wwPDB X-ray Structure Validation Summary Report (i)

Jun 15, 2024 – 07:56 PM EDT

PDB ID	:	4U5Z
Title	:	Trichodysplasia spinulosa-associated polyomavirus (TSPyV) VP1
Authors	:	Stroh, L.J.; Stehle, T.
Deposited on	:	2014-07-26
Resolution	:	2.10 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:

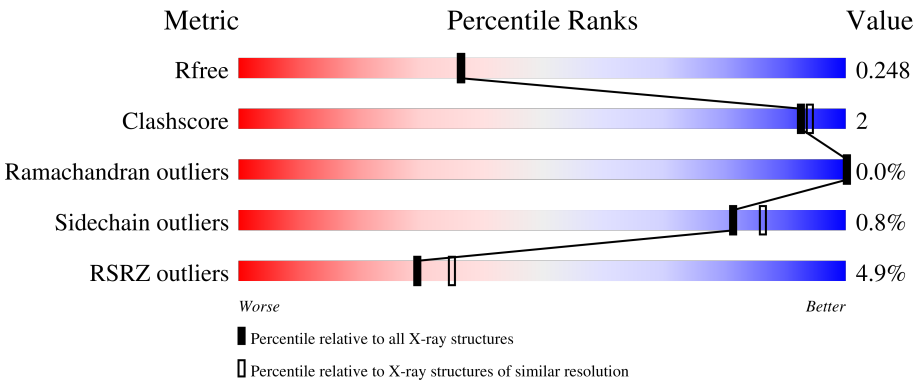
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.10 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}	130704	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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	280	5%, 92%, • •
1	B	280	4%, 90%, 5% 5%
1	C	280	2%, 92%, 5% •
1	D	280	5%, 90%, • 6%
1	E	280	4%, 89%, 5% 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	280	6% 90% 5% 5%
1	G	280	4% 88% 7% 5%
1	H	280	4% 90% • 6%
1	I	280	8% 90% • 7%
1	J	280	5% 89% 6% 6%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 20926 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 Structural protein VP1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	271	2096	1318	352	412	14	0	0	0
1	B	266	2063	1298	346	406	13	0	0	0
1	C	270	2092	1314	351	413	14	0	1	0
1	D	263	2017	1274	338	393	12	0	0	0
1	E	265	2045	1291	341	401	12	0	0	0
1	F	265	2028	1275	340	400	13	0	0	0
1	G	265	2030	1280	340	397	13	0	0	0
1	H	264	2038	1281	341	403	13	0	0	0
1	I	260	1974	1245	327	390	12	0	0	0
1	J	264	1992	1257	331	392	12	0	0	0

There are 60 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	24	GLY	-	expression tag	UNP E2ESL7
A	25	SER	-	expression tag	UNP E2ESL7
A	26	HIS	-	expression tag	UNP E2ESL7
A	27	MET	-	expression tag	UNP E2ESL7
A	28	ALA	-	expression tag	UNP E2ESL7
A	29	SER	-	expression tag	UNP E2ESL7
B	24	GLY	-	expression tag	UNP E2ESL7
B	25	SER	-	expression tag	UNP E2ESL7
B	26	HIS	-	expression tag	UNP E2ESL7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	27	MET	-	expression tag	UNP E2ESL7
B	28	ALA	-	expression tag	UNP E2ESL7
B	29	SER	-	expression tag	UNP E2ESL7
C	24	GLY	-	expression tag	UNP E2ESL7
C	25	SER	-	expression tag	UNP E2ESL7
C	26	HIS	-	expression tag	UNP E2ESL7
C	27	MET	-	expression tag	UNP E2ESL7
C	28	ALA	-	expression tag	UNP E2ESL7
C	29	SER	-	expression tag	UNP E2ESL7
D	24	GLY	-	expression tag	UNP E2ESL7
D	25	SER	-	expression tag	UNP E2ESL7
D	26	HIS	-	expression tag	UNP E2ESL7
D	27	MET	-	expression tag	UNP E2ESL7
D	28	ALA	-	expression tag	UNP E2ESL7
D	29	SER	-	expression tag	UNP E2ESL7
E	24	GLY	-	expression tag	UNP E2ESL7
E	25	SER	-	expression tag	UNP E2ESL7
E	26	HIS	-	expression tag	UNP E2ESL7
E	27	MET	-	expression tag	UNP E2ESL7
E	28	ALA	-	expression tag	UNP E2ESL7
E	29	SER	-	expression tag	UNP E2ESL7
F	24	GLY	-	expression tag	UNP E2ESL7
F	25	SER	-	expression tag	UNP E2ESL7
F	26	HIS	-	expression tag	UNP E2ESL7
F	27	MET	-	expression tag	UNP E2ESL7
F	28	ALA	-	expression tag	UNP E2ESL7
F	29	SER	-	expression tag	UNP E2ESL7
G	24	GLY	-	expression tag	UNP E2ESL7
G	25	SER	-	expression tag	UNP E2ESL7
G	26	HIS	-	expression tag	UNP E2ESL7
G	27	MET	-	expression tag	UNP E2ESL7
G	28	ALA	-	expression tag	UNP E2ESL7
G	29	SER	-	expression tag	UNP E2ESL7
H	24	GLY	-	expression tag	UNP E2ESL7
H	25	SER	-	expression tag	UNP E2ESL7
H	26	HIS	-	expression tag	UNP E2ESL7
H	27	MET	-	expression tag	UNP E2ESL7
H	28	ALA	-	expression tag	UNP E2ESL7
H	29	SER	-	expression tag	UNP E2ESL7
I	24	GLY	-	expression tag	UNP E2ESL7
I	25	SER	-	expression tag	UNP E2ESL7
I	26	HIS	-	expression tag	UNP E2ESL7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
I	27	MET	-	expression tag	UNP E2ESL7
I	28	ALA	-	expression tag	UNP E2ESL7
I	29	SER	-	expression tag	UNP E2ESL7
J	24	GLY	-	expression tag	UNP E2ESL7
J	25	SER	-	expression tag	UNP E2ESL7
J	26	HIS	-	expression tag	UNP E2ESL7
J	27	MET	-	expression tag	UNP E2ESL7
J	28	ALA	-	expression tag	UNP E2ESL7
J	29	SER	-	expression tag	UNP E2ESL7

- Molecule 2 is water.

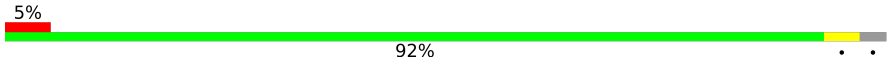
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	86	86	86	0	0
			Total	O		
2	B	89	89	89	0	0
			Total	O		
2	C	88	88	88	0	0
			Total	O		
2	D	48	48	48	0	0
			Total	O		
2	E	89	89	89	0	0
			Total	O		
2	F	26	26	26	0	0
			Total	O		
2	G	41	41	41	0	0
			Total	O		
2	H	34	34	34	0	0
			Total	O		
2	I	17	17	17	0	0
			Total	O		
2	J	33	33	33	0	0

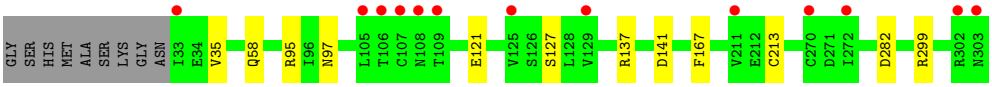
3 Residue-property plots (i)

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: Structural protein VP1

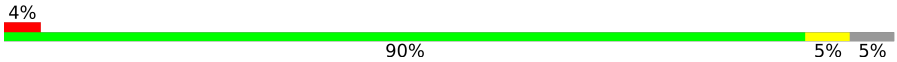
Chain A:

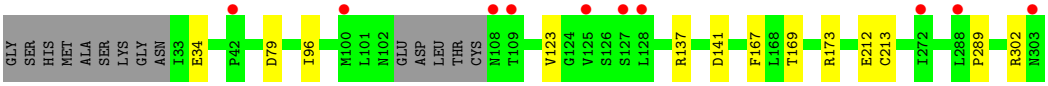




• Molecule 1: Structural protein VP1

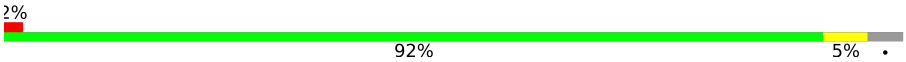
Chain B:

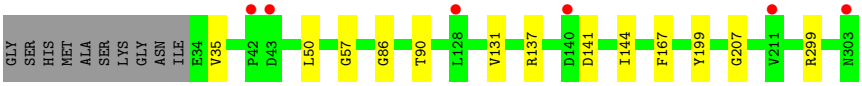




• Molecule 1: Structural protein VP1

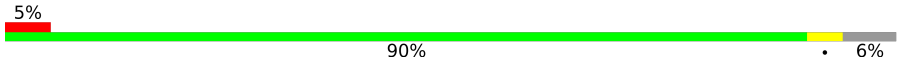
Chain C:

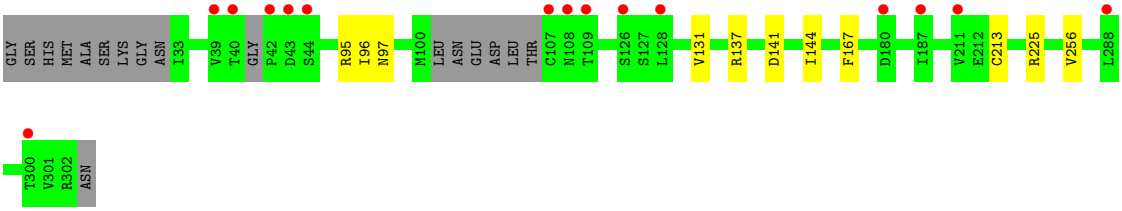




• Molecule 1: Structural protein VP1

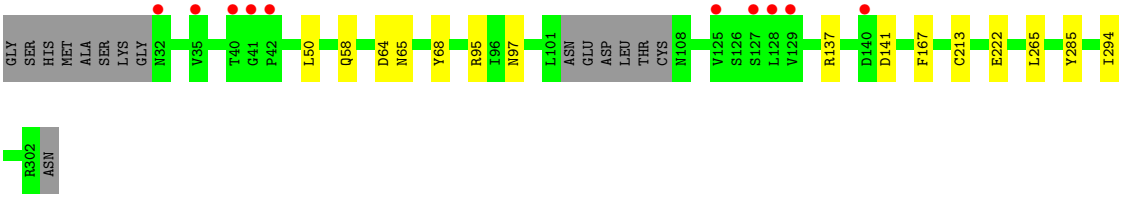
Chain D:





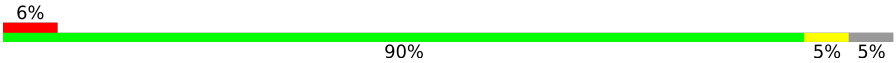
• Molecule 1: Structural protein VP1

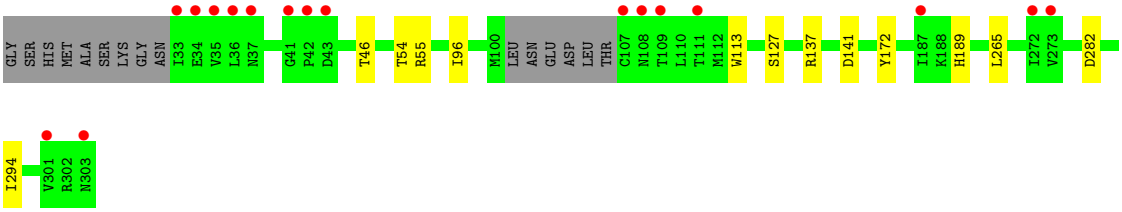
Chain E:



• Molecule 1: Structural protein VP1

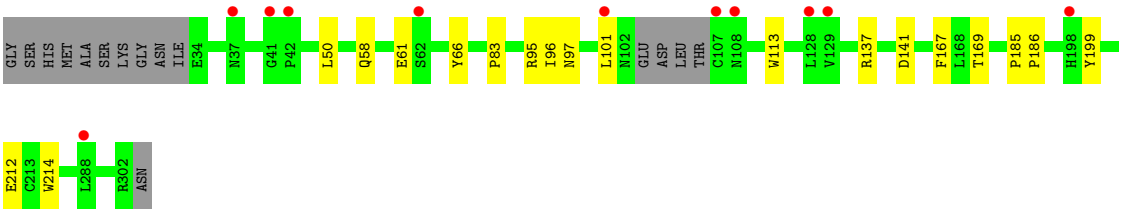
Chain F:





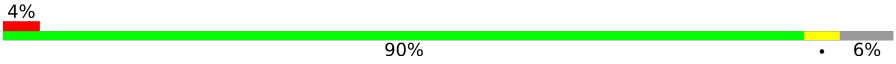
• Molecule 1: Structural protein VP1

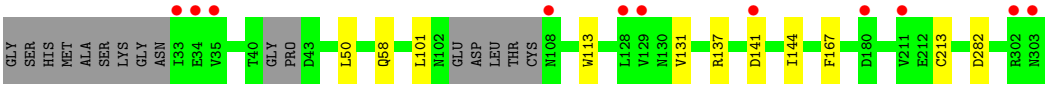
Chain G:



• Molecule 1: Structural protein VP1

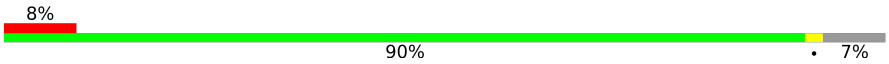
Chain H:

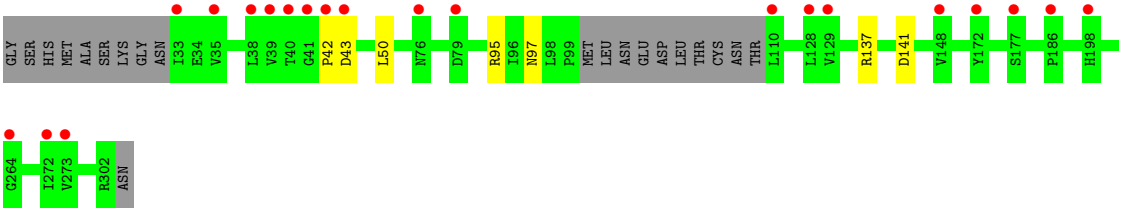




• Molecule 1: Structural protein VP1

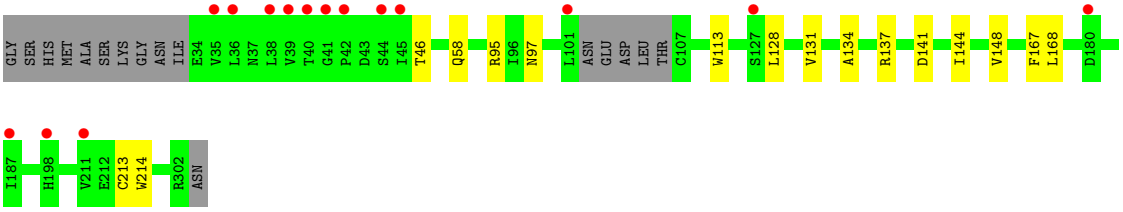
Chain I:





• Molecule 1: Structural protein VP1

Chain J:


GLY
SER
HIS
MET
ALA
SER
LYS
GLY
ASN
ILE
E34
V35
L36
N37
L38
V39
T40
G41
P42
D43
S44
I45
T46
Q58
R95
I96
N97
L101
ASN
GLU
ASP
LEU
THR
C107
W113
S127
L128
V131
A134
R137
D141
I144
V148
F167
L168
D180

I187
H198
V211
E212
C213
W214
R302
ASN

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	64.10Å 153.64Å 143.98Å 90.00° 91.83° 90.00°	Depositor
Resolution (Å)	29.57 – 2.10 29.57 – 2.10	Depositor EDS
% Data completeness (in resolution range)	98.3 (29.57-2.10) 98.3 (29.57-2.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.56 (at 2.10Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, R_{free}	0.205 , 0.249 0.207 , 0.248	Depositor DCC
R_{free} test set	7970 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	37.0	Xtriage
Anisotropy	0.523	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 36.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.25$	Xtriage
Estimated twinning fraction	0.064 for h,-k,-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	20926	wwPDB-VP
Average B, all atoms (Å^2)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.17% 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 (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	A	0.47	0/2141	0.67	0/2912
1	B	0.48	0/2107	0.68	0/2864
1	C	0.48	0/2140	0.68	0/2912
1	D	0.42	0/2060	0.63	0/2801
1	E	0.46	0/2089	0.65	0/2842
1	F	0.40	0/2072	0.62	0/2821
1	G	0.42	0/2074	0.65	0/2824
1	H	0.39	0/2080	0.61	0/2829
1	I	0.35	0/2018	0.56	0/2754
1	J	0.39	0/2036	0.61	0/2778
All	All	0.43	0/20817	0.64	0/28337

There are no bond length outliers.

There are no bond angle outliers.

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	A	2096	0	2040	5	0
1	B	2063	0	1999	7	0
1	C	2092	0	2025	6	0
1	D	2017	0	1945	6	0
1	E	2045	0	1985	7	0
1	F	2028	0	1945	8	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	G	2030	0	1958	9	0
1	H	2038	0	1956	4	0
1	I	1974	0	1872	3	0
1	J	1992	0	1877	7	0
2	A	86	0	0	0	0
2	B	89	0	0	0	0
2	C	88	0	0	0	0
2	D	48	0	0	0	0
2	E	89	0	0	0	0
2	F	26	0	0	0	0
2	G	41	0	0	0	0
2	H	34	0	0	0	0
2	I	17	0	0	0	0
2	J	33	0	0	0	0
All	All	20926	0	19602	60	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 2.

The worst 5 of 60 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:B:137:ARG:HG3	1:B:141:ASP:HA	1.72	0.71
1:E:222:GLU:H	1:E:222:GLU:CD	2.06	0.58
1:F:137:ARG:HG3	1:F:141:ASP:HA	1.88	0.55
1:F:265:LEU:HD21	1:F:294:ILE:HD13	1.87	0.54
1:G:95:ARG:NH1	1:G:97:ASN:OD1	2.39	0.54

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	269/280 (96%)	259 (96%)	10 (4%)	0	100	100
1	B	262/280 (94%)	253 (97%)	9 (3%)	0	100	100
1	C	269/280 (96%)	259 (96%)	10 (4%)	0	100	100
1	D	257/280 (92%)	248 (96%)	9 (4%)	0	100	100
1	E	261/280 (93%)	253 (97%)	8 (3%)	0	100	100
1	F	261/280 (93%)	251 (96%)	10 (4%)	0	100	100
1	G	261/280 (93%)	249 (95%)	12 (5%)	0	100	100
1	H	258/280 (92%)	247 (96%)	11 (4%)	0	100	100
1	I	256/280 (91%)	245 (96%)	11 (4%)	0	100	100
1	J	260/280 (93%)	246 (95%)	13 (5%)	1 (0%)	34	32
All	All	2614/2800 (93%)	2510 (96%)	103 (4%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	J	134	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.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	236/246 (96%)	233 (99%)	3 (1%)	69	75
1	B	231/246 (94%)	230 (100%)	1 (0%)	91	94
1	C	235/246 (96%)	234 (100%)	1 (0%)	91	94
1	D	222/246 (90%)	222 (100%)	0	100	100
1	E	228/246 (93%)	226 (99%)	2 (1%)	78	84
1	F	225/246 (92%)	224 (100%)	1 (0%)	91	94
1	G	225/246 (92%)	221 (98%)	4 (2%)	59	65
1	H	227/246 (92%)	224 (99%)	3 (1%)	69	75
1	I	216/246 (88%)	215 (100%)	1 (0%)	88	92

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	J	214/246 (87%)	213 (100%)	1 (0%)	88 92
All	All	2259/2460 (92%)	2242 (99%)	17 (1%)	81 86

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	H	282	ASP
1	J	58	GLN
1	F	282	ASP
1	G	50	LEU
1	G	58	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

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	271/280 (96%)	0.06	13 (4%)	30	36	26, 34, 59, 94	0
1	B	266/280 (95%)	-0.01	10 (3%)	40	46	26, 33, 51, 71	0
1	C	270/280 (96%)	-0.08	6 (2%)	62	66	27, 35, 51, 70	0
1	D	263/280 (93%)	0.12	15 (5%)	23	29	29, 39, 64, 93	0
1	E	265/280 (94%)	-0.03	10 (3%)	40	46	25, 35, 56, 86	0
1	F	265/280 (94%)	0.30	17 (6%)	19	24	32, 44, 69, 106	0
1	G	265/280 (94%)	0.13	11 (4%)	36	42	30, 41, 60, 87	0
1	H	264/280 (94%)	0.10	11 (4%)	36	42	31, 45, 63, 77	0
1	I	260/280 (92%)	0.46	21 (8%)	12	15	37, 52, 78, 99	0
1	J	264/280 (94%)	0.23	15 (5%)	23	29	33, 46, 74, 107	0
All	All	2653/2800 (94%)	0.13	129 (4%)	29	35	25, 41, 66, 107	0

The worst 5 of 129 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	I	110	LEU	7.0
1	A	106	THR	6.4
1	F	107	CYS	6.2
1	J	42	PRO	5.3
1	I	42	PRO	4.8

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 monosaccharides 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.