wwPDB X-ray Structure Validation Summary Report (i)

Oct 8, 2024 – 02:12 PM JST

PDB ID	:	8WQK
Title	:	Crystal structure of the C-terminal domain of PEDV nucleocapsid protein
Authors	:	Zhang, J.T.; Jiao, Z.; Liang, R.; Shi, Y.J.; Peng, G.Q.
Deposited on	:	2023-10-11
Resolution	:	2.88 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 (i)) were used in the production of this report:

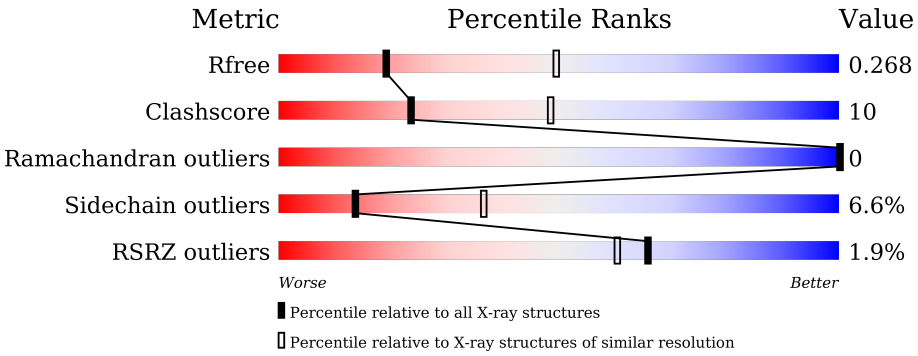
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

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 2.88 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.



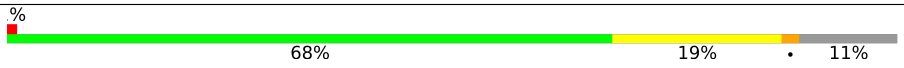
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	3316 (2.90-2.86)
Clashscore	180529	3609 (2.90-2.86)
Ramachandran outliers	177936	3529 (2.90-2.86)
Sidechain outliers	177891	3532 (2.90-2.86)
RSRZ outliers	164620	3319 (2.90-2.86)

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	114	% 78% 10% • 10%
1	B	114	4% 65% 18% • 13%
1	C	114	3% 71% 15% • 11%
1	D	114	67% 19% • 12%
1	E	114	2% 61% 24% • 14%
1	F	114	3% 57% 25% • 15%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	114	
1	H	114	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 6192 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 Nucleoprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	103	Total	C	N	O	S	0	0	0
			791	504	134	151	2			
1	B	99	Total	C	N	O	S	0	0	0
			756	485	124	145	2			
1	C	101	Total	C	N	O	S	0	0	0
			772	492	129	149	2			
1	D	100	Total	C	N	O	S	0	0	0
			761	488	125	146	2			
1	E	98	Total	C	N	O	S	0	0	0
			746	477	125	142	2			
1	F	97	Total	C	N	O	S	0	0	0
			733	470	118	143	2			
1	G	101	Total	C	N	O	S	0	0	0
			769	492	126	149	2			
1	H	101	Total	C	N	O	S	0	0	0
			769	492	126	149	2			

There are 88 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	initiating methionine	UNP Q07499
A	104	LEU	-	expression tag	UNP Q07499
A	105	GLU	-	expression tag	UNP Q07499
A	106	HIS	-	expression tag	UNP Q07499
A	107	HIS	-	expression tag	UNP Q07499
A	108	HIS	-	expression tag	UNP Q07499
A	109	HIS	-	expression tag	UNP Q07499
A	110	HIS	-	expression tag	UNP Q07499
A	111	HIS	-	expression tag	UNP Q07499
A	112	HIS	-	expression tag	UNP Q07499
A	113	HIS	-	expression tag	UNP Q07499
B	0	MET	-	initiating methionine	UNP Q07499
B	104	LEU	-	expression tag	UNP Q07499

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	105	GLU	-	expression tag	UNP Q07499
B	106	HIS	-	expression tag	UNP Q07499
B	107	HIS	-	expression tag	UNP Q07499
B	108	HIS	-	expression tag	UNP Q07499
B	109	HIS	-	expression tag	UNP Q07499
B	110	HIS	-	expression tag	UNP Q07499
B	111	HIS	-	expression tag	UNP Q07499
B	112	HIS	-	expression tag	UNP Q07499
B	113	HIS	-	expression tag	UNP Q07499
C	0	MET	-	initiating methionine	UNP Q07499
C	104	LEU	-	expression tag	UNP Q07499
C	105	GLU	-	expression tag	UNP Q07499
C	106	HIS	-	expression tag	UNP Q07499
C	107	HIS	-	expression tag	UNP Q07499
C	108	HIS	-	expression tag	UNP Q07499
C	109	HIS	-	expression tag	UNP Q07499
C	110	HIS	-	expression tag	UNP Q07499
C	111	HIS	-	expression tag	UNP Q07499
C	112	HIS	-	expression tag	UNP Q07499
C	113	HIS	-	expression tag	UNP Q07499
D	0	MET	-	initiating methionine	UNP Q07499
D	104	LEU	-	expression tag	UNP Q07499
D	105	GLU	-	expression tag	UNP Q07499
D	106	HIS	-	expression tag	UNP Q07499
D	107	HIS	-	expression tag	UNP Q07499
D	108	HIS	-	expression tag	UNP Q07499
D	109	HIS	-	expression tag	UNP Q07499
D	110	HIS	-	expression tag	UNP Q07499
D	111	HIS	-	expression tag	UNP Q07499
D	112	HIS	-	expression tag	UNP Q07499
D	113	HIS	-	expression tag	UNP Q07499
E	0	MET	-	initiating methionine	UNP Q07499
E	104	LEU	-	expression tag	UNP Q07499
E	105	GLU	-	expression tag	UNP Q07499
E	106	HIS	-	expression tag	UNP Q07499
E	107	HIS	-	expression tag	UNP Q07499
E	108	HIS	-	expression tag	UNP Q07499
E	109	HIS	-	expression tag	UNP Q07499
E	110	HIS	-	expression tag	UNP Q07499
E	111	HIS	-	expression tag	UNP Q07499
E	112	HIS	-	expression tag	UNP Q07499
E	113	HIS	-	expression tag	UNP Q07499

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	0	MET	-	initiating methionine	UNP Q07499
F	104	LEU	-	expression tag	UNP Q07499
F	105	GLU	-	expression tag	UNP Q07499
F	106	HIS	-	expression tag	UNP Q07499
F	107	HIS	-	expression tag	UNP Q07499
F	108	HIS	-	expression tag	UNP Q07499
F	109	HIS	-	expression tag	UNP Q07499
F	110	HIS	-	expression tag	UNP Q07499
F	111	HIS	-	expression tag	UNP Q07499
F	112	HIS	-	expression tag	UNP Q07499
F	113	HIS	-	expression tag	UNP Q07499
G	0	MET	-	initiating methionine	UNP Q07499
G	104	LEU	-	expression tag	UNP Q07499
G	105	GLU	-	expression tag	UNP Q07499
G	106	HIS	-	expression tag	UNP Q07499
G	107	HIS	-	expression tag	UNP Q07499
G	108	HIS	-	expression tag	UNP Q07499
G	109	HIS	-	expression tag	UNP Q07499
G	110	HIS	-	expression tag	UNP Q07499
G	111	HIS	-	expression tag	UNP Q07499
G	112	HIS	-	expression tag	UNP Q07499
G	113	HIS	-	expression tag	UNP Q07499
H	0	MET	-	initiating methionine	UNP Q07499
H	104	LEU	-	expression tag	UNP Q07499
H	105	GLU	-	expression tag	UNP Q07499
H	106	HIS	-	expression tag	UNP Q07499
H	107	HIS	-	expression tag	UNP Q07499
H	108	HIS	-	expression tag	UNP Q07499
H	109	HIS	-	expression tag	UNP Q07499
H	110	HIS	-	expression tag	UNP Q07499
H	111	HIS	-	expression tag	UNP Q07499
H	112	HIS	-	expression tag	UNP Q07499
H	113	HIS	-	expression tag	UNP Q07499

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	26	Total O 26 26	0	0
2	B	15	Total O 15 15	0	0
2	C	7	Total O 7 7	0	0

Continued on next page...

Continued from previous page...

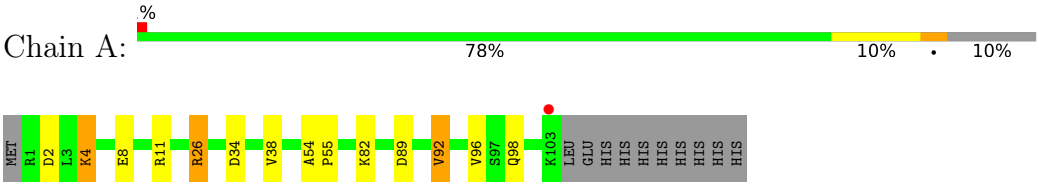
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	D	18	Total O 18 18	0	0
2	E	16	Total O 16 16	0	0
2	F	3	Total O 3 3	0	0
2	G	9	Total O 9 9	0	0
2	H	1	Total O 1 1	0	0

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.

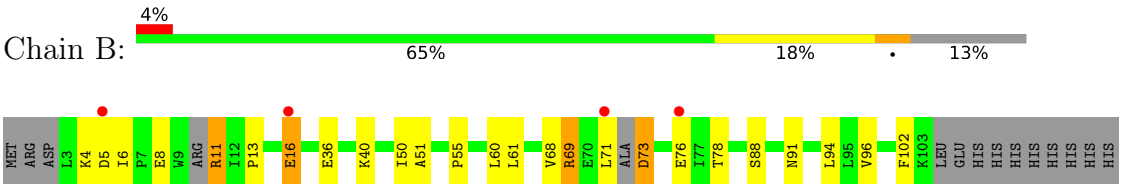
• Molecule 1: Nucleoprotein

Chain A:



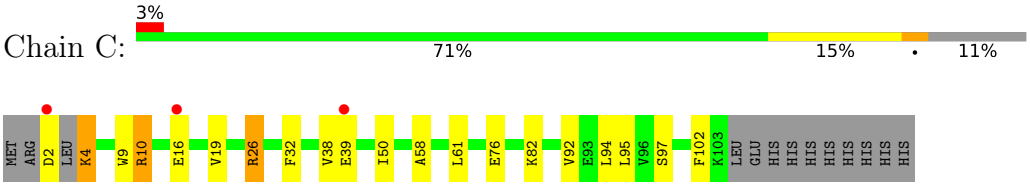
• Molecule 1: Nucleoprotein

Chain B:



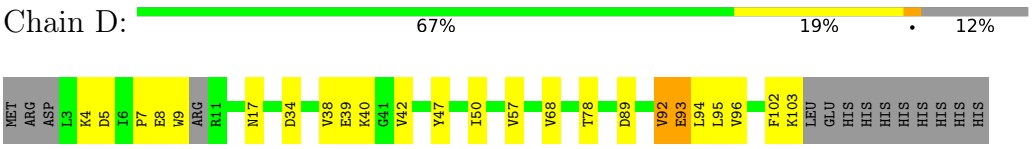
• Molecule 1: Nucleoprotein

Chain C:



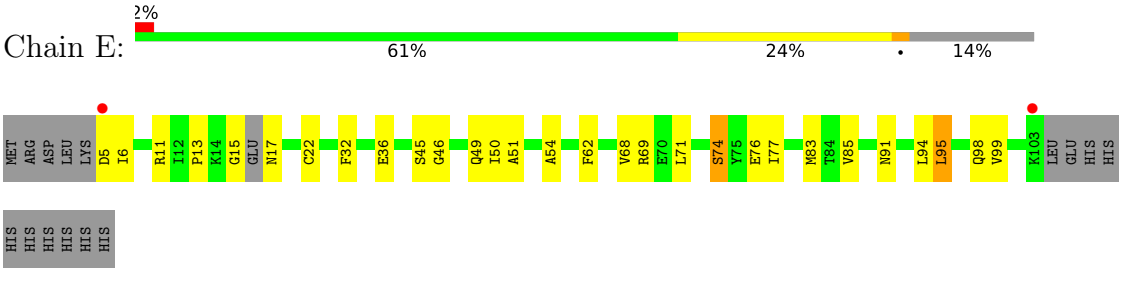
• Molecule 1: Nucleoprotein

Chain D:



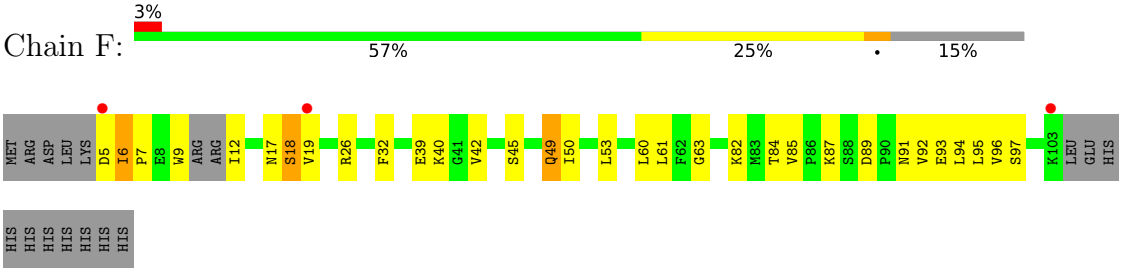
• Molecule 1: Nucleoprotein

Chain E:



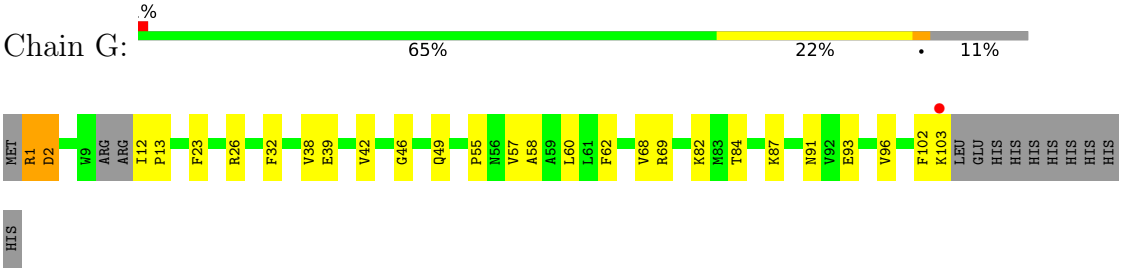
• Molecule 1: Nucleoprotein

Chain F:



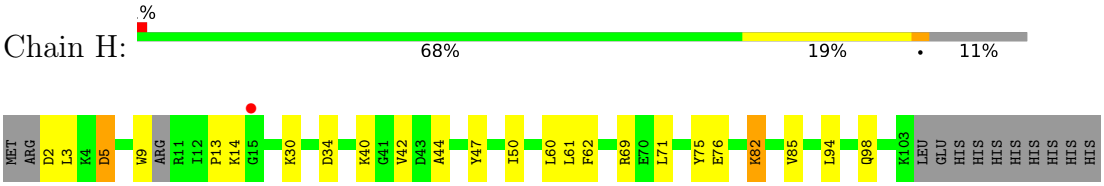
• Molecule 1: Nucleoprotein

Chain G:



• Molecule 1: Nucleoprotein

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	52.92Å 93.50Å 97.20Å 90.00° 97.37° 90.00°	Depositor
Resolution (Å)	26.24 – 2.88 26.24 – 2.88	Depositor EDS
% Data completeness (in resolution range)	94.0 (26.24-2.88) 94.0 (26.24-2.88)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.55 (at 2.51Å)	Xtriage
Refinement program	PHENIX (1.20.1_4487: ???)	Depositor
R, R_{free}	0.220 , 0.275 0.219 , 0.268	Depositor DCC
R_{free} test set	1116 reflections (5.22%)	wwPDB-VP
Wilson B-factor (Å^2)	46.4	Xtriage
Anisotropy	0.575	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 44.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	6192	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.58% 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	0.49	0/808	0.71	0/1094
1	B	0.51	0/771	0.68	0/1042
1	C	0.54	0/788	0.83	2/1066 (0.2%)
1	D	0.61	1/777 (0.1%)	0.81	0/1052
1	E	0.52	0/762	0.76	0/1032
1	F	0.58	1/749 (0.1%)	0.81	0/1016
1	G	0.57	0/785	0.84	1/1063 (0.1%)
1	H	0.57	0/785	0.84	1/1063 (0.1%)
All	All	0.55	2/6225 (0.0%)	0.79	4/8428 (0.0%)

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	B	0	2
1	C	0	1
All	All	0	3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	93	GLU	CD-OE1	5.99	1.32	1.25
1	F	6	ILE	CB-CG2	-5.31	1.36	1.52

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	76	GLU	CA-CB-CG	5.87	126.31	113.40
1	H	14	LYS	CA-CB-CG	5.34	125.15	113.40
1	G	2	ASP	CB-CG-OD2	-5.10	113.71	118.30
1	C	10	ARG	CG-CD-NE	-5.07	101.16	111.80

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	11	ARG	Sidechain
1	B	69	ARG	Sidechain
1	C	26	ARG	Sidechain

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	791	0	778	11	0
1	B	756	0	738	21	1
1	C	772	0	750	20	0
1	D	761	0	744	14	1
1	E	746	0	727	21	0
1	F	733	0	707	24	0
1	G	769	0	751	25	0
1	H	769	0	748	18	0
2	A	26	0	0	0	0
2	B	15	0	0	1	0
2	C	7	0	0	0	0
2	D	18	0	0	1	0
2	E	16	0	0	2	0
2	F	3	0	0	0	0
2	G	9	0	0	0	0
2	H	1	0	0	0	0
All	All	6192	0	5943	120	1

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 10.

The worst 5 of 120 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:H:5:ASP:N	1:H:5:ASP:OD1	2.00	0.94
1:F:50:ILE:HG23	1:F:95:LEU:HD13	1.58	0.86

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:10:ARG:HH22	1:E:54:ALA:N	1.76	0.83
1:F:60:LEU:HD11	1:H:60:LEU:HD11	1.68	0.76
1:E:50:ILE:HG23	1:E:95:LEU:HD12	1.71	0.72

All (1) 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:B:69:ARG:NH1	1:D:8:GLU:OE1[1_455]	1.97	0.23

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	101/114 (89%)	99 (98%)	2 (2%)	0	100	100
1	B	93/114 (82%)	88 (95%)	5 (5%)	0	100	100
1	C	98/114 (86%)	93 (95%)	5 (5%)	0	100	100
1	D	96/114 (84%)	93 (97%)	3 (3%)	0	100	100
1	E	94/114 (82%)	90 (96%)	4 (4%)	0	100	100
1	F	93/114 (82%)	87 (94%)	6 (6%)	0	100	100
1	G	97/114 (85%)	92 (95%)	5 (5%)	0	100	100
1	H	97/114 (85%)	93 (96%)	4 (4%)	0	100	100
All	All	769/912 (84%)	735 (96%)	34 (4%)	0	100	100

There are no Ramachandran outliers to report.

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	82/93 (88%)	78 (95%)	4 (5%)	21 50
1	B	79/93 (85%)	75 (95%)	4 (5%)	20 48
1	C	80/93 (86%)	76 (95%)	4 (5%)	20 49
1	D	79/93 (85%)	71 (90%)	8 (10%)	6 18
1	E	77/93 (83%)	70 (91%)	7 (9%)	7 22
1	F	76/93 (82%)	69 (91%)	7 (9%)	7 22
1	G	80/93 (86%)	76 (95%)	4 (5%)	20 49
1	H	80/93 (86%)	76 (95%)	4 (5%)	20 49
All	All	633/744 (85%)	591 (93%)	42 (7%)	14 36

5 of 42 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	17	ASN
1	G	69	ARG
1	F	18	SER
1	F	82	LYS
1	G	93	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	17	ASN
1	B	65	ASN
1	G	17	ASN
1	G	49	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 oligosaccharides 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

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	103/114 (90%)	-0.29	1 (0%) 79 75	24, 31, 44, 57	0
1	B	99/114 (86%)	-0.06	4 (4%) 43 37	26, 34, 63, 75	0
1	C	101/114 (88%)	-0.03	3 (2%) 52 47	27, 43, 57, 74	0
1	D	100/114 (87%)	-0.14	0 100 100	20, 37, 61, 83	0
1	E	98/114 (85%)	0.08	2 (2%) 64 59	27, 38, 64, 75	0
1	F	97/114 (85%)	0.48	3 (3%) 51 46	31, 60, 83, 98	0
1	G	101/114 (88%)	0.01	1 (0%) 79 75	26, 45, 64, 79	0
1	H	101/114 (88%)	0.23	1 (0%) 79 75	32, 49, 75, 88	0
All	All	800/912 (87%)	0.03	15 (1%) 66 60	20, 41, 71, 98	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	5	ASP	2.9
1	C	2	ASP	2.7
1	B	5	ASP	2.7
1	F	103	LYS	2.6
1	F	19	VAL	2.6

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.