



wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 31, 2016 – 10:30 PM GMT

PDB ID : 1TUE
Title : The X-ray Structure of the Papillomavirus Helicase in Complex with its Molecular Matchmaker E2
Authors : Abbate, E.A.; Berger, J.M.; Botchan, M.R.
Deposited on : 2004-06-24
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
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk26865

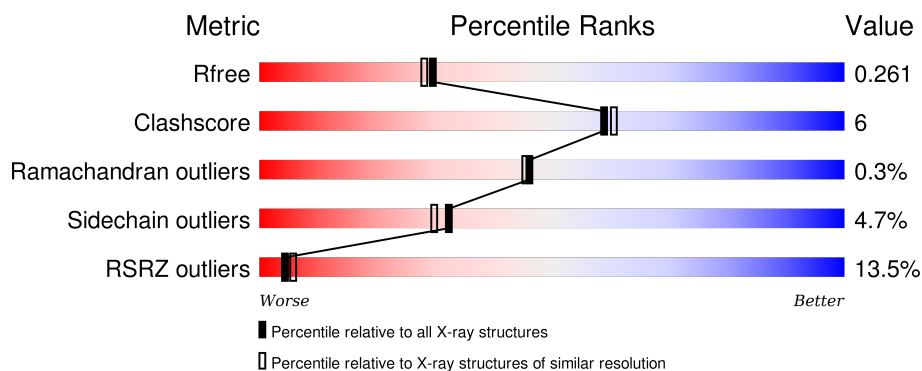
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 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}	91344	3939 (2.10-2.10)
Clashscore	102246	4460 (2.10-2.10)
Ramachandran outliers	100387	4413 (2.10-2.10)
Sidechain outliers	100360	4414 (2.10-2.10)
RSRZ outliers	91569	3948 (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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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 $\leq 5\%$. 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	212	<div> <div>5%</div> <div>77%</div> <div>15%</div> <div>• 5%</div> </div>
1	D	212	<div> <div>5%</div> <div>69%</div> <div>23%</div> <div>•• 5%</div> </div>
1	F	212	<div> <div>24%</div> <div>75%</div> <div>13%</div> <div>• 11%</div> </div>
1	H	212	<div> <div>12%</div> <div>72%</div> <div>17%</div> <div>• 10%</div> </div>
1	K	212	<div> <div>8%</div> <div>70%</div> <div>19%</div> <div>• 9%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	M	212	
2	B	218	
2	E	218	
2	G	218	
2	J	218	
2	L	218	
2	Q	218	

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 20186 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 Replication protein E1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	201	Total	C	N	O	S	0	0	0
			1645	1071	272	292	10			
1	D	202	Total	C	N	O	S	0	0	0
			1668	1087	280	291	10			
1	F	188	Total	C	N	O	S	0	0	0
			1551	1010	258	273	10			
1	H	191	Total	C	N	O	S	0	0	0
			1573	1025	261	277	10			
1	K	192	Total	C	N	O	S	0	0	0
			1584	1033	263	278	10			
1	M	196	Total	C	N	O	S	0	0	0
			1614	1054	266	284	10			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	420	GLY	-	EXPRESSION TAG	UNP P06789
A	421	SER	-	EXPRESSION TAG	UNP P06789
A	422	GLU	-	EXPRESSION TAG	UNP P06789
A	423	PHE	-	EXPRESSION TAG	UNP P06789
A	424	GLY	-	EXPRESSION TAG	UNP P06789
A	425	SER	-	EXPRESSION TAG	UNP P06789
A	426	GLY	-	EXPRESSION TAG	UNP P06789
A	427	SER	-	EXPRESSION TAG	UNP P06789
D	420	GLY	-	EXPRESSION TAG	UNP P06789
D	421	SER	-	EXPRESSION TAG	UNP P06789
D	422	GLU	-	EXPRESSION TAG	UNP P06789
D	423	PHE	-	EXPRESSION TAG	UNP P06789
D	424	GLY	-	EXPRESSION TAG	UNP P06789
D	425	SER	-	EXPRESSION TAG	UNP P06789
D	426	GLY	-	EXPRESSION TAG	UNP P06789
D	427	SER	-	EXPRESSION TAG	UNP P06789
F	420	GLY	-	EXPRESSION TAG	UNP P06789

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	421	SER	-	EXPRESSION TAG	UNP P06789
F	422	GLU	-	EXPRESSION TAG	UNP P06789
F	423	PHE	-	EXPRESSION TAG	UNP P06789
F	424	GLY	-	EXPRESSION TAG	UNP P06789
F	425	SER	-	EXPRESSION TAG	UNP P06789
F	426	GLY	-	EXPRESSION TAG	UNP P06789
F	427	SER	-	EXPRESSION TAG	UNP P06789
H	420	GLY	-	EXPRESSION TAG	UNP P06789
H	421	SER	-	EXPRESSION TAG	UNP P06789
H	422	GLU	-	EXPRESSION TAG	UNP P06789
H	423	PHE	-	EXPRESSION TAG	UNP P06789
H	424	GLY	-	EXPRESSION TAG	UNP P06789
H	425	SER	-	EXPRESSION TAG	UNP P06789
H	426	GLY	-	EXPRESSION TAG	UNP P06789
H	427	SER	-	EXPRESSION TAG	UNP P06789
K	420	GLY	-	EXPRESSION TAG	UNP P06789
K	421	SER	-	EXPRESSION TAG	UNP P06789
K	422	GLU	-	EXPRESSION TAG	UNP P06789
K	423	PHE	-	EXPRESSION TAG	UNP P06789
K	424	GLY	-	EXPRESSION TAG	UNP P06789
K	425	SER	-	EXPRESSION TAG	UNP P06789
K	426	GLY	-	EXPRESSION TAG	UNP P06789
K	427	SER	-	EXPRESSION TAG	UNP P06789
M	420	GLY	-	EXPRESSION TAG	UNP P06789
M	421	SER	-	EXPRESSION TAG	UNP P06789
M	422	GLU	-	EXPRESSION TAG	UNP P06789
M	423	PHE	-	EXPRESSION TAG	UNP P06789
M	424	GLY	-	EXPRESSION TAG	UNP P06789
M	425	SER	-	EXPRESSION TAG	UNP P06789
M	426	GLY	-	EXPRESSION TAG	UNP P06789
M	427	SER	-	EXPRESSION TAG	UNP P06789

- Molecule 2 is a protein called Regulatory protein E2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	194	Total	C	N	O	S	0	0	0
			1599	1015	265	310	9			
2	E	196	Total	C	N	O	S	0	0	0
			1607	1020	266	313	8			
2	G	196	Total	C	N	O	S	0	0	0
			1608	1020	266	314	8			
2	J	194	Total	C	N	O	S	0	0	0
			1592	1010	264	310	8			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	L	193	Total	C	N	O	S	0	0	0
			1584	1006	262	308	8			
2	Q	194	Total	C	N	O	S	0	0	0
			1592	1010	264	310	8			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-2	GLY	-	EXPRESSION TAG	UNP Q80B71
B	-1	SER	-	EXPRESSION TAG	UNP Q80B71
B	0	HIS	-	EXPRESSION TAG	UNP Q80B71
E	-2	GLY	-	EXPRESSION TAG	UNP Q80B71
E	-1	SER	-	EXPRESSION TAG	UNP Q80B71
E	0	HIS	-	EXPRESSION TAG	UNP Q80B71
G	-2	GLY	-	EXPRESSION TAG	UNP Q80B71
G	-1	SER	-	EXPRESSION TAG	UNP Q80B71
G	0	HIS	-	EXPRESSION TAG	UNP Q80B71
J	-2	GLY	-	EXPRESSION TAG	UNP Q80B71
J	-1	SER	-	EXPRESSION TAG	UNP Q80B71
J	0	HIS	-	EXPRESSION TAG	UNP Q80B71
L	-2	GLY	-	EXPRESSION TAG	UNP Q80B71
L	-1	SER	-	EXPRESSION TAG	UNP Q80B71
L	0	HIS	-	EXPRESSION TAG	UNP Q80B71
Q	-2	GLY	-	EXPRESSION TAG	UNP Q80B71
Q	-1	SER	-	EXPRESSION TAG	UNP Q80B71
Q	0	HIS	-	EXPRESSION TAG	UNP Q80B71

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	132	Total	O	0	0
			132	132		
3	B	97	Total	O	0	0
			97	97		
3	D	109	Total	O	0	0
			109	109		
3	E	98	Total	O	0	0
			98	98		
3	F	67	Total	O	0	0
			67	67		
3	G	97	Total	O	0	0
			97	97		

Continued on next page...

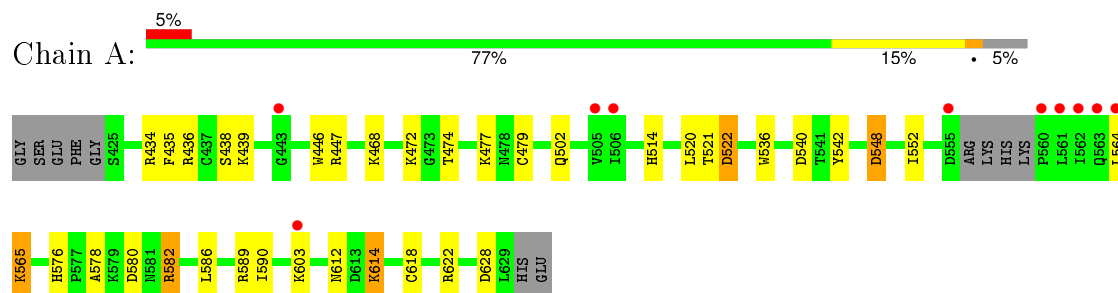
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	H	87	Total 87	O 87	0	0
3	J	54	Total 54	O 54	0	0
3	K	112	Total 112	O 112	0	0
3	L	32	Total 32	O 32	0	0
3	M	55	Total 55	O 55	0	0
3	Q	29	Total 29	O 29	0	0

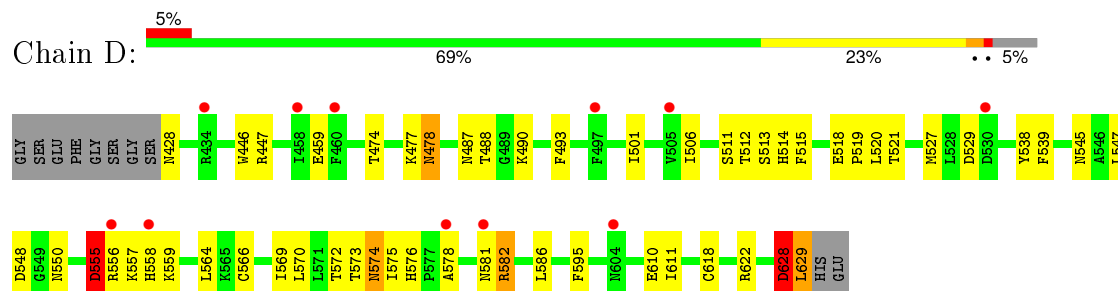
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

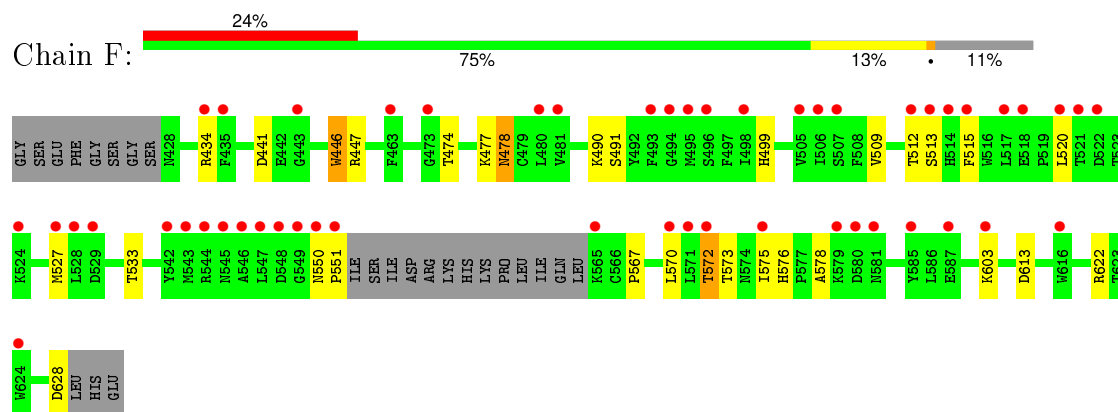
- Molecule 1: Replication protein E1



- Molecule 1: Replication protein E1



- Molecule 1: Replication protein E1



- Molecule 1: Replication protein E1



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	82.30Å 88.75Å 375.03Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.10 29.79 – 2.10	Depositor EDS
% Data completeness (in resolution range)	99.0 (30.00-2.10) 99.0 (29.79-2.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.05	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.40 (at 2.10Å)	Xtriage
Refinement program	REFMAC 5.2.0003	Depositor
R, R_{free}	0.219 , 0.263 0.218 , 0.261	Depositor DCC
R_{free} test set	8003 reflections (5.28%)	DCC
Wilson B-factor (Å ²)	34.4	Xtriage
Anisotropy	0.155	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 47.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	0 of 159542 reflections	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	20186	wwPDB-VP
Average B, all atoms (Å ²)	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.51% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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	0.47	0/1697	0.70	3/2304 (0.1%)
1	D	0.45	0/1722	0.70	4/2338 (0.2%)
1	F	0.40	0/1602	0.67	1/2175 (0.0%)
1	H	0.44	0/1624	0.68	5/2205 (0.2%)
1	K	0.44	0/1635	0.68	4/2220 (0.2%)
1	M	0.40	0/1664	0.67	5/2258 (0.2%)
2	B	0.41	0/1643	0.69	3/2229 (0.1%)
2	E	0.41	0/1650	0.70	3/2240 (0.1%)
2	G	0.39	0/1651	0.69	3/2240 (0.1%)
2	J	0.36	0/1635	0.66	3/2218 (0.1%)
2	L	0.35	0/1626	0.66	4/2204 (0.2%)
2	Q	0.34	0/1635	0.66	2/2218 (0.1%)
All	All	0.41	0/19784	0.68	40/26849 (0.1%)

There are no bond length outliers.

The worst 5 of 40 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	130	ASP	CB-CG-OD2	6.74	124.37	118.30
2	B	146	ASP	CB-CG-OD2	6.29	123.96	118.30
1	H	540	ASP	CB-CG-OD2	6.27	123.94	118.30
1	H	522	ASP	CB-CG-OD2	6.12	123.81	118.30
2	J	130	ASP	CB-CG-OD2	6.04	123.74	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	1645	0	1599	20	0
1	D	1668	0	1632	33	0
1	F	1551	0	1495	17	0
1	H	1573	0	1522	23	0
1	K	1584	0	1536	24	0
1	M	1614	0	1571	22	0
2	B	1599	0	1496	18	0
2	E	1607	0	1504	11	0
2	G	1608	0	1502	24	0
2	J	1592	0	1487	9	0
2	L	1584	0	1480	15	0
2	Q	1592	0	1487	9	0
3	A	132	0	0	2	0
3	B	97	0	0	0	0
3	D	109	0	0	7	0
3	E	98	0	0	0	0
3	F	67	0	0	0	0
3	G	97	0	0	4	0
3	H	87	0	0	1	0
3	J	54	0	0	2	0
3	K	112	0	0	5	0
3	L	32	0	0	1	0
3	M	55	0	0	0	0
3	Q	29	0	0	1	0
All	All	20186	0	18311	219	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 6.

The worst 5 of 219 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:F:490:LYS:HE2	1:F:572:THR:HG23	1.32	1.11
2:G:159:HIS:HD2	2:G:198:ILE:CD1	1.65	1.08
2:L:55:ILE:HG21	2:L:58:LEU:HD21	1.41	1.02
1:M:520:LEU:CD2	1:M:526:ALA:HB2	1.91	1.00
1:H:518:GLU:HG3	1:H:519:PRO:HD3	1.44	0.97

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	197/212 (93%)	194 (98%)	3 (2%)	0	100	100
1	D	200/212 (94%)	190 (95%)	8 (4%)	2 (1%)	19	13
1	F	184/212 (87%)	177 (96%)	7 (4%)	0	100	100
1	H	187/212 (88%)	181 (97%)	6 (3%)	0	100	100
1	K	188/212 (89%)	186 (99%)	2 (1%)	0	100	100
1	M	190/212 (90%)	187 (98%)	3 (2%)	0	100	100
2	B	192/218 (88%)	185 (96%)	5 (3%)	2 (1%)	19	13
2	E	194/218 (89%)	190 (98%)	3 (2%)	1 (0%)	34	30
2	G	194/218 (89%)	187 (96%)	6 (3%)	1 (0%)	34	30
2	J	192/218 (88%)	187 (97%)	5 (3%)	0	100	100
2	L	189/218 (87%)	180 (95%)	9 (5%)	0	100	100
2	Q	192/218 (88%)	186 (97%)	5 (3%)	1 (0%)	34	30
All	All	2299/2580 (89%)	2230 (97%)	62 (3%)	7 (0%)	46	45

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	184	GLY
2	E	194	GLY
2	G	197	VAL
2	Q	185	ASN
2	B	181	GLU

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	182/191 (95%)	174 (96%)	8 (4%)	35	33
1	D	184/191 (96%)	172 (94%)	12 (6%)	21	17
1	F	170/191 (89%)	162 (95%)	8 (5%)	32	30
1	H	173/191 (91%)	166 (96%)	7 (4%)	38	38
1	K	174/191 (91%)	165 (95%)	9 (5%)	29	25
1	M	178/191 (93%)	168 (94%)	10 (6%)	26	22
2	B	171/192 (89%)	162 (95%)	9 (5%)	28	25
2	E	172/192 (90%)	165 (96%)	7 (4%)	37	36
2	G	172/192 (90%)	163 (95%)	9 (5%)	29	25
2	J	170/192 (88%)	164 (96%)	6 (4%)	43	44
2	L	169/192 (88%)	164 (97%)	5 (3%)	48	51
2	Q	170/192 (88%)	163 (96%)	7 (4%)	37	36
All	All	2085/2298 (91%)	1988 (95%)	97 (5%)	32	30

5 of 97 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	603	LYS
1	H	446	TRP
1	M	603	LYS
2	G	11	ARG
2	G	75	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 41 such sidechains are listed below:

Mol	Chain	Res	Type
2	G	159	HIS
1	H	499	HIS
2	Q	35	GLN
2	G	185	ASN
2	G	195	ASN

5.3.3 RNA ⓘ

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 carbohydrates 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	201/212 (94%)	0.41	10 (4%) 32 41	16, 25, 48, 68	0
1	D	202/212 (95%)	0.58	11 (5%) 29 38	20, 29, 44, 58	0
1	F	188/212 (88%)	1.50	51 (27%) 1 1	26, 42, 73, 96	0
1	H	191/212 (90%)	0.88	26 (13%) 4 6	20, 33, 53, 73	0
1	K	192/212 (90%)	0.59	16 (8%) 14 19	20, 33, 55, 62	0
1	M	196/212 (92%)	0.67	17 (8%) 13 17	30, 40, 56, 70	0
2	B	194/218 (88%)	0.63	20 (10%) 9 12	18, 32, 76, 83	0
2	E	196/218 (89%)	0.44	10 (5%) 32 40	21, 30, 49, 69	0
2	G	196/218 (89%)	0.50	13 (6%) 22 29	25, 36, 49, 71	0
2	J	194/218 (88%)	1.07	32 (16%) 2 3	26, 47, 67, 81	0
2	L	193/218 (88%)	1.36	53 (27%) 1 1	26, 52, 98, 101	0
2	Q	194/218 (88%)	1.62	56 (28%) 1 1	30, 60, 96, 101	0
All	All	2337/2580 (90%)	0.85	315 (13%) 4 6	16, 36, 75, 101	0

The worst 5 of 315 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	551	PRO	11.1
2	J	147	ALA	8.9
1	F	546	ALA	8.6
1	F	550	ASN	8.4
1	H	554	ILE	8.2

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 carbohydrates 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.