wwPDB X-ray Structure Validation Summary Report (i)

Sep 14, 2023 – 10:02 AM EDT

PDB ID	:	2GP1
Title	:	Bacteriophage HK97 Prohead II crystal structure
Authors	:	Gertsman, I.; Gan, L.; Johnson, J.E.
Deposited on	:	2006-04-15
Resolution	:	5.20 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 (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

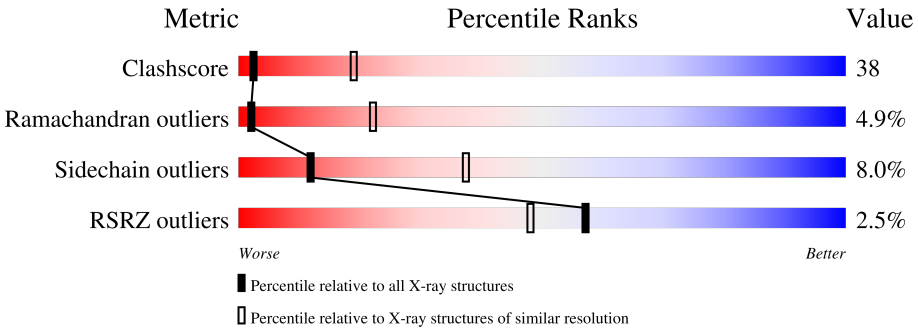
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 5.20 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))
Clashscore	141614	1006 (6.56-3.84)
Ramachandran outliers	138981	1173 (6.60-3.80)
Sidechain outliers	138945	1148 (6.60-3.80)
RSRZ outliers	127900	1008 (6.64-3.74)

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 of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. 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 <=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	282	3% 42% 37% 9% • 10%
1	B	282	3% 52% 33% • • 10%
1	C	282	3% 49% 35% 5% 10%
1	D	282	% 45% 41% • 10%
1	E	282	3% 49% 35% 6% 10%
1	F	282	% 47% 37% 6% • 10%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	282	

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 13797 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 Major capsid protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	254	Total	C	N	O	S	0	0	0
			1971	1233	345	383	10			
1	B	254	Total	C	N	O	S	0	0	0
			1971	1233	345	383	10			
1	C	254	Total	C	N	O	S	0	0	0
			1971	1233	345	383	10			
1	D	254	Total	C	N	O	S	0	0	0
			1971	1233	345	383	10			
1	E	254	Total	C	N	O	S	0	0	0
			1971	1233	345	383	10			
1	F	254	Total	C	N	O	S	0	0	0
			1971	1233	345	383	10			
1	G	254	Total	C	N	O	S	0	0	0
			1971	1233	345	383	10			

There are 7 discrepancies between the modelled and reference sequences:

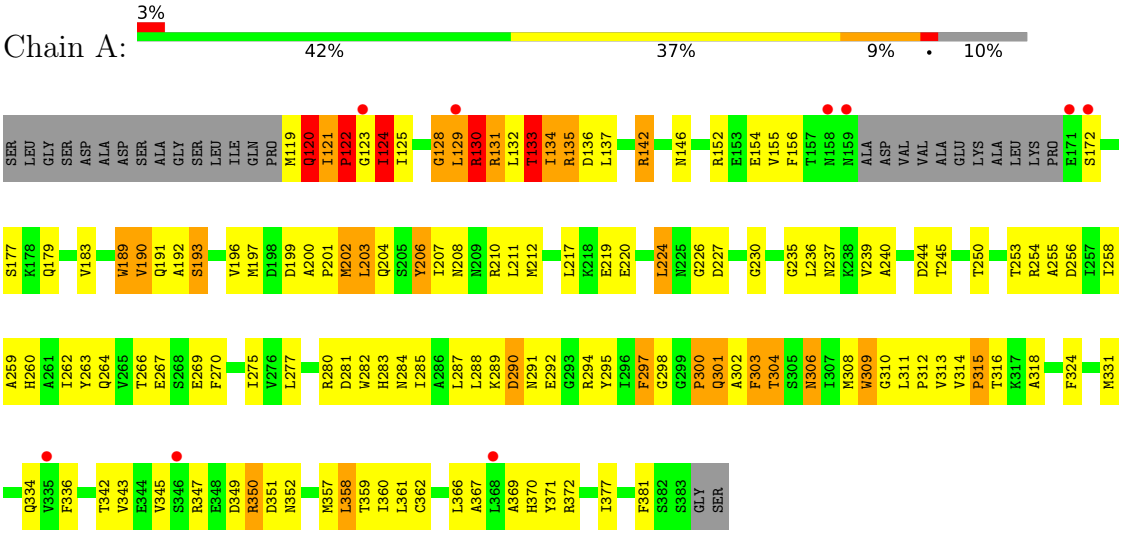
Chain	Residue	Modelled	Actual	Comment	Reference
A	336	PHE	TRP	engineered mutation	UNP P49861
B	336	PHE	TRP	engineered mutation	UNP P49861
C	336	PHE	TRP	engineered mutation	UNP P49861
D	336	PHE	TRP	engineered mutation	UNP P49861
E	336	PHE	TRP	engineered mutation	UNP P49861
F	336	PHE	TRP	engineered mutation	UNP P49861
G	336	PHE	TRP	engineered mutation	UNP P49861

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

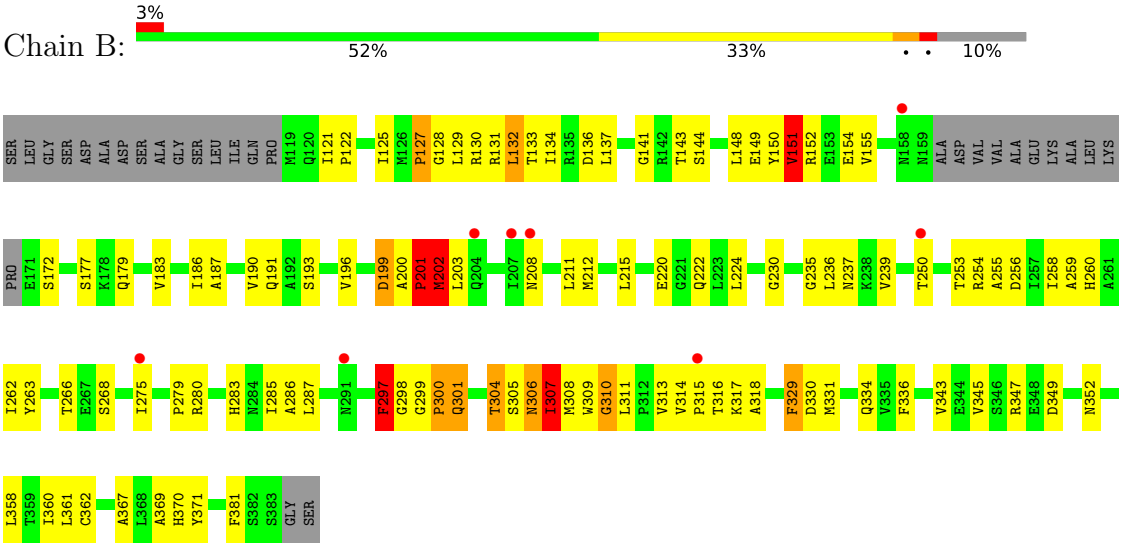
• Molecule 1: Major capsid protein

Chain A:



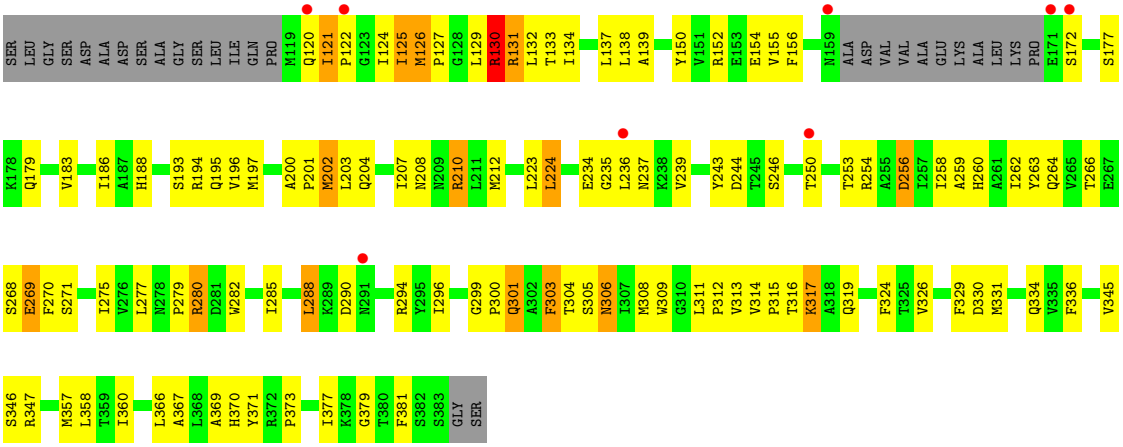
• Molecule 1: Major capsid protein

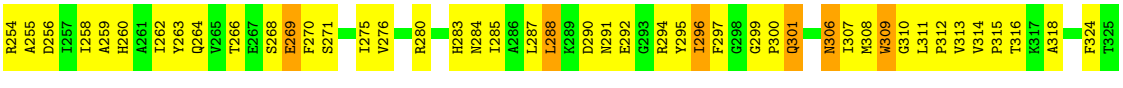
Chain B:

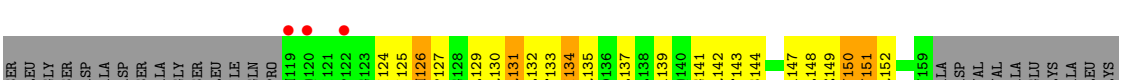


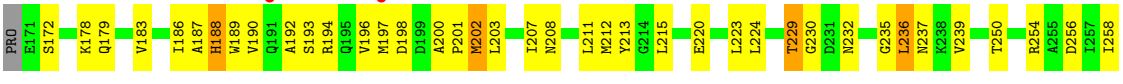
• Molecule 1: Major capsid protein

Chain C:



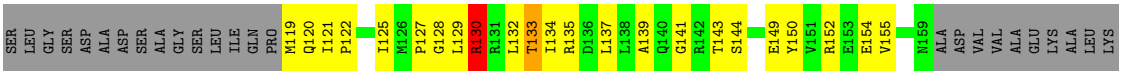


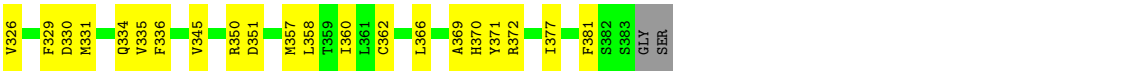


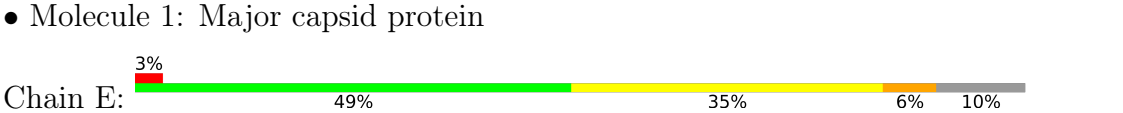


• Molecule 1: Major capsid protein

Chain D:





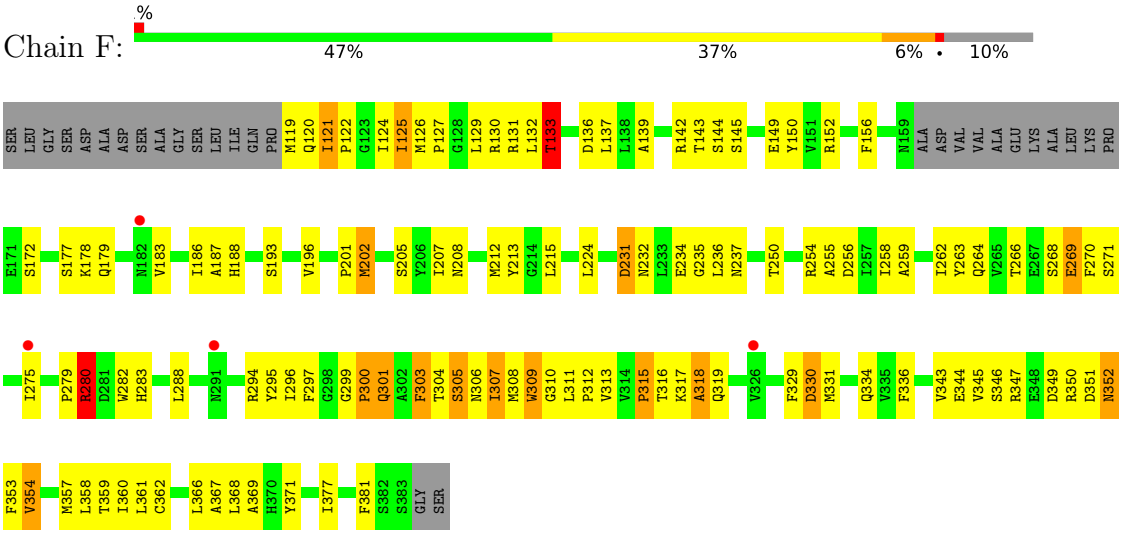


• Molecule 1: Major capsid protein

Chain E:

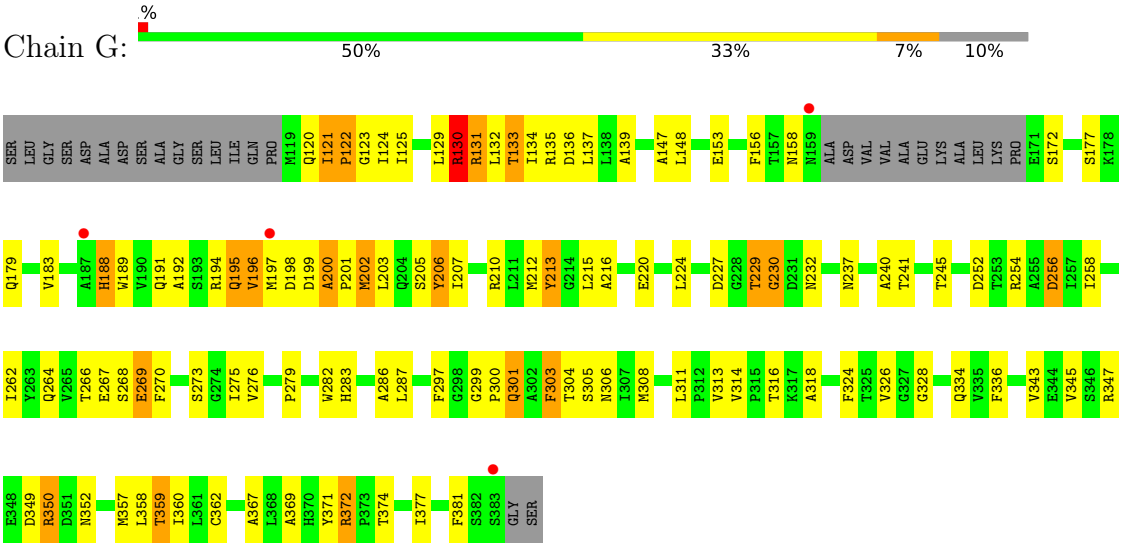
• Molecule 1: Major capsid protein

Chain F:



• Molecule 1: Major capsid protein

Chain G:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 3	Depositor
Cell constants a, b, c, α, β, γ	706.94Å 706.94Å 706.94Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 5.20 29.98 – 5.20	Depositor EDS
% Data completeness (in resolution range)	(Not available) (30.00-5.20) 84.2 (29.98-5.20)	Depositor EDS
R_{merge}	0.18	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.45 (at 5.34Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.410 , (Not available) 0.371 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	124.3	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 52.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.37$, $< L^2 > = 0.20$	Xtriage
Estimated twinning fraction	0.076 for l,-k,h	Xtriage
F_o,F_c correlation	0.70	EDS
Total number of atoms	13797	wwPDB-VP
Average B, all atoms (Å^2)	111.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.16% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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	A	0.55	4/2004 (0.2%)	0.79	6/2715 (0.2%)
1	B	0.35	0/2004	0.64	1/2715 (0.0%)
1	C	0.35	0/2004	0.62	0/2715
1	D	0.36	0/2004	0.63	0/2715
1	E	0.36	0/2004	0.61	0/2715
1	F	0.61	2/2004 (0.1%)	0.68	3/2715 (0.1%)
1	G	0.36	0/2004	0.65	0/2715
All	All	0.43	6/14028 (0.0%)	0.66	10/19005 (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	A	0	1
1	G	0	1
All	All	0	2

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	231	ASP	C-N	-21.70	0.84	1.34
1	A	133	THR	N-CA	-7.58	1.31	1.46
1	A	120	GLN	C-N	7.04	1.50	1.34
1	F	280	ARG	N-CA	6.10	1.58	1.46
1	A	128	GLY	N-CA	-5.97	1.37	1.46

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	122	PRO	CA-N-CD	-9.35	98.41	111.50
1	A	132	LEU	C-N-CA	-7.34	103.34	121.70

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	123	GLY	C-N-CA	7.02	139.26	121.70
1	F	318	ALA	N-CA-C	-5.91	95.03	111.00
1	A	133	THR	CA-C-N	-5.89	104.24	117.20

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	206	TYR	Sidechain
1	G	213	TYR	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	A	1971	0	1937	223	0
1	B	1971	0	1938	142	0
1	C	1971	0	1938	174	0
1	D	1971	0	1937	180	0
1	E	1971	0	1938	198	0
1	F	1971	0	1936	195	0
1	G	1971	0	1938	127	0
All	All	13797	0	13562	1026	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 38.

The worst 5 of 1026 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:A:280:ARG:HH22	1:B:262:ILE:CD1	1.21	1.51
1:E:279:PRO:CB	1:F:266:THR:HG21	1.51	1.39
1:A:280:ARG:NH2	1:B:262:ILE:HD13	1.37	1.37
1:C:279:PRO:CB	1:D:266:THR:HG21	1.55	1.36
1:D:202:MET:SD	1:E:143:THR:HA	1.71	1.30

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
1	A	250/282 (89%)	198 (79%)	37 (15%)	15 (6%)	1 18
1	B	250/282 (89%)	201 (80%)	32 (13%)	17 (7%)	1 16
1	C	250/282 (89%)	212 (85%)	30 (12%)	8 (3%)	4 29
1	D	250/282 (89%)	205 (82%)	33 (13%)	12 (5%)	2 22
1	E	250/282 (89%)	206 (82%)	34 (14%)	10 (4%)	3 25
1	F	250/282 (89%)	201 (80%)	39 (16%)	10 (4%)	3 25
1	G	250/282 (89%)	200 (80%)	36 (14%)	14 (6%)	2 20
All	All	1750/1974 (89%)	1423 (81%)	241 (14%)	86 (5%)	2 22

5 of 86 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	122	PRO
1	A	124	ILE
1	A	135	ARG
1	A	193	SER
1	A	350	ARG

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
1	A	211/231 (91%)	188 (89%)	23 (11%)	6 25
1	B	211/231 (91%)	198 (94%)	13 (6%)	18 45
1	C	211/231 (91%)	191 (90%)	20 (10%)	8 29
1	D	211/231 (91%)	200 (95%)	11 (5%)	23 49
1	E	211/231 (91%)	195 (92%)	16 (8%)	13 39
1	F	211/231 (91%)	195 (92%)	16 (8%)	13 39
1	G	211/231 (91%)	192 (91%)	19 (9%)	9 32
All	All	1477/1617 (91%)	1359 (92%)	118 (8%)	12 37

5 of 118 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	133	THR
1	G	206	TYR
1	E	172	SER
1	G	205	SER
1	G	120	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 46 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	208	ASN
1	F	237	ASN
1	E	232	ASN
1	F	146	ASN
1	F	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 monosaccharides 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)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	F	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	F	231:ASP	C	232:ASN	N	0.84

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, 95^{th} 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	254/282 (90%)	0.08	9 (3%)	44	36	47, 95, 95, 95	0
1	B	254/282 (90%)	0.14	8 (3%)	49	39	54, 114, 114, 114	0
1	C	254/282 (90%)	0.12	8 (3%)	49	39	57, 122, 122, 122	0
1	D	254/282 (90%)	0.01	3 (1%)	79	70	50, 103, 103, 103	0
1	E	254/282 (90%)	0.13	8 (3%)	49	39	54, 114, 114, 114	0
1	F	254/282 (90%)	0.14	4 (1%)	72	62	51, 104, 104, 104	0
1	G	254/282 (90%)	0.03	4 (1%)	72	62	61, 132, 132, 132	0
All	All	1778/1974 (90%)	0.09	44 (2%)	57	48	47, 114, 132, 132	0

The worst 5 of 44 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	171	GLU	5.0
1	G	383	SER	4.9
1	B	291	ASN	4.7
1	A	159	ASN	4.6
1	B	207	ILE	3.7

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

There are no monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.