



wwPDB X-ray Structure Validation Summary Report ⓘ

Apr 7, 2016 – 08:03 AM EDT

PDB ID : 5IUD
Title : Human DNA polymerase alpha in binary complex with a DNA:DNA template-primer
Authors : Coloma, J.; Aggarwal, A.K.
Deposited on : 2016-03-17
Resolution : 3.30 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
<http://wwpdb.org/validation/2016/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
Mogul : unknown
Xtriage (Phenix) : 1.9-1692
EDS : rb-20027107
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20027107

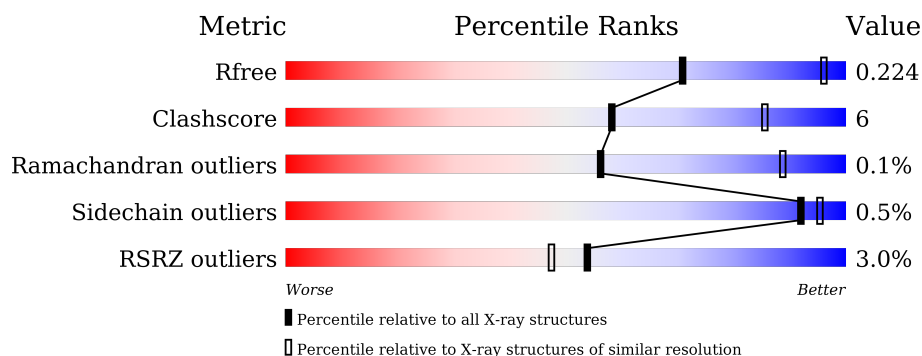
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.30 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}	91344	2060 (3.40-3.20)
Clashscore	102246	1058 (3.38-3.22)
Ramachandran outliers	100387	1038 (3.38-3.22)
Sidechain outliers	100360	1037 (3.38-3.22)
RSRZ outliers	91569	2070 (3.40-3.20)

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	A	918	<div> <div></div> <div>81% 12% 6%</div> </div>
1	D	918	<div> <div>3%</div> <div>83% 12% 5%</div> </div>
1	G	918	<div> <div>3%</div> <div>79% 15% 6%</div> </div>
1	J	918	<div> <div>4%</div> <div>80% 13% 6%</div> </div>
2	B	16	<div> <div></div> <div>56% 38% 6%</div> </div>
2	E	16	<div> <div></div> <div>63% 38%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	H	16	 56% 25% 19%
2	K	16	 56% 38% 6%
3	C	13	 85% 15%
3	F	13	 8% 85% 15%
3	I	13	 77% 23%
3	L	13	 77% 23%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 28207 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 protein called DNA polymerase alpha catalytic subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	859	Total	C	N	O	S	0	0	0
			6480	4177	1082	1182	39			
1	D	871	Total	C	N	O	S	0	0	0
			6580	4231	1103	1206	40			
1	G	863	Total	C	N	O	S	0	0	0
			6456	4151	1075	1189	41			
1	J	861	Total	C	N	O	S	0	0	0
			6391	4110	1079	1165	37			

- Molecule 2 is a DNA chain called DNA template.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	15	Total	C	N	O	P	0	0	0
			320	150	66	89	15			
2	E	16	Total	C	N	O	P	0	0	0
			341	160	71	94	16			
2	H	13	Total	C	N	O	P	0	0	0
			275	130	59	74	12			
2	K	15	Total	C	N	O	P	0	0	0
			317	150	66	87	14			

- Molecule 3 is a DNA chain called DNA primer.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	13	Total	C	N	O	P	0	0	0
			251	123	39	77	12			
3	F	13	Total	C	N	O	P	0	0	0
			251	123	39	77	12			
3	I	10	Total	C	N	O	P	0	0	0
			194	94	29	61	10			
3	L	13	Total	C	N	O	P	0	0	0
			251	123	39	77	12			

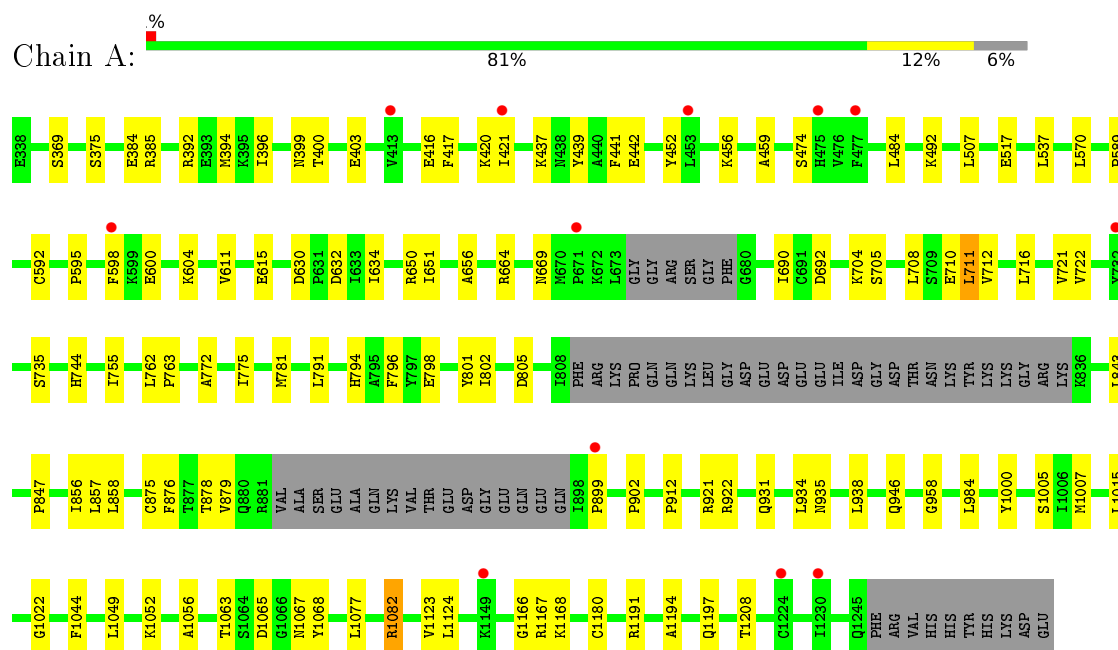
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	31	Total 31	O 31	0	0
4	B	2	Total 2	O 2	0	0
4	D	35	Total 35	O 35	0	0
4	E	3	Total 3	O 3	0	0
4	G	16	Total 16	O 16	0	0
4	H	1	Total 1	O 1	0	0
4	J	11	Total 11	O 11	0	0
4	L	1	Total 1	O 1	0	0

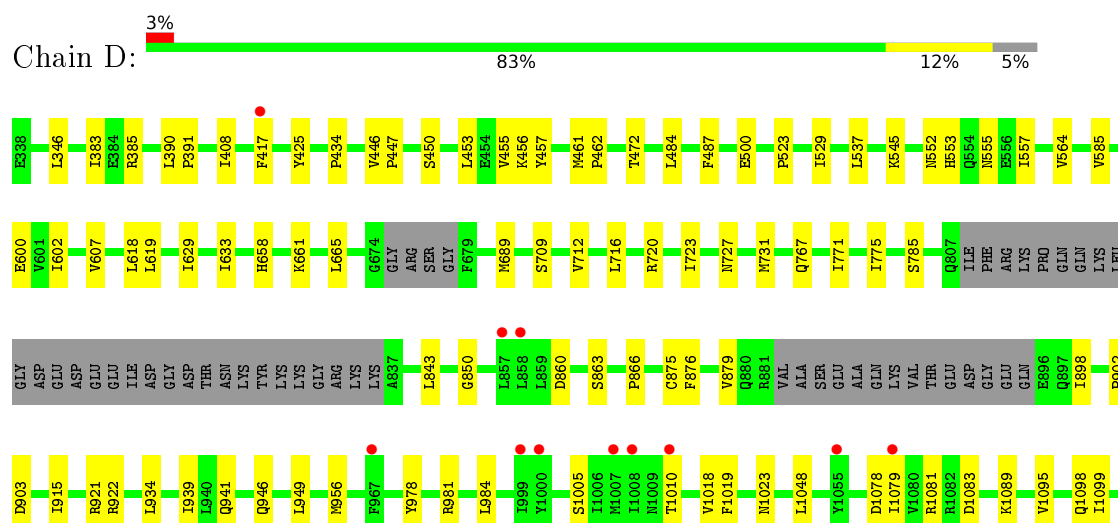
3 Residue-property plots

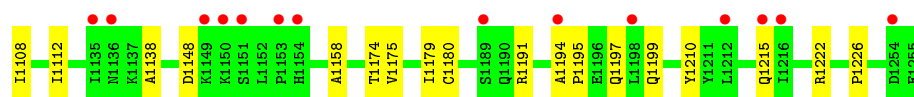
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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 polymerase alpha catalytic subunit

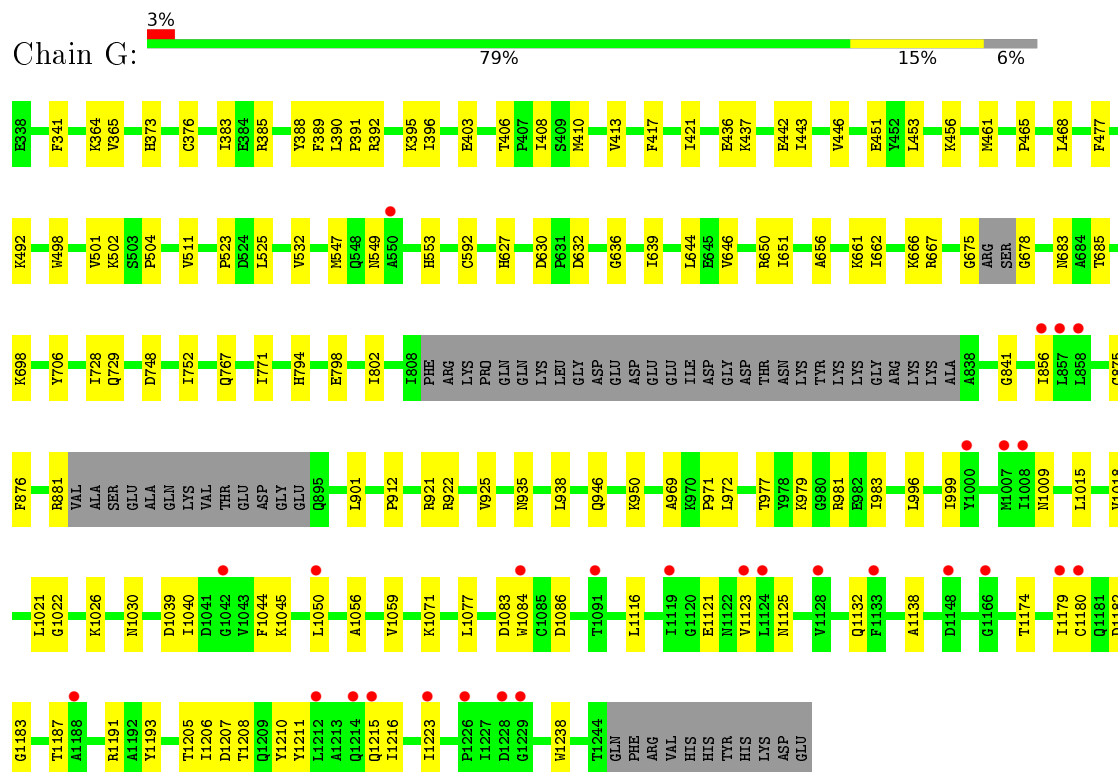


- Molecule 1: DNA polymerase alpha catalytic subunit

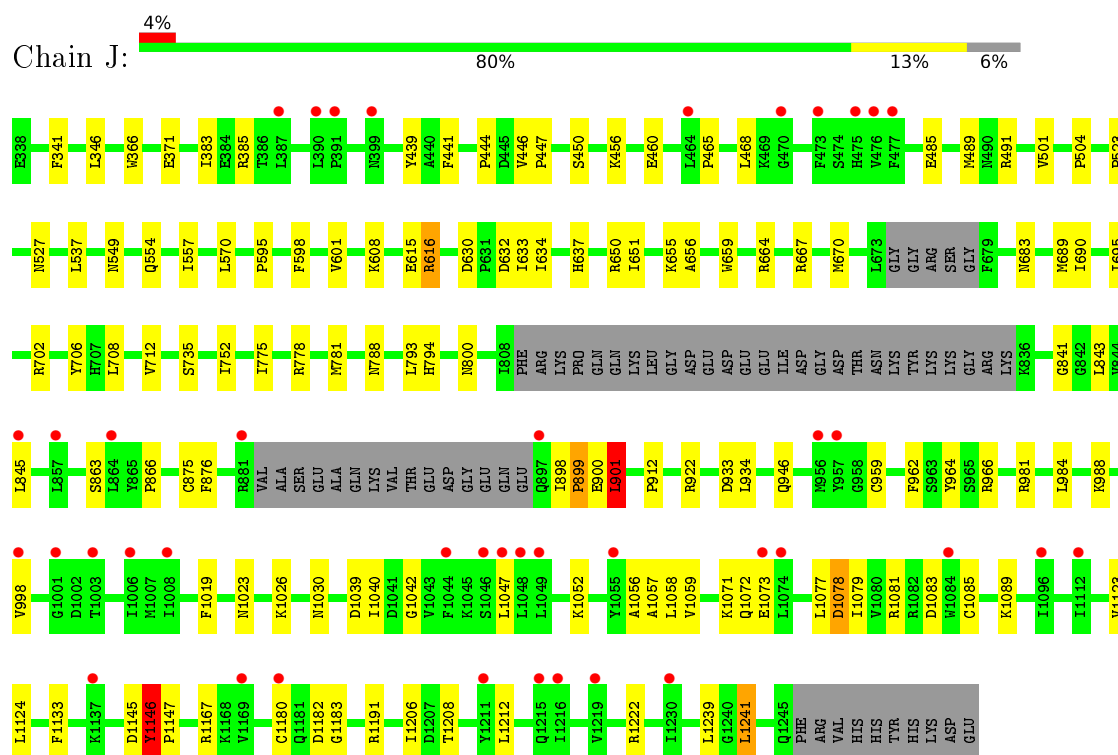




• Molecule 1: DNA polymerase alpha catalytic subunit

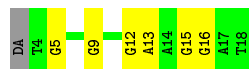


• Molecule 1: DNA polymerase alpha catalytic subunit



- Molecule 2: DNA template

Chain B:  56% 38% 6%



- Molecule 2: DNA template

Chain E:  63% 38%



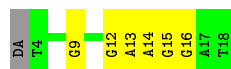
- Molecule 2: DNA template

Chain H:  56% 25% 19%




- Molecule 2: DNA template

Chain K:  56% 38% 6%




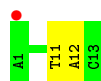
- Molecule 3: DNA primer

Chain C:  85% 15%



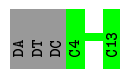
- Molecule 3: DNA primer

Chain F:  8% 85% 15%




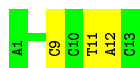
- Molecule 3: DNA primer

Chain I:  77% 23%



- Molecule 3: DNA primer

Chain L:  77% 23%



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	137.04Å 132.03Å 163.14Å 90.00° 109.13° 90.00°	Depositor
Resolution (Å)	64.74 – 3.30 86.20 – 3.30	Depositor EDS
% Data completeness (in resolution range)	99.4 (64.74-3.30) 93.4 (86.20-3.30)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.37 (at 3.33Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.183 , 0.228 0.179 , 0.224	Depositor DCC
R_{free} test set	3863 reflections (5.00%)	DCC
Wilson B-factor (Å ²)	87.0	Xtriage
Anisotropy	0.130	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.28 , 72.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	1 of 82362 reflections (0.001%)	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	28207	wwPDB-VP
Average B, all atoms (Å ²)	108.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.89% 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, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

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	A	0.24	0/6620	0.40	0/9032
1	D	0.24	0/6726	0.41	0/9177
1	G	0.24	0/6595	0.42	0/9000
1	J	0.26	0/6527	0.41	1/8908 (0.0%)
2	B	0.52	0/361	0.86	0/558
2	E	0.50	0/385	0.87	0/595
2	H	0.52	0/311	0.80	0/481
2	K	0.52	0/358	0.85	0/554
3	C	0.58	0/278	0.97	0/424
3	F	0.55	0/278	0.95	0/424
3	I	0.53	0/214	0.94	0/325
3	L	0.61	0/278	0.94	0/424
All	All	0.28	0/28931	0.48	1/39902 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	J	901	LEU	CA-CB-CG	5.20	127.27	115.30

There are no chirality outliers.

There are no planarity outliers.

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	A	6480	0	6180	71	0
1	D	6580	0	6225	65	0
1	G	6456	0	6089	84	0
1	J	6391	0	5989	81	0
2	B	320	0	169	5	0
2	E	341	0	180	6	0
2	H	275	0	147	2	0
2	K	317	0	170	5	0
3	C	251	0	147	1	0
3	F	251	0	147	1	0
3	I	194	0	112	0	0
3	L	251	0	147	4	0
4	A	31	0	0	3	0
4	B	2	0	0	0	0
4	D	35	0	0	0	0
4	E	3	0	0	0	0
4	G	16	0	0	2	0
4	H	1	0	0	0	0
4	J	11	0	0	2	0
4	L	1	0	0	0	0
All	All	28207	0	25702	313	0

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 6.

The worst 5 of 313 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)
1:G:396:ILE:HG13	1:G:403:GLU:HB3	1.58	0.85
1:G:395:LYS:HB2	1:G:408:ILE:HD11	1.61	0.81
1:J:1146:TYR:OH	3:L:9:DC:OP1	2.06	0.71
1:G:1180:CYS:O	1:G:1191:ARG:NH1	2.25	0.69
1:G:1086:ASP:HB3	1:G:1132:GLN:HA	1.75	0.68

There are no symmetry-related clashes.

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	
1	A	851/918 (93%)	810 (95%)	41 (5%)	0	100	100
1	D	863/918 (94%)	826 (96%)	37 (4%)	0	100	100
1	G	855/918 (93%)	812 (95%)	41 (5%)	2 (0%)	52	85
1	J	853/918 (93%)	804 (94%)	47 (6%)	2 (0%)	52	85
All	All	3422/3672 (93%)	3252 (95%)	166 (5%)	4 (0%)	56	89

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	G	406	THR
1	G	1187	THR
1	J	899	PRO
1	J	1146	TYR

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	
1	A	656/824 (80%)	653 (100%)	3 (0%)	92	95
1	D	667/824 (81%)	666 (100%)	1 (0%)	95	98
1	G	647/824 (78%)	643 (99%)	4 (1%)	90	95
1	J	625/824 (76%)	619 (99%)	6 (1%)	82	91
All	All	2595/3296 (79%)	2581 (100%)	14 (0%)	92	95

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	901	LEU
1	G	1015	LEU
1	J	1078	ASP
1	G	630	ASP
1	J	901	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	D	1103	GLN
1	G	1122	ASN
1	G	1125	ASN

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 ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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	A	859/918 (93%)	-0.03	12 (1%)	78	73	42, 93, 164, 222	0
1	D	871/918 (94%)	0.03	25 (2%)	55	49	42, 103, 167, 224	0
1	G	863/918 (94%)	0.08	28 (3%)	51	44	49, 111, 184, 248	0
1	J	861/918 (93%)	0.15	41 (4%)	34	28	45, 110, 181, 233	0
2	B	15/16 (93%)	-0.17	0	100	100	63, 88, 125, 166	0
2	E	16/16 (100%)	-0.22	0	100	100	95, 134, 210, 235	0
2	H	13/16 (81%)	-0.09	0	100	100	101, 174, 270, 272	0
2	K	15/16 (93%)	-0.32	0	100	100	80, 119, 196, 201	0
3	C	13/13 (100%)	-0.16	0	100	100	81, 90, 99, 105	0
3	F	13/13 (100%)	0.14	1 (7%)	16	13	94, 123, 224, 228	0
3	I	10/13 (76%)	-0.00	0	100	100	117, 133, 215, 245	0
3	L	13/13 (100%)	-0.27	0	100	100	86, 109, 242, 258	0
All	All	3562/3788 (94%)	0.05	107 (3%)	54	47	42, 104, 178, 272	0

The worst 5 of 107 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	1128	VAL	7.6
1	G	1148	ASP	5.3
1	G	1008	ILE	4.6
1	J	399	ASN	4.3
1	J	1047	LEU	4.3

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

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.