



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 13, 2024 – 12:31 pm BST

PDB ID : 2JJD  
Title : Protein Tyrosine Phosphatase, Receptor Type, E isoform  
Authors : Elkins, J.M.; Ugochukwu, E.; Alfano, I.; Barr, A.J.; Bunkoczi, G.; King, O.N.F.; Filippakopoulos, P.; Savitsky, P.; Salah, E.; Pike, A.; Johansson, C.; Das, S.; Burgess-Brown, N.A.; Gileadi, O.; von Delft, F.; Arrowsmith, C.H.; Bountra, C.; Edwards, A.M.; Knapp, S.  
Deposited on : 2008-03-31  
Resolution : 3.20 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	:	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

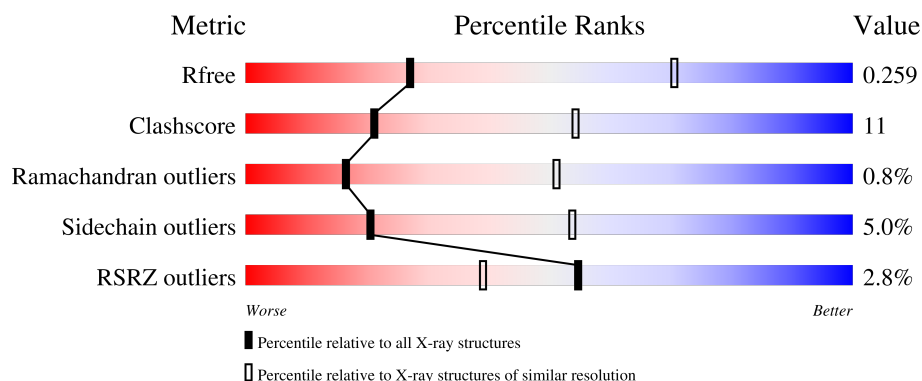
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.20 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	1370 (3.20-3.20)
Clashscore	180529	1497 (3.20-3.20)
Ramachandran outliers	177936	1479 (3.20-3.20)
Sidechain outliers	177891	1478 (3.20-3.20)
RSRZ outliers	164620	1371 (3.20-3.20)

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	599	<div> <div>2%</div> <div>70% 18% • 10%</div> </div>
1	B	599	<div> <div>2%</div> <div>71% 17% • 10%</div> </div>
1	C	599	<div> <div>3%</div> <div>70% 19% • 10%</div> </div>
1	D	599	<div> <div>4%</div> <div>71% 18% • 10%</div> </div>
1	E	599	<div> <div>2%</div> <div>74% 15% • 10%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	F	599	<div><div></div><div>3%</div><div>72%</div><div>16%</div><div>• 10%</div></div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 23750 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 RECEPTOR-TYPE TYROSINE-PROTEIN PHOSPHATASE EPSILON.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	537	Total	C	N	O	S	0	0	0
			3973	2574	689	689	21			
1	B	541	Total	C	N	O	S	0	0	0
			3972	2581	692	677	22			
1	C	541	Total	C	N	O	S	0	0	0
			3970	2582	690	676	22			
1	D	539	Total	C	N	O	S	0	0	0
			3936	2554	690	671	21			
1	E	541	Total	C	N	O	S	0	0	0
			3950	2569	688	672	21			
1	F	542	Total	C	N	O	S	0	0	0
			3937	2560	692	664	21			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	581	GLY	GLU	cloning artifact	UNP P23469
B	581	GLY	GLU	cloning artifact	UNP P23469
C	581	GLY	GLU	cloning artifact	UNP P23469
D	581	GLY	GLU	cloning artifact	UNP P23469
E	581	GLY	GLU	cloning artifact	UNP P23469
F	581	GLY	GLU	cloning artifact	UNP P23469

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	2	Total	Cl	0	0
			2	2		
2	B	3	Total	Cl	0	0
			3	3		
2	C	1	Total	Cl	0	0
			1	1		

*Continued on next page...*

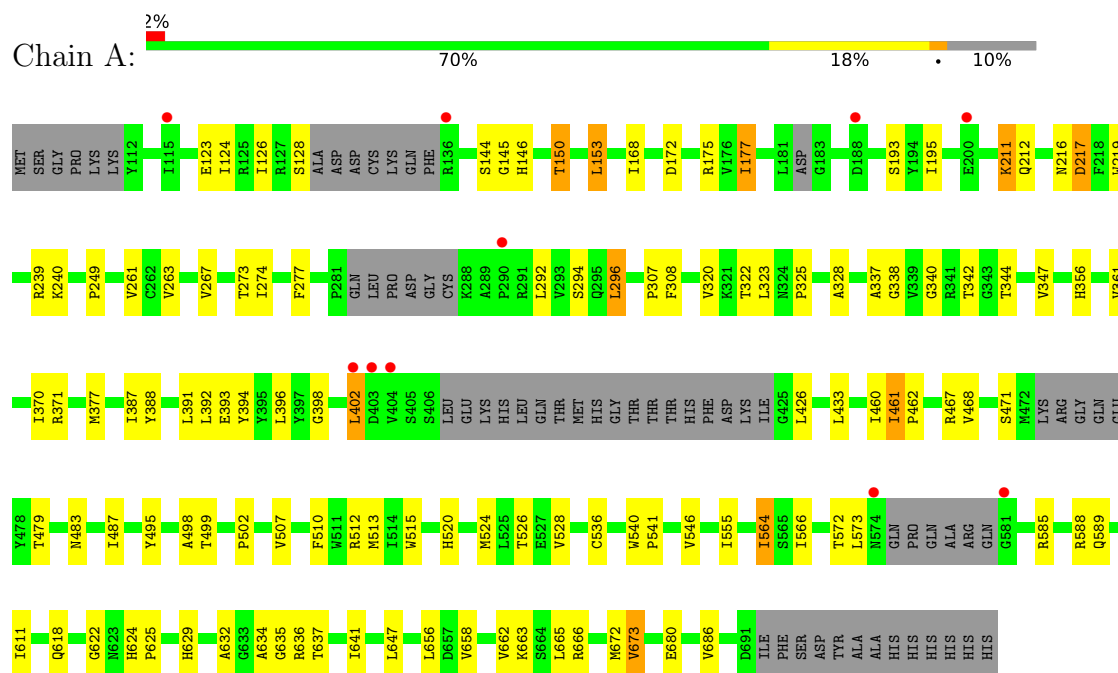
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	D	2	Total 2	Cl 2	0	0
2	E	2	Total 2	Cl 2	0	0
2	F	2	Total 2	Cl 2	0	0

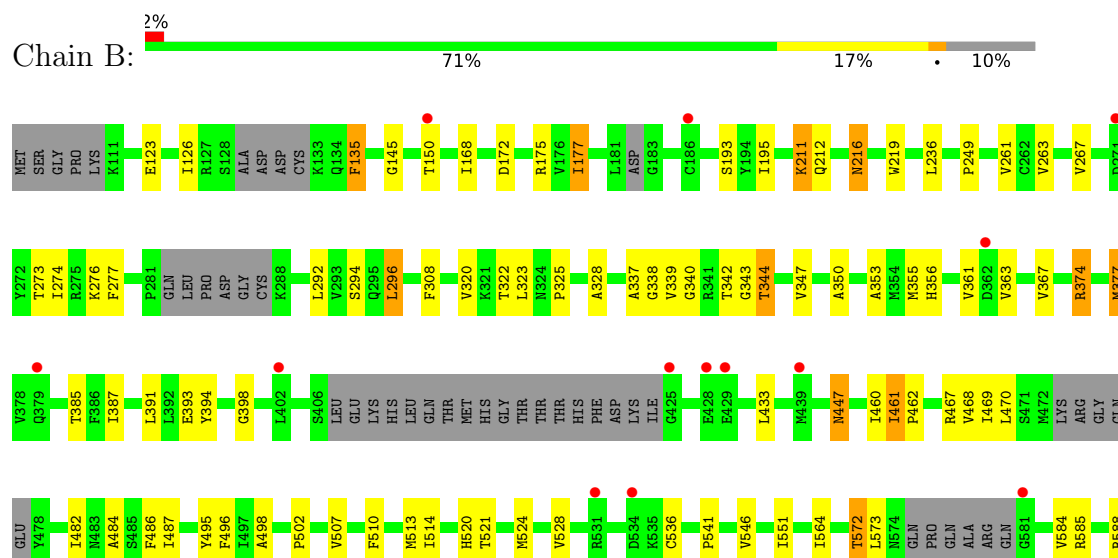
### 3 Residue-property plots

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: RECEPTOR-TYPE TYROSINE-PROTEIN PHOSPHATASE EPSILON

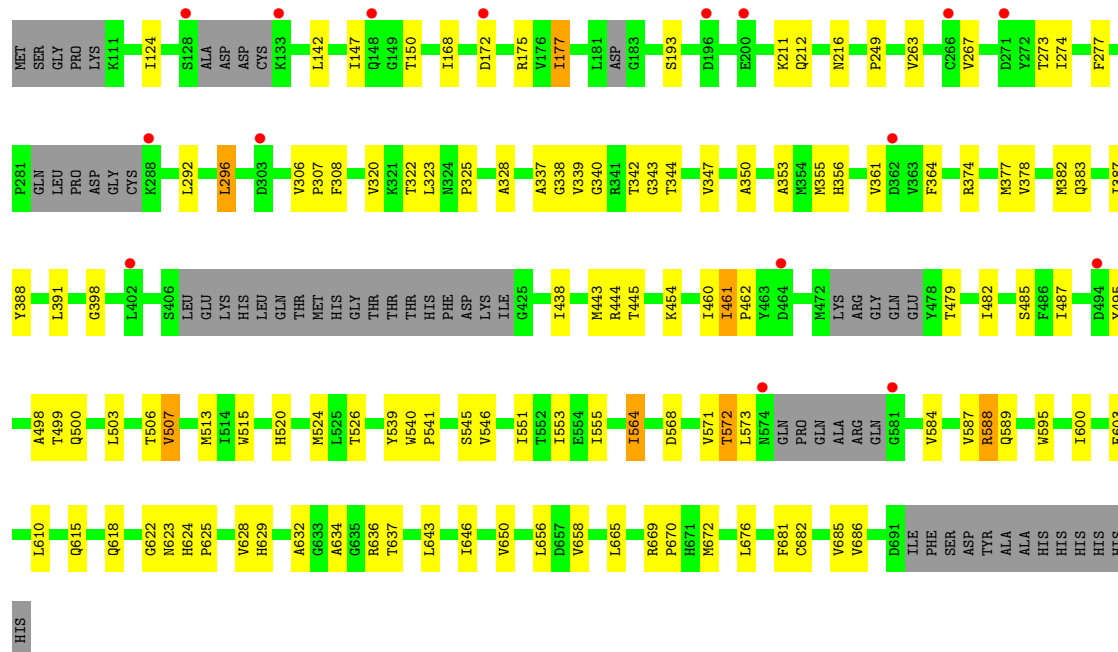


#### • Molecule 1: RECEPTOR-TYPE TYROSINE-PROTEIN PHOSPHATASE EPSILON

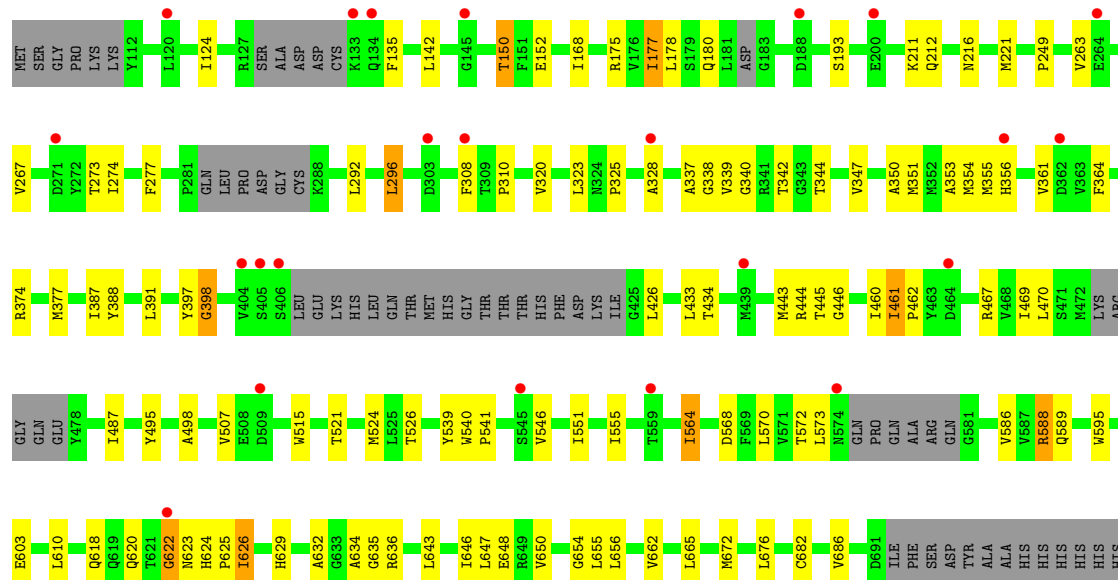




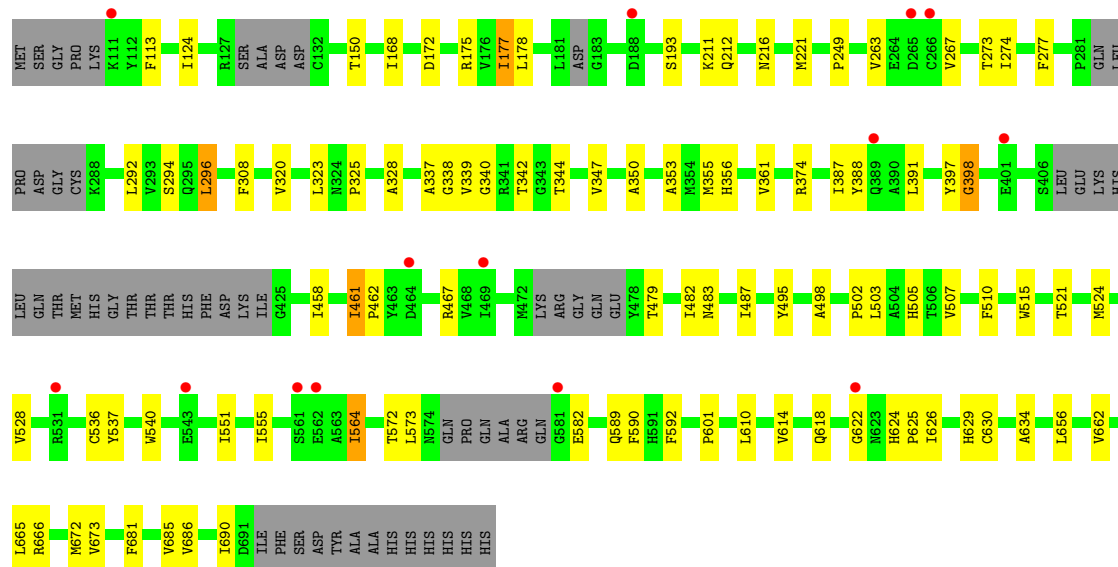
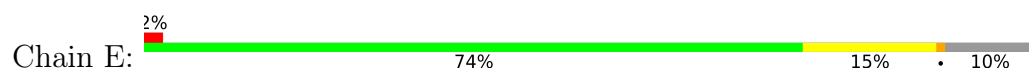
● Molecule 1: RECEPTOR-TYPE TYROSINE-PROTEIN PHOSPHATASE EPSILON



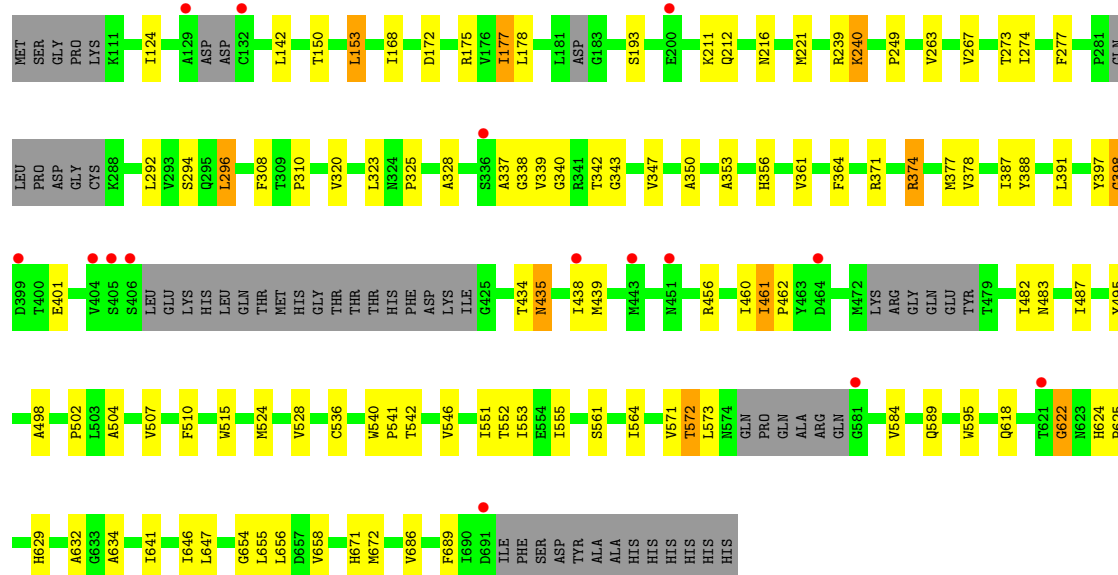
● Molecule 1: RECEPTOR-TYPE TYROSINE-PROTEIN PHOSPHATASE EPSILON



● Molecule 1: RECEPTOR-TYPE TYROSINE-PROTEIN PHOSPHATASE EPSILON



• Molecule 1: RECEPTOR-TYPE TYROSINE-PROTEIN PHOSPHATASE EPSILON





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	126.02Å 123.62Å 219.12Å 90.00° 91.13° 90.00°	Depositor
Resolution (Å)	218.22 – 3.20 218.22 – 3.20	Depositor EDS
% Data completeness (in resolution range)	98.7 (218.22-3.20) 98.7 (218.22-3.20)	Depositor EDS
$R_{merge}$	0.14	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.62 (at 3.18Å)	Xtriage
Refinement program	REFMAC 5.4.0066	Depositor
R, $R_{free}$	0.221 , 0.256 0.226 , 0.259	Depositor DCC
$R_{free}$ test set	5498 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	76.1	Xtriage
Anisotropy	0.555	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 94.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	0.026 for k,h,-l 0.027 for -k,-h,-l 0.038 for h,-k,-l	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	23750	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	62.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.98% 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:  
CL

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	0/4072	0.59	1/5570 (0.0%)
1	B	0.48	0/4074	0.58	0/5575
1	C	0.47	0/4071	0.57	0/5570
1	D	0.44	0/4036	0.55	0/5522
1	E	0.49	0/4051	0.58	0/5544
1	F	0.47	0/4037	0.56	0/5525
All	All	0.47	0/24341	0.57	1/33306 (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	C	0	1
1	F	0	1
All	All	0	2

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	673	VAL	N-CA-C	-5.12	97.19	111.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	622	GLY	Peptide

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Group
1	F	622	GLY	Peptide

## 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	3973	0	3575	91	0
1	B	3972	0	3557	73	0
1	C	3970	0	3551	94	0
1	D	3936	0	3491	88	0
1	E	3950	0	3506	68	0
1	F	3937	0	3491	74	0
2	A	2	0	0	0	0
2	B	3	0	0	0	0
2	C	1	0	0	0	0
2	D	2	0	0	0	0
2	E	2	0	0	0	0
2	F	2	0	0	0	0
All	All	23750	0	21171	484	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 11.

The worst 5 of 484 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:656:LEU:HD11	1:E:686:VAL:HG11	1.27	1.17
1:C:656:LEU:CD1	1:C:686:VAL:HG11	1.86	1.05
1:D:656:LEU:HD11	1:D:686:VAL:HG11	1.41	1.02
1:F:656:LEU:HD11	1:F:686:VAL:HG11	1.46	0.97
1:C:564:ILE:HG21	1:C:610:LEU:HD13	1.44	0.96

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	523/599 (87%)	479 (92%)	40 (8%)	4 (1%)	16	51
1	B	527/599 (88%)	486 (92%)	37 (7%)	4 (1%)	16	51
1	C	527/599 (88%)	493 (94%)	31 (6%)	3 (1%)	22	57
1	D	525/599 (88%)	484 (92%)	37 (7%)	4 (1%)	16	51
1	E	527/599 (88%)	492 (93%)	32 (6%)	3 (1%)	22	57
1	F	528/599 (88%)	498 (94%)	23 (4%)	7 (1%)	10	41
All	All	3157/3594 (88%)	2932 (93%)	200 (6%)	25 (1%)	16	51

5 of 25 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	623	ASN
1	A	398	GLY
1	B	398	GLY
1	C	398	GLY
1	D	398	GLY

### 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	358/535 (67%)	339 (95%)	19 (5%)	19	52
1	B	352/535 (66%)	331 (94%)	21 (6%)	16	48
1	C	348/535 (65%)	333 (96%)	15 (4%)	25	57

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	D	340/535 (64%)	322 (95%)	18 (5%)	19	52
1	E	341/535 (64%)	327 (96%)	14 (4%)	26	59
1	F	336/535 (63%)	320 (95%)	16 (5%)	21	55
All	All	2075/3210 (65%)	1972 (95%)	103 (5%)	20	54

5 of 103 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	177	ILE
1	D	623	ASN
1	F	542	THR
1	D	193	SER
1	D	467	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 27 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	216	ASN
1	D	589	GLN
1	F	435	ASN
1	D	538	GLN
1	D	623	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 oligosaccharides in this entry.

## 5.6 Ligand geometry

Of 12 ligands modelled in this entry, 12 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

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 [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	537/599 (89%)	0.13	10 (1%) 66 50	57, 63, 66, 69	0
1	B	541/599 (90%)	0.18	14 (2%) 57 41	57, 63, 66, 69	0
1	C	541/599 (90%)	0.25	16 (2%) 52 37	57, 63, 66, 69	0
1	D	539/599 (89%)	0.35	23 (4%) 40 27	57, 63, 66, 69	0
1	E	541/599 (90%)	0.22	14 (2%) 57 41	57, 63, 66, 69	0
1	F	542/599 (90%)	0.28	15 (2%) 55 39	57, 63, 66, 69	0
All	All	3241/3594 (90%)	0.24	92 (2%) 55 39	57, 63, 66, 69	0

The worst 5 of 92 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	200	GLU	4.6
1	D	133	LYS	4.5
1	F	581	GLY	4.4
1	A	581	GLY	4.3
1	D	362	ASP	3.9

### 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( $\text{\AA}^2$ )	Q<0.9
2	CL	B	1694	1/1	0.95	0.10	82,82,82,82	0
2	CL	F	1692	1/1	0.96	0.09	68,68,68,68	0
2	CL	A	1692	1/1	0.97	0.09	65,65,65,65	0
2	CL	D	1692	1/1	0.97	0.10	78,78,78,78	0
2	CL	E	1692	1/1	0.97	0.12	66,66,66,66	0
2	CL	B	1693	1/1	0.97	0.10	74,74,74,74	0
2	CL	D	1693	1/1	0.98	0.06	40,40,40,40	0
2	CL	C	1692	1/1	0.98	0.05	43,43,43,43	0
2	CL	B	1692	1/1	0.98	0.05	41,41,41,41	0
2	CL	E	1693	1/1	0.99	0.05	40,40,40,40	0
2	CL	A	1693	1/1	0.99	0.05	51,51,51,51	0
2	CL	F	1693	1/1	0.99	0.04	43,43,43,43	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.