Full wwPDB X-ray Structure Validation Report (i)

Jun 25, 2024 – 04:08 AM EDT

PDB ID	:	6G3K
Title	:	MHC A02 Allele presenting MTSAIGILPV
Authors	:	Rizkallah, P.J.; Sewell, A.K.
Deposited on	:	2018-03-26
Resolution	:	2.90 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 (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:

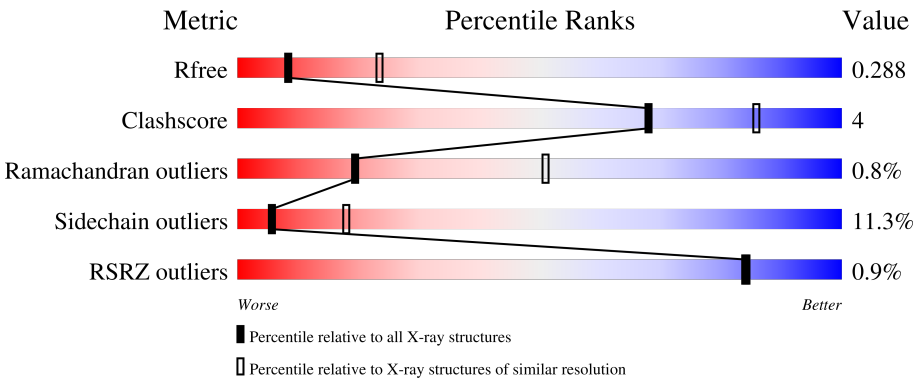
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.90 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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	276	.% 82% 15% .
1	D	276	83% 15% .
2	B	100	.% 79% 20% .
2	E	100	.% 79% 18% .
3	C	10	60% 40%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	10	70% 30%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 6359 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 HLA class I histocompatibility antigen, A-2 alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	276	2264	1414	413	428	9	0	1	0
1	D	276	2254	1408	410	427	9	0	0	0

- Molecule 2 is a protein called Beta-2-microglobulin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	100	837	533	141	159	4	0	0	0
2	E	100	837	533	141	159	4	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	MET	-	initiating methionine	UNP P61769
E	0	MET	-	initiating methionine	UNP P61769

- Molecule 3 is a protein called ILE-THR-SER-GLY-ILE-GLY-VAL-LEU-PRO-VAL.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	C	10	67	44	10	13	0	0	0
3	F	10	67	44	10	13	0	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	9	9	9	0	0

Continued on next page...

Continued from previous page...

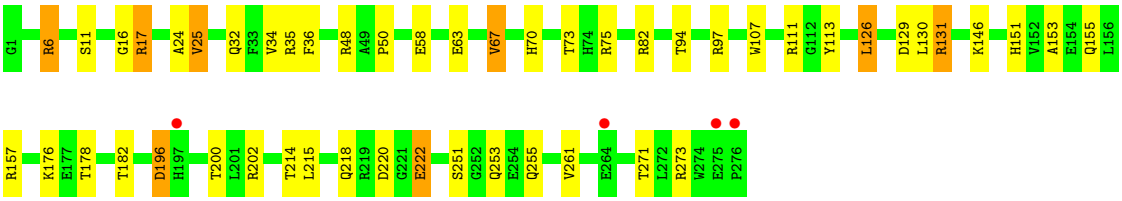
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	B	4	4	4	0	0
			Total	O		
4	D	12	12	12	0	0
			Total	O		
4	E	7	7	7	0	0
			Total	O		
4	F	1	1	1	0	0

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.

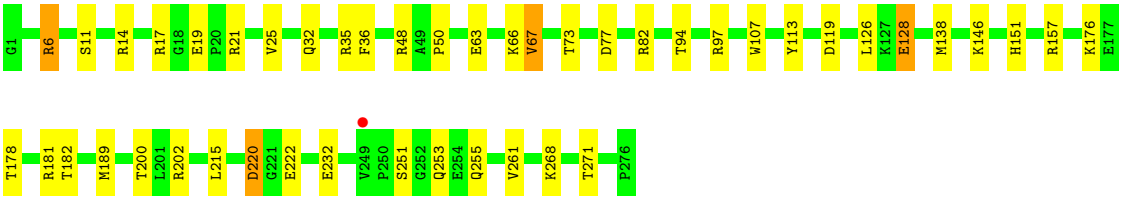
• Molecule 1: HLA class I histocompatibility antigen, A-2 alpha chain

Chain A:



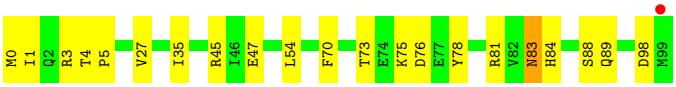
• Molecule 1: HLA class I histocompatibility antigen, A-2 alpha chain

Chain D:



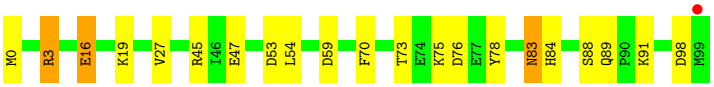
• Molecule 2: Beta-2-microglobulin

Chain B:



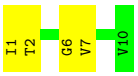
• Molecule 2: Beta-2-microglobulin

Chain E:



• Molecule 3: ILE-THR-SER-GLY-ILE-GLY-VAL-LEU-PRO-VAL

Chain C:



- Molecule 3: ILE-THR-SER-GLY-ILE-GLY-VAL-LEU-PRO-VAL

Chain F:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	119.98Å 171.09Å 50.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.58 – 2.90 44.97 – 2.66	Depositor EDS
% Data completeness (in resolution range)	91.4 (50.58-2.90) 89.1 (44.97-2.66)	Depositor EDS
R_{merge}	0.17	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.70 (at 2.65Å)	Xtriage
Refinement program	REFMAC 5.8.0189	Depositor
R, R_{free}	0.236 , 0.287 0.239 , 0.288	Depositor DCC
R_{free} test set	1370 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	68.9	Xtriage
Anisotropy	0.180	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 24.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6359	wwPDB-VP
Average B, all atoms (Å^2)	73.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 54.07 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 3.8494e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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	0.88	0/2331	0.99	4/3164 (0.1%)
1	D	0.91	1/2320 (0.0%)	1.00	4/3149 (0.1%)
2	B	0.91	0/860	1.04	2/1162 (0.2%)
2	E	0.97	1/860 (0.1%)	1.02	1/1162 (0.1%)
3	C	0.84	0/67	1.08	0/90
3	F	0.75	0/67	1.01	0/90
All	All	0.90	2/6505 (0.0%)	1.00	11/8817 (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

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	E	53	ASP	CB-CG	6.11	1.64	1.51
1	D	128	GLU	CG-CD	5.62	1.60	1.51

All (11) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	97	ARG	NE-CZ-NH2	5.95	123.28	120.30
1	D	181	ARG	NE-CZ-NH1	-5.83	117.39	120.30
1	A	131	ARG	N-CA-C	5.73	126.48	111.00
1	A	202	ARG	NE-CZ-NH1	5.61	123.11	120.30
1	A	75	ARG	NE-CZ-NH1	5.58	123.09	120.30
1	A	97	ARG	NE-CZ-NH2	5.49	123.04	120.30
2	B	45	ARG	NE-CZ-NH1	5.37	122.99	120.30
1	D	181	ARG	NE-CZ-NH2	5.26	122.93	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	81	ARG	NE-CZ-NH2	-5.24	117.68	120.30
2	E	45	ARG	NE-CZ-NH1	5.19	122.89	120.30
1	D	119	ASP	CB-CG-OD1	5.10	122.89	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	130	LEU	Peptide

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	2264	0	2109	20	2
1	D	2254	0	2103	15	1
2	B	837	0	803	7	0
2	E	837	0	803	6	1
3	C	67	0	78	3	0
3	F	67	0	78	4	0
4	A	9	0	0	0	0
4	B	4	0	0	0	0
4	D	12	0	0	0	0
4	E	7	0	0	1	0
4	F	1	0	0	0	0
All	All	6359	0	5974	47	2

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 4.

All (47) 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:25:VAL:HG22	1:A:32:GLN:NE2	1.94	0.82
1:A:129:ASP:O	1:A:131:ARG:HD3	1.86	0.75

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:83:ASN:HD22	2:B:84:HIS:H	1.39	0.71
2:B:83:ASN:HD22	2:B:84:HIS:N	1.89	0.71
2:E:83:ASN:HD22	2:E:84:HIS:N	1.89	0.70
1:A:25:VAL:HG22	1:A:32:GLN:HE22	1.58	0.69
2:E:83:ASN:HD22	2:E:84:HIS:H	1.43	0.66
1:A:111:ARG:HD2	1:A:113:TYR:OH	2.02	0.58
1:A:16:GLY:O	1:A:17:ARG:HB2	2.03	0.58
1:A:129:ASP:O	1:A:131:ARG:CD	2.53	0.54
1:D:21:ARG:HH11	1:D:21:ARG:HG2	1.73	0.53
1:D:215:LEU:HD23	1:D:261:VAL:HG22	1.91	0.53
1:A:126:LEU:HD22	1:A:131:ARG:HD2	1.92	0.52
1:D:25:VAL:HB	1:D:32:GLN:HE21	1.74	0.52
1:A:36:PHE:CD2	1:A:67:VAL:HG11	2.45	0.52
1:A:215:LEU:HD23	1:A:261:VAL:HG22	1.92	0.51
1:D:215:LEU:CD2	1:D:261:VAL:HG22	2.40	0.51
1:A:215:LEU:CD2	1:A:261:VAL:HG22	2.41	0.50
1:D:189:MET:HG3	1:D:202:ARG:O	2.12	0.50
1:D:36:PHE:CD2	1:D:67:VAL:HG11	2.47	0.50
1:D:63:GLU:OE1	3:F:2:THR:HG22	2.14	0.48
1:A:35:ARG:HG2	1:A:48:ARG:HG2	1.96	0.47
2:B:1:ILE:HG22	1:D:14:ARG:NH1	2.30	0.47
2:E:3:ARG:NH1	4:E:101:HOH:O	2.47	0.46
2:E:76:ASP:HB3	2:E:78:TYR:CE2	2.50	0.46
2:B:76:ASP:HB3	2:B:78:TYR:CE1	2.51	0.46
1:D:66:LYS:HE3	3:F:2:THR:HG23	2.00	0.44
1:A:131:ARG:HB2	1:A:153:ALA:HB1	1.99	0.43
1:D:35:ARG:HG2	1:D:48:ARG:HG2	1.99	0.43
2:E:16:GLU:HG2	2:E:19:LYS:CG	2.48	0.43
1:A:155:GLN:HE21	3:C:6:GLY:H	1.66	0.43
2:B:27:VAL:HG11	2:B:35:ILE:HD13	2.01	0.43
1:D:25:VAL:HB	1:D:32:GLN:NE2	2.33	0.43
1:A:70:HIS:NE2	3:C:7:VAL:HG11	2.33	0.43
1:A:131:ARG:CB	1:A:153:ALA:HB1	2.49	0.42
1:A:6:ARG:NH2	1:A:113:TYR:CE1	2.87	0.42
1:D:6:ARG:NH2	1:D:113:TYR:CE1	2.88	0.42
1:A:24:ALA:O	1:A:25:VAL:HG23	2.19	0.42
1:D:21:ARG:HG2	1:D:21:ARG:NH1	2.35	0.41
1:A:63:GLU:OE1	3:C:2:THR:HG22	2.20	0.41
1:A:196:ASP:OD1	1:A:196:ASP:N	2.54	0.41
2:B:1:ILE:HD12	2:B:1:ILE:O	2.21	0.41
1:D:77:ASP:OD1	3:F:9:PRO:HA	2.21	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:218:GLN:HA	1:A:222:GLU:O	2.21	0.40
2:B:4:THR:OG1	2:B:5:PRO:CD	2.69	0.40
2:E:27:VAL:O	2:E:27:VAL:HG23	2.22	0.40
1:D:63:GLU:CD	3:F:2:THR:HG22	2.42	0.40

All (2) 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:A:17:ARG:NH2	2:E:59:ASP:O[1_554]	2.11	0.09
1:A:273:ARG:NH2	1:D:220:ASP:OD2[3_547]	2.12	0.08

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	275/276 (100%)	250 (91%)	23 (8%)	2 (1%)	22	54
1	D	274/276 (99%)	252 (92%)	20 (7%)	2 (1%)	22	54
2	B	98/100 (98%)	92 (94%)	5 (5%)	1 (1%)	15	45
2	E	98/100 (98%)	93 (95%)	4 (4%)	1 (1%)	15	45
3	C	8/10 (80%)	7 (88%)	1 (12%)	0	100	100
3	F	8/10 (80%)	7 (88%)	1 (12%)	0	100	100
All	All	761/772 (99%)	701 (92%)	54 (7%)	6 (1%)	19	51

All (6) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	17	ARG
1	A	17	ARG
2	B	88	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	E	88	SER
1	D	50	PRO
1	A	50	PRO

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	233/232 (100%)	207 (89%)	26 (11%)	6	18
1	D	232/232 (100%)	206 (89%)	26 (11%)	6	18
2	B	95/95 (100%)	85 (90%)	10 (10%)	7	21
2	E	95/95 (100%)	83 (87%)	12 (13%)	4	13
3	C	8/8 (100%)	7 (88%)	1 (12%)	4	14
3	F	8/8 (100%)	7 (88%)	1 (12%)	4	14
All	All	671/670 (100%)	595 (89%)	76 (11%)	6	18

All (76) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	6	ARG
1	A	11	SER
1	A	25	VAL
1	A	34	VAL
1	A	58	GLU
1	A	67	VAL
1	A	73	THR
1	A	82	ARG
1	A	94	THR
1	A	107	TRP
1	A	126	LEU
1	A	146	LYS
1	A	151	HIS
1	A	157	ARG
1	A	176	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	178	THR
1	A	182	THR
1	A	196	ASP
1	A	200	THR
1	A	214	THR
1	A	220	ASP
1	A	222	GLU
1	A	251	SER
1	A	253	GLN
1	A	255	GLN
1	A	271	THR
2	B	0	MET
2	B	3	ARG
2	B	47	GLU
2	B	54	LEU
2	B	70	PHE
2	B	73	THR
2	B	75	LYS
2	B	83	ASN
2	B	89	GLN
2	B	98	ASP
3	C	1	ILE
1	D	6	ARG
1	D	11	SER
1	D	19	GLU
1	D	67	VAL
1	D	73	THR
1	D	82	ARG
1	D	94	THR
1	D	107	TRP
1	D	126	LEU
1	D	128	GLU
1	D	138	MET
1	D	146	LYS
1	D	151	HIS
1	D	157	ARG
1	D	176	LYS
1	D	178	THR
1	D	182	THR
1	D	200	THR
1	D	220	ASP
1	D	222	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	232	GLU
1	D	251	SER
1	D	253	GLN
1	D	255	GLN
1	D	268	LYS
1	D	271	THR
2	E	0	MET
2	E	3	ARG
2	E	16	GLU
2	E	47	GLU
2	E	54	LEU
2	E	70	PHE
2	E	73	THR
2	E	75	LYS
2	E	83	ASN
2	E	89	GLN
2	E	91	LYS
2	E	98	ASP
3	F	1	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (16) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	32	GLN
1	A	54	GLN
1	A	155	GLN
1	A	253	GLN
2	B	31	HIS
2	B	83	ASN
2	B	89	GLN
1	D	32	GLN
1	D	54	GLN
1	D	86	ASN
1	D	115	GLN
1	D	155	GLN
1	D	253	GLN
2	E	31	HIS
2	E	83	ASN
2	E	89	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)

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	276/276 (100%)	-0.13	4 (1%)	75	75	39, 75, 127, 162	0
1	D	276/276 (100%)	-0.18	1 (0%)	92	93	40, 73, 121, 140	0
2	B	100/100 (100%)	-0.36	1 (1%)	82	82	33, 61, 97, 112	0
2	E	100/100 (100%)	-0.38	1 (1%)	82	82	30, 58, 94, 108	0
3	C	10/10 (100%)	0.02	0	100	100	59, 67, 70, 75	0
3	F	10/10 (100%)	0.06	0	100	100	56, 66, 69, 69	0
All	All	772/772 (100%)	-0.20	7 (0%)	84	84	30, 69, 119, 162	0

All (7) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	197	HIS	5.5
1	A	276	PRO	4.1
1	D	249	VAL	3.0
1	A	275	GLU	2.4
2	B	99	MET	2.4
2	E	99	MET	2.3
1	A	264	GLU	2.0

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.