Full wwPDB X-ray Structure Validation Report ⓘ

May 19, 2021 – 06:02 pm BST

PDB ID	:	7AEJ
Title	:	Crystal structure of asymmetric HIV-1 gp41 containing all membrane anchors
Authors	:	Caillat, C.; Guilligay, D.; Weissenhorn, W.
Deposited on	:	2020-09-17
Resolution	:	3.80 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 ⓘ symbol.

The following versions of software and data (see references ①) were used in the production of this report:

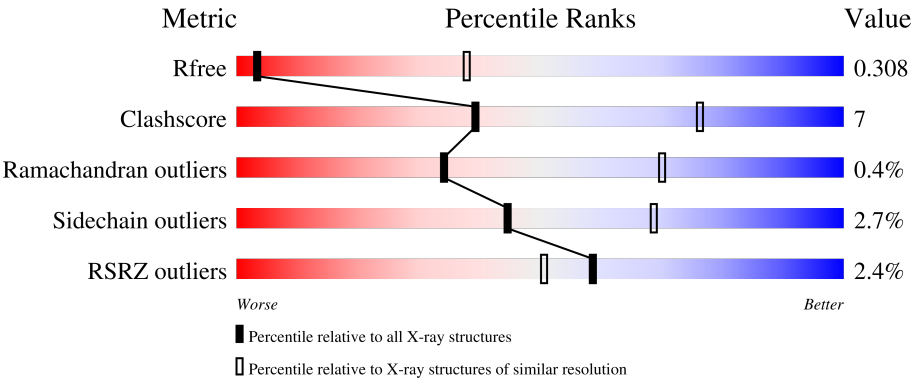
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.18
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.18

1 Overall quality at a glance (i)

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

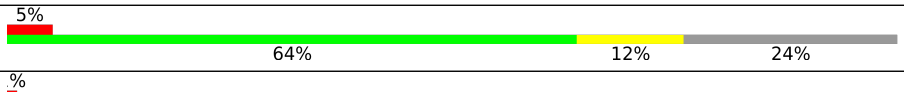
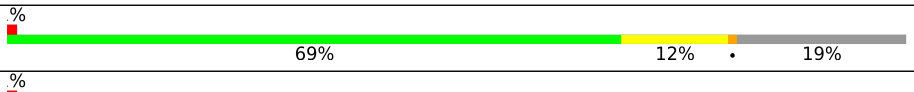
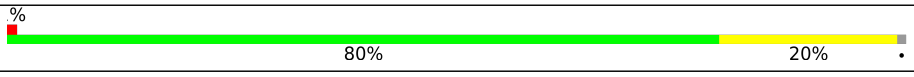
The reported resolution of this entry is 3.80 Å.

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}	130704	1212 (4.00-3.60)
Clashscore	141614	1288 (4.00-3.60)
Ramachandran outliers	138981	1243 (4.00-3.60)
Sidechain outliers	138945	1237 (4.00-3.60)
RSRZ outliers	127900	1121 (4.00-3.60)

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	192	
1	B	192	
1	C	192	
2	D	122	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 4440 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 Envelope glycoprotein gp160,Envelope glycoprotein gp160.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	156	Total	C	N	O	S	0	0	0
			1263	813	216	230	4			
1	A	129	Total	C	N	O	S	0	0	0
			1067	685	184	194	4			
1	B	145	Total	C	N	O	S	0	0	0
			1182	765	203	210	4			

There are 66 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	501	GLY	-	expression tag	UNP B2CPZ5
C	502	ALA	-	expression tag	UNP B2CPZ5
C	503	MET	-	expression tag	UNP B2CPZ5
C	504	ASP	-	expression tag	UNP B2CPZ5
C	505	TYR	-	expression tag	UNP B2CPZ5
C	506	LYS	-	expression tag	UNP B2CPZ5
C	507	ASP	-	expression tag	UNP B2CPZ5
C	508	ASP	-	expression tag	UNP B2CPZ5
C	509	ASP	-	expression tag	UNP B2CPZ5
C	510	ASP	-	expression tag	UNP B2CPZ5
C	511	LYS	-	expression tag	UNP B2CPZ5
C	519	LEU	PHE	conflict	UNP B2CPZ5
C	621	GLY	-	linker	UNP B2CPZ5
C	622	ALA	-	linker	UNP B2CPZ5
C	623	MET	-	linker	UNP B2CPZ5
C	624	ASP	-	linker	UNP B2CPZ5
C	625	ASP	-	linker	UNP B2CPZ5
C	626	ASP	-	linker	UNP B2CPZ5
C	627	ASP	-	linker	UNP B2CPZ5
C	628	LYS	-	linker	UNP B2CPZ5
C	717	ARG	PHE	engineered mutation	UNP B2CPZ5
C	718	ARG	GLN	engineered mutation	UNP B2CPZ5
A	501	GLY	-	expression tag	UNP B2CPZ5

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	502	ALA	-	expression tag	UNP B2CPZ5
A	503	MET	-	expression tag	UNP B2CPZ5
A	504	ASP	-	expression tag	UNP B2CPZ5
A	505	TYR	-	expression tag	UNP B2CPZ5
A	506	LYS	-	expression tag	UNP B2CPZ5
A	507	ASP	-	expression tag	UNP B2CPZ5
A	508	ASP	-	expression tag	UNP B2CPZ5
A	509	ASP	-	expression tag	UNP B2CPZ5
A	510	ASP	-	expression tag	UNP B2CPZ5
A	511	LYS	-	expression tag	UNP B2CPZ5
A	519	LEU	PHE	conflict	UNP B2CPZ5
A	621	GLY	-	linker	UNP B2CPZ5
A	622	ALA	-	linker	UNP B2CPZ5
A	623	MET	-	linker	UNP B2CPZ5
A	624	ASP	-	linker	UNP B2CPZ5
A	625	ASP	-	linker	UNP B2CPZ5
A	626	ASP	-	linker	UNP B2CPZ5
A	627	ASP	-	linker	UNP B2CPZ5
A	628	LYS	-	linker	UNP B2CPZ5
A	717	ARG	PHE	engineered mutation	UNP B2CPZ5
A	718	ARG	GLN	engineered mutation	UNP B2CPZ5
B	501	GLY	-	expression tag	UNP B2CPZ5
B	502	ALA	-	expression tag	UNP B2CPZ5
B	503	MET	-	expression tag	UNP B2CPZ5
B	504	ASP	-	expression tag	UNP B2CPZ5
B	505	TYR	-	expression tag	UNP B2CPZ5
B	506	LYS	-	expression tag	UNP B2CPZ5
B	507	ASP	-	expression tag	UNP B2CPZ5
B	508	ASP	-	expression tag	UNP B2CPZ5
B	509	ASP	-	expression tag	UNP B2CPZ5
B	510	ASP	-	expression tag	UNP B2CPZ5
B	511	LYS	-	expression tag	UNP B2CPZ5
B	519	LEU	PHE	conflict	UNP B2CPZ5
B	621	GLY	-	linker	UNP B2CPZ5
B	622	ALA	-	linker	UNP B2CPZ5
B	623	MET	-	linker	UNP B2CPZ5
B	624	ASP	-	linker	UNP B2CPZ5
B	625	ASP	-	linker	UNP B2CPZ5
B	626	ASP	-	linker	UNP B2CPZ5
B	627	ASP	-	linker	UNP B2CPZ5
B	628	LYS	-	linker	UNP B2CPZ5
B	717	ARG	PHE	engineered mutation	UNP B2CPZ5

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	718	ARG	GLN	engineered mutation	UNP B2CPZ5

- Molecule 2 is a protein called 2H10.

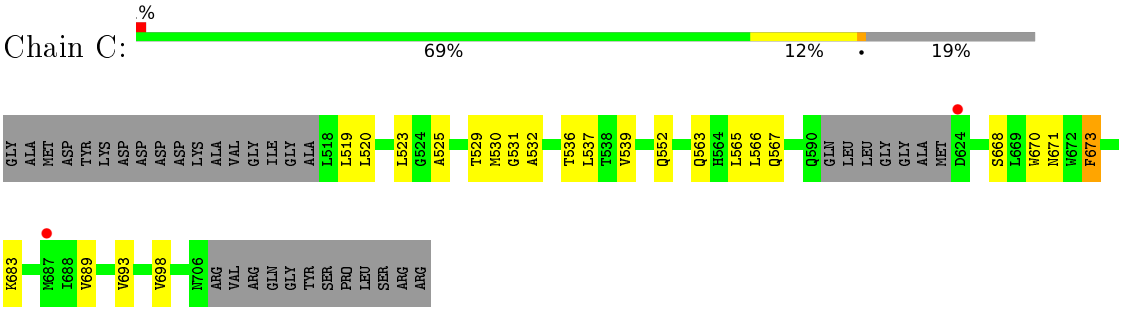
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	121	Total	C	N	O	S	0	0	0
			928	572	169	182	5			

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.

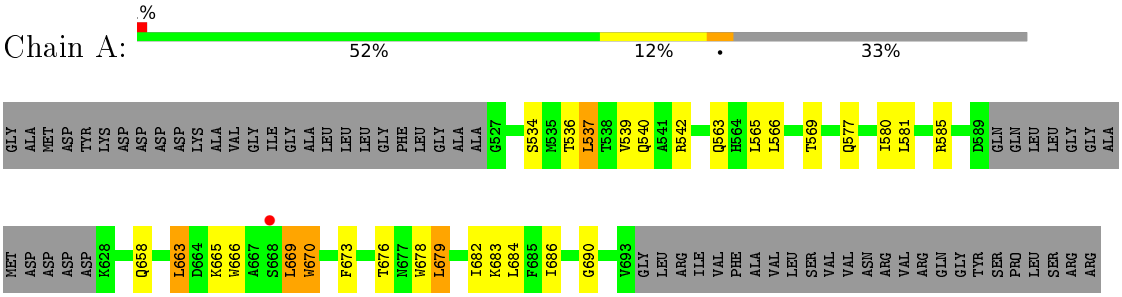
- Molecule 1: Envelope glycoprotein gp160,Envelope glycoprotein gp160

Chain C:



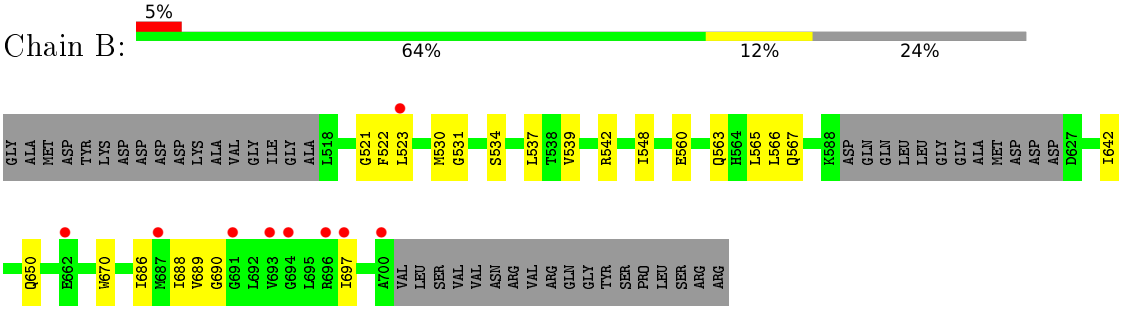
- Molecule 1: Envelope glycoprotein gp160,Envelope glycoprotein gp160

Chain A:



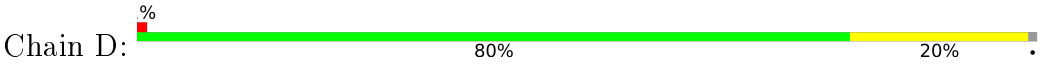
- Molecule 1: Envelope glycoprotein gp160,Envelope glycoprotein gp160

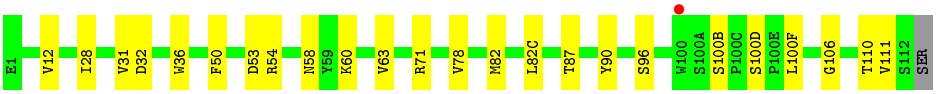
Chain B:



- Molecule 2: 2H10

Chain D:




E1
V12
I28
V31
D32
W36
F50
D53
R54
N58
Y59
K60
V63
R71
V78
M82
L82C
T87
Y90
S96
W100
S100A
S100B
P100C
S100D
P100E
L100F
G106
T110
V111
S112
SER

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	96.75Å 101.41Å 234.42Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.56 – 3.80 49.56 – 3.28	Depositor EDS
% Data completeness (in resolution range)	78.3 (49.56-3.80) 62.0 (49.56-3.28)	Depositor EDS
R_{merge}	0.23	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.27 (at 3.25Å)	Xtriage
Refinement program	PHENIX 1.18.2_3874+SVN	Depositor
R, R_{free}	0.265 , 0.308 0.266 , 0.308	Depositor DCC
R_{free} test set	665 reflections (5.94%)	wwPDB-VP
Wilson B-factor (Å^2)	75.8	Xtriage
Anisotropy	0.430	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 40.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.26$	Xtriage
Estimated twinning fraction	0.057 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.85	EDS
Total number of atoms	4440	wwPDB-VP
Average B, all atoms (Å^2)	91.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.40% 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	0.35	0/1086	0.51	0/1470
1	B	0.23	0/1203	0.35	0/1628
1	C	0.26	0/1284	0.41	0/1740
2	D	0.26	0/944	0.48	0/1276
All	All	0.28	0/4517	0.44	0/6114

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

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	1067	0	1074	26	0
1	B	1182	0	1204	21	0
1	C	1263	0	1275	27	0
2	D	928	0	919	13	0
All	All	4440	0	4472	66	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 7.

All (66) 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:669:LEU:HG	1:A:670:TRP:H	1.33	0.92
1:A:669:LEU:HG	1:A:670:TRP:N	1.92	0.84
1:A:683:LYS:O	1:A:686:ILE:HG22	1.80	0.82
2:D:53:ASP:HA	2:D:71:ARG:HH12	1.53	0.72
2:D:96:SER:HB2	2:D:100(F):LEU:HB3	1.76	0.68
1:C:537:LEU:HD13	1:A:542:ARG:HE	1.59	0.66
2:D:82:MET:HB3	2:D:82(C):LEU:HD21	1.77	0.65
1:A:537:LEU:O	1:A:540:GLN:HB3	1.97	0.64
1:C:673:PHE:HD1	1:C:673:PHE:H	1.48	0.62
1:A:686:ILE:HD11	1:B:521:GLY:HA2	1.82	0.62
1:C:565:LEU:HD23	1:A:566:LEU:HD13	1.81	0.61
1:A:565:LEU:HD23	1:B:566:LEU:HD13	1.82	0.61
1:C:563:GLN:HE21	1:C:567:GLN:HE21	1.48	0.60
1:A:678:TRP:O	1:A:682:ILE:HG12	2.01	0.60
1:A:669:LEU:CG	1:A:670:TRP:H	2.09	0.59
2:D:60:LYS:HB3	2:D:63:VAL:HG22	1.86	0.56
1:C:698:VAL:HG21	1:B:697:ILE:HB	1.87	0.56
1:C:693:VAL:HG23	1:B:690:GLY:HA3	1.89	0.55
1:C:566:LEU:HD13	1:B:565:LEU:HD23	1.89	0.54
1:C:529:THR:HG23	1:A:679:LEU:HB3	1.90	0.54
1:B:563:GLN:HE22	1:B:567:GLN:HE21	1.56	0.53
1:C:519:LEU:HD23	1:C:519:LEU:H	1.72	0.53
1:C:683:LYS:HB2	1:B:522:PHE:HZ	1.74	0.53
1:B:560:GLU:OE2	1:B:650:GLN:NE2	2.39	0.52
1:B:539:VAL:HG22	1:B:542:ARG:HH21	1.75	0.52
1:C:670:TRP:HE1	1:B:537:LEU:HD21	1.74	0.51
1:A:539:VAL:O	1:A:542:ARG:HB2	2.11	0.50
2:D:87:THR:HG23	2:D:110:THR:HA	1.93	0.50
1:C:531:GLY:HA2	1:B:534:SER:HB2	1.94	0.49
1:C:525:ALA:HB1	1:A:682:ILE:HB	1.94	0.49
1:A:663:LEU:C	1:A:665:LYS:H	2.16	0.49
1:C:523:LEU:HD12	1:B:523:LEU:HD22	1.93	0.49
1:C:531:GLY:HA2	1:B:534:SER:CB	2.43	0.48
2:D:28:ILE:HG22	2:D:31:VAL:HG21	1.95	0.48
2:D:36:TRP:HE1	2:D:78:VAL:HG12	1.77	0.48
1:C:532:ALA:HB2	1:B:670:TRP:HH2	1.79	0.48
1:B:563:GLN:NE2	1:B:567:GLN:HE21	2.12	0.48
1:C:563:GLN:NE2	1:C:567:GLN:HE21	2.12	0.48
1:C:673:PHE:CD1	1:C:673:PHE:N	2.83	0.46
2:D:32:ASP:HB2	2:D:96:SER:HA	1.97	0.46
1:C:566:LEU:HG	1:A:566:LEU:HD21	1.98	0.46
1:A:690:GLY:O	1:B:688:ILE:HD11	2.16	0.46

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:D:82(C):LEU:HB3	2:D:111:VAL:HG21	1.96	0.46
1:A:669:LEU:CG	1:A:670:TRP:N	2.66	0.46
1:C:668:SER:O	1:C:671:ASN:ND2	2.49	0.46
1:C:689:VAL:O	1:C:693:VAL:HG22	2.16	0.45
1:A:563:GLN:O	1:A:566:LEU:HB2	2.16	0.45
1:A:577:GLN:O	1:A:580:ILE:HG22	2.15	0.45
1:C:552:GLN:NE2	1:B:548:ILE:HG23	2.32	0.45
1:A:679:LEU:HD12	1:A:679:LEU:HA	1.83	0.45
2:D:12:VAL:HG11	2:D:82(C):LEU:HD12	1.98	0.45
1:A:581:LEU:HD21	1:A:585:ARG:HH21	1.82	0.44
2:D:50:PHE:CE1	2:D:58:ASN:HB3	2.52	0.44
1:B:686:ILE:HA	1:B:689:VAL:HG12	1.99	0.44
1:C:525:ALA:CB	1:A:682:ILE:HB	2.48	0.43
1:C:536:THR:HA	1:C:539:VAL:HG22	2.01	0.43
1:A:566:LEU:O	1:A:569:THR:OG1	2.27	0.43
1:A:683:LYS:O	1:A:686:ILE:CG2	2.59	0.42
1:A:563:GLN:HA	1:A:566:LEU:HD12	2.01	0.42
1:B:563:GLN:OE1	1:B:642:ILE:HG21	2.20	0.41
2:D:90:TYR:O	2:D:106:GLY:HA2	2.20	0.41
2:D:100(B):SER:OG	2:D:100(D):SER:O	2.39	0.41
1:C:673:PHE:HD1	1:C:673:PHE:N	2.17	0.41
1:C:530:MET:HB3	1:B:531:GLY:HA3	2.03	0.40
1:A:686:ILE:HD11	1:B:521:GLY:CA	2.50	0.40
1:C:519:LEU:HG	1:C:520:LEU:HD12	2.03	0.40

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	125/192 (65%)	117 (94%)	6 (5%)	2 (2%)	9 44

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	B	141/192 (73%)	140 (99%)	1 (1%)	0	100 100
1	C	152/192 (79%)	146 (96%)	6 (4%)	0	100 100
2	D	119/122 (98%)	116 (98%)	3 (2%)	0	100 100
All	All	537/698 (77%)	519 (97%)	16 (3%)	2 (0%)	34 70

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	669	LEU
1	A	673	PHE

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	115/162 (71%)	105 (91%)	10 (9%)	10 38
1	B	125/162 (77%)	124 (99%)	1 (1%)	81 89
1	C	135/162 (83%)	134 (99%)	1 (1%)	84 91
2	D	104/105 (99%)	103 (99%)	1 (1%)	76 86
All	All	479/591 (81%)	466 (97%)	13 (3%)	44 69

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	673	PHE
1	A	534	SER
1	A	536	THR
1	A	537	LEU
1	A	658	GLN
1	A	663	LEU
1	A	666	TRP
1	A	670	TRP
1	A	676	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	679	LEU
1	A	684	LEU
1	B	530	MET
2	D	54	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	567	GLN
1	C	637	ASN
1	A	562	GLN
1	A	567	GLN
1	B	567	GLN

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 monosaccharides in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

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

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	129/192 (67%)	-0.17	1 (0%)	86	81	45, 85, 140, 176	0
1	B	145/192 (75%)	0.13	9 (6%)	20	16	46, 94, 147, 170	0
1	C	156/192 (81%)	0.06	2 (1%)	77	70	34, 95, 156, 183	0
2	D	121/122 (99%)	-0.06	1 (0%)	86	81	53, 71, 121, 170	0
All	All	551/698 (78%)	-0.00	13 (2%)	59	50	34, 82, 151, 183	0

All (13) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	668	SER	3.5
1	B	687	MET	3.0
1	B	694	GLY	3.0
1	B	697	ILE	2.8
1	B	691	GLY	2.7
1	B	693	VAL	2.5
1	B	700	ALA	2.5
2	D	100	TRP	2.4
1	B	696	ARG	2.3
1	C	687	MET	2.3
1	B	662	GLU	2.2
1	C	624	ASP	2.1
1	B	523	LEU	2.1

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.