



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 8, 2018 – 05:45 pm 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 X-ray Structure Validation Summary Report for a publicly released PDB entry.

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

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

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

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	trunk30967
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk30967

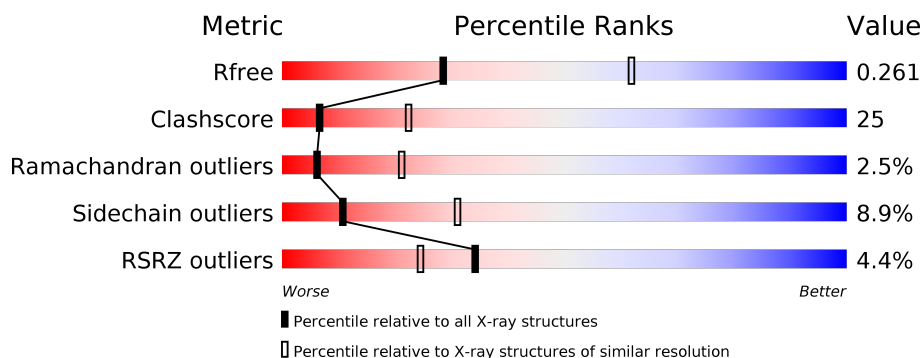
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

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}	111664	2792 (2.80-2.80)
Clashscore	122126	3209 (2.80-2.80)
Ramachandran outliers	120053	3158 (2.80-2.80)
Sidechain outliers	120020	3160 (2.80-2.80)
RSRZ outliers	108989	2726 (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. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	E	40	<div> <div>8%</div> <div>28%</div> <div>45%</div> <div>5%</div> <div>23%</div> </div>
1	G	40	<div> <div>3%</div> <div>30%</div> <div>38%</div> <div>5%</div> <div>28%</div> </div>
2	F	43	<div> <div>9%</div> <div>42%</div> <div>35%</div> <div>•</div> <div>21%</div> </div>
2	H	43	<div> <div>5%</div> <div>16%</div> <div>49%</div> <div>9%</div> <div>26%</div> </div>
3	A	320	<div> <div>5%</div> <div>50%</div> <div>40%</div> <div>7%</div> <div>•</div> </div>
3	B	320	<div> <div>3%</div> <div>58%</div> <div>36%</div> <div>6%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	C	320	<div><div></div><div>6%</div><div>52%</div><div>40%</div><div>7%</div><div></div></div>
3	D	320	<div><div></div><div>2%</div><div>58%</div><div>35%</div><div>7%</div><div></div></div>

2 Entry composition

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

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
1	E	31	Total	C	N	O	P	0	0	0
			632	300	120	182	30			
1	G	29	Total	C	N	O	P	0	0	0
			592	281	112	171	28			

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

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

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

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

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

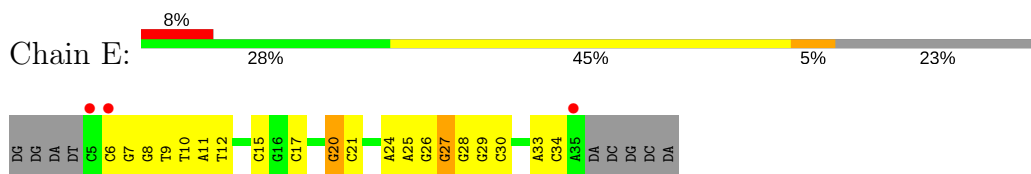
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	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
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

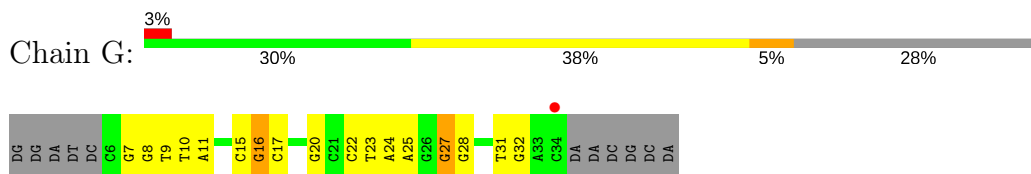
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes 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.

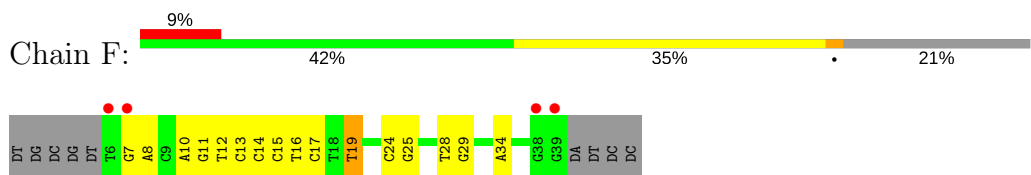
- Molecule 1: DNA (31-MER)



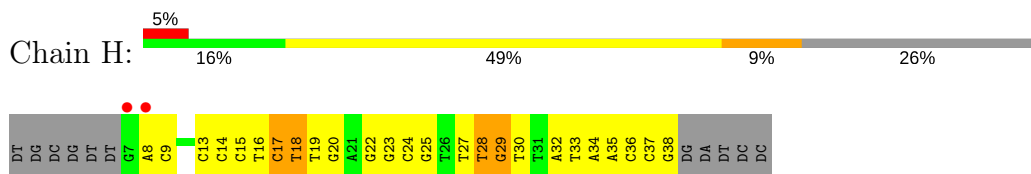
- Molecule 1: DNA (31-MER)



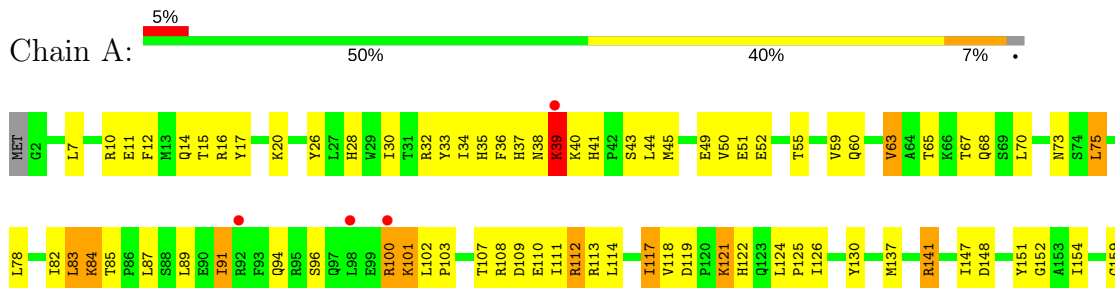
- Molecule 2: DNA (34-MER)

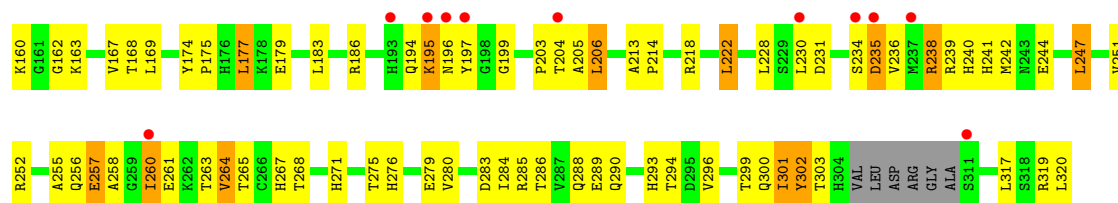


- Molecule 2: DNA (34-MER)

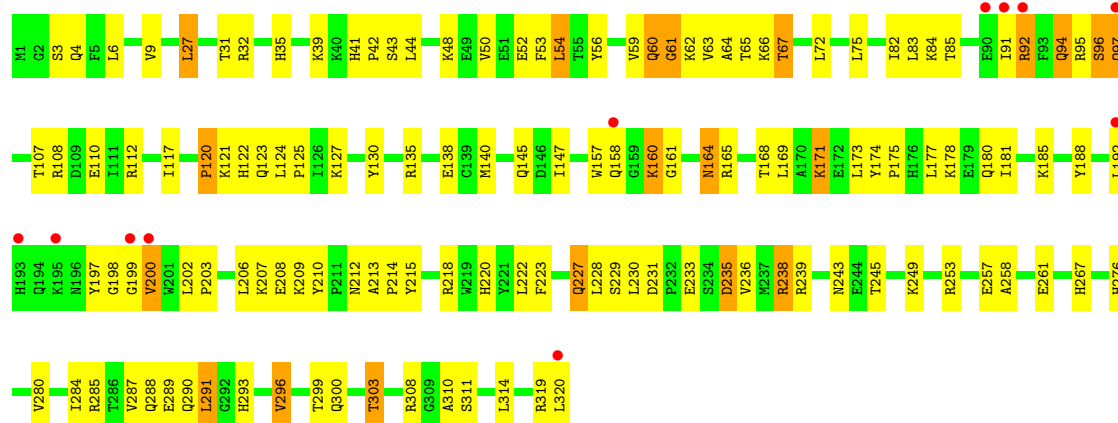


- Molecule 3: site-specific recombinase IntI4

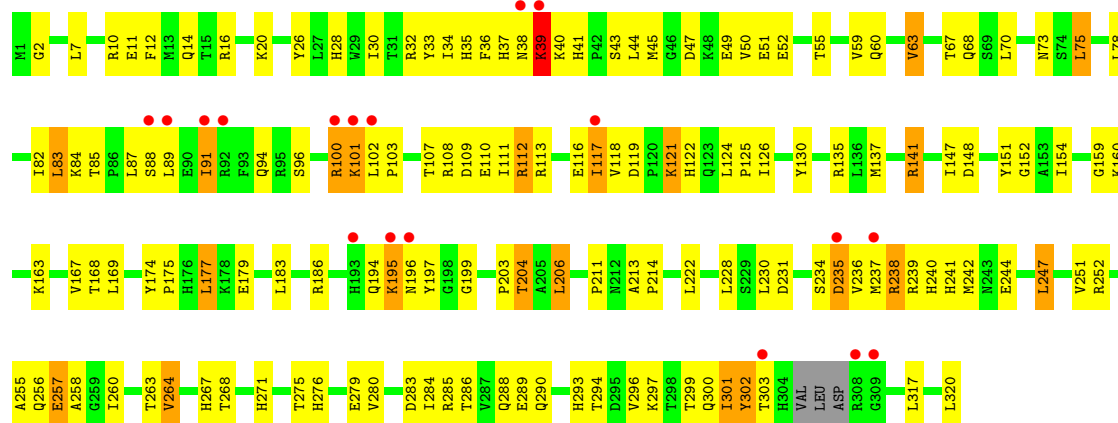




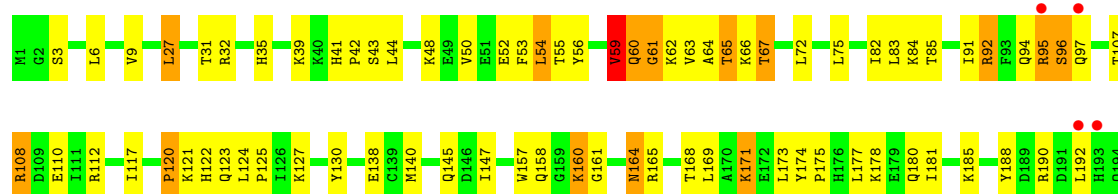
• Molecule 3: site-specific recombinase IntI4

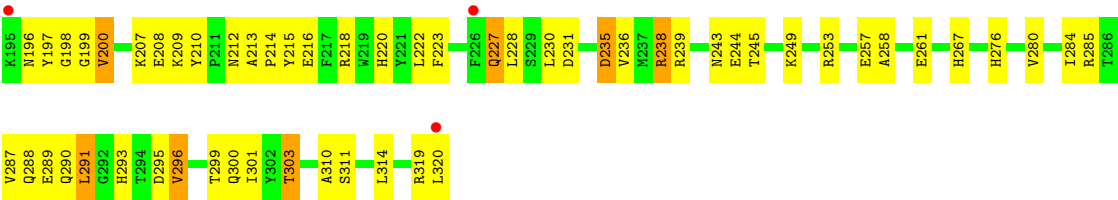


• Molecule 3: site-specific recombinase IntI4



• Molecule 3: site-specific recombinase IntI4





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)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.234 , 0.262 0.234 , 0.261	Depositor DCC
R_{free} test set	3214 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å ²)	59.5	Xtriage
Anisotropy	0.221	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 56.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
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

Xtriage'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.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

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 Too-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 hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

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

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 25.

The worst 5 of 633 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:B:35:HIS:ND1	3:B:39:LYS:HD3	1.38	1.36

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:B:35:HIS:CE1	3:B:39:LYS:HD3	1.81	1.13
3:B:35:HIS:CE1	3:B:39:LYS:CD	2.32	1.10
3:D:65:THR:HG21	3:D:95:ARG:HD2	1.28	1.08
1:E:9:DT:H2"	1:E:10:DT:H5"	1.35	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	Interatomic distance (Å)	Clash overlap (Å)
3:A:35:HIS:CE1	3:A:35:HIS:CE1[3_656]	1.78	0.42

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

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%)	4	15
3	B	318/320 (99%)	282 (89%)	31 (10%)	5 (2%)	11	34
3	C	313/320 (98%)	276 (88%)	26 (8%)	11 (4%)	4	13
3	D	318/320 (99%)	280 (88%)	33 (10%)	5 (2%)	11	34
All	All	1258/1280 (98%)	1108 (88%)	119 (10%)	31 (2%)	6	21

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 [i](#)

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%)	14	37
3	B	287/287 (100%)	259 (90%)	28 (10%)	9	26
3	C	284/287 (99%)	262 (92%)	22 (8%)	14	38
3	D	287/287 (100%)	258 (90%)	29 (10%)	8	24
All	All	1140/1148 (99%)	1039 (91%)	101 (9%)	11	31

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 [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

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

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

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%)	0.26	3 (9%) 8 4	38, 62, 154, 171	0
1	G	29/40 (72%)	0.13	1 (3%) 45 35	33, 63, 132, 160	0
2	F	34/43 (79%)	0.10	4 (11%) 4 2	35, 58, 164, 175	0
2	H	32/43 (74%)	0.09	2 (6%) 20 12	35, 57, 119, 169	0
3	A	313/320 (97%)	0.11	15 (4%) 30 21	30, 60, 107, 125	0
3	B	320/320 (100%)	-0.02	11 (3%) 45 35	27, 54, 93, 131	0
3	C	317/320 (99%)	0.24	18 (5%) 24 15	30, 61, 108, 125	0
3	D	320/320 (100%)	-0.02	7 (2%) 62 52	26, 54, 93, 131	0
All	All	1396/1446 (96%)	0.08	61 (4%) 34 24	26, 57, 108, 175	0

The worst 5 of 61 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	308	ARG	7.2
3	C	309	GLY	6.0
3	A	98	LEU	5.3
1	E	5	DC	5.1
3	B	320	LEU	4.3

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 [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.