



wwPDB X-ray Structure Validation Summary Report i

Feb 28, 2014 – 11:12 AM GMT

PDB ID : 2A3V
Title : Structural basis for broad DNA-specificity in integron recombination
Authors : MacDonald, D.; Demarre, G.; Bouvier, M.; Mazel, D.; Gopaul, D.N.
Deposited on : 2005-06-27
Resolution : 2.80 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

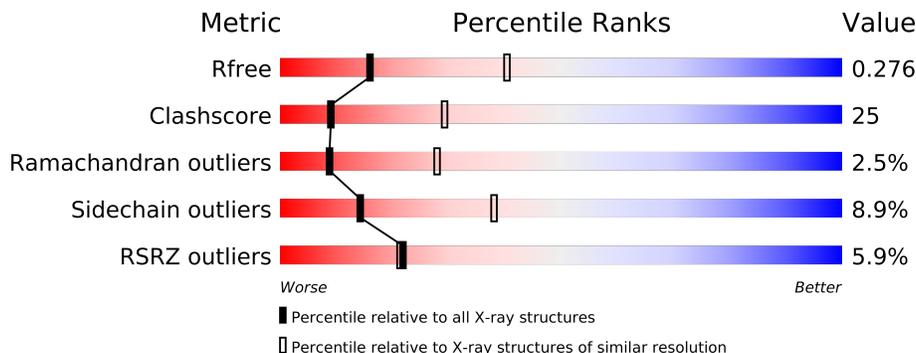
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.15 2013
Xtriage (Phenix) : dev-1323
EDS : stable22639
Percentile statistics : 21963
Refmac : 5.8.0049
CCP4 : 6.3.0 (Settle)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : stable22683

1 Overall quality at a glance

The reported resolution of this entry is 2.80 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	66092	1799 (2.80-2.80)
Clashscore	79885	2295 (2.80-2.80)
Ramachandran outliers	78287	2252 (2.80-2.80)
Sidechain outliers	78261	2254 (2.80-2.80)
RSRZ outliers	66119	1802 (2.80-2.80)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	E	40	
1	G	40	
2	F	43	
2	H	43	
3	A	320	
3	B	320	
3	C	320	
3	D	320	

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 13236 atoms, of which 0 are hydrogen and 0 are deuterium.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a DNA chain called DNA (31-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	E	31	Total 632	C 300	N 120	O 182	P 30	0	0	0
1	G	29	Total 592	C 281	N 112	O 171	P 28	0	0	0

- Molecule 2 is a DNA chain called DNA (34-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	F	34	Total 693	C 332	N 121	O 207	P 33	0	0	0
2	H	32	Total 651	C 312	N 114	O 194	P 31	0	0	0

- Molecule 3 is a protein called site-specific recombinase IntI4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	313	Total 2592	C 1662	N 465	O 457	S 8	0	0	0
3	B	320	Total 2643	C 1693	N 475	O 466	S 9	0	0	0
3	C	317	Total 2620	C 1678	N 472	O 461	S 9	0	0	0
3	D	320	Total 2643	C 1693	N 475	O 466	S 9	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2	GLY	LYS	ENGINEERED	GB 9657688
B	2	GLY	LYS	ENGINEERED	GB 9657688
C	2	GLY	LYS	ENGINEERED	GB 9657688
D	2	GLY	LYS	ENGINEERED	GB 9657688

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	36	Total O 36 36	0	0
4	B	40	Total O 40 40	0	0
4	C	32	Total O 32 32	0	0
4	D	41	Total O 41 41	0	0
4	E	7	Total O 7 7	0	0
4	F	4	Total O 4 4	0	0
4	G	6	Total O 6 6	0	0
4	H	4	Total O 4 4	0	0

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: DNA (31-MER)

Chain E: 



- Molecule 1: DNA (31-MER)

Chain G: 



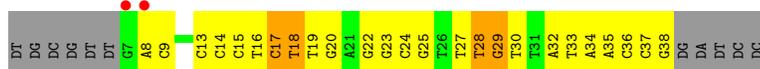
- Molecule 2: DNA (34-MER)

Chain F: 



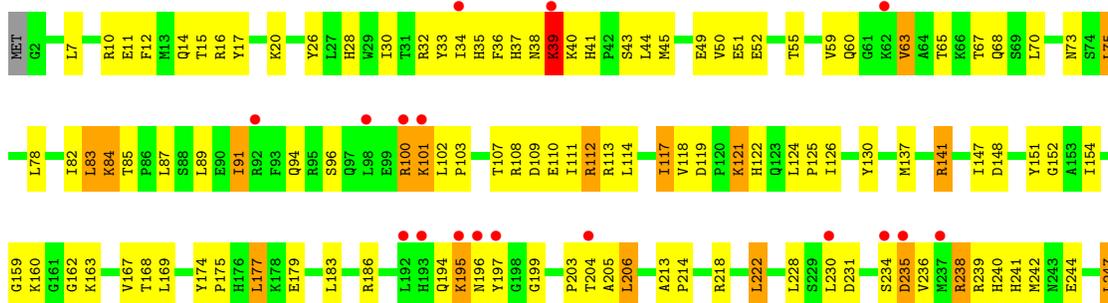
- Molecule 2: DNA (34-MER)

Chain H: 



- Molecule 3: site-specific recombinase IntI4

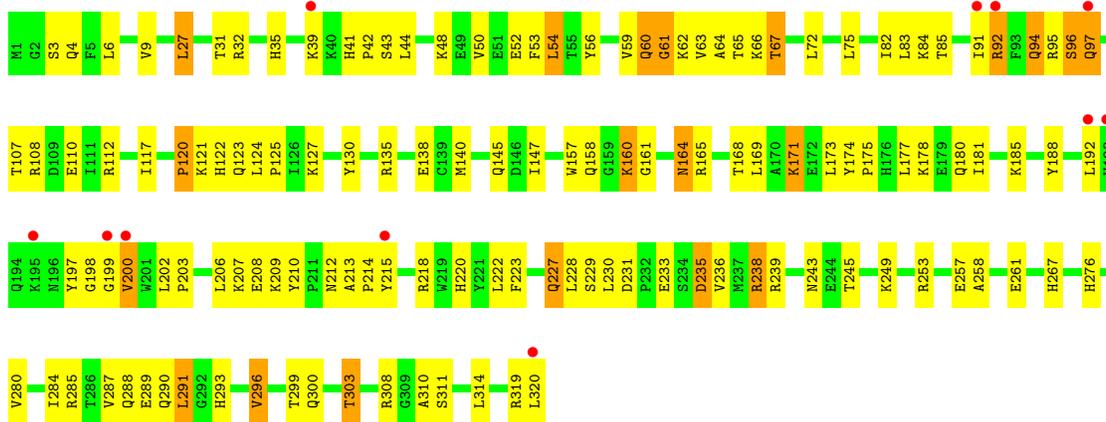
Chain A: 





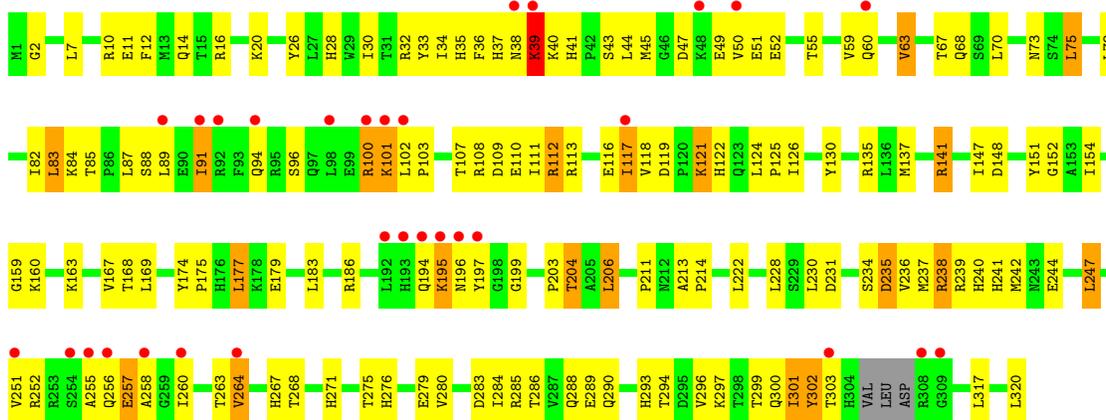
• Molecule 3: site-specific recombinase IntI4

Chain B:



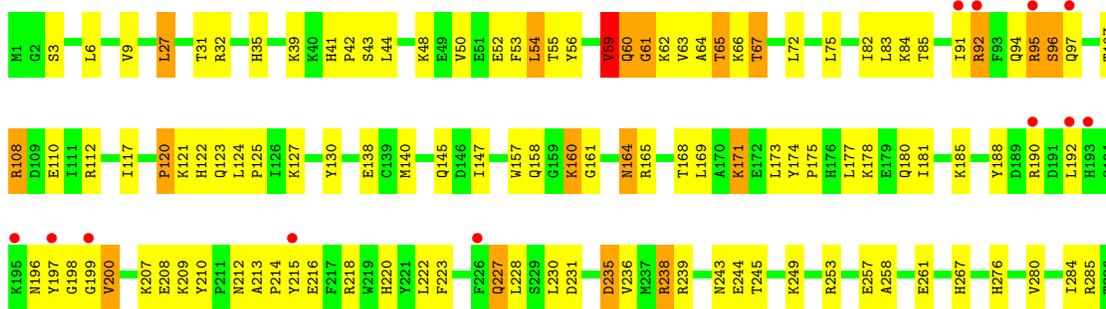
• Molecule 3: site-specific recombinase IntI4

Chain C:



• Molecule 3: site-specific recombinase IntI4

Chain D:





4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	149.90Å 170.20Å 209.40Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	44.60 – 2.80 44.59 – 2.80	Depositor EDS
% Data completeness (in resolution range)	96.8 (44.60-2.80) 96.4 (44.59-2.80)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	0.08	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.04 (at 2.81Å)	Xtrriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.234 , 0.262 0.249 , 0.276	Depositor DCC
R_{free} test set	3211 reflections (5.04%)	DCC
Wilson B-factor (Å ²)	59.5	Xtrriage
Anisotropy	0.221	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 33.9	EDS
Estimated twinning fraction	No twinning to report.	Xtrriage
L-test for twinning	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Outliers	0 of 63712 reflections	Xtrriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	13236	wwPDB-VP
Average B, all atoms (Å ²)	64.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.16% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

5 Model quality i

5.1 Standard geometry i

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	E	0.58	0/709	0.74	0/1092
1	G	0.61	0/664	0.79	0/1023
2	F	0.59	0/775	0.79	0/1195
2	H	0.65	0/728	0.92	3/1122 (0.3%)
3	A	0.45	0/2652	0.64	3/3578 (0.1%)
3	B	0.51	0/2704	0.76	3/3649 (0.1%)
3	C	0.45	0/2680	0.64	0/3614
3	D	0.51	0/2704	0.76	7/3649 (0.2%)
All	All	0.51	0/13616	0.73	16/18922 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	E	0	2
1	G	0	3
2	F	0	3
2	H	0	3
All	All	0	11

There are no bond length outliers.

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	96	SER	CA-C-N	-8.96	97.50	117.20
3	A	39	LYS	CA-C-N	-7.41	100.91	117.20
2	H	28	DT	OP1-P-OP2	-6.91	109.23	119.60
3	B	94	GLN	C-N-CA	-6.60	105.21	121.70
3	B	96	SER	CA-C-N	-6.57	102.76	117.20

There are no chirality outliers.

5 of 11 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	E	20	DG	Sidechain
1	E	27	DG	Sidechain
2	F	17	DC	Sidechain
2	F	19	DT	Sidechain
2	F	29	DG	Sidechain

5.2 Close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	632	0	348	35	0
1	G	592	0	326	15	0
2	F	693	0	387	20	0
2	H	651	0	364	40	0
3	A	2592	0	2637	147	1
3	B	2643	0	2695	144	0
3	C	2620	0	2670	155	0
3	D	2643	0	2695	138	0
4	A	36	0	0	5	0
4	B	40	0	0	3	0
4	C	32	0	0	7	0
4	D	41	0	0	5	0
4	E	7	0	0	4	0
4	F	4	0	0	1	0
4	G	6	0	0	0	0
4	H	4	0	0	0	0
All	All	13236	0	12122	633	1

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 25.

The worst 5 of 633 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
3:B:35:HIS:ND1	3:B:39:LYS:HD3	1.38	1.36
3:B:35:HIS:CE1	3:B:39:LYS:HD3	1.81	1.13

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Distance(Å)	Clash(Å)
3:B:35:HIS:CE1	3:B:39:LYS:CD	2.32	1.10
1:E:9:DT:H2"	1:E:10:DT:H5"	1.35	1.08
3:D:65:THR:HG21	3:D:95:ARG:HD2	1.28	1.08

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
3:A:35:HIS:CE1	3:A:35:HIS:CE1[3_656]	1.78	0.42

5.3 Torsion angles

5.3.1 Protein backbone

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution. The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
3	A	309/320 (97%)	270 (87%)	29 (9%)	10 (3%)	6	20
3	B	318/320 (99%)	282 (89%)	31 (10%)	5 (2%)	14	44
3	C	313/320 (98%)	276 (88%)	26 (8%)	11 (4%)	6	18
3	D	318/320 (99%)	280 (88%)	33 (10%)	5 (2%)	14	44
All	All	1258/1280 (98%)	1108 (88%)	119 (10%)	31 (2%)	9	28

5 of 31 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	39	LYS
3	A	84	LYS
3	A	117	ILE
3	A	235	ASP
3	B	84	LYS

5.3.2 Protein sidechains

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of

similar resolution. The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
3	A	282/287 (98%)	260 (92%)	22 (8%)	18	45
3	B	287/287 (100%)	259 (90%)	28 (10%)	12	32
3	C	284/287 (99%)	262 (92%)	22 (8%)	18	45
3	D	287/287 (100%)	258 (90%)	29 (10%)	11	30
All	All	1140/1148 (99%)	1039 (91%)	101 (9%)	14	38

5 of 101 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	B	289	GLU
3	C	101	LYS
3	D	235	ASP
3	B	296	VAL
3	C	63	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 54 such sidechains are listed below:

Mol	Chain	Res	Type
3	B	290	GLN
3	C	180	GLN
3	D	220	HIS
3	B	293	HIS
3	C	68	GLN

5.3.3 RNA ⓘ

There are no RNA chains in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data i

6.1 Protein, DNA and RNA chains i

In the following table, the column labelled ‘#RSRZ > 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q < 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	E	31/40 (77%)	0.13	2 (6%) 18 17	38, 62, 154, 171	0
1	G	29/40 (72%)	0.03	1 (3%) 43 44	33, 63, 132, 160	0
2	F	34/43 (79%)	0.09	3 (8%) 10 8	35, 58, 164, 175	0
2	H	32/43 (74%)	0.01	2 (6%) 19 18	35, 57, 119, 169	0
3	A	313/320 (97%)	0.29	21 (6%) 17 16	30, 60, 107, 125	0
3	B	320/320 (100%)	0.09	11 (3%) 43 44	27, 54, 93, 131	0
3	C	317/320 (99%)	0.48	30 (9%) 8 7	30, 61, 108, 125	0
3	D	320/320 (100%)	0.11	13 (4%) 35 36	26, 54, 93, 131	0
All	All	1396/1446 (96%)	0.23	83 (5%) 22 21	26, 57, 108, 175	0

The worst 5 of 83 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	308	ARG	11.1
3	C	309	GLY	7.0
3	C	193	HIS	6.2
3	A	98	LEU	6.1
3	B	320	LEU	5.6

6.2 Non-standard residues in protein, DNA, RNA chains i

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates i

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.