



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 5, 2016 – 03:49 AM EDT

PDB ID : 5K81
Title : Crystal Structure of a Primate APOBEC3G N-Terminal Domain
Authors : Xiao, X.; Li, S.-X.; Yang, H.; Chen, X.S.
Deposited on : 2016-05-27
Resolution : 2.01 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-20027939
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-20027939

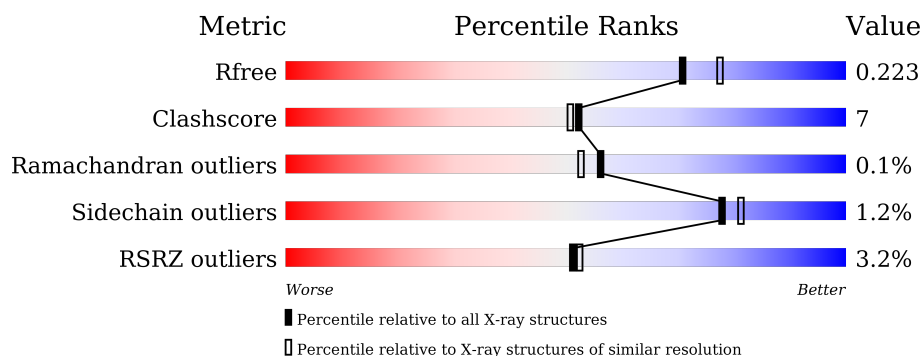
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.01 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	6249 (2.00-2.00)
Clashscore	102246	7340 (2.00-2.00)
Ramachandran outliers	100387	7248 (2.00-2.00)
Sidechain outliers	100360	7247 (2.00-2.00)
RSRZ outliers	91569	6262 (2.00-2.00)

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	196	<div> <div>2%</div> <div> <div></div> <div>81%</div> <div>16%</div> <div>• •</div> </div> </div>
1	B	196	<div> <div>3%</div> <div> <div></div> <div>84%</div> <div>13%</div> <div>• •</div> </div> </div>
1	C	196	<div> <div>%</div> <div> <div></div> <div>83%</div> <div>14%</div> <div>• •</div> </div> </div>
1	D	196	<div> <div>5%</div> <div> <div></div> <div>81%</div> <div>13%</div> <div>• 5%</div> </div> </div>
1	E	196	<div> <div>4%</div> <div> <div></div> <div>79%</div> <div>15%</div> <div>• 5%</div> </div> </div>
1	F	196	<div> <div>5%</div> <div> <div></div> <div>81%</div> <div>13%</div> <div>• 5%</div> </div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	ZN	B	401	-	-	-	X
2	ZN	E	401	-	-	-	X

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 10140 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 Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G, Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	191	Total	C	N	O	S	0	0	0
			1603	1045	279	268	11			
1	B	191	Total	C	N	O	S	0	0	0
			1603	1045	279	268	11			
1	C	191	Total	C	N	O	S	0	0	0
			1603	1045	279	268	11			
1	D	186	Total	C	N	O	S	0	0	0
			1561	1018	272	261	10			
1	E	186	Total	C	N	O	S	0	0	0
			1561	1018	272	261	10			
1	F	186	Total	C	N	O	S	0	0	0
			1561	1018	272	261	10			

There are 54 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	GLY	-	expression tag	UNP M1GSK9
A	-3	PRO	-	expression tag	UNP M1GSK9
A	-2	ALA	-	expression tag	UNP M1GSK9
A	-1	GLY	-	expression tag	UNP M1GSK9
A	0	SER	-	expression tag	UNP M1GSK9
A	143	ALA	-	linker	UNP M1GSK9
A	144	GLU	-	linker	UNP M1GSK9
A	145	ALA	-	linker	UNP M1GSK9
A	146	GLY	-	linker	UNP M1GSK9
B	-4	GLY	-	expression tag	UNP M1GSK9
B	-3	PRO	-	expression tag	UNP M1GSK9
B	-2	ALA	-	expression tag	UNP M1GSK9
B	-1	GLY	-	expression tag	UNP M1GSK9
B	0	SER	-	expression tag	UNP M1GSK9
B	143	ALA	-	linker	UNP M1GSK9
B	144	GLU	-	linker	UNP M1GSK9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	145	ALA	-	linker	UNP M1GSK9
B	146	GLY	-	linker	UNP M1GSK9
C	-4	GLY	-	expression tag	UNP M1GSK9
C	-3	PRO	-	expression tag	UNP M1GSK9
C	-2	ALA	-	expression tag	UNP M1GSK9
C	-1	GLY	-	expression tag	UNP M1GSK9
C	0	SER	-	expression tag	UNP M1GSK9
C	143	ALA	-	linker	UNP M1GSK9
C	144	GLU	-	linker	UNP M1GSK9
C	145	ALA	-	linker	UNP M1GSK9
C	146	GLY	-	linker	UNP M1GSK9
D	-4	GLY	-	expression tag	UNP M1GSK9
D	-3	PRO	-	expression tag	UNP M1GSK9
D	-2	ALA	-	expression tag	UNP M1GSK9
D	-1	GLY	-	expression tag	UNP M1GSK9
D	0	SER	-	expression tag	UNP M1GSK9
D	143	ALA	-	linker	UNP M1GSK9
D	144	GLU	-	linker	UNP M1GSK9
D	145	ALA	-	linker	UNP M1GSK9
D	146	GLY	-	linker	UNP M1GSK9
E	-4	GLY	-	expression tag	UNP M1GSK9
E	-3	PRO	-	expression tag	UNP M1GSK9
E	-2	ALA	-	expression tag	UNP M1GSK9
E	-1	GLY	-	expression tag	UNP M1GSK9
E	0	SER	-	expression tag	UNP M1GSK9
E	143	ALA	-	linker	UNP M1GSK9
E	144	GLU	-	linker	UNP M1GSK9
E	145	ALA	-	linker	UNP M1GSK9
E	146	GLY	-	linker	UNP M1GSK9
F	-4	GLY	-	expression tag	UNP M1GSK9
F	-3	PRO	-	expression tag	UNP M1GSK9
F	-2	ALA	-	expression tag	UNP M1GSK9
F	-1	GLY	-	expression tag	UNP M1GSK9
F	0	SER	-	expression tag	UNP M1GSK9
F	143	ALA	-	linker	UNP M1GSK9
F	144	GLU	-	linker	UNP M1GSK9
F	145	ALA	-	linker	UNP M1GSK9
F	146	GLY	-	linker	UNP M1GSK9

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	D	1	Total 1	Zn 1	0	0
2	E	1	Total 1	Zn 1	0	0
2	B	1	Total 1	Zn 1	0	0
2	C	1	Total 1	Zn 1	0	0
2	A	1	Total 1	Zn 1	0	0
2	F	1	Total 1	Zn 1	0	0

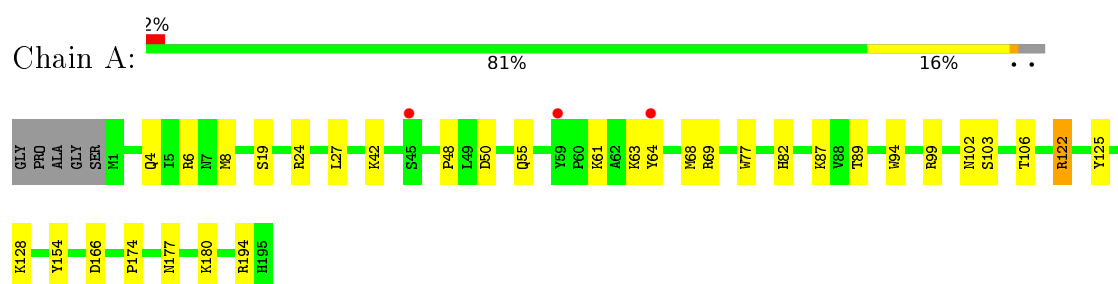
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	121	Total 121	O 121	0	0
3	B	117	Total 117	O 117	0	0
3	C	125	Total 125	O 125	0	0
3	D	89	Total 89	O 89	0	0
3	E	99	Total 99	O 99	0	0
3	F	91	Total 91	O 91	0	0

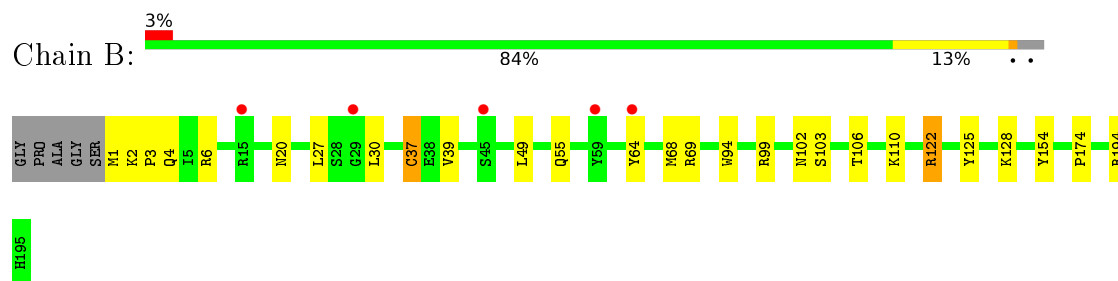
3 Residue-property plots [i](#)

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

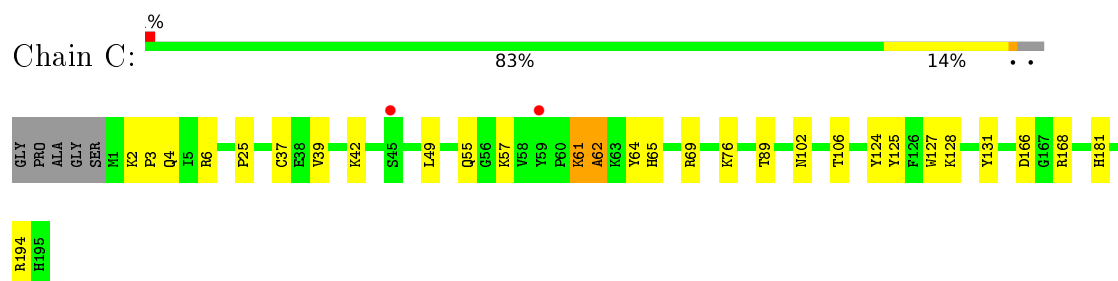
- Molecule 1: Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G, Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G



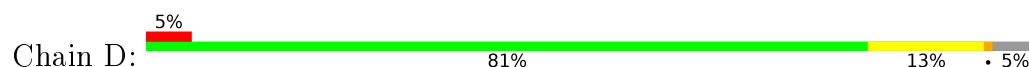
- Molecule 1: Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G, Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G

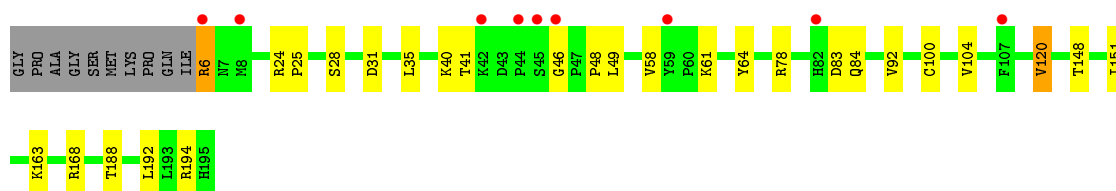


- Molecule 1: Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G, Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G

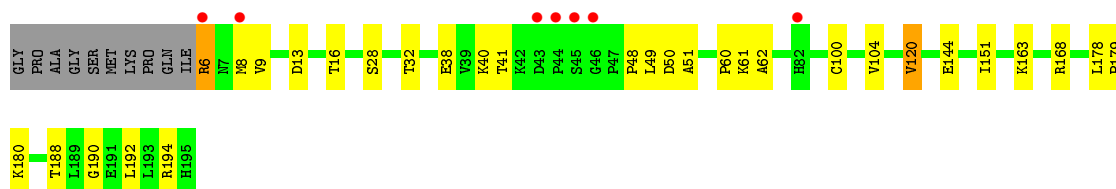
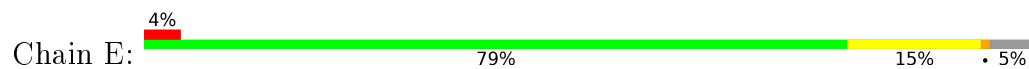


- Molecule 1: Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G, Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G

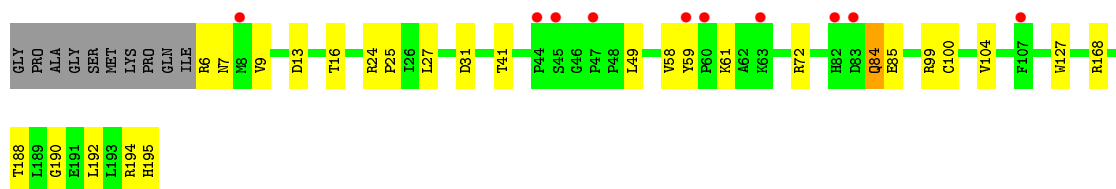
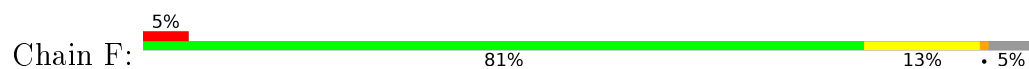




- Molecule 1: Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G, Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G



- Molecule 1: Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G, Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	106.21 Å 81.30 Å 106.30 Å 90.00° 120.03° 90.00°	Depositor
Resolution (Å)	46.01 – 2.01 46.01 – 2.01	Depositor EDS
% Data completeness (in resolution range)	94.8 (46.01-2.01) 94.4 (46.01-2.01)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.63 (at 2.01 Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, R_{free}	0.188 , 0.223 0.185 , 0.223	Depositor DCC
R_{free} test set	1966 reflections (1.98%)	DCC
Wilson B-factor (Å ²)	28.6	Xtriage
Anisotropy	0.294	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 33.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.469 for -h-l,k,h 0.469 for l,k,-h-l 0.036 for l,-k,h 0.035 for h,-k,-h-l 0.035 for -h-l,-k,l	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	10140	wwPDB-VP
Average B, all atoms (Å ²)	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.10% 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 ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	0.48	0/1659	0.60	2/2252 (0.1%)
1	B	0.49	1/1659 (0.1%)	0.61	2/2252 (0.1%)
1	C	0.49	0/1659	0.57	0/2252
1	D	0.47	0/1616	0.54	0/2196
1	E	0.48	0/1616	0.56	0/2196
1	F	0.47	0/1616	0.56	0/2196
All	All	0.48	1/9825 (0.0%)	0.58	4/13344 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	37	CYS	CB-SG	-5.04	1.73	1.81

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	122	ARG	NE-CZ-NH1	-10.18	115.21	120.30
1	A	122	ARG	NE-CZ-NH1	-9.98	115.31	120.30
1	B	122	ARG	NE-CZ-NH2	8.01	124.30	120.30
1	A	122	ARG	NE-CZ-NH2	6.57	123.58	120.30

There are no chirality outliers.

There are no planarity outliers.

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	A	1603	0	1587	27	1
1	B	1603	0	1587	19	1
1	C	1603	0	1587	21	2
1	D	1561	0	1536	21	1
1	E	1561	0	1536	29	1
1	F	1561	0	1537	18	1
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
2	E	1	0	0	0	0
2	F	1	0	0	0	0
3	A	121	0	0	6	0
3	B	117	0	0	3	1
3	C	125	0	0	4	2
3	D	89	0	0	6	0
3	E	99	0	0	7	0
3	F	91	0	0	4	2
All	All	10140	0	9370	127	6

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 7.

The worst 5 of 127 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:C:4:GLN:OE1	1:C:6:ARG:NH1	1.93	1.01
1:B:4:GLN:OE1	1:B:6:ARG:NH1	1.96	0.97
1:A:4:GLN:OE1	1:A:6:ARG:NH1	2.01	0.93
1:E:6:ARG:NH2	1:E:50:ASP:O	2.10	0.84
1:F:7:ASN:OD1	3:F:501:HOH:O	1.99	0.79

The worst 5 of 6 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)
1:C:64:TYR:OH	1:D:168:ARG:NH1[1_556]	1.70	0.50
1:A:64:TYR:OH	1:E:168:ARG:NH1[1_454]	2.01	0.19
1:B:64:TYR:OH	1:F:168:ARG:NH1[2_555]	2.04	0.16

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:76:LYS:NZ	3:F:502:HOH:O[2_555]	2.18	0.02
3:C:592:HOH:O	3:F:566:HOH:O[2_555]	2.19	0.01

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	
1	A	189/196 (96%)	187 (99%)	2 (1%)	0	100	100
1	B	189/196 (96%)	186 (98%)	3 (2%)	0	100	100
1	C	189/196 (96%)	185 (98%)	3 (2%)	1 (0%)	34	26
1	D	184/196 (94%)	181 (98%)	3 (2%)	0	100	100
1	E	184/196 (94%)	181 (98%)	3 (2%)	0	100	100
1	F	184/196 (94%)	182 (99%)	2 (1%)	0	100	100
All	All	1119/1176 (95%)	1102 (98%)	16 (1%)	1 (0%)	56	53

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	62	ALA

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.

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	173/175 (99%)	172 (99%)	1 (1%)	90	93
1	B	173/175 (99%)	172 (99%)	1 (1%)	90	93
1	C	173/175 (99%)	171 (99%)	2 (1%)	78	81
1	D	168/175 (96%)	166 (99%)	2 (1%)	78	81
1	E	168/175 (96%)	165 (98%)	3 (2%)	66	69
1	F	168/175 (96%)	165 (98%)	3 (2%)	66	69
All	All	1023/1050 (97%)	1011 (99%)	12 (1%)	78	81

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	120	VAL
1	E	6	ARG
1	F	6	ARG
1	D	6	ARG
1	E	120	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	F	7	ASN

5.3.3 RNA ⓘ

There are no RNA molecules 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

Of 6 ligands modelled in this entry, 6 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	A	191/196 (97%)	0.00	3 (1%) 74 75	18, 32, 56, 68	0
1	B	191/196 (97%)	0.02	5 (2%) 59 60	18, 32, 55, 69	0
1	C	191/196 (97%)	0.02	2 (1%) 84 84	18, 32, 58, 67	0
1	D	186/196 (94%)	0.24	9 (4%) 34 36	19, 36, 69, 87	0
1	E	186/196 (94%)	0.22	7 (3%) 44 45	18, 35, 68, 91	0
1	F	186/196 (94%)	0.26	10 (5%) 29 31	18, 35, 69, 88	0
All	All	1131/1176 (96%)	0.13	36 (3%) 51 52	18, 34, 64, 91	0

The worst 5 of 36 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	6	ARG	5.4
1	F	44	PRO	4.9
1	E	45	SER	4.5
1	D	59	TYR	4.3
1	F	59	TYR	3.9

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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å ²)	Q<0.9
2	ZN	B	401	1/1	1.00	0.14	7.65	24,24,24,24	0
2	ZN	E	401	1/1	0.98	0.13	3.27	37,37,37,37	0
2	ZN	A	401	1/1	0.99	0.13	0.99	26,26,26,26	0
2	ZN	D	401	1/1	0.99	0.12	0.24	36,36,36,36	0
2	ZN	C	401	1/1	0.99	0.11	0.15	26,26,26,26	0
2	ZN	F	401	1/1	0.99	0.12	0.12	36,36,36,36	0

6.5 Other polymers [i](#)

There are no such residues in this entry.