WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 20, 2024 – 03:27 PM EST

PDB ID	:	4LO7
Title	:	HA70(D3)-HA17-HA33
Authors	:	Lee, K.; Gu, S.; Jin, L.; Le, T.T.; Cheng, L.W.; Strotmeier, J.; Kruel, A.M.; Yao, G.; Perry, K.; Rummel, A.; Jin, R.
Deposited on	:	2013-07-12
Resolution	:	3.73 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 ⓘ 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 ⓘ) were used in the production of this report:

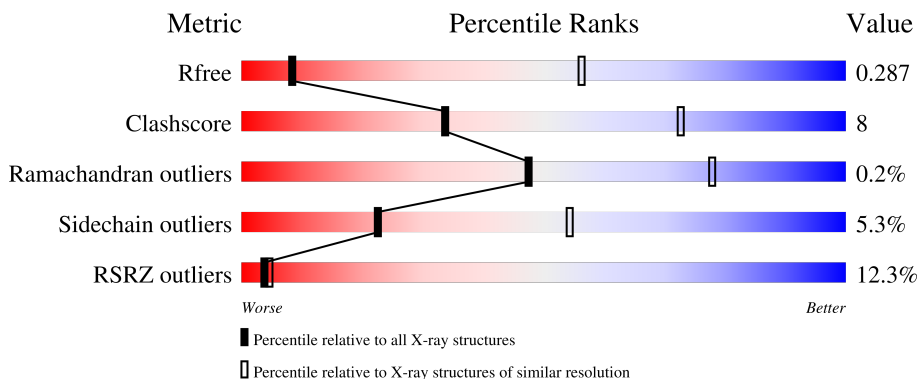|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1   Overall quality at a glance ⓘ

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

The reported resolution of this entry is 3.73 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.
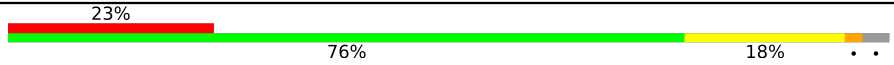
Metric	Percentile Ranks	Value
Rfree		0.287
Clashscore		8
Ramachandran outliers		0.2%
Sidechain outliers		5.3%
RSRZ outliers		12.3%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	1001 (3.90-3.58)
Clashscore	141614	1063 (3.90-3.58)
Ramachandran outliers	138981	1027 (3.90-3.58)
Sidechain outliers	138945	1023 (3.90-3.58)
RSRZ outliers	127900	1006 (3.92-3.56)

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	254	14% / 72% 15% · 10%
1	E	254	4% / 76% 13% · 10%
2	B	296	5% / 84% 11% ··
2	D	296	22% / 78% 16% ··
2	F	296	11% / 85% 10% ··

Continued on next page...

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	H	296	23% ▬▬▬▬ 76% 18% ··
3	C	147	2% ▬ 79% 15% · 5%
3	G	147	.% ▬ 80% 14% · 5%

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 15326 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 HA-70.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	228	Total	C	N	O	S	12	0	0
			1852	1171	315	365	1			
1	E	229	Total	C	N	O	S	12	0	0
			1860	1177	316	366	1			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	373	GLY	-	expression tag	UNP Q8KHU9
A	374	PRO	-	expression tag	UNP Q8KHU9
A	375	LEU	-	expression tag	UNP Q8KHU9
A	376	GLY	-	expression tag	UNP Q8KHU9
A	377	SER	-	expression tag	UNP Q8KHU9
E	373	GLY	-	expression tag	UNP Q8KHU9
E	374	PRO	-	expression tag	UNP Q8KHU9
E	375	LEU	-	expression tag	UNP Q8KHU9
E	376	GLY	-	expression tag	UNP Q8KHU9
E	377	SER	-	expression tag	UNP Q8KHU9

- Molecule 2 is a protein called HA-33.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	286	Total	C	N	O	S	0	1	0
			2334	1477	399	455	3			
2	D	286	Total	C	N	O	S	0	1	0
			2334	1477	399	455	3			
2	F	286	Total	C	N	O	S	0	1	0
			2334	1477	399	455	3			
2	H	286	Total	C	N	O	S	0	1	0
			2334	1477	399	455	3			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	294	PRO	-	expression tag	UNP Q45871
B	295	GLY	-	expression tag	UNP Q45871
B	296	SER	-	expression tag	UNP Q45871
B	297	ALA	-	expression tag	UNP Q45871
D	294	PRO	-	expression tag	UNP Q45871
D	295	GLY	-	expression tag	UNP Q45871
D	296	SER	-	expression tag	UNP Q45871
D	297	ALA	-	expression tag	UNP Q45871
F	294	PRO	-	expression tag	UNP Q45871
F	295	GLY	-	expression tag	UNP Q45871
F	296	SER	-	expression tag	UNP Q45871
F	297	ALA	-	expression tag	UNP Q45871
H	294	PRO	-	expression tag	UNP Q45871
H	295	GLY	-	expression tag	UNP Q45871
H	296	SER	-	expression tag	UNP Q45871
H	297	ALA	-	expression tag	UNP Q45871

- Molecule 3 is a protein called HA-17.

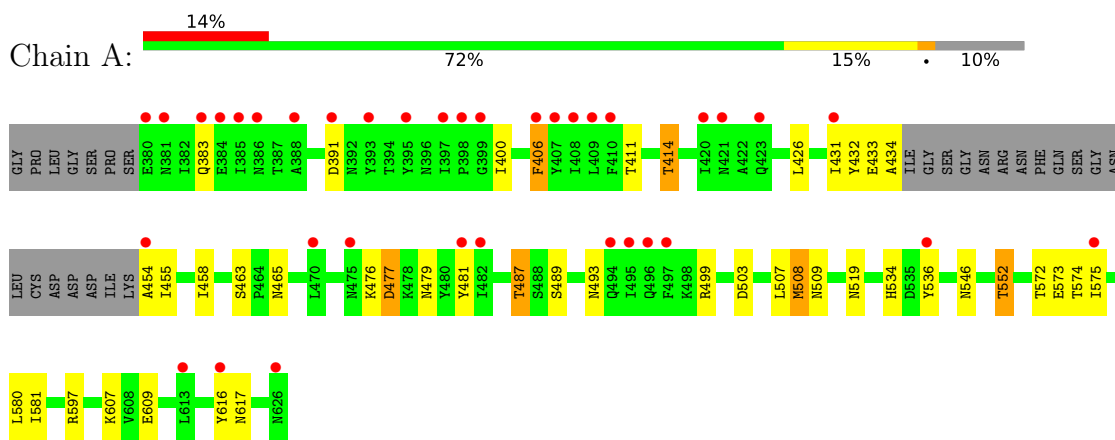| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	139	Total	C	N	O	S	0	0	0
			1139	735	181	218	5			
3	G	139	Total	C	N	O	S	0	0	0
			1139	735	181	218	5			

There are 4 discrepancies between the modelled and reference sequences:

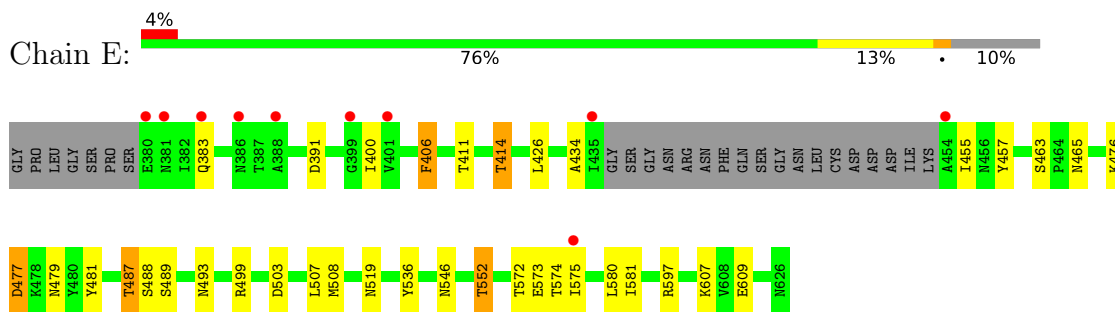| Chain	Residue	Modelled	Actual	Comment	Reference
C	0	GLY	-	expression tag	UNP Q45878
C	1	PRO	-	expression tag	UNP Q45878
G	0	GLY	-	expression tag	UNP Q45878
G	1	PRO	-	expression tag	UNP Q45878

3 Residue-property plots ⓘ

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.
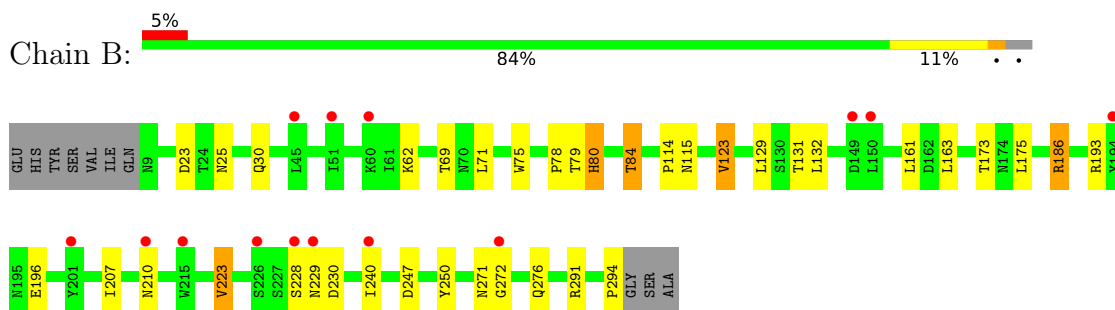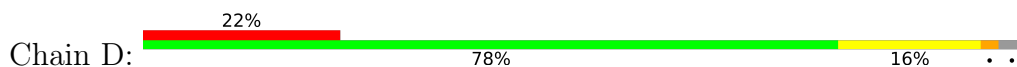
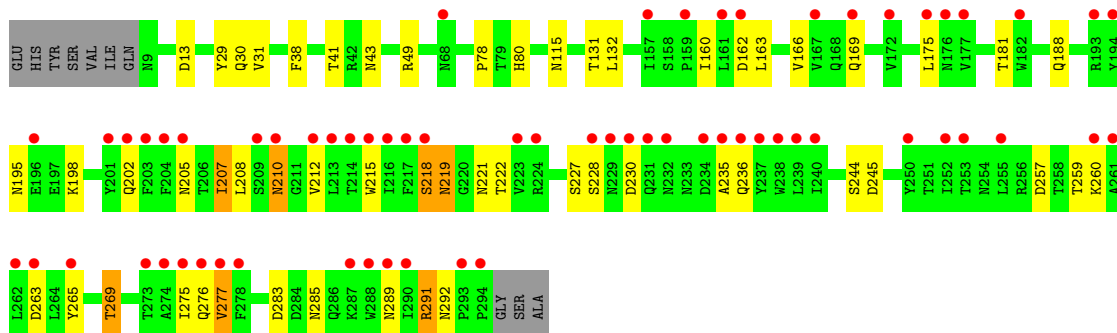• Molecule 1: HA-70

Chain A:

• Molecule 1: HA-70

Chain E:

• Molecule 2: HA-33

Chain B:

• Molecule 2: HA-33

Chain D:

GLU HIS TYR SER VAL ILE GLN N9 D13 Y29 Q30 V31 F38 T41 R42 N43 R49 N68 P78 T79 H80 N115 T131 L132 I157 S158 P159 I160 L161 D162 L163 V166 V167 Q168 Q169 V172 L175 N176 V177 T181 W182 Q188 R193 Y194

N195 E196 E197 K198 Y201 Q202 F203 F204 N205 T206 I207 L208 S209 N210 G211 V212 L213 T214 W215 I216 F217 S218 N219 G220 N221 T222 V223 R224 S227 S228 N229 D230 Q231 N232 N233 D234 A235 Q236 Y237 W238 L239 I240 S244 D245 Y250 T251 I252 T253 N254 L255 R256 D257 T258 T259 K260 A261

L262 D263 L264 Y265 T269 T273 A274 I275 Q276 V277 F278 D283 D284 N285 Q286 K287 W288 N289 I290 R291 N292 P293 P294 GLY SER ALA

• Molecule 2: HA-33

Chain F:

GLU HIS TYR SER VAL ILE GLN N9 D23 T24 N25 Q30 K62 T69 N70 L71 W75 H80 T84 D100 N115 V123 L129 L132 N133 N134 L161 D162 L163 V172 L175 N176 V177 N178 L179 R186 W190 R193 E196 F203

I207 N210 G211 V212 L213 T214 W215 I216 F217 S218 N221 T222 V223 R224 V225 S226 S227 S228 N229 D230 N233 Q236 I240 D247 Y250 T251 I252 R256 K260 A261 L262 T273 A274 I275 Q276 V277 F278 R291 P294 GLY SER ALA

• Molecule 2: HA-33

Chain H:

GLU HIS TYR SER VAL ILE GLN N9 D13 F28 Y29 Q30 V31 F38 T41 R42 N43 R49 I66 P78 T79 H80 L98 Y111 P114 N115 T131 L132 Y145 C155 K156 I157 S158 P159 I160 L161 D162 L163 V166 V167 Q168 Q169 L175 N176 V177

N178 T181 N187 Q188 K189 W190 T191 I192 R193 Y194 N195 E196 E197 K198 Y201 Q202 F203 F204 N205 T206 I207 L208 S209 N210 G211 V212 L213 T214 W215 I216 F217 S218 N219 G220 N221 T222 V223 R224 S227 S228 N229 D230 A235 Q236 Y237 I240 N241 P242 V243 S244 D245 T253 N254

L255 R256 D257 T258 T259 K260 A261 L262 D263 L264 Y265 G266 T269 A274 I275 Q276 V277 F278 N279 Y280 H281 G282 D283 D284 N285 Q286 K287 W288 N289 I290 R291 N292 P293 P294 GLY SER ALA

• Molecule 3: HA-17

Chain C:

GLY PRO SER VAL GLU ARG THR PHE L8 N26 Y49 R54 C55 F56 K66 F75 L78 D79 W87 I90 K91 I92 L102 N106 E107 L108 D109 Y110 A111 W112 L122 P125 L128 L129 F132 D133 I134 I146

• Molecule 3: HA-17

Chain G:

GLY PRO SER VAL GLU ARG THR PHE L8 N14 N26 Y49 R54 C55 F56 K57 I58 K66 F75 L78 D79 I90 K91 I92 L102 M106 E107 Y110 L122 S123 L128 L129 I134 I146

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	159.59Å 187.49Å 120.43Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.20 – 3.73 49.20 – 3.73	Depositor EDS
% Data completeness (in resolution range)	98.7 (49.20-3.73) 99.5 (49.20-3.73)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.17 (at 3.77Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.2_869)	Depositor
R, R_{free}	0.268 , 0.293 0.263 , 0.287	Depositor DCC
R_{free} test set	1907 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	111.9	Xtriage
Anisotropy	0.447	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 97.3	EDS
L-test for twinning[2]	$< |L| > = 0.42, < L^2 > = 0.25$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	15326	wwPDB-VP
Average B, all atoms (Å^2)	169.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.40% 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	0.24	0/1886	0.45	0/2570
1	E	0.24	0/1894	0.45	0/2581
2	B	0.49	0/2393	0.59	1/3267 (0.0%)
2	D	0.46	0/2393	0.61	1/3267 (0.0%)
2	F	0.49	0/2393	0.59	1/3267 (0.0%)
2	H	0.47	0/2393	0.61	1/3267 (0.0%)
3	C	0.50	0/1168	0.62	0/1587
3	G	0.50	0/1168	0.62	0/1587
All	All	0.44	0/15688	0.57	4/21393 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	H	230	ASP	CB-CG-OD2	5.20	122.98	118.30
2	F	230	ASP	CB-CG-OD2	5.20	122.97	118.30
2	D	230	ASP	CB-CG-OD2	5.18	122.97	118.30
2	B	230	ASP	CB-CG-OD2	5.16	122.94	118.30

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	1852	0	1807	61	9

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	1860	0	1818	59	4
2	B	2334	0	2250	46	7
2	D	2334	0	2250	35	3
2	F	2334	0	2250	41	0
2	H	2334	0	2250	42	3
3	C	1139	0	1097	57	0
3	G	1139	0	1097	50	0
All	All	15326	0	14819	255	13

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 8.

The worst 5 of 255 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:575:ILE:HG12	3:G:49:TYR:CE1	1.62	1.33
1:E:575:ILE:HD13	3:G:56:PHE:CZ	1.65	1.31
1:A:575:ILE:HG12	3:C:49:TYR:CE1	1.74	1.22
2:B:294:PRO:HB3	2:F:23:ASP:OD2	1.40	1.17
1:E:575:ILE:CG1	3:G:49:TYR:CE1	2.27	1.16

The worst 5 of 13 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:616:TYR:CB	2:B:272:GLY:O[3_456]	1.30	0.90
1:A:616:TYR:CD2	2:B:272:GLY:CA[3_456]	1.34	0.86
2:B:173:THR:O	1:E:481:TYR:OH[4_546]	1.55	0.65
1:A:455:ILE:O	1:E:455:ILE:O[2_455]	1.65	0.55
1:A:616:TYR:CB	2:B:272:GLY:C[3_456]	1.86	0.34

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	224/254 (88%)	212 (95%)	12 (5%)	0	100	100
1	E	225/254 (89%)	213 (95%)	11 (5%)	1 (0%)	34	69
2	B	285/296 (96%)	276 (97%)	9 (3%)	0	100	100
2	D	285/296 (96%)	269 (94%)	15 (5%)	1 (0%)	34	69
2	F	285/296 (96%)	277 (97%)	8 (3%)	0	100	100
2	H	285/296 (96%)	270 (95%)	14 (5%)	1 (0%)	34	69
3	C	137/147 (93%)	133 (97%)	4 (3%)	0	100	100
3	G	137/147 (93%)	133 (97%)	4 (3%)	0	100	100
All	All	1863/1986 (94%)	1783 (96%)	77 (4%)	3 (0%)	47	78

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	D	265	TYR
2	H	265	TYR
1	E	488	SER

5.3.2 Protein sidechains ⓘ

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	211/232 (91%)	197 (93%)	14 (7%)	16	48
1	E	212/232 (91%)	198 (93%)	14 (7%)	16	48
2	B	263/270 (97%)	253 (96%)	10 (4%)	33	61
2	D	263/270 (97%)	245 (93%)	18 (7%)	16	47
2	F	263/270 (97%)	253 (96%)	10 (4%)	33	61
2	H	263/270 (97%)	245 (93%)	18 (7%)	16	47
3	C	130/138 (94%)	126 (97%)	4 (3%)	40	65
3	G	130/138 (94%)	126 (97%)	4 (3%)	40	65

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
All	All	1735/1820 (95%)	1643 (95%)	92 (5%)	22	54

5 of 92 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	519	ASN
3	G	78	LEU
1	E	552	THR
2	F	186	ARG
2	H	160	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
2	D	86	GLN
1	E	392	ASN
2	H	86	GLN
1	A	534	HIS
1	A	392	ASN

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, 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(Å²)	Q<0.9
1	A	228/254 (89%)	0.81	36 (15%)	2	2	99, 151, 222, 268	3 (1%)
1	E	229/254 (90%)	0.22	10 (4%)	34	29	101, 147, 213, 274	3 (1%)
2	B	286/296 (96%)	0.48	14 (4%)	29	26	85, 132, 183, 234	0
2	D	286/296 (96%)	1.15	64 (22%)	0	0	99, 207, 325, 398	0
2	F	286/296 (96%)	0.61	34 (11%)	4	5	93, 145, 203, 256	0
2	H	286/296 (96%)	1.09	69 (24%)	0	0	94, 195, 353, 422	0
3	C	139/147 (94%)	0.18	3 (2%)	62	55	107, 154, 194, 221	0
3	G	139/147 (94%)	0.19	2 (1%)	75	70	109, 147, 192, 208	0
All	All	1879/1986 (94%)	0.66	232 (12%)	4	4	85, 153, 303, 422	6 (0%)

The worst 5 of 232 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	202	GLN	15.5
2	D	215	TRP	8.5
2	D	237	TYR	7.6
1	A	381	ASN	7.6
2	H	169	GLN	7.6

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.