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

Christopher J. Ritacco, Thomas A. Steitz, Jimin Wang

Online Supporting Information

Contents: 5 Sections, 3 Tables, 10 Figures, and 4 additional references.

S1. Isomorphous and non-isomorphous differences

There was a difficulty in properly scaling native-derivative pairs with each other due to the relatively low resolution compounded with the severe anisotropy. This resulted in variations of the calculated isomorphous differences depending on the scaling strategy and reference data set. Such variations by more than 6% were seen commonly in many derivative-native pairings. The isomorphous differences for the same osmate-native pair was reduced from 26.6% to 24.5% when the native data set was scaled to the derivative instead of the conventional derivative-to-native scaling. Among the potential derivatives, the osmate data set had the stronger signal in the *l* direction and the higher overall resolution around 7 Å than many other derivcatives (Table 1). In the native-to-derivative scaling, the actual resolution of this Os derivative might have been better than that of the corresponding native as indicated from the spread of relative anisotropic B-factors of (−26, −26, −90) $Å^2$. Large isomorphous differences may in part result from an inability of proper scaling due to severe anisotropy and inherent non-isomorphism among crystals, requiring accurate refinement of the relative anisotropic B-factors between the data sets. However, the refinement of the relative anisotropic B-factors at resolution of 7.5 to 10 Å was very unstable. Moreover, the unusually high solvent content could also have contributed to the difficulty in scaling, because the measured intensities would be highly modulated by variations of solvent contrast. This solvent-contrast variation cannot be removed through scaling procedures alone, and also contributes to the apparent non-isomorphism among crystals (Fig. S3).

S2. Systematic attempts to generate complete heavy atom coordinates

The Gin-DNA complex and the *apo* Gin (Ritacco et al., 2013) synaptic complex show striking similarities: (i) both were in hexagonal space groups, (ii) they had nearly identical hexagonal *a* and *b* unit cell dimensions, (iii) they had similar Harker peaks in difference Patterson maps for the same set of heavy atom derivatives (Fig. 2, Fig S2), and; (iv) the same cysteine residues were available for binding HA and remained solvent exposed even in the presence of DNA. From these observations, we assumed that the packing of two complexes in the hexagonal lattice would be similar, and that DNA duplex would expand the lattice in the *z* direction. However, the two crystals had the three key differences that favored successful determination of the previous structure by SIR/SAD (and/or by molecular

replacement): (i) the resolution of the diffraction pattern was much higher than that of this complex, (ii) there were much stronger anomalous signals along *z*-axis (*l* direction) and; (iii) the non-isomorphism was so small that every possible derivative-native pair produced interpretable SIR/SAD maps.

In one attempt, we started with the partial solution of the *x* and *y* coordinates of the heavy compound PIP calculated directly from peaks in anomalous difference Patterson maps (Fig. 2CD). We then systematically varied the *z*-coordinate with an increment of 0.01 units from 0 to 1 and refined the trial heavy atom structure in both the enantiomeric space groups $P6_422$ and $P6_222$ using various SAD phasing programs. However, all attempts failed to yield interpretable electron density maps or significant discriminatory statistics such as figure of merit. It is likely that that our failure to stably refine the z-coordinate was in part due to the poor quality of the diffraction in the *l* direction as well as low occupancy (see below).

Two Os binding sites were identified in the Os derivative. A strong site was located at C24 in the Core of Gin and a weak site was located at H179 in the DBD+D (Fig. 3A). The inclusion of the isomorphous native data was critical for obtaining the complete coordinates of the HA because anomalous signals alone failed (Fig. 2). In retrospect, we re-examined the occupancies of all HA after we obtained the best experimental phases for each derivative through non-isomorphous cross-crystal averaging. A surprising finding is that the occupancies of the PIP cluster compound were far less than 5%, even though the derivative had outstanding Harker peaks in the anomalous difference Patterson maps (Fig. 2D). These low occupancies explain in part why we initially failed to determine to the complete coordinates of the PIP HA sub-structure. The orientation of the Gin-DNA complex in the unit cell places the residue H179 at the lattice interface along the c-axis. If the binding to the H179 site caused lattice distortions, severe heterogeneity (anisotropy) along this direction would be introduced. According to our modeling, the binding of a PIP molecule, which is much larger than Os, to H179 would likely induce major distortions to the crystal lattice and destroy the crystal diffraction. Large differences between *Native* 1 data and the PIP derivative data (32%) and between *Native* 2 and the PIP derivative data (38%) were due to severe non-isomorphous differences, not due to high-occupancy PIP (Fig. 2) as we had falsely concluded on the basis of outstanding anomalous difference Patterson maps in early stages of the structure determination.

S3. Severe anisotropy impaired low-resolution structure refinement

In addition to problems with data processing, diffraction anisotropy had a severe impact on structure refinement. After manually fitting individual domains, namely, Core, E helix, and DNA duplex (without DBD), into the initial electron density maps derived from SIR/SAD, phase-restrained rigid-body refinement was carried out against *Native* 2 using REFMAC (Murshudov *et al.*, 1997). No matter how well the domains were fitted and refitted into the experimental maps, the crystallographic and cross-validation R-factors were always all above 70% at the start and end of refinement. The refinement of domains kept bringing them away from the appropriate places in the density as a result of failed,

divergent refinement; these statistics are poorer than a random placement of a completely unrelated structure (Srinivasan & Parthasarathy, 1976).

The reason for the unusually high R-factors might be that the refinement tried to simultaneously determine overall anisotropic B-factors of the model relative to the observed data as well as the level of bulk solvent. This is a difficult task for an incomplete and imperfect model as we had left out the DBD in the DNA duplex region. This unusual behavior of refinement may also explain why all molecular replacement attempts using various homology models failed (Table S1, S2). Anisotropy correction should result in a refined model that is more isotropic after restoring the intensities for reflections along the *l* direction. We note that anisotropic correction at such a low resolution may not be accurate, because the refined model against anisotropy-corrected data remains anisotropic. Moreover, the anisotropy correction transfers anisotropy from intensities I to associated standard deviations σ_I, but does not actually remove anisotropy, when σ_I was used as a weighting parameter in the maximal likelihood refinement. This in part explains that the refined model retains some anisotropy. Using the atomic model, the residual anisotropic B-factor spread was refined to (16.3, 16.3, −32.6) $Å^2$ between the calculated and observed structure factors and this value varied with the refinement cycle due to refinement instability at such low resolution. Finally, the Gin-DNA complex occupied only 12% of the unit cell volume, an error in determination of envelop such as the unknown DBD+D in the early stage could severely affect bulk solvent correction, which could also destabilize the structure refinement. The final solvent model has the mean density of 0.2 e/$Å^3$ and B-factor of 262 $Å^2$.

S4. Asymmetric distribution of domain B-factors and its relationship with anisotropy and crystal packing

In the Gin-DNA complex, distribution of domain-averaged B-factors appears to be correlated with the observed diffraction anisotropy even though individual atomic B-factors have not been reliably refined. The average B-factors for the Core, DNA duplex, and DBD were 315, 355, and 405 Å2, respectively, with an overall value of 335 Å2 for the entire structure. The lattice contacts in the *a* and *b* planes are largely between the Core (low B-factor domains) and contacts along *c* direction are largely between the DBD+D (high B-factor domains).

Overall packing in the lattice resembles beads (dimeric catalytic core domains) on strings (DNA duplexes) whereas beads are far away from each other. When the anisotropically-corrected, observed intensity of reflections were spherically averaged in finite resolution shells, we could actually see the transformation details of these beads with the corresponding radius of gyration as if they were in small-angle and wide-angle X-ray solution scattering (Fig. S10). Addition of heavy atoms changed the transformation as it changed the radius of gyration. Thus, the amplitudes of observed structure factors alone at low resolution could potentially reveal the underlying structural features in the lattice in this particular case due to a high degree of reciprocal space oversampling that results in statistically sufficient numbers of reflections in each finite resolution shells.

S5. Implications of the observed DNA conformation on DNA cleavage by Gin resolvase

A striking finding is that the substrate DNA duplex bound to each dimeric Gin in this Gin-DNA complex is straight. In contrast, the DNA duplex bound to the similar dimeric γδ resolvase is bent about 60° (Yang & Steitz, 1995). The bending in γδ resolvase occurs in two perpendicular directions in an asymmetric manner (Fig. 5), one direction curves towards the dyad while maintaining the dyad, and the other twists around the dyad while destroying it (Fig. 5B). While it is still unclear whether cleavage is sequential or concerted, if two catalytic subunits cleave two opposite strands of the DNA duplex in concert, the dyad symmetry relating the two enzyme-substrate complexes should be maintained. The observed DNA conformation in our complex as well as the modeling of a pre-cleavage tetramer (Kamtekar *et al.*, 2006) suggest that this Gin-DNA complex might be more catalytically relevant than the bent duplex in the previous γδ resolvase (Keenholtz *et al*., 2011; Yang & Steitz, 1995). With the observed Gin-DNA complex obeying the dyad symmetry in this study, a question arises as to whether a straight DNA duplex could fit into a catalytically active complex as a bent DNA duplex as previously modeled (Keenholtz *et al*., 2011). We repeated their modeling using our straight DNA duplex to show that it indeed fits equivalently well to, or better than, the bent DNA duplex (Fig. S9). In either straight or bent DNA duplex, the P-P distance of cleavage centers of the DNA duplexes that interact with the E helices remain unchanged. Therefore, it is possible that the enzyme-substrate complex as observed in the Gin-DNA complex in this study is a step closer to the catalytically active complex than the asymmetric enzyme-substrate dimeric complex in the γδ resolvase. Additionally, the ability of the Gin to reach an active state is consistent with the "hit-and-run" hypothesis of its activation by Fis.

Table S1. Retrospect molecular replacement test runs[a]

Models	RF Z-score[b]	TF Z-score[b]
Apo-Gin Core Model	Failed[c]	Failed[c]
Apo-Gin Core (B=30)	Failed[c]	Failed[c]
Hin-DNA Model	Failed[c]	Failed[c]
γδ Resolvase-DNA model	Failed[c]	Failed[c]
Core (residues 2-99)	3.4	6.4
Core (B=30)	2.8	5.9
Core (B=300)	3.4	8.8
DNA-Binding Domain+DNA	2.0	9.6
DNA-Binding Domain+DNA (B=30)	2.8	5.9
DNA-Binding Domain+DNA (B=300)	2.9	5.7
Complete Model	2.5	26.0
Complete Model (B=30)	3.2	15.3
Complete Model (B=300)	3.2	15.3

a Molecular replacement was tried using both PHASER (McCoy et al., 2007) and Patterson correlation methods as implemented in CNS (Brünger *et al.*, 1998). While partial solutions were obtained, they were unable to be refined convincingly (ST-3). Initial attempts at molecular replacement were performed using the apo-Gin as the search model (Ritacco *et al.*, 2013), specifically, the Core domain as it was expected to be nearly identical in both structures based on previous results from γδ resolvase (Kamtekar *et al.*, 2006, Li *et al.*, 2005, Yang & Steitz, 1995). The rmsD of the main chain atoms of the structure of the apo-Gin to the structure of the DNA bound complex deviates 1.9 Å. This was within an acceptable range for a successful molecular replacement search model. In addition, the structures of Sin and synaptic structures of resolvase were also used as models without success (Keenholtz *et al.*, 2011). After the structure was solved, we were baffled as to why a molecular replacement solution could not be found. We tested the ability of PHASER to find models that we know should be found during a molecular replacement search. Based on the values of the Z-Scores from PHASER it is apparent that convincing scores are achieved only using the complete final refined model. It is worth mentioning that calculating a 2Fo-Fc maps phased from a correctly positioned Core from the *apo* structure of Gin (only 34% of the scattering mass of the asymmetric unit) failed to reveal the missing DBD-DNA features, which differed from the corresponding dimer of the resolvase-DNA complex.

b Values are defined and generated using PHASER (McCoy *et al.*, 2007).

c A solution was not found in at least the top 5 results of the PHASER search.

Table S2. Refinement R-factors of Complete and Partial Models[a]

Models	$-F_{solvent}$	$+F_{solvent}$
Complete Model	26.6%	24.9%
Complete Model (B=30)	34.3%	32.2%
Core (residues 2-99)	50.1%	51.0%*
Core (B=30)	52.7%	50.4%
DNA-Binding Domain+DNA	47.5%	46.3%
DNA-Binding Domain+DNA (B=30)	48.5%	49.0%*

a For the incomplete models, an addition of bulk solvent correction ($+F_{solvent}$) to model amplitudes does not always lower the crystallographic R-factors. The percentage of the scattering mass is 34% and 54%, for the catalytic domain and the DNA-binding domain plus DNA, respectively.

Table S3. Data processing statistics of native data sets

Native	*a*=*b* (Å)	*c* (Å)	Resolution (Å)	R_{merge} (%)[a]	R_{iso} (%)[b]
1	119.80	343.40	5.0	11.7	-
2	119.81	342.14	8.0	11.2	13.5
3	119.95	343.00	5.51	9.7	32.7
4	120.54	344.63	5.52	8.0	20.2

a $R_{merge}=<\Sigma_{hkl}\Sigma_j|I_j(hkl)- <I(hkl)|>/<I(hkl)>$, merging intensity statistics for all symmetry mates.
b R_{iso} are non-isomorphous differences relative *Native* 1 data set on amplitude. Note that the extent of non-isomorphous differences are not in proportion to the change in unit cell parameters.

Figure S1. Flowchart and scheme of computational procedures for phase determination and improvement used in this study.

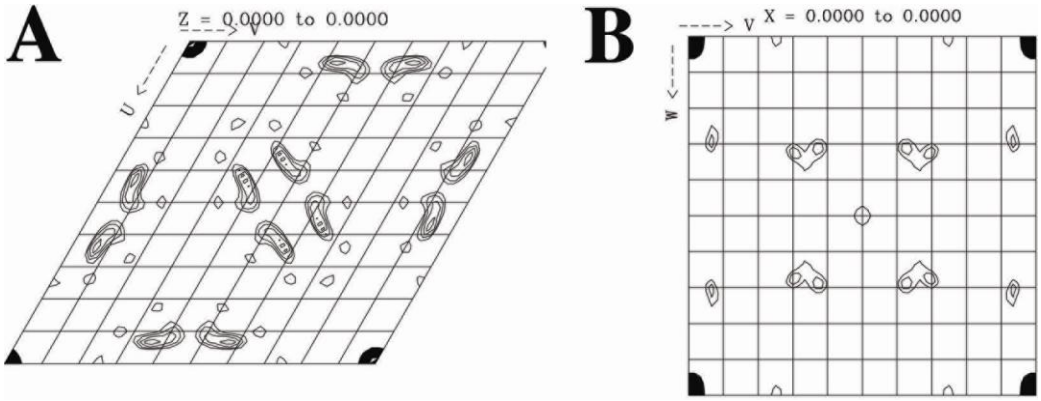

Figure S2. Anomalous difference Patterson of di-platinum PIP derivative at 6.0 Å of resolution of the previous Apo Gin structure in $P6_222$ (Ritacco et al., 2013) in two perpendicular Harker sections show interpretable features for comparison with Figure 2. Each section is contoured with 0.5 σ increments starting at 1.5 σ.

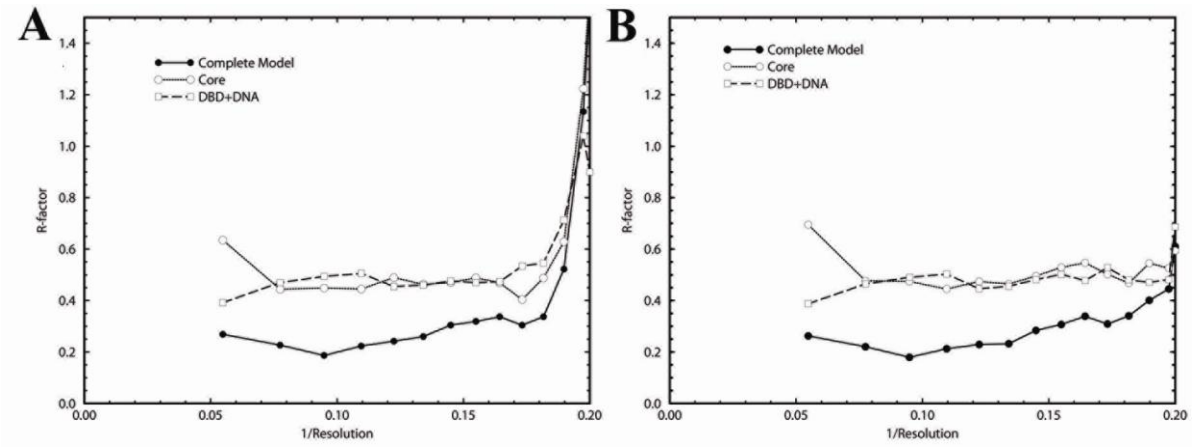

Figure S3. Contribution of bulk solvent variations to calculated structure factors. (A) Bulk solvent level was set to be zero. (B) The bulk solvent level was refined by REFMAC (Murshudov *et al.*, 1997) and was included in the structure factor calculation. Each graph plots the crystallographic R-Factors as a function of resolution for the complete model (filled circles), the NTD (open circles), and DNA-binding domain with DNA (open squares). Note that without bulk solvent correction, R-factors for the highest resolution bin were as high as 140% (A). The effects on the lower resolution were much smaller than on the highest resolution so that bulk solvent correction does not simply affect the scaling between data sets. Thus, solvent-contrast variations cannot be completely removed at such low resolution by simple scaling and partially explain why we had severe non-isomorphism. This is also a reason why we can extract additional phases from non-isomorphous native data sets.

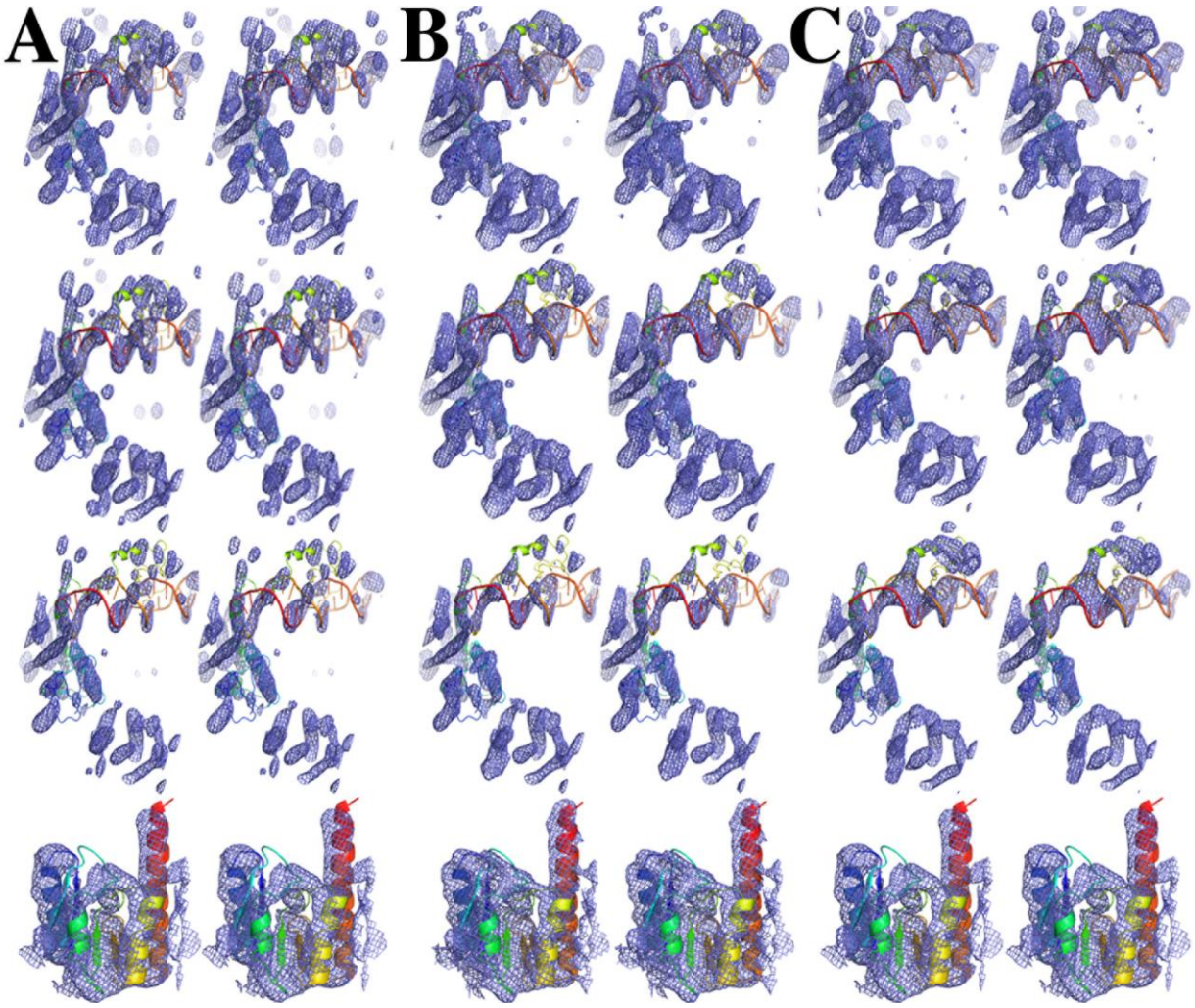

Figure S4. Experimental maps of non-isomorphous cross-crystal averaging in stereodiagram. (A-C) Experimental maps are the initial SIR/SAD phasing, and step one and step two of the two-steps averaging, respectively. Maps in row 1 through 3 are contoured at 1, 1.5, and 2.5 σ for this Gin-DNA complex, respectively. Maps in row 4 are contoured at 1 σ for the previous *apo* Gin structure (3UJ3, Ritacco et al. 2013). Maps in row 1 are identical to ones in Figure 3 and included here for comparison.

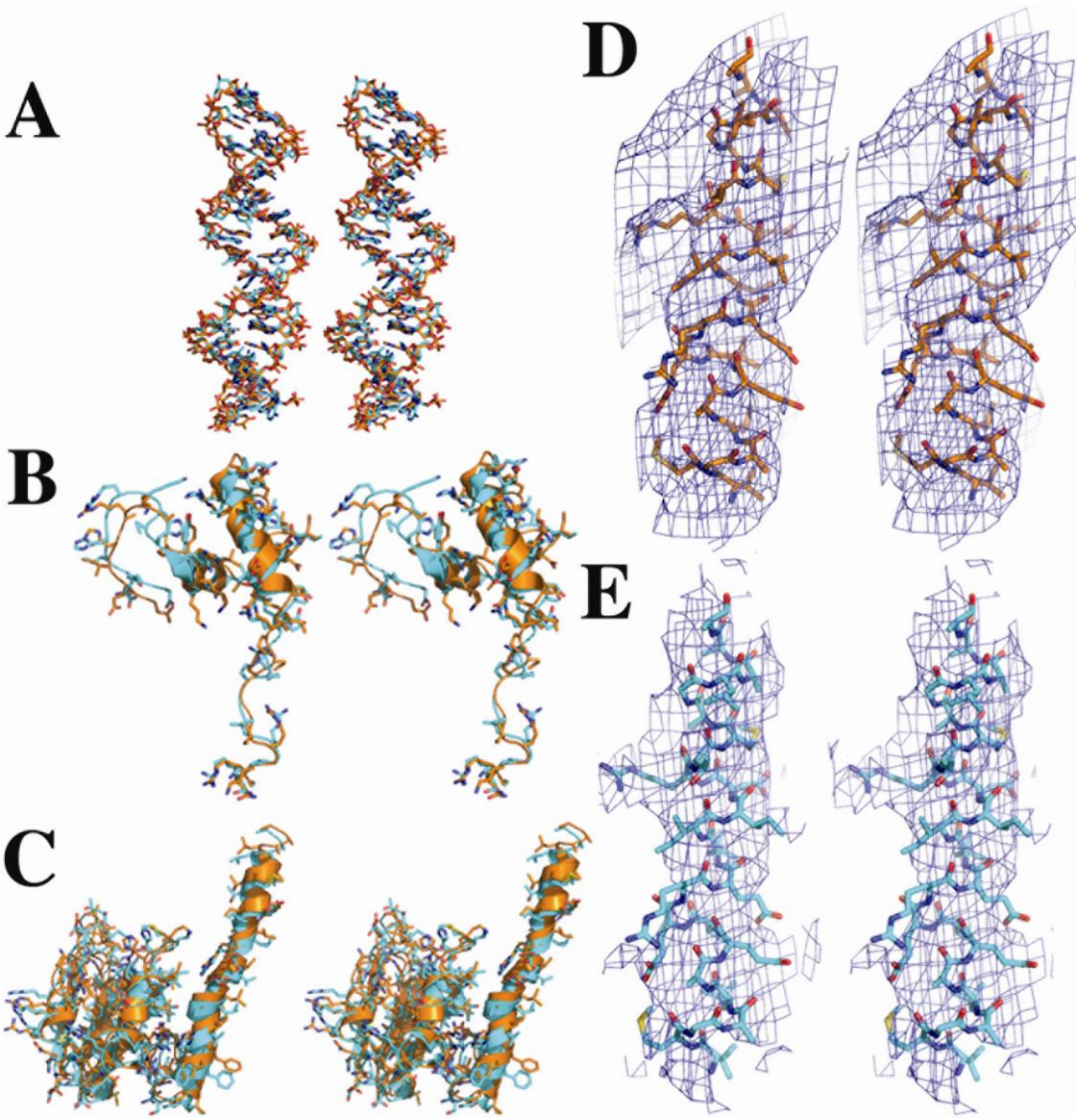

Figure S5. Comparison of the Refmac (orange) and CNS/DEN (blue) refined models in stereodiagram. (A) Major differences between the two refinements lie at the ends of DNA strands with some rotations, and at the DNA-binding domain (B), as well as both the E-helix and Core (C). Differences are largely at side chain conformations, whereas the helix-turn-helix and the majority of the loops remain unchanged in the two models. Similar trends are also present in the Core domain, although the extent of differences is reduced. There was a translational difference in positioning of helices E domains in the two models. (D,E) Corresponding 2Fo-Fc maps superimposed onto the helices E contoured at 1 σ where Fo and Fc denote the observed and calculated amplitudes.

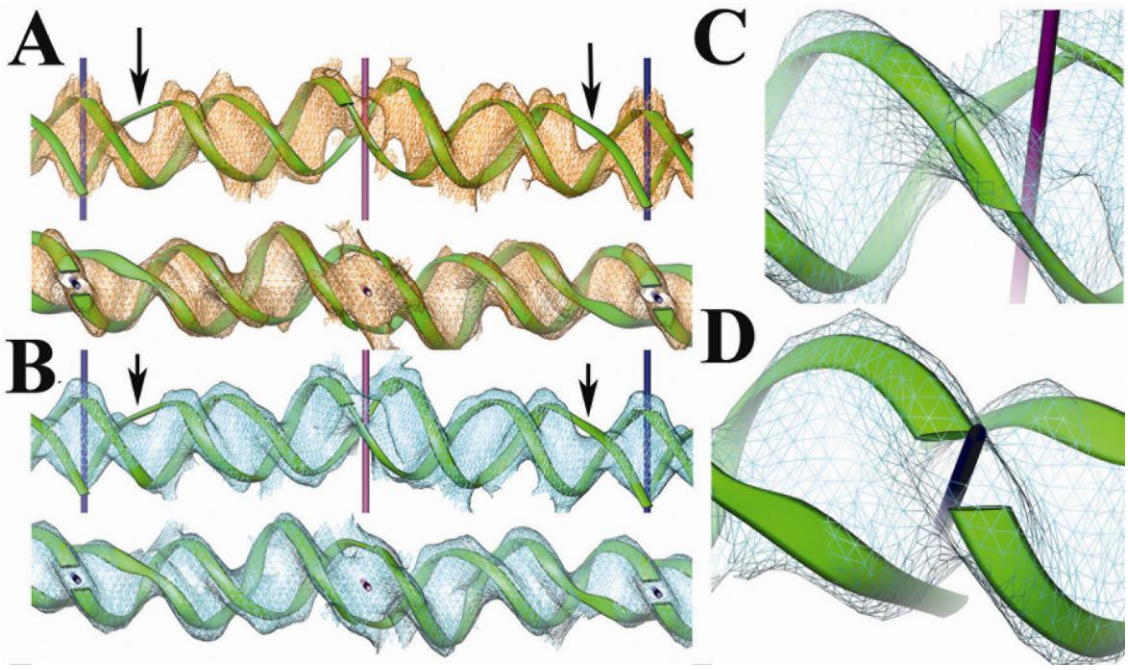

Figure S6. Comparison of model-derived maps with the best experimental maps for the Gin-DNA complex. (A) A experimental map from step-two averaging contoured at 0.9 σ is superimposed onto the model of the DNA and extended across two asymmetric units. Crystallographic dyads along *z*-axis (blue) and the middle point of the short diagonal axis (magenta) are shown. Arrows indicate electron density gaps for the end of DNA duplexes. (B-D) The final model-phased 2Fo-Fc maps contoured at 0.9 σ are superimposed onto the refined model.

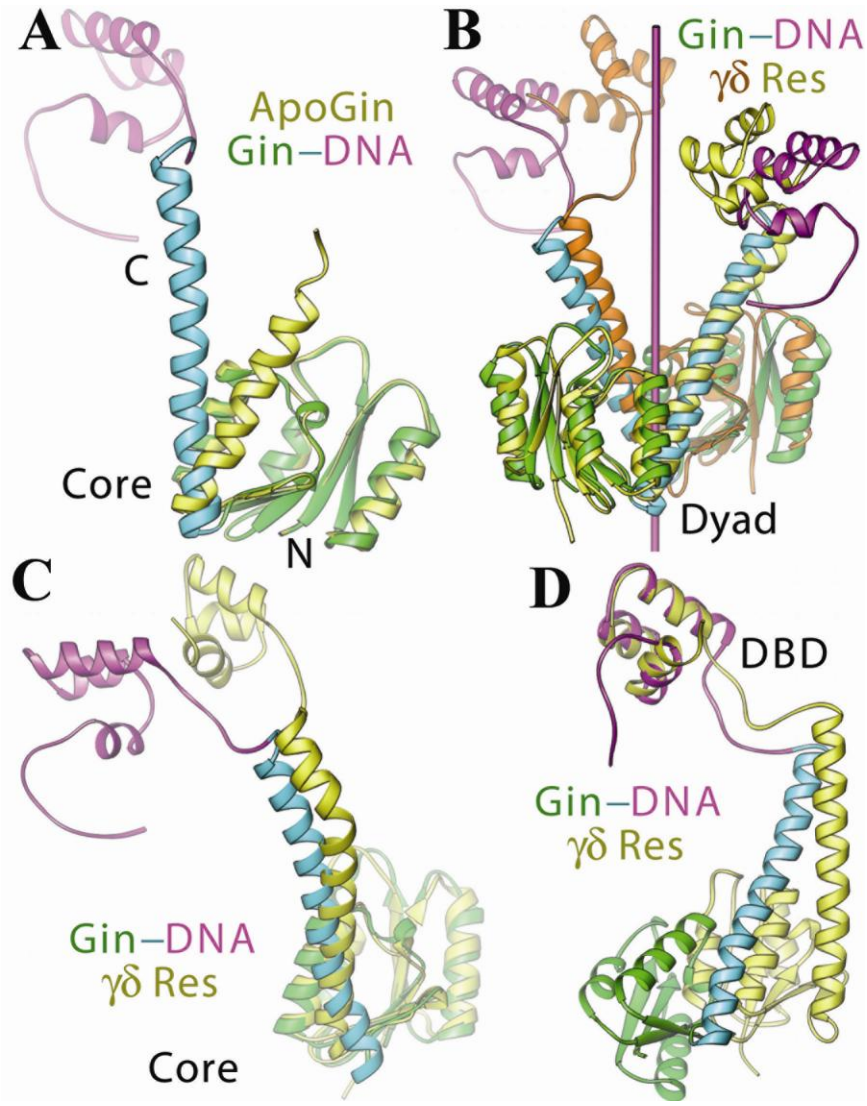

Figure S7. Conformation of Gin in this Gin-DNA complex. (A) A pair-wise alignment of the catalytic domains of the Gin-DNA structure (green, cyan, and magenta) with the *apo* structure of Gin (yellow). The E helix in the DNA bound complex has not changed to a catalytic/synaptic conformation. (B) A pair-wise alignment of the dyad relating the Core domain of the symmetric dimeric Gin (green, cyan, and magenta) with the asymmetric dimeric γδ resolvase (yellow and golden). (C/D) Comparison of a wild type Gin (green, cyan, magenta) subunit with a wild type γδ resolvase subunit (yellow), after a pair-wise alignment of the Core domains (C) or the DNA binding domains (D). The secondary structure arrangement within these two domains is identical, even though their relationship to each other is very different.

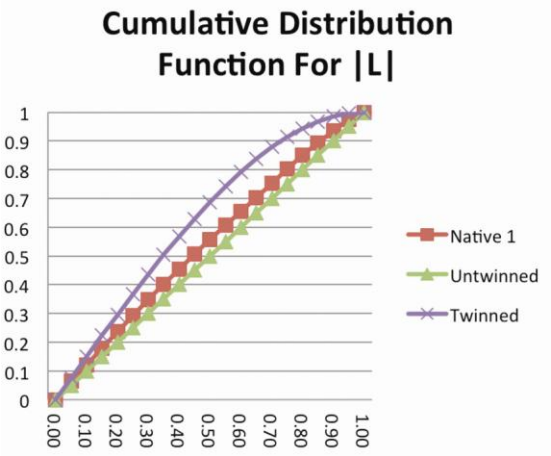

Figure S8. Graphical representation of results of the twinning statistical tests of Native 1 data shows no evidence of perfect merohedry twinning.

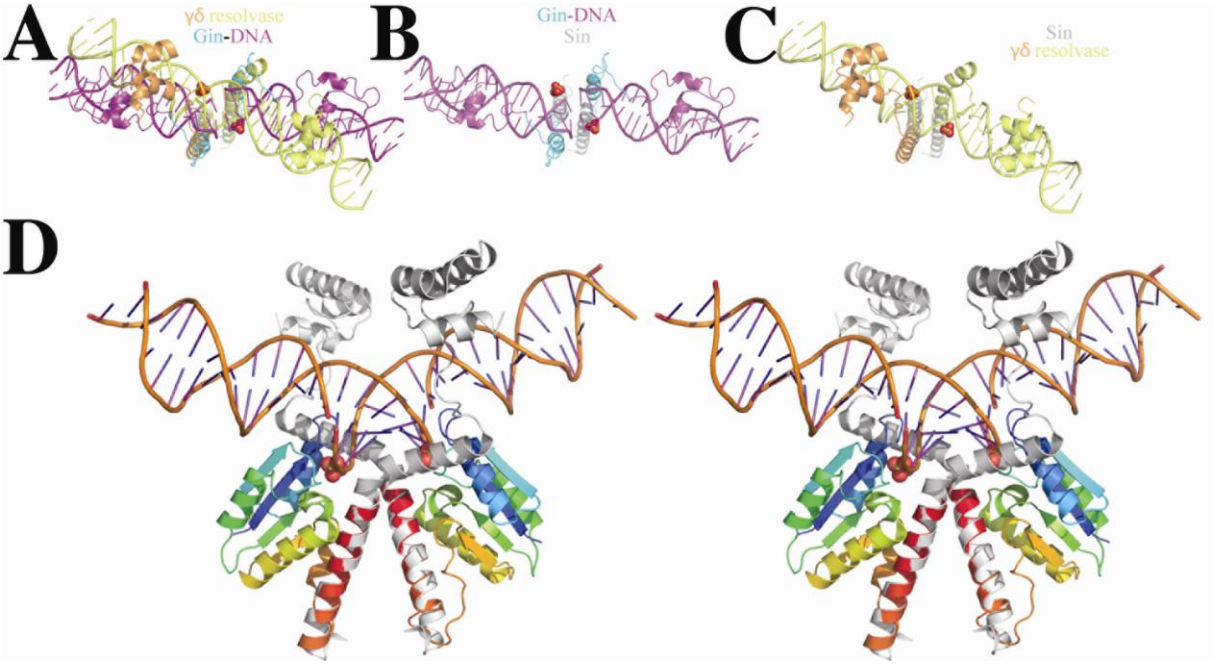

Figure S9. Models of a DNA duplex and the E helices in various recombinases. (A) The synaptic structure of Sin (grey) provided a high-resolution snapshot of the organization of the serine recombinase active site during catalysis. The location of a sulfate ion (red/yellow spheres) in the complex structure specifically suggests the relative locations of the scissile phosphates during cleavage (Keenholtz et al, 2011). Aligning the Gin-DNA complex (cyan/magenta) with the structure of γδ resolvase (yellow/orange) using the method previously described (Keenholtz et al, 2011) shows that the overall orientation of the pair of E helices is nearly identical. This implies that these recombinases enforce the same structural constraints on the internal base pairs of the DNA. (B/C) In both the Gin (B) and the γδ resolvase (C), the duplex of DNA is superimposed onto the complex of Sin. In each case, the scissile

phosphates of the DNA duplexes are close to the location of bound sulfate ions. In standard B-form DNA duplex, the distances between the expected scissile phosphates range from 13-16 Å across the minor groove. The distance between the sulfate positions across subunits in the Sin structure is 18 Å. Thus, the recombinases must widen the minor groove to meet the geometric requirements of catalysis as predicted by the Sin structure. In addition, both panel B and panel C show that compared with the projection of the E helices in the Gin and resolvase, the arrangement of the E helices in the Sin tetramer appears not to be compatible for binding of an un-cleaved DNA substrate. Alternatively, the distance of the bound sulfates in the Sin tetramer may be irrelevant to the geometry of the un-cleaved DNA substrate. (D) Stereodiagram of the DNA duplex modeled onto dimeric resolvase as well as on the Sin tetramer as described previously (Keenholtz *et al.*, 2011). In addition, the E Helices from a regulatory structure of Sin (2ROQ) bound to DNA (grey) have been modeled onto the catalytic E Helices of Sin (rainbow colors) (Mouw *et al.*, 2008). It is apparent that the E Helices and DNA are bound in significantly different states in the two structures and that the formation of various complexes relies on the flexibility of the E Helix (Kamtekar *et al.*, 2006). In order for the complex to reach a state where the DNA is bound in a manner that aligns the catalytic residues and scissile phosphates, the E helices must rotate by 90° at their C-terminus. In this modeling, the C-terminal of the E helices project directly into the minor groove of the DNA duplex. This insertion would stretch the minor groove and increase the distance between the scissile phosphates allowing them to potentially reach the 18-Å distance between the sulfate ions in the catalytic Sin structure.

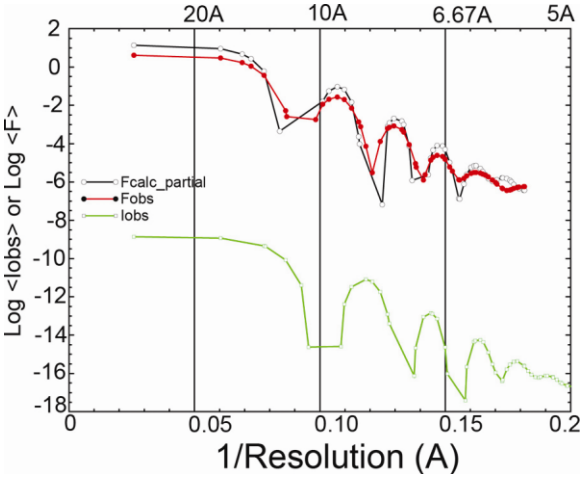

Figure S10. The logarithm of the observed intensity (green) and amplitude (red) are plotted a function of 1/resolution as well as the logarithm of the calculated amplitudes (black) using the partial structure of the Core alone.

Supporting references

Kamtekar, S., Ho, R., Cocco, M., Li, W., Wenwieser, S., Boocock, M., Grindley, N. & Steitz, T. (2006). *Proc Natl Acad Sci U S A* **103**, 10642-10647.

Keenholtz, R. A., Rowland, S. J., Boocock, M. R., Stark, W. M. & Rice, P. A. (2011). *Structure* **19**, 799-809.

Mouw, K. W., Rowland, S. J., Gajjar, M. M., Boocock, M. R., Stark, W. M. & Rice, P. A. (2008). *Mol Cell* **30**, 145-155.

Srinivasan, R. & Parthasarathy, S. (1976). *Some statistical applications in X-ray crystallography*, 1st ed. Oxford ; New York: Pergamon Press.