



# wwPDB X-ray Structure Validation Summary Report

Jun 27, 2022 – 02:33 PM EDT

PDB ID : 7T7I  
Title : EBV nuclear egress complex  
Authors : Thorsen, M.K.; Heldwein, E.E.  
Deposited on : 2021-12-15  
Resolution : 3.97 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](mailto: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 : 2.29  
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) : 2.29

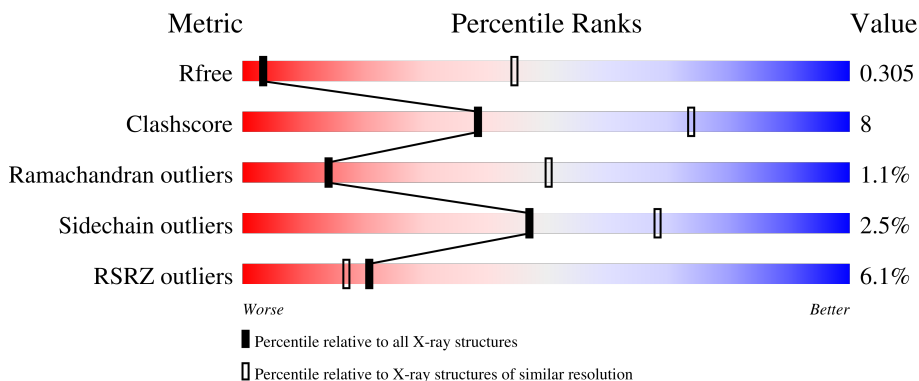
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.97 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.



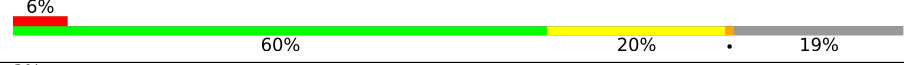

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1039 (4.26-3.70)
Clashscore	141614	1099 (4.26-3.70)
Ramachandran outliers	138981	1061 (4.26-3.70)
Sidechain outliers	138945	1053 (4.26-3.70)
RSRZ outliers	127900	1021 (4.30-3.66)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	200	 2% 90% 7%
1	C	200	 2% 83% 14%
1	E	200	 4% 90% 6%
1	G	200	 3% 84% 13%
1	I	200	 16% 61% 16% 24%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	B	257	
2	D	257	
2	F	257	
2	H	257	
2	J	257	

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 32335 atoms, of which 16193 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 Nuclear egress protein 2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	194	3030	956	1519	260	281	14	0	0	0
1	C	194	3030	956	1519	260	281	14	0	0	0
1	E	194	3058	964	1536	261	282	15	0	2	0
1	G	194	3044	960	1528	260	281	15	0	1	0
1	I	153	2446	781	1228	205	221	11	0	0	0

There are 25 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	GLY	-	expression tag	UNP P03185
A	-3	PRO	-	expression tag	UNP P03185
A	-2	LEU	-	expression tag	UNP P03185
A	-1	GLY	-	expression tag	UNP P03185
A	0	SER	-	expression tag	UNP P03185
C	-4	GLY	-	expression tag	UNP P03185
C	-3	PRO	-	expression tag	UNP P03185
C	-2	LEU	-	expression tag	UNP P03185
C	-1	GLY	-	expression tag	UNP P03185
C	0	SER	-	expression tag	UNP P03185
E	-4	GLY	-	expression tag	UNP P03185
E	-3	PRO	-	expression tag	UNP P03185
E	-2	LEU	-	expression tag	UNP P03185
E	-1	GLY	-	expression tag	UNP P03185
E	0	SER	-	expression tag	UNP P03185
G	-4	GLY	-	expression tag	UNP P03185
G	-3	PRO	-	expression tag	UNP P03185
G	-2	LEU	-	expression tag	UNP P03185
G	-1	GLY	-	expression tag	UNP P03185

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
G	0	SER	-	expression tag	UNP P03185
I	-4	GLY	-	expression tag	UNP P03185
I	-3	PRO	-	expression tag	UNP P03185
I	-2	LEU	-	expression tag	UNP P03185
I	-1	GLY	-	expression tag	UNP P03185
I	0	SER	-	expression tag	UNP P03185

- Molecule 2 is a protein called Nuclear egress protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				S
2	B	239	Total	C	H	N	O	S	0	0	0
			3736	1195	1867	314	348	12			
2	D	242	Total	C	H	N	O	S	0	0	0
			3791	1211	1897	319	352	12			
2	F	208	Total	C	H	N	O	S	0	0	0
			3227	1038	1611	269	298	11			
2	H	241	Total	C	H	N	O	S	0	0	0
			3775	1206	1888	318	351	12			
2	J	203	Total	C	H	N	O	S	0	0	0
			3190	1024	1600	263	291	12			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	62	GLY	-	expression tag	UNP P0CK47
B	63	PRO	-	expression tag	UNP P0CK47
B	64	GLY	-	expression tag	UNP P0CK47
B	65	SER	-	expression tag	UNP P0CK47
D	62	GLY	-	expression tag	UNP P0CK47
D	63	PRO	-	expression tag	UNP P0CK47
D	64	GLY	-	expression tag	UNP P0CK47
D	65	SER	-	expression tag	UNP P0CK47
F	62	GLY	-	expression tag	UNP P0CK47
F	63	PRO	-	expression tag	UNP P0CK47
F	64	GLY	-	expression tag	UNP P0CK47
F	65	SER	-	expression tag	UNP P0CK47
H	62	GLY	-	expression tag	UNP P0CK47
H	63	PRO	-	expression tag	UNP P0CK47
H	64	GLY	-	expression tag	UNP P0CK47
H	65	SER	-	expression tag	UNP P0CK47
J	62	GLY	-	expression tag	UNP P0CK47
J	63	PRO	-	expression tag	UNP P0CK47

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
J	64	GLY	-	expression tag	UNP P0CK47
J	65	SER	-	expression tag	UNP P0CK47

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total Zn 1 1	0	0
3	D	1	Total Zn 1 1	0	0
3	F	1	Total Zn 1 1	0	0
3	H	1	Total Zn 1 1	0	0
3	J	1	Total Zn 1 1	0	0

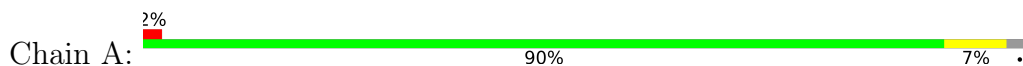
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	F	1	Total O 1 1	0	0
4	G	1	Total O 1 1	0	0
4	I	1	Total O 1 1	0	0

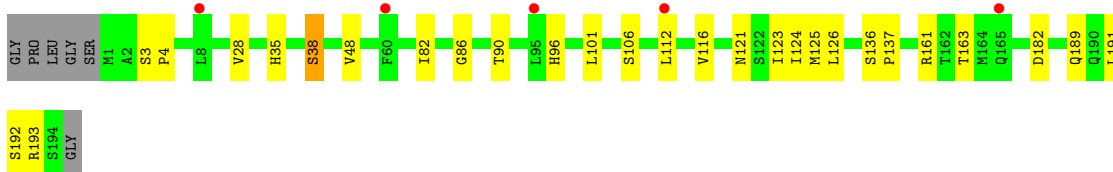
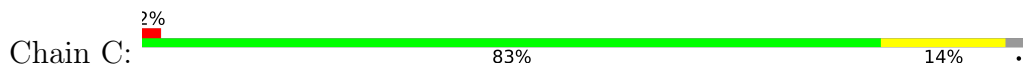
### 3 Residue-property plots [i](#)

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.

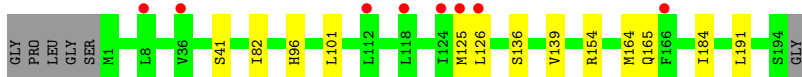
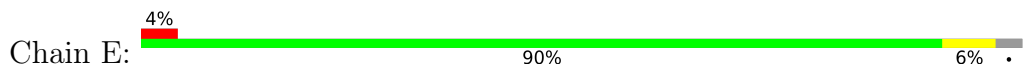
- Molecule 1: Nuclear egress protein 2



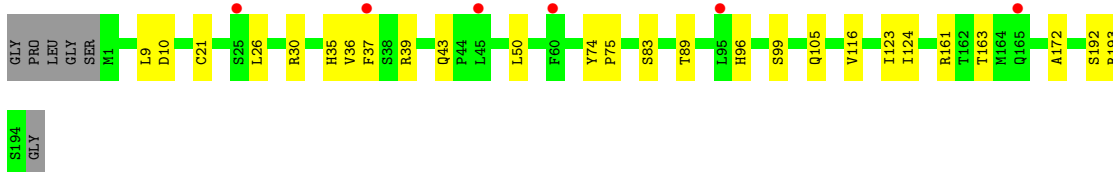
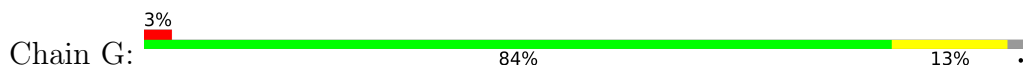
- Molecule 1: Nuclear egress protein 2



- Molecule 1: Nuclear egress protein 2

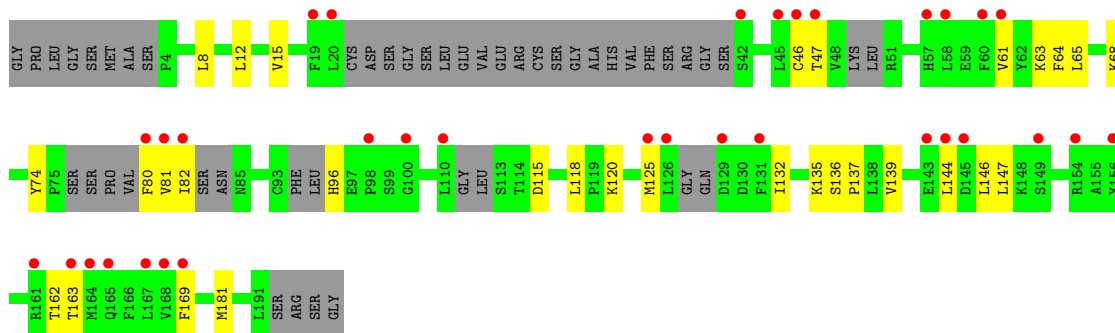


- Molecule 1: Nuclear egress protein 2

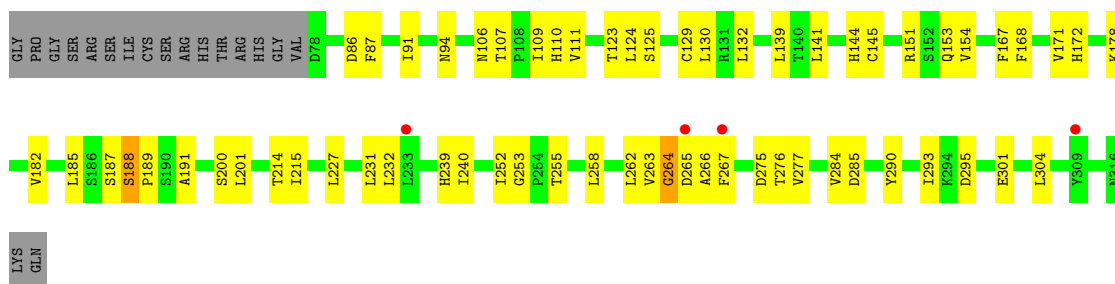


- Molecule 1: Nuclear egress protein 2

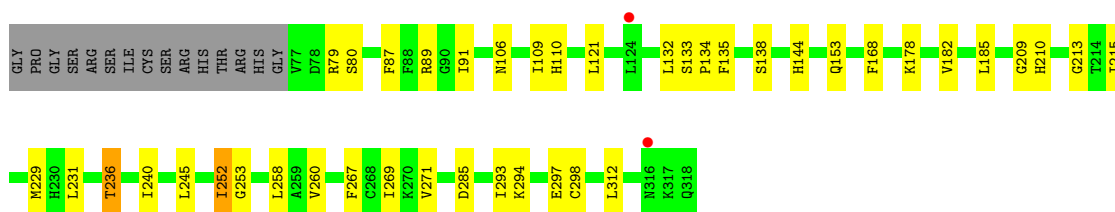
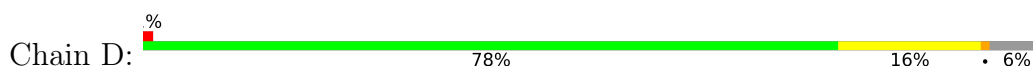




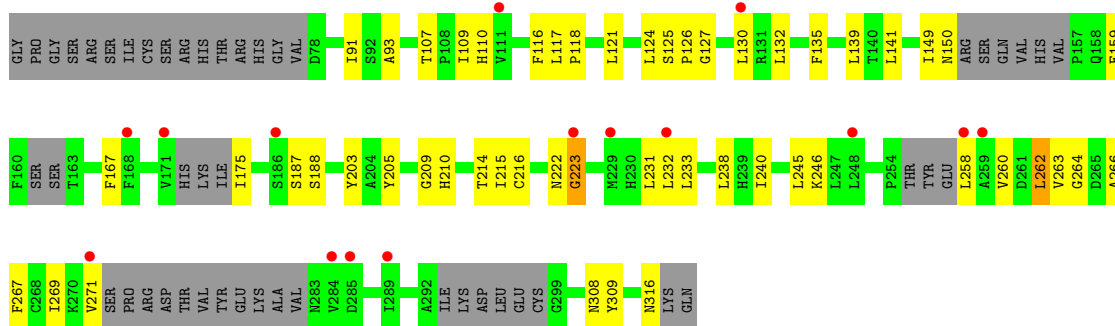
- Molecule 2: Nuclear egress protein 1



- Molecule 2: Nuclear egress protein 1

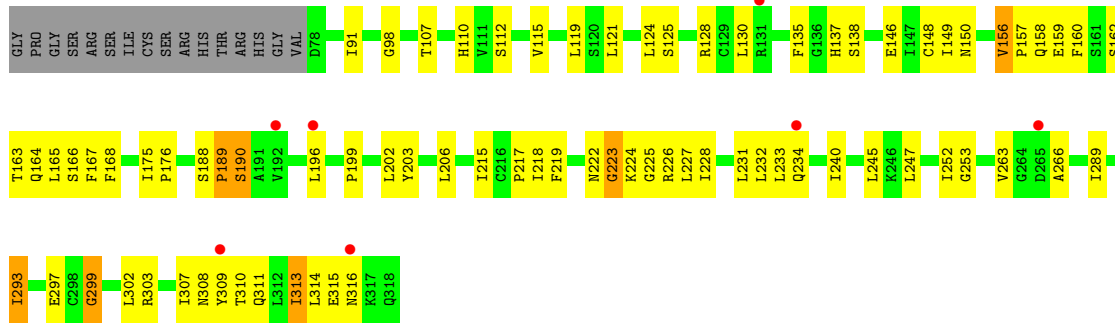


- Molecule 2: Nuclear egress protein 1

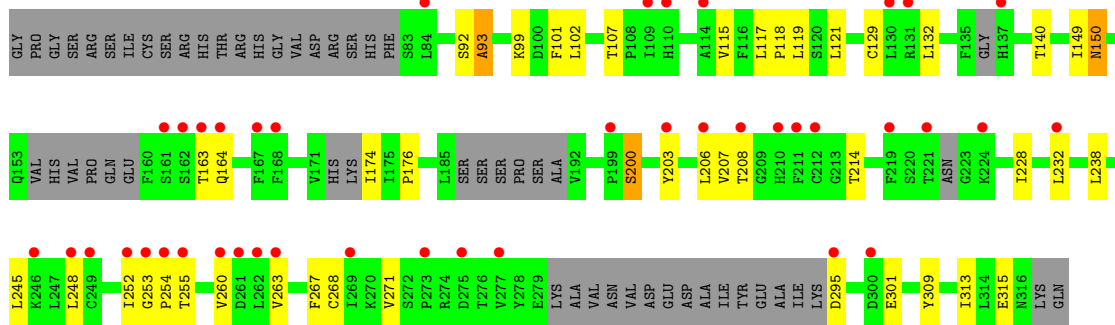


- Molecule 2: Nuclear egress protein 1





• Molecule 2: Nuclear egress protein 1



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	238.15Å 238.15Å 137.53Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	106.52 – 3.97 106.52 – 3.97	Depositor EDS
% Data completeness (in resolution range)	99.5 (106.52-3.97) 99.5 (106.52-3.97)	Depositor EDS
$R_{merge}$	0.28	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.40 (at 4.01Å)	Xtrriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, $R_{free}$	0.270 , 0.308 0.268 , 0.305	Depositor DCC
$R_{free}$ test set	2000 reflections (5.79%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	166.6	Xtrriage
Anisotropy	0.315	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 140.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.43$ , $\langle L^2 \rangle = 0.25$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	32335	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	219.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.22% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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

Bond lengths and bond angles in the following residue types are not validated in this section:  
ZN

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	0.28	0/1540	0.53	0/2078
1	C	0.27	0/1540	0.52	0/2078
1	E	0.27	0/1557	0.50	0/2100
1	G	0.26	0/1548	0.52	0/2088
1	I	0.24	0/1236	0.45	0/1659
2	B	0.28	0/1910	0.58	0/2592
2	D	0.30	0/1935	0.57	0/2625
2	F	0.28	0/1646	0.56	0/2224
2	H	0.28	0/1928	0.56	0/2615
2	J	0.25	0/1618	0.46	0/2186
All	All	0.27	0/16458	0.53	0/22245

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	1511	1519	1521	6	0
1	C	1511	1519	1521	27	2
1	E	1522	1536	1538	13	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	G	1516	1528	1530	19	0
1	I	1218	1228	1228	16	1
2	B	1869	1867	1867	33	2
2	D	1894	1897	1897	25	2
2	F	1616	1611	1611	44	1
2	H	1887	1888	1888	54	0
2	J	1590	1600	1600	31	0
3	B	1	0	0	0	0
3	D	1	0	0	0	0
3	F	1	0	0	0	0
3	H	1	0	0	0	0
3	J	1	0	0	0	0
4	F	1	0	0	0	0
4	G	1	0	0	0	0
4	I	1	0	0	0	0
All	All	16142	16193	16201	244	4

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 244 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)
2:H:289:ILE:O	2:H:293:ILE:HG23	1.80	0.82
2:F:238:LEU:HD12	2:F:267:PHE:CZ	2.17	0.79
2:H:175:ILE:HG22	2:H:176:PRO:HD3	1.64	0.78
2:F:127:GLY:O	2:F:141:LEU:HD12	1.84	0.77
1:A:90:THR:HG21	2:H:135:PHE:O	1.85	0.77

All (4) 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)
2:B:172:HIS:HA	2:D:135:PHE:HE1[7_555]	1.35	0.25
2:B:151:ARG:NH2	1:C:189:GLN:O[8_554]	2.09	0.11
1:C:136:SER:OG	2:D:80:SER:OG[8_554]	2.12	0.08
2:F:159:GLU:OE2	1:I:120:LYS:NZ[4_444]	2.16	0.04

## 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	192/200 (96%)	179 (93%)	11 (6%)	2 (1%)	15	52
1	C	192/200 (96%)	186 (97%)	6 (3%)	0	100	100
1	E	194/200 (97%)	188 (97%)	6 (3%)	0	100	100
1	G	193/200 (96%)	186 (96%)	7 (4%)	0	100	100
1	I	137/200 (68%)	124 (90%)	13 (10%)	0	100	100
2	B	237/257 (92%)	211 (89%)	21 (9%)	5 (2%)	7	38
2	D	240/257 (93%)	218 (91%)	20 (8%)	2 (1%)	19	57
2	F	194/257 (76%)	179 (92%)	11 (6%)	4 (2%)	7	38
2	H	239/257 (93%)	214 (90%)	17 (7%)	8 (3%)	4	30
2	J	189/257 (74%)	174 (92%)	14 (7%)	1 (0%)	29	66
All	All	2007/2285 (88%)	1859 (93%)	126 (6%)	22 (1%)	14	50

5 of 22 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	188	SER
2	H	225	GLY
2	J	93	ALA
1	A	104	GLY
2	B	123	THR

### 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	174/177 (98%)	169 (97%)	5 (3%)	42	64
1	C	174/177 (98%)	171 (98%)	3 (2%)	60	78
1	E	176/177 (99%)	176 (100%)	0	100	100
1	G	175/177 (99%)	175 (100%)	0	100	100
1	I	140/177 (79%)	136 (97%)	4 (3%)	42	64
2	B	215/230 (94%)	209 (97%)	6 (3%)	43	65
2	D	218/230 (95%)	211 (97%)	7 (3%)	39	62
2	F	185/230 (80%)	180 (97%)	5 (3%)	44	66
2	H	217/230 (94%)	210 (97%)	7 (3%)	39	62
2	J	184/230 (80%)	174 (95%)	10 (5%)	22	50
All	All	1858/2035 (91%)	1811 (98%)	47 (2%)	47	68

5 of 47 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	H	215	ILE
1	I	163	THR
2	H	252	ILE
2	H	313	ILE
2	J	129	CYS

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

Mol	Chain	Res	Type
1	A	43	GLN
2	B	153	GLN
2	D	251	ASN
2	H	251	ASN
2	H	316	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 [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 5 ligands modelled in this entry, 5 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	194/200 (97%)	0.29	5 (2%) 56 46	127, 149, 210, 277	0
1	C	194/200 (97%)	0.48	5 (2%) 56 46	129, 157, 209, 259	0
1	E	194/200 (97%)	0.34	8 (4%) 37 30	148, 184, 228, 277	0
1	G	194/200 (97%)	0.17	6 (3%) 49 38	149, 178, 216, 273	0
1	I	153/200 (76%)	1.18	33 (21%) 0 1	207, 240, 284, 310	0
2	B	239/257 (92%)	0.37	4 (1%) 70 60	131, 188, 227, 250	0
2	D	242/257 (94%)	0.28	2 (0%) 86 79	136, 204, 243, 271	0
2	F	208/257 (80%)	0.23	15 (7%) 15 12	170, 239, 280, 318	0
2	H	241/257 (93%)	0.32	7 (2%) 51 41	148, 189, 231, 309	0
2	J	203/257 (78%)	0.97	41 (20%) 1 1	212, 255, 291, 325	0
All	All	2062/2285 (90%)	0.44	126 (6%) 21 17	127, 195, 270, 325	0

The worst 5 of 126 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	I	144	LEU	12.9
1	I	80	PHE	8.7
1	I	126	LEU	7.8
2	F	171	VAL	5.8
2	F	258	LEU	5.4

### 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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
3	ZN	H	401	1/1	0.90	0.15	145,145,145,145	0
3	ZN	J	401	1/1	0.93	0.11	205,205,205,205	0
3	ZN	D	401	1/1	0.97	0.29	140,140,140,140	0
3	ZN	F	401	1/1	0.97	0.10	166,166,166,166	0
3	ZN	B	401	1/1	0.98	0.23	125,125,125,125	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.