wwPDB X-ray Structure Validation Summary Report (i)

Apr 29, 2025 – 03:24 AM EDT

PDB ID	:	3E8K / pdb_00003e8k
Title	:	Crystal structure of HK97 Prohead II
Authors	:	Gertsman, I.; Speir, J.; Johnson, J.E.
Deposited on	:	2008-08-20
Resolution	:	3.65 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 (1)) were used in the production of this report:

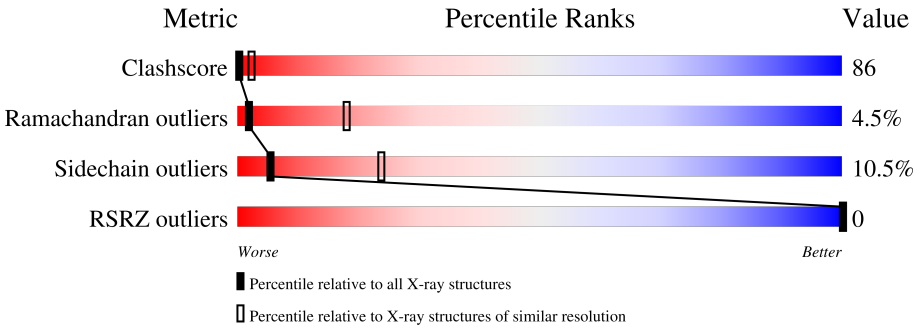
MolProbity	:	4-5-2 with Phenix2.0rc1
Xtriage (Phenix)	:	2.0rc1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.006 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.43.1

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 3.65 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	180529	1328 (3.80-3.52)
Ramachandran outliers	177936	1306 (3.80-3.52)
Sidechain outliers	177891	1303 (3.80-3.52)
RSRZ outliers	164620	1260 (3.80-3.52)

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	273	29% 51% 12% • 6%
1	B	273	26% 48% 16% • 8%
1	C	273	29% 51% 10% • 9%
1	D	273	30% 51% 11% • 7%
1	E	273	22% 57% 12% • 7%
1	F	273	19% 54% 15% • 9%
1	G	273	27% 49% 13% • 10%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 13590 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	256	Total	C	N	O	S	0	0	0
			1978	1238	346	384	10			
1	B	250	Total	C	N	O	S	0	0	0
			1934	1209	339	377	9			
1	C	248	Total	C	N	O	S	0	0	0
			1918	1198	337	375	8			
1	D	254	Total	C	N	O	S	0	0	0
			1961	1228	343	381	9			
1	E	255	Total	C	N	O	S	0	0	0
			1970	1233	345	383	9			
1	F	248	Total	C	N	O	S	0	0	0
			1918	1198	337	375	8			
1	G	247	Total	C	N	O	S	0	0	0
			1911	1193	336	374	8			

There are 35 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	159	ALA	-	linker	UNP P49861
A	160	PRO	-	linker	UNP P49861
A	161	GLY	-	linker	UNP P49861
A	162	ASP	-	linker	UNP P49861
A	336	PHE	TRP	engineered mutation	UNP P49861
B	159	ALA	-	linker	UNP P49861
B	160	PRO	-	linker	UNP P49861
B	161	GLY	-	linker	UNP P49861
B	162	ASP	-	linker	UNP P49861
B	336	PHE	TRP	engineered mutation	UNP P49861
C	159	ALA	-	linker	UNP P49861
C	160	PRO	-	linker	UNP P49861
C	161	GLY	-	linker	UNP P49861
C	162	ASP	-	linker	UNP P49861
C	336	PHE	TRP	engineered mutation	UNP P49861

Continued on next page...

Continued from previous page...

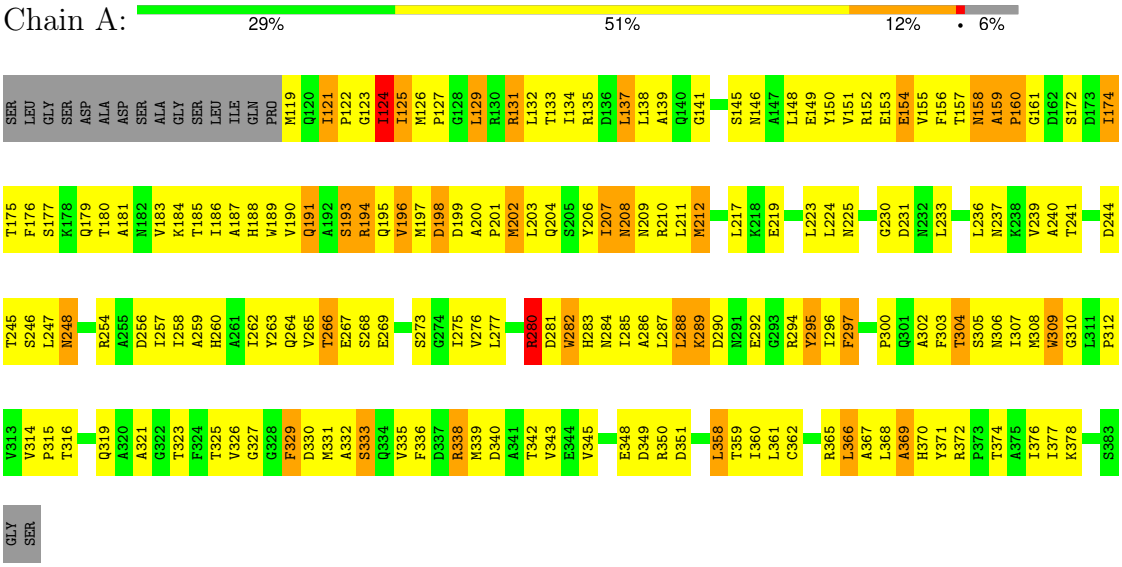
Chain	Residue	Modelled	Actual	Comment	Reference
D	159	ALA	-	linker	UNP P49861
D	160	PRO	-	linker	UNP P49861
D	161	GLY	-	linker	UNP P49861
D	162	ASP	-	linker	UNP P49861
D	336	PHE	TRP	engineered mutation	UNP P49861
E	159	ALA	-	linker	UNP P49861
E	160	PRO	-	linker	UNP P49861
E	161	GLY	-	linker	UNP P49861
E	162	ASP	-	linker	UNP P49861
E	336	PHE	TRP	engineered mutation	UNP P49861
F	159	ALA	-	linker	UNP P49861
F	160	PRO	-	linker	UNP P49861
F	161	GLY	-	linker	UNP P49861
F	162	ASP	-	linker	UNP P49861
F	336	PHE	TRP	engineered mutation	UNP P49861
G	159	ALA	-	linker	UNP P49861
G	160	PRO	-	linker	UNP P49861
G	161	GLY	-	linker	UNP P49861
G	162	ASP	-	linker	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.

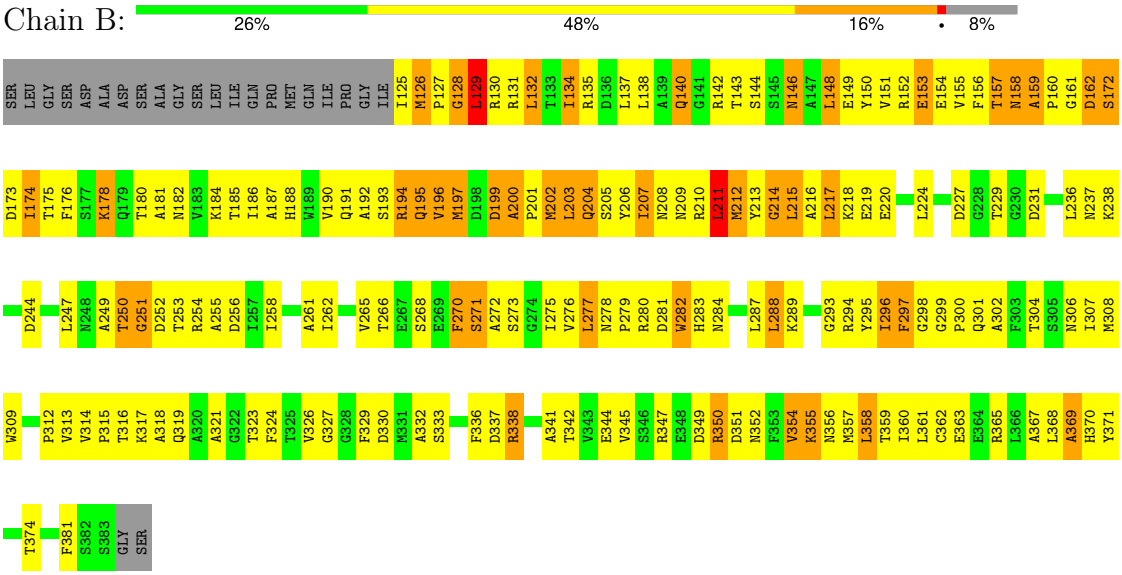
• Molecule 1: Major capsid protein

Chain A:



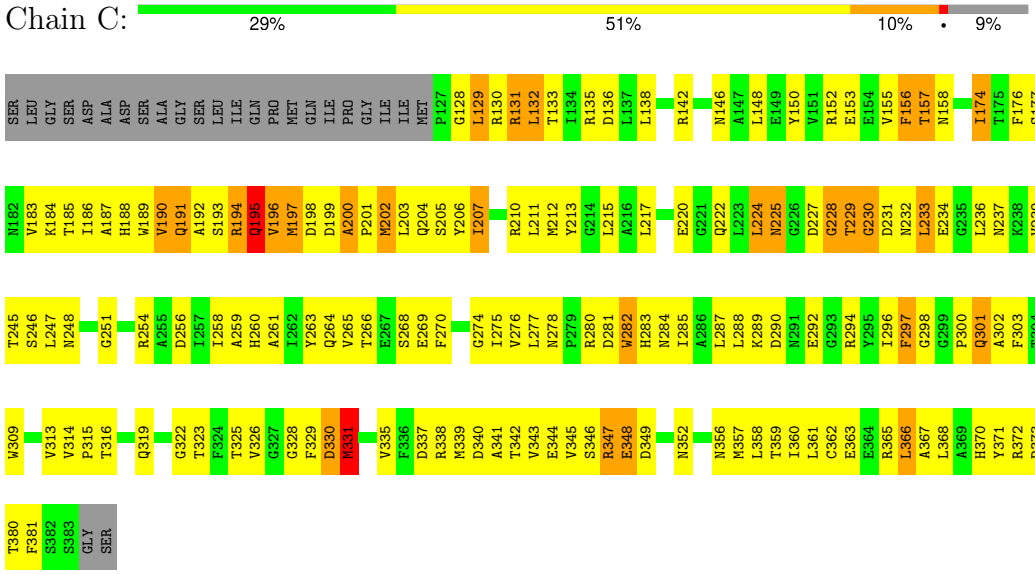
• Molecule 1: Major capsid protein

Chain B:



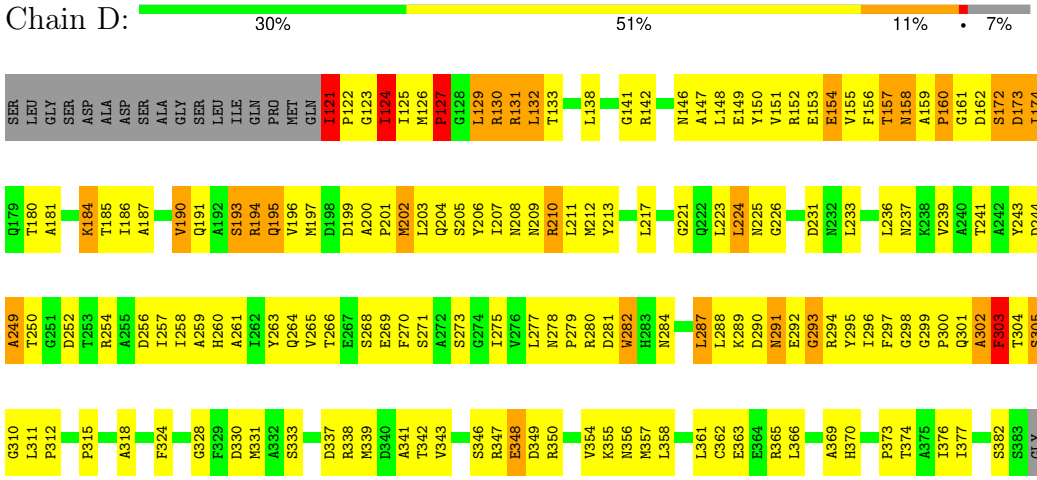
• Molecule 1: Major capsid protein

Chain C:



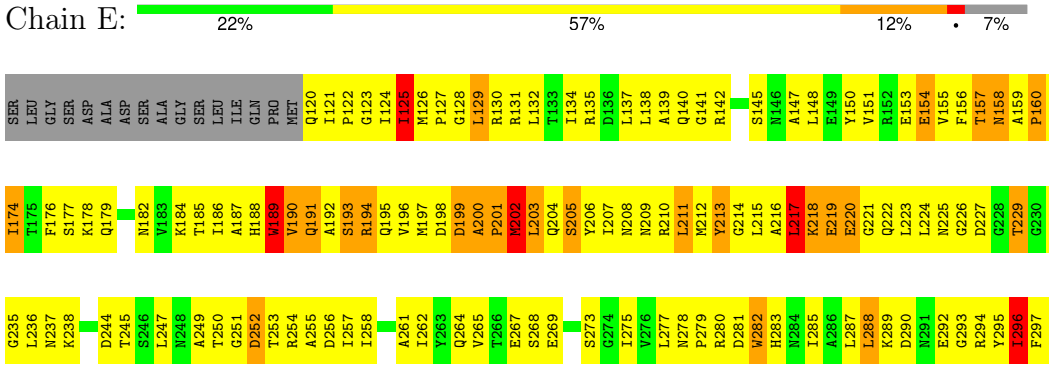
• Molecule 1: Major capsid protein

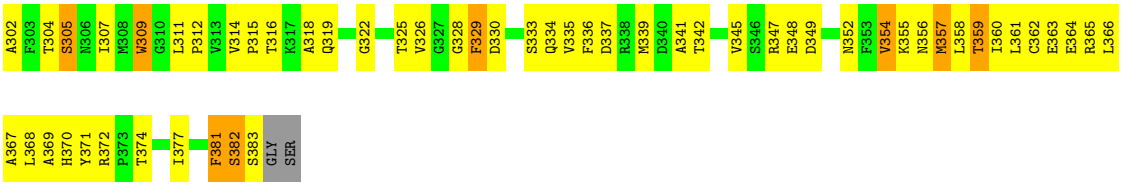
Chain D:



• Molecule 1: Major capsid protein

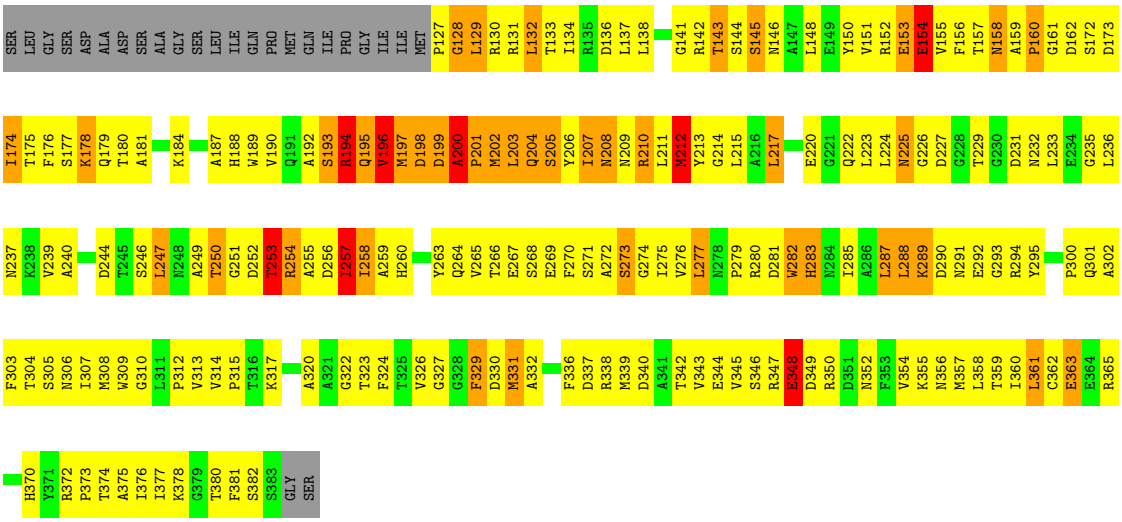
Chain E:





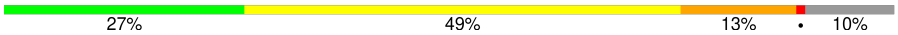
• Molecule 1: Major capsid protein

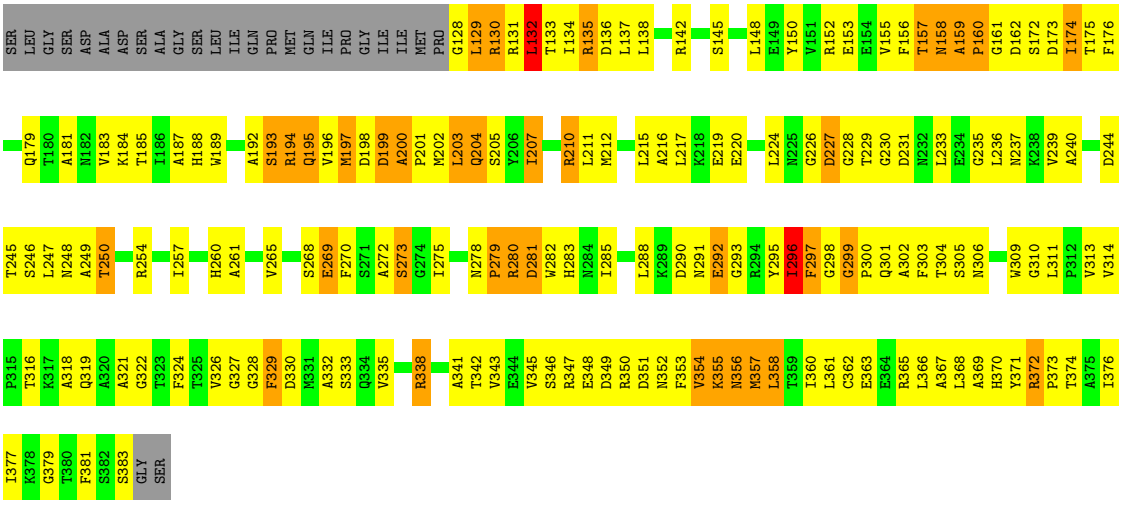
Chain F:



• Molecule 1: Major capsid protein

Chain G:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	553.03Å 574.39Å 587.36Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	35.00 – 3.65 35.00 – 3.65	Depositor EDS
% Data completeness (in resolution range)	64.8 (35.00-3.65) 54.3 (35.00-3.65)	Depositor EDS
R_{merge}	0.18	Depositor
R_{sym}	0.18	Depositor
$< I/\sigma(I) >$ [1]	1.97 (at 3.66Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.366 , (Not available) 0.311 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	72.5	Xtriage
Anisotropy	0.013	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.23 , 0.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.23$, $< L^2 > = 0.09$	Xtriage
Estimated twinning fraction	0.276 for -h,-l,-k	Xtriage
F_o,F_c correlation	0.51	EDS
Total number of atoms	13590	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.58% 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	0.59	3/2013 (0.1%)	0.99	13/2730 (0.5%)
1	B	0.56	3/1968 (0.2%)	1.17	22/2669 (0.8%)
1	C	0.49	1/1952 (0.1%)	0.95	6/2647 (0.2%)
1	D	0.49	2/1996 (0.1%)	1.03	19/2708 (0.7%)
1	E	0.56	2/2005 (0.1%)	1.07	16/2720 (0.6%)
1	F	0.57	0/1952	1.15	20/2647 (0.8%)
1	G	0.52	0/1944	1.19	24/2636 (0.9%)
All	All	0.54	11/13830 (0.1%)	1.08	120/18757 (0.6%)

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

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	329	PHE	C-N	-9.49	1.21	1.33
1	A	212	MET	C-N	-9.44	1.21	1.33
1	A	338	ARG	C-N	8.97	1.44	1.33
1	E	288	LEU	C-N	7.97	1.44	1.33
1	D	288	LEU	C-N	-7.91	1.22	1.33

The worst 5 of 120 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	G	195	GLN	N-CA-C	-17.61	91.46	111.71
1	G	207	ILE	N-CA-C	15.32	125.16	110.30
1	D	250	THR	N-CA-C	14.47	127.09	111.03
1	G	228	GLY	N-CA-C	13.27	133.63	115.32

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	202	MET	N-CA-C	13.11	125.27	110.97

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	369	ALA	Mainchain

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	1978	0	1946	320	0
1	B	1934	0	1897	362	0
1	C	1918	0	1878	354	0
1	D	1961	0	1929	347	0
1	E	1970	0	1937	389	0
1	F	1918	0	1878	372	0
1	G	1911	0	1870	300	0
All	All	13590	0	13335	2329	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 86.

The worst 5 of 2329 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:E:127:PRO:HB2	1:E:213:TYR:CD1	1.32	1.62
1:D:125:ILE:HD11	1:D:208:ASN:CG	1.32	1.54
1:D:125:ILE:HD11	1:D:208:ASN:ND2	1.26	1.45
1:C:196:VAL:HG12	1:C:203:LEU:CD1	1.44	1.45
1:B:201:PRO:O	1:B:204:GLN:CB	1.67	1.41

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	254/273 (93%)	195 (77%)	49 (19%)	10 (4%)	2 21
1	B	248/273 (91%)	189 (76%)	48 (19%)	11 (4%)	2 19
1	C	246/273 (90%)	188 (76%)	47 (19%)	11 (4%)	2 18
1	D	252/273 (92%)	209 (83%)	34 (14%)	9 (4%)	3 22
1	E	253/273 (93%)	190 (75%)	49 (19%)	14 (6%)	1 15
1	F	246/273 (90%)	192 (78%)	40 (16%)	14 (6%)	1 15
1	G	245/273 (90%)	199 (81%)	36 (15%)	10 (4%)	2 20
All	All	1744/1911 (91%)	1362 (78%)	303 (17%)	79 (4%)	2 18

5 of 79 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	124	ILE
1	A	208	ASN
1	B	200	ALA
1	B	297	PHE
1	C	131	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	211/223 (95%)	188 (89%)	23 (11%)	5 23
1	B	206/223 (92%)	186 (90%)	20 (10%)	6 26

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	C	204/223 (92%)	184 (90%)	20 (10%)	6 26
1	D	209/223 (94%)	192 (92%)	17 (8%)	9 33
1	E	210/223 (94%)	188 (90%)	22 (10%)	5 24
1	F	204/223 (92%)	169 (83%)	35 (17%)	1 10
1	G	203/223 (91%)	188 (93%)	15 (7%)	11 35
All	All	1447/1561 (93%)	1295 (90%)	152 (10%)	5 24

5 of 152 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	197	MET
1	G	245	THR
1	F	210	ARG
1	F	289	LYS
1	G	357	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 58 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	284	ASN
1	G	283	HIS
1	D	225	ASN
1	G	260	HIS
1	F	306	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 oligosaccharides 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, 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	256/273 (93%)	-1.58	0	100	100	1, 28, 53, 72	0
1	B	250/273 (91%)	-1.58	0	100	100	1, 31, 59, 78	0
1	C	248/273 (90%)	-1.52	0	100	100	7, 33, 59, 69	0
1	D	254/273 (93%)	-1.56	0	100	100	1, 32, 54, 79	0
1	E	255/273 (93%)	-1.57	0	100	100	5, 34, 57, 90	0
1	F	248/273 (90%)	-1.60	0	100	100	1, 31, 54, 75	0
1	G	247/273 (90%)	-1.55	0	100	100	3, 31, 57, 87	0
All	All	1758/1911 (91%)	-1.57	0	100	100	1, 31, 57, 90	0

There are no RSRZ outliers to report.

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.