



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 03:31 am GMT

PDB ID : 1XHZ
Title : Phi29 DNA polymerase, orthorhombic crystal form, ssDNA complex
Authors : Kamtekar, S.; Berman, A.J.; Wang, J.; Lazaro, J.M.; de Vega, M.; Blanco, L.; Salas, M.; Steitz, T.A.
Deposited on : 2004-09-21
Resolution : 2.70 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
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	recalc28949

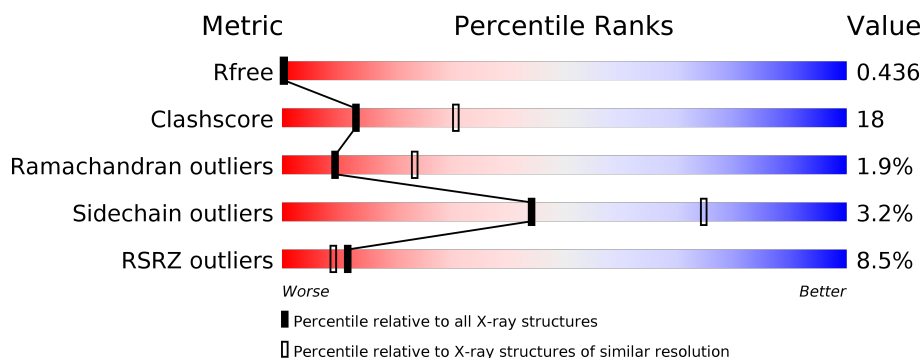
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.70 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}	100719	2259 (2.70-2.70)
Clashscore	112137	2590 (2.70-2.70)
Ramachandran outliers	110173	2550 (2.70-2.70)
Sidechain outliers	110143	2550 (2.70-2.70)
RSRZ outliers	101464	2275 (2.70-2.70)

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	5	<div> <div>100%</div> <div>100%</div> </div>
1	F	5	<div> <div>40%</div> <div>20%</div> <div>20%</div> <div>60%</div> </div>
1	G	5	<div> <div>100%</div> <div>100%</div> </div>
1	H	5	<div> <div>100%</div> <div>100%</div> </div>
2	A	575	<div> <div>10%</div> <div>61%</div> <div>35%</div> <div>..</div> </div>
2	B	575	<div> <div>6%</div> <div>64%</div> <div>34%</div> <div>..</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	C	575	<div><div>8%</div><div><div></div><div>66%</div><div>29%</div><div></div></div><div>...</div></div>
2	D	575	<div><div>7%</div><div><div></div><div>69%</div><div>29%</div><div></div></div><div>..</div></div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 19512 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 5'-D(*TP*TP*TP*TP*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	E	5	Total	C	N	O	P	0	0	0
			97	50	10	33	4			
1	F	2	Total	C	N	O	P	0	0	0
			37	20	4	12	1			
1	G	5	Total	C	N	O	P	0	0	0
			97	50	10	33	4			
1	H	5	Total	C	N	O	P	0	0	0
			97	50	10	33	4			

- Molecule 2 is a protein called DNA polymerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	A	571	Total	C	N	O	S	0	0	0
			4668	3041	754	852	21			
2	B	571	Total	C	N	O	S	0	0	0
			4668	3041	754	852	21			
2	C	571	Total	C	N	O	S	0	0	0
			4668	3041	754	852	21			
2	D	571	Total	C	N	O	S	0	0	0
			4668	3041	754	852	21			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	12	ALA	ASP	ENGINEERED	UNP P03680
A	66	ALA	ASP	ENGINEERED	UNP P03680
B	12	ALA	ASP	ENGINEERED	UNP P03680
B	66	ALA	ASP	ENGINEERED	UNP P03680
C	12	ALA	ASP	ENGINEERED	UNP P03680
C	66	ALA	ASP	ENGINEERED	UNP P03680
D	12	ALA	ASP	ENGINEERED	UNP P03680
D	66	ALA	ASP	ENGINEERED	UNP P03680

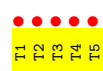
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	124	Total 124	O 124	0	0
3	B	129	Total 129	O 129	0	0
3	C	140	Total 140	O 140	0	0
3	D	116	Total 116	O 116	0	0
3	G	1	Total 1	O 1	0	0
3	H	2	Total 2	O 2	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.

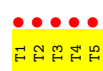
- Molecule 1: 5'-D(*TP*TP*TP*TP*T)-3'



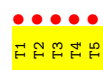
- Molecule 1: 5'-D(*TP*TP*TP*TP*T)-3'



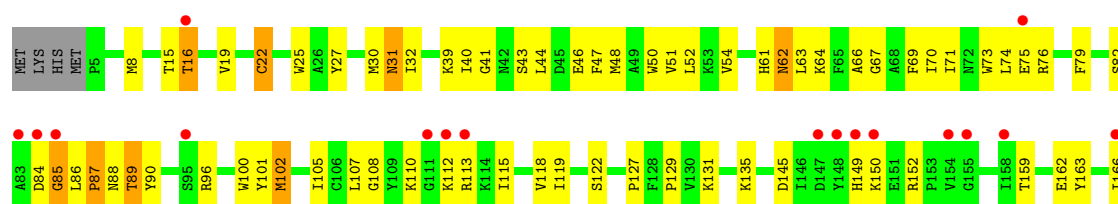
- Molecule 1: 5'-D(*TP*TP*TP*TP*T)-3'

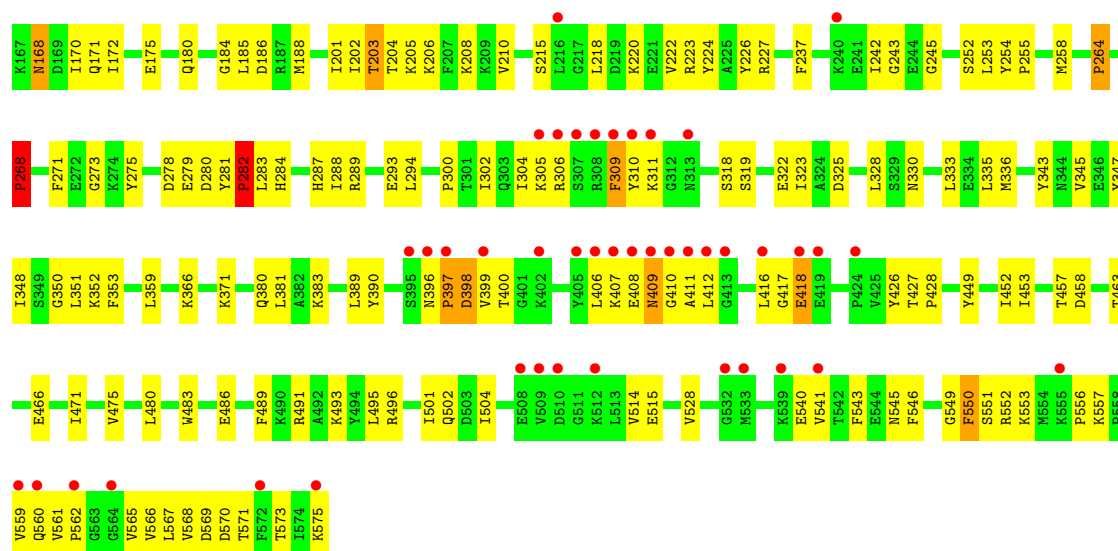


- Molecule 1: 5'-D(*TP*TP*TP*TP*T)-3'

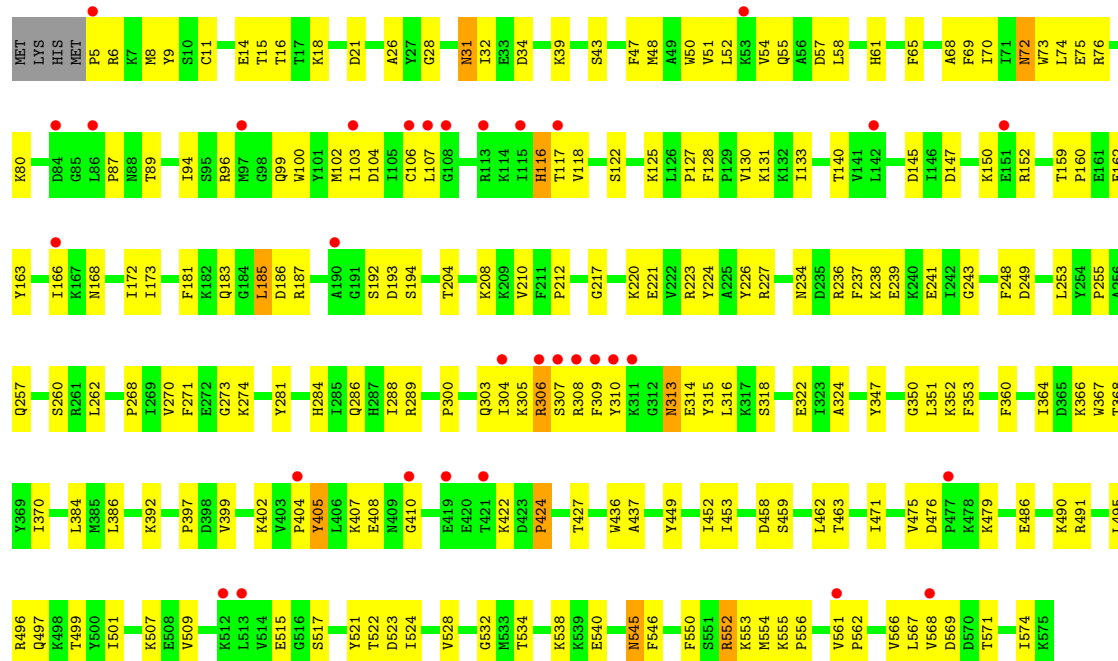


- Molecule 2: DNA polymerase

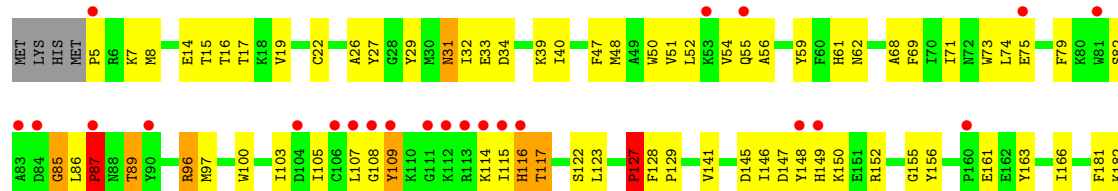


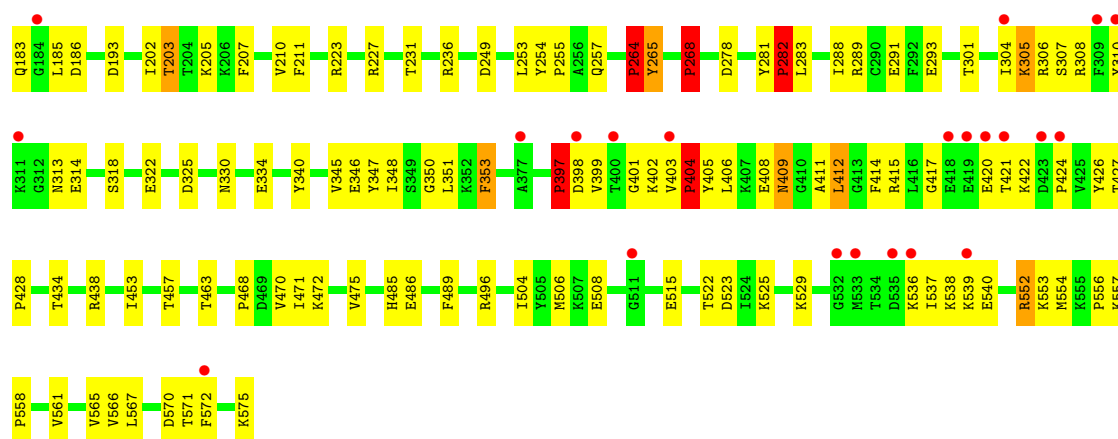


• Molecule 2: DNA polymerase

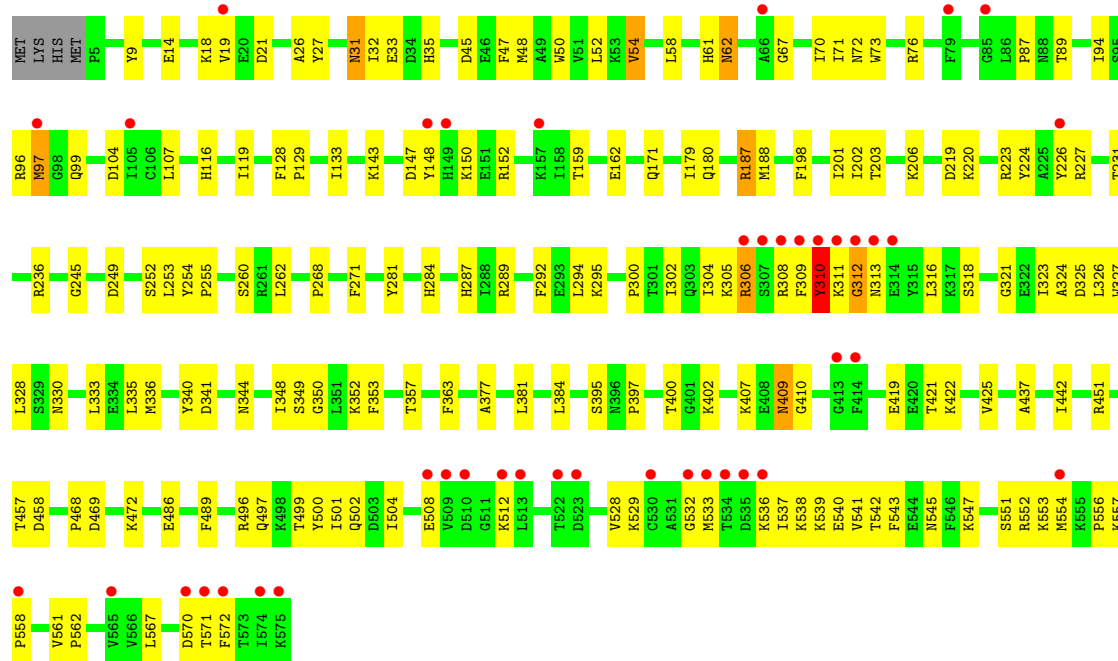


• Molecule 2: DNA polymerase





• Molecule 2: DNA polymerase



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	96.66Å 150.40Å 198.32Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.99 – 2.70 48.33 – 2.34	Depositor EDS
% Data completeness (in resolution range)	97.7 (19.99-2.70) 98.5 (48.33-2.34)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.82 (at 2.34Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.219 , 0.268 0.301 , 0.436	Depositor DCC
R_{free} test set	1799 reflections (2.25%)	DCC
Wilson B-factor (Å ²)	42.9	Xtriage
Anisotropy	0.250	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 46.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.88	EDS
Total number of atoms	19512	wwPDB-VP
Average B, all atoms (Å ²)	61.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.64% 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 [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	0.34	0/106	0.79	0/162
1	F	0.32	0/40	0.80	0/60
1	G	0.62	0/106	0.86	0/162
1	H	0.46	0/106	0.78	0/162
2	A	0.41	0/4788	0.69	4/6459 (0.1%)
2	B	0.39	0/4788	0.64	0/6459
2	C	0.43	0/4788	0.78	7/6459 (0.1%)
2	D	0.40	0/4788	0.65	0/6459
All	All	0.41	0/19510	0.70	11/26382 (0.0%)

There are no bond length outliers.

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	268	PRO	CA-N-CD	-16.41	88.52	111.50
2	C	264	PRO	CA-N-CD	-13.88	92.07	111.50
2	C	404	PRO	CA-N-CD	-13.55	92.53	111.50
2	A	397	PRO	CA-N-CD	-11.68	95.15	111.50
2	C	127	PRO	CA-N-CD	-11.42	95.51	111.50

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	E	97	0	62	10	0
1	F	37	0	26	1	0
1	G	97	0	62	13	0
1	H	97	0	62	12	0
2	A	4668	0	4676	189	0
2	B	4668	0	4676	149	0
2	C	4668	0	4676	195	0
2	D	4668	0	4676	152	0
3	A	124	0	0	4	0
3	B	129	0	0	6	0
3	C	140	0	0	2	0
3	D	116	0	0	6	0
3	G	1	0	0	0	0
3	H	2	0	0	0	0
All	All	19512	0	18916	692	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 18.

The worst 5 of 692 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)
2:A:264:PRO:HD3	2:A:283:LEU:CD1	1.41	1.47
2:A:264:PRO:CD	2:A:283:LEU:HD12	1.60	1.28
2:A:264:PRO:CD	2:A:283:LEU:CD1	2.18	1.18
1:G:3:DT:H5'	2:C:129:PRO:HG3	1.24	1.13
2:B:453:ILE:HD11	2:B:463:THR:HG23	1.20	1.12

There are no symmetry-related clashes.

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	
2	A	569/575 (99%)	506 (89%)	47 (8%)	16 (3%)	6	14
2	B	569/575 (99%)	513 (90%)	51 (9%)	5 (1%)	20	46
2	C	569/575 (99%)	518 (91%)	39 (7%)	12 (2%)	8	21
2	D	569/575 (99%)	519 (91%)	39 (7%)	11 (2%)	9	23
All	All	2276/2300 (99%)	2056 (90%)	176 (8%)	44 (2%)	9	23

5 of 44 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	A	62	ASN
2	A	85	GLY
2	A	309	PHE
2	A	408	GLU
2	A	185	LEU

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	
2	A	502/506 (99%)	487 (97%)	15 (3%)	46	76
2	B	502/506 (99%)	486 (97%)	16 (3%)	44	75
2	C	502/506 (99%)	479 (95%)	23 (5%)	31	61
2	D	502/506 (99%)	492 (98%)	10 (2%)	60	86
All	All	2008/2024 (99%)	1944 (97%)	64 (3%)	44	75

5 of 64 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	318	SER
2	C	96	ARG
2	D	188	MET
2	B	545	ASN
2	C	31	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 31 such sidechains are listed below:

Mol	Chain	Res	Type
2	C	35	HIS
2	C	171	GLN
2	D	409	ASN
2	C	55	GLN
2	C	313	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 [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	5/5 (100%)	11.51	5 (100%) 0 0	98, 113, 125, 129	0
1	F	2/5 (40%)	8.29	2 (100%) 0 0	128, 128, 128, 130	0
1	G	5/5 (100%)	9.33	5 (100%) 0 0	72, 77, 81, 98	0
1	H	5/5 (100%)	8.82	5 (100%) 0 0	65, 76, 100, 112	0
2	A	571/575 (99%)	0.77	60 (10%) 7 5	31, 65, 116, 130	0
2	B	571/575 (99%)	0.70	32 (5%) 25 23	29, 56, 96, 130	0
2	C	571/575 (99%)	0.65	45 (7%) 13 11	25, 54, 110, 130	0
2	D	571/575 (99%)	0.63	42 (7%) 15 13	28, 51, 86, 130	0
All	All	2301/2320 (99%)	0.75	196 (8%) 11 9	25, 56, 111, 130	0

The worst 5 of 196 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	5	DT	15.2
1	E	2	DT	12.4
1	E	3	DT	12.2
1	E	4	DT	10.0
1	G	1	DT	9.9

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.