



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 02:33 AM GMT

PDB ID : 2HMI
Title : HIV-1 REVERSE TRANSCRIPTASE/FRAGMENT OF FAB 28/DNA COMPLEX
Authors : Ding, J.; Arnold, E.
Deposited on : 1998-04-10
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
<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 : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
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) : trunk26865

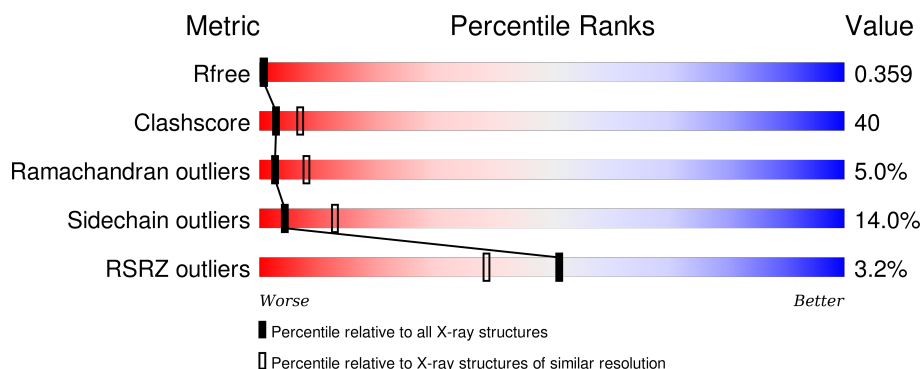
1 Overall quality at a glance ⓘ

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}	91344	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (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	19	<div> <div>5%</div> <div>95%</div> <div>5%</div> </div>
2	F	18	<div> <div>94%</div> <div>6%</div> </div>
3	A	558	<div> <div>5%</div> <div>41%</div> <div>49%</div> <div>9%</div> <div>•</div> </div>
4	B	430	<div> <div>3%</div> <div>40%</div> <div>50%</div> <div>10%</div> </div>
5	C	214	<div> <div>%</div> <div>38%</div> <div>52%</div> <div>10%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
6	D	220	<div><div><div>%</div><div><div></div><div>42%</div><div>48%</div><div>10%</div></div></div></div>

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 11647 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 (5'-D(*AP*TP*GP*GP*CP*GP*CP*CP*GP*AP*AP*CP*AP*GP*GP*GP*AP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	E	19	Total	C	N	O	P	0	0	0
			390	184	80	108	18			

- Molecule 2 is a DNA chain called DNA (5'-D(*GP*TP*CP*CP*CP*TP*GP*TP*TP*CP*GP*GP*GP*CP*GP*CP*CP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	F	18	Total	C	N	O	P	0	0	0
			363	173	64	109	17			

- Molecule 3 is a protein called SUBUNIT OF V-1 REVERSE TRANSCRIPTASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	558	Total	C	N	O	S	0	0	0
			4219	2728	703	781	7			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	280	SER	CYS	ENGINEERED	UNP P03366

- Molecule 4 is a protein called HISUBUNIT OF V-1 REVERSE TRANSCRIPTASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	B	430	Total	C	N	O	S	0	0	0
			3411	2216	568	620	7			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	280	SER	CYS	ENGINEERED	UNP P03366

- Molecule 5 is a protein called FAB FRAGMENT OF MONOCLONAL ANTIBODY 28.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	C	214	Total	C	N	O	S	0	0	0
			1616	1010	256	343	7			

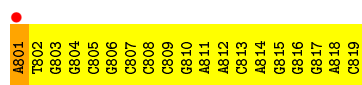
- Molecule 6 is a protein called FAB FRAGMENT OF MONOCLONAL ANTIBODY 28.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	D	220	Total	C	N	O	S	0	0	0
			1648	1037	270	333	8			

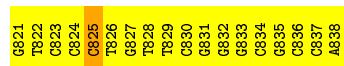
3 Residue-property plots

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.

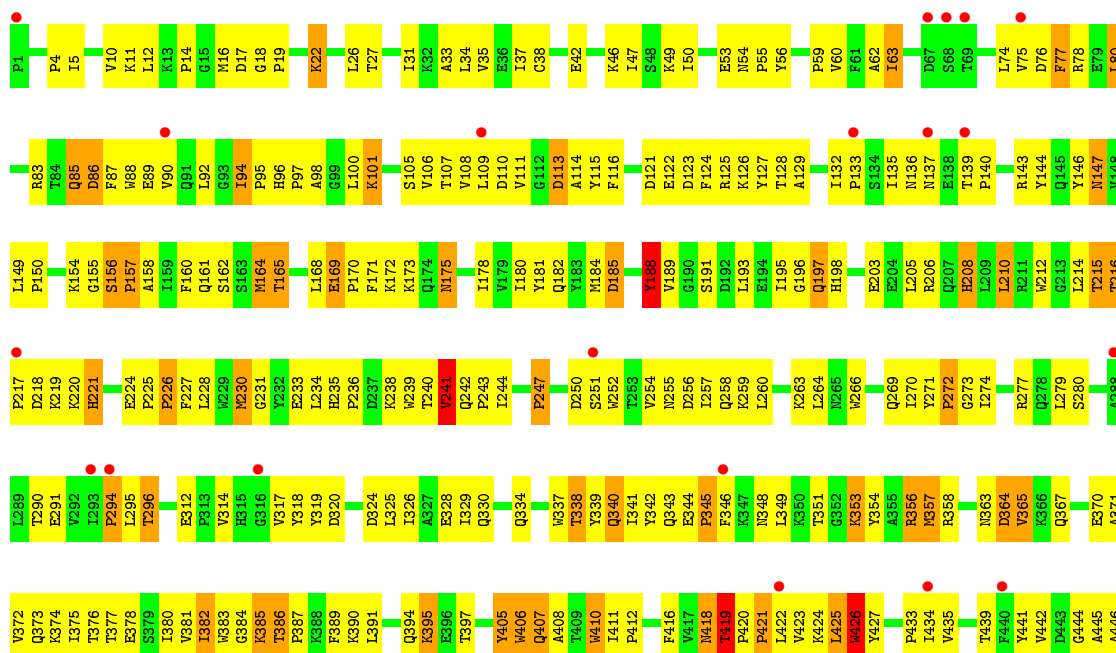
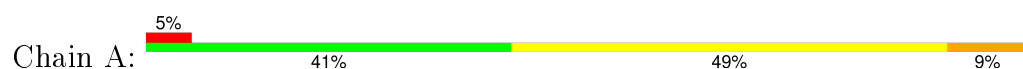
- Molecule 1: DNA (5'-D(*AP*TP*GP*GP*CP*GP*CP*CP*GP*AP*AP*CP*AP*GP*GP*GP*AP*C)-3')

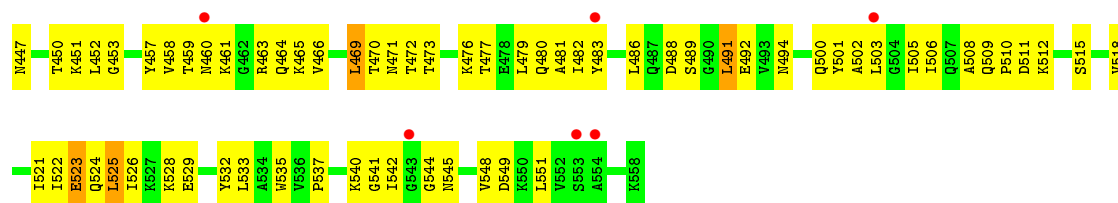


- Molecule 2: DNA (5'-D(*GP*TP*CP*CP*CP*TP*GP*TP*TP*CP*GP*GP*GP*CP*GP*CP*CP*A)-3')

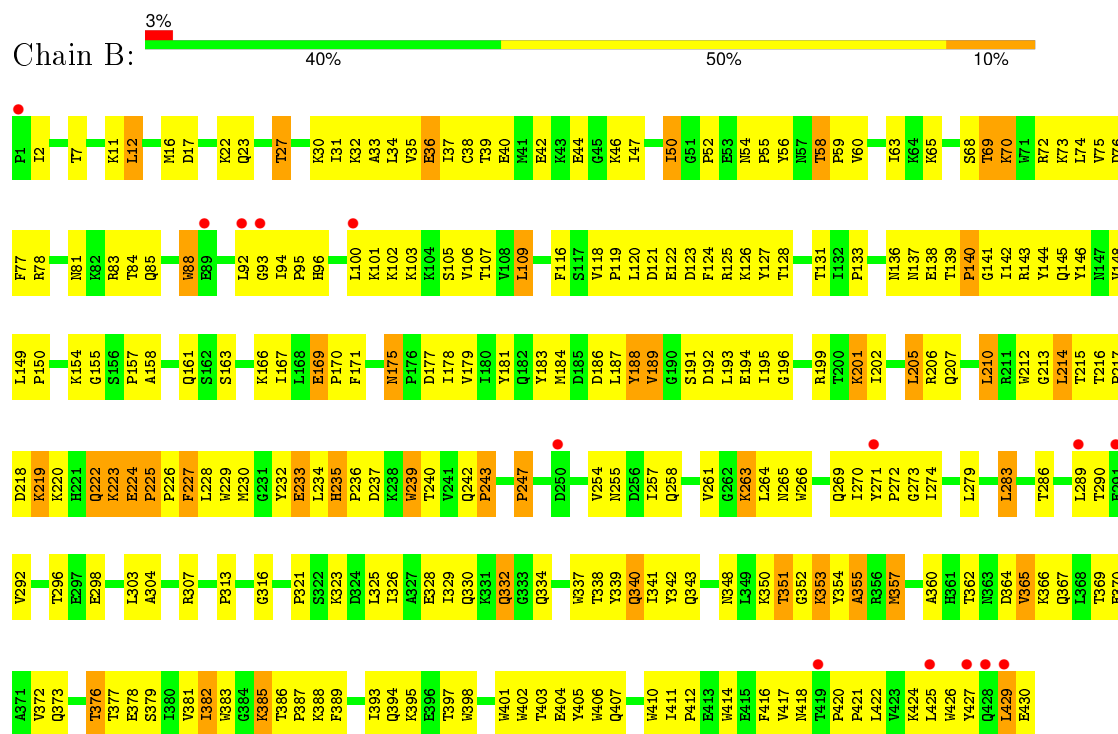


- Molecule 3: SUBUNIT OF V-1 REVERSE TRANSCRIPTASE

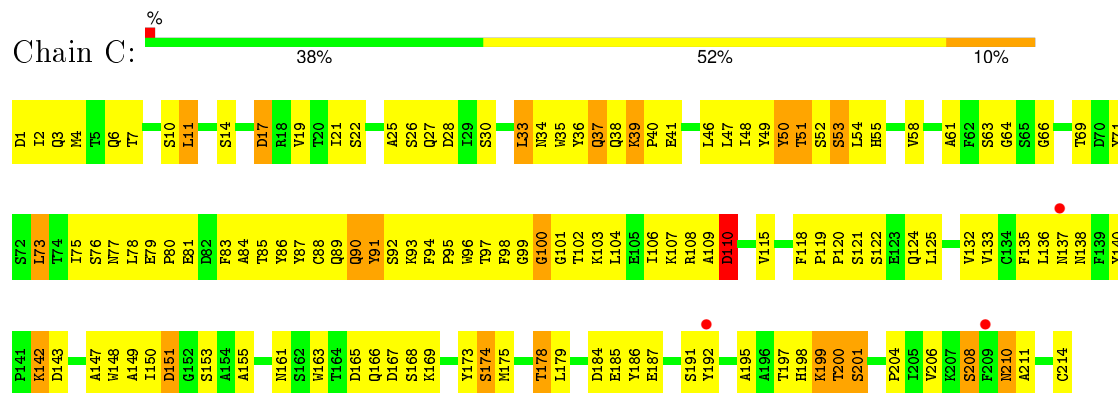




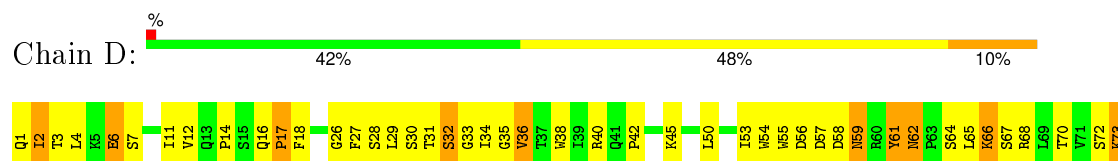
• Molecule 4: HISUBUNIT OF V-1 REVERSE TRANSCRIPTASE

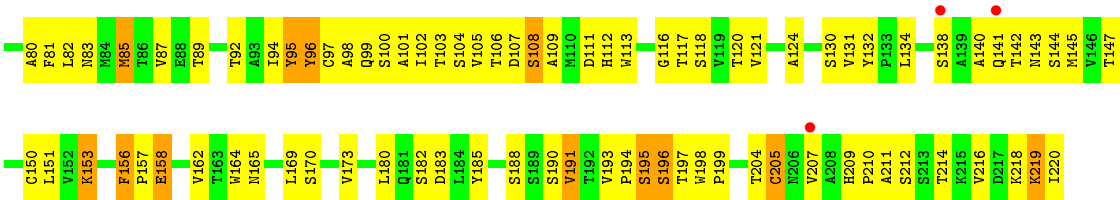


• Molecule 5: FAB FRAGMENT OF MONOCLONAL ANTIBODY 28



• Molecule 6: FAB FRAGMENT OF MONOCLONAL ANTIBODY 28





4 Data and refinement statistics

Property	Value	Source
Space group	P 32 1 2	Depositor
Cell constants a, b, c, α , β , γ	169.00Å 169.00Å 221.00Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	8.00 – 2.80 15.02 – 2.81	Depositor EDS
% Data completeness (in resolution range)	74.8 (8.00-2.80) 86.9 (15.02-2.81)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.96 (at 2.81Å)	Xtriage
Refinement program	X-PLOR 3.1	Depositor
R, R_{free}	0.271 , 0.352 0.287 , 0.359	Depositor DCC
R_{free} test set	3737 reflections (5.09%)	DCC
Wilson B-factor (Å ²)	39.7	Xtriage
Anisotropy	0.298	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.19 , 14.1	EDS
Estimated twinning fraction	0.069 for -h,-k,l	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.39$, $\langle L^2 \rangle = 0.21$	Xtriage
Outliers	0 of 76187 reflections	Xtriage
F_o, F_c correlation	0.81	EDS
Total number of atoms	11647	wwPDB-VP
Average B, all atoms (Å ²)	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.12% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical 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

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	0.62	0/439	0.84	0/676
2	F	0.49	0/405	0.84	1/623 (0.2%)
3	A	0.50	0/4330	0.47	2/5925 (0.0%)
4	B	0.63	0/3510	0.51	3/4783 (0.1%)
5	C	0.53	0/1654	0.50	0/2256
6	D	0.64	0/1691	0.53	0/2320
All	All	0.57	0/12029	0.53	6/16583 (0.0%)

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	1
2	F	1	0
3	A	0	1
4	B	0	1
5	C	0	2
All	All	1	5

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	B	243	PRO	N-CA-CB	6.05	110.56	103.30
4	B	247	PRO	N-CA-CB	5.96	110.45	103.30
3	A	247	PRO	N-CA-CB	5.95	110.44	103.30
4	B	313	PRO	N-CA-CB	5.74	110.19	103.30
3	A	294	PRO	N-CA-CB	5.74	110.19	103.30

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	F	825	DC	C3'

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	188	TYR	Sidechain
4	B	188	TYR	Sidechain
5	C	87	TYR	Sidechain
5	C	91	TYR	Sidechain
1	E	801	DA	Sidechain

5.2 Too-close contacts

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	390	0	212	53	0
2	F	363	0	204	80	0
3	A	4219	0	3981	321	0
4	B	3411	0	3298	267	0
5	C	1616	0	1517	109	0
6	D	1648	0	1602	115	0
All	All	11647	0	10814	897	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 40.

The worst 5 of 897 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:F:825:DC:C2'	2:F:826:DT:H5'	1.68	1.23
2:F:833:DG:H2''	2:F:834:DC:H5'	1.23	1.18
2:F:828:DT:H2'	2:F:829:DT:H71	1.32	1.12
1:E:802:DT:H2''	1:E:803:DG:H5'	1.31	1.10
2:F:823:DC:H2''	2:F:824:DC:H5''	1.27	1.09

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	
3	A	556/558 (100%)	428 (77%)	104 (19%)	24 (4%)	3	10
4	B	428/430 (100%)	337 (79%)	70 (16%)	21 (5%)	3	8
5	C	212/214 (99%)	171 (81%)	29 (14%)	12 (6%)	2	6
6	D	218/220 (99%)	173 (79%)	32 (15%)	13 (6%)	2	5
All	All	1414/1422 (99%)	1109 (78%)	235 (17%)	70 (5%)	3	8

5 of 70 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	63	ILE
3	A	85	GLN
3	A	136	ASN
3	A	247	PRO
3	A	251	SER

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	414/498 (83%)	356 (86%)	58 (14%)	4	13
4	B	350/392 (89%)	305 (87%)	45 (13%)	5	16
5	C	182/182 (100%)	152 (84%)	30 (16%)	3	8
6	D	191/191 (100%)	165 (86%)	26 (14%)	5	14
All	All	1137/1263 (90%)	978 (86%)	159 (14%)	4	13

5 of 159 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	B	169	GLU
4	B	283	LEU
6	D	100	SER
4	B	189	VAL
4	B	219	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 23 such sidechains are listed below:

Mol	Chain	Res	Type
4	B	54	ASN
4	B	255	ASN
6	D	41	GLN
4	B	222	GLN
4	B	334	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 ⓘ

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	19/19 (100%)	0.19	1 (5%) 30 20	54, 67, 83, 83	0
2	F	18/18 (100%)	-0.03	0 100 100	60, 70, 79, 82	0
3	A	558/558 (100%)	0.37	26 (4%) 35 24	3, 47, 71, 79	0
4	B	430/430 (100%)	0.29	14 (3%) 50 38	3, 21, 65, 91	0
5	C	214/214 (100%)	0.31	3 (1%) 78 69	7, 34, 61, 70	0
6	D	220/220 (100%)	0.19	3 (1%) 78 69	3, 23, 43, 68	0
All	All	1459/1459 (100%)	0.30	47 (3%) 51 39	3, 35, 68, 91	0

The worst 5 of 47 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	B	427	TYR	9.5
3	A	294	PRO	7.1
3	A	554	ALA	5.1
4	B	92	LEU	4.7
4	B	289	LEU	4.6

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 [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.