. A., Grimes, J. M., Oksanen, H. M., Bamford, J. K. H., Bamford, D. H. & Stuart, D. I. (2011). *Acta Cryst.* D**67**, 228–232.
Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*. New York: Academic Press.
Brünger, 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.
Carson, M. (1997). *Methods Enzymol.* **277**, 493–505.
Carter, C. W., Crumley, K. V., Coleman, D. E., Hage, F. & Bricogne, G. (1990). *Acta Cryst.* A**46**, 57–68.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
DeLaBarre, B. & Brunger, A. T. (2006). *Acta Cryst.* D**62**, 923–932.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Feng, J. A., Johnson, R. C. & Dickerson, R. E. (1994). *Science*, **263**, 248–355.
Freymann, D., Down, J., Carrington, M., Roditi, I., Turner, M. & Wiley, D. (1990). *J. Mol. Biol.* **216**, 141–160.
Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Law, J. (1973). *Soc. Stud. Sci.* **3**, 275–303.
Li, W., Kamtekar, S., Xiong, Y., Sarkis, G. J., Grindley, N. D. & Steitz, T. A. (2005). *Science*, **309**, 1210–1215.
Li, W. & Li, F. (2011). *Structure*, **19**, 155–161.
Liu, Z.-J., Chen, L., Wu, D., Ding, W., Zhang, H., Zhou, W., Fu, Z.-Q. & Wang, B.-C. (2011). *Acta Cryst.* A**67**, 544–549.
Lunin, V. Y., Lunina, N. L., Petrova, T. E., Skovoroda, T. P., Urzhumtsev, A. G. & Podjarny, A. D. (2000). *Acta Cryst.* D**56**, 1223–1232.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Mertens, G., Klippel, A., Fuss, H., Blöcker, H., Frank, R. & Kahmann, R. (1988). *EMBO J.* **7**, 1219–1227.
Miao, J., Sayre, D. & Chapman, H. N. (1998). *J. Opt. Soc. Am.* **15**, 1662.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 303–326.
Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.
Ritacco, C. J., Kamtekar, S., Wang, J. & Steitz, T. A. (2013). *Nucleic Acids Res.* **41**, 2673–2682.
Saper, M. A., Bjorkman, P. J. & Wiley, D. C. (1991). *J. Mol. Biol.* **219**, 277–319.
Sayre, D. (2002). *Struct. Chem.* **13**, 81–96.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Schröder, G. F., Brunger, A. T. & Levitt, M. (2007). *Structure*, **15**, 1630–1641.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Sheldrick, G. M. (2010). *Acta Cryst.* D**66**, 479–485.
Su, J., Li, Y., Shaw, N., Zhou, W., Zhang, M., Xu, H., Wang, B.-C. & Liu, Z.-J. (2010). *Protein Cell*, **1**, 453–458.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Wang, J. (2010). *Acta Cryst.* D**66**, 988–1000.
Wang, J. & Boisvert, D. C. (2003). *J. Mol. Biol.* **327**, 843–855.
Wang, J., Hartling, J. A. & Flanagan, J. M. (1998). *J. Struct. Biol.* **124**, 151–163.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Yang, W. & Steitz, T. A. (1995). *Cell*, **82**, 193–207.