



wwPDB X-ray Structure Validation Summary Report

Jun 12, 2024 – 05:21 PM EDT

PDB ID : 3CLW
Title : Crystal structure of conserved exported protein from *Bacteroides fragilis*
Authors : Bonanno, J.B.; Rutter, M.; Bain, K.T.; Chang, S.; Ozyurt, K.; Smith, D.; Wasserman, S.; Sauder, J.M.; Burley, S.K.; Almo, S.C.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on : 2008-03-20
Resolution : 2.20 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

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) : 1.20.1
EDS : 2.36.2
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) : 2.36.2

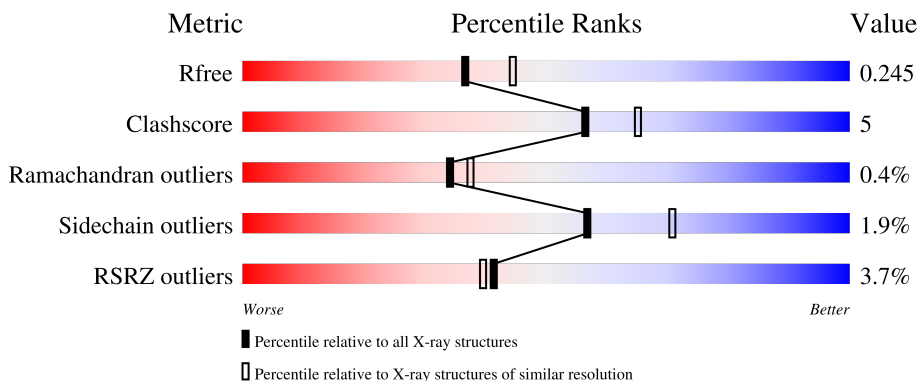
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.20 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}	130704	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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 ≥ 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	507	 88% 8% ..
1	B	507	 2% 82% 14% ..
1	C	507	 11% 80% 16% .
1	D	507	 6% 84% 12% ..
1	E	507	 87% 9% ..

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	507	 <p>A horizontal bar chart representing the quality of chain. The bar is divided into four segments: a small red segment at the beginning labeled '2%', a large green segment labeled '85%', a yellow segment labeled '11%', and a small grey segment at the end. There are two small black dots to the right of the grey segment.</p>

2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 24381 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 Conserved exported protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	492	3918	2496	662	742	18	0	1	0
1	B	492	3895	2483	656	738	18	0	0	0
1	C	491	3897	2483	657	739	18	0	1	0
1	D	491	3888	2478	656	736	18	0	0	0
1	E	491	3901	2488	656	739	18	0	2	0
1	F	492	3938	2509	666	745	18	0	1	0

There are 66 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	24	MET	-	expression tag	UNP Q5LF82
A	25	SER	-	expression tag	UNP Q5LF82
A	26	LEU	-	expression tag	UNP Q5LF82
A	523	GLU	-	expression tag	UNP Q5LF82
A	524	GLY	-	expression tag	UNP Q5LF82
A	525	HIS	-	expression tag	UNP Q5LF82
A	526	HIS	-	expression tag	UNP Q5LF82
A	527	HIS	-	expression tag	UNP Q5LF82
A	528	HIS	-	expression tag	UNP Q5LF82
A	529	HIS	-	expression tag	UNP Q5LF82
A	530	HIS	-	expression tag	UNP Q5LF82
B	24	MET	-	expression tag	UNP Q5LF82
B	25	SER	-	expression tag	UNP Q5LF82
B	26	LEU	-	expression tag	UNP Q5LF82
B	523	GLU	-	expression tag	UNP Q5LF82
B	524	GLY	-	expression tag	UNP Q5LF82
B	525	HIS	-	expression tag	UNP Q5LF82

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	526	HIS	-	expression tag	UNP Q5LF82
B	527	HIS	-	expression tag	UNP Q5LF82
B	528	HIS	-	expression tag	UNP Q5LF82
B	529	HIS	-	expression tag	UNP Q5LF82
B	530	HIS	-	expression tag	UNP Q5LF82
C	24	MET	-	expression tag	UNP Q5LF82
C	25	SER	-	expression tag	UNP Q5LF82
C	26	LEU	-	expression tag	UNP Q5LF82
C	523	GLU	-	expression tag	UNP Q5LF82
C	524	GLY	-	expression tag	UNP Q5LF82
C	525	HIS	-	expression tag	UNP Q5LF82
C	526	HIS	-	expression tag	UNP Q5LF82
C	527	HIS	-	expression tag	UNP Q5LF82
C	528	HIS	-	expression tag	UNP Q5LF82
C	529	HIS	-	expression tag	UNP Q5LF82
C	530	HIS	-	expression tag	UNP Q5LF82
D	24	MET	-	expression tag	UNP Q5LF82
D	25	SER	-	expression tag	UNP Q5LF82
D	26	LEU	-	expression tag	UNP Q5LF82
D	523	GLU	-	expression tag	UNP Q5LF82
D	524	GLY	-	expression tag	UNP Q5LF82
D	525	HIS	-	expression tag	UNP Q5LF82
D	526	HIS	-	expression tag	UNP Q5LF82
D	527	HIS	-	expression tag	UNP Q5LF82
D	528	HIS	-	expression tag	UNP Q5LF82
D	529	HIS	-	expression tag	UNP Q5LF82
D	530	HIS	-	expression tag	UNP Q5LF82
E	24	MET	-	expression tag	UNP Q5LF82
E	25	SER	-	expression tag	UNP Q5LF82
E	26	LEU	-	expression tag	UNP Q5LF82
E	523	GLU	-	expression tag	UNP Q5LF82
E	524	GLY	-	expression tag	UNP Q5LF82
E	525	HIS	-	expression tag	UNP Q5LF82
E	526	HIS	-	expression tag	UNP Q5LF82
E	527	HIS	-	expression tag	UNP Q5LF82
E	528	HIS	-	expression tag	UNP Q5LF82
E	529	HIS	-	expression tag	UNP Q5LF82
E	530	HIS	-	expression tag	UNP Q5LF82
F	24	MET	-	expression tag	UNP Q5LF82
F	25	SER	-	expression tag	UNP Q5LF82
F	26	LEU	-	expression tag	UNP Q5LF82
F	523	GLU	-	expression tag	UNP Q5LF82

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	524	GLY	-	expression tag	UNP Q5LF82
F	525	HIS	-	expression tag	UNP Q5LF82
F	526	HIS	-	expression tag	UNP Q5LF82
F	527	HIS	-	expression tag	UNP Q5LF82
F	528	HIS	-	expression tag	UNP Q5LF82
F	529	HIS	-	expression tag	UNP Q5LF82
F	530	HIS	-	expression tag	UNP Q5LF82

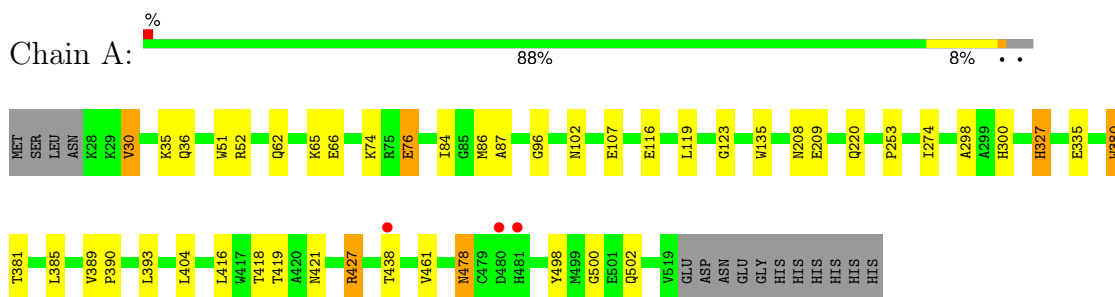
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	220	Total 220	O 220	0	0
2	B	132	Total 132	O 132	0	0
2	C	56	Total 56	O 56	0	0
2	D	130	Total 130	O 130	0	0
2	E	219	Total 219	O 219	0	0
2	F	187	Total 187	O 187	0	0

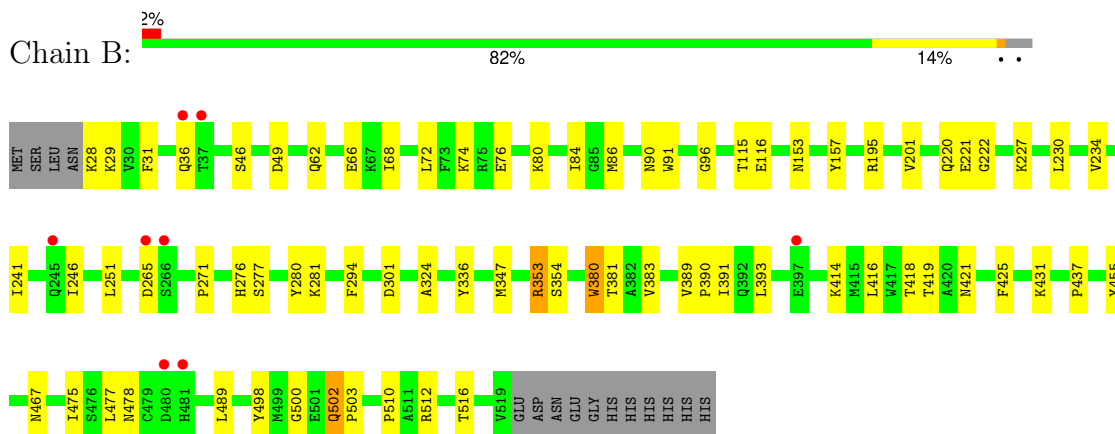
3 Residue-property plots [i](#)

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.

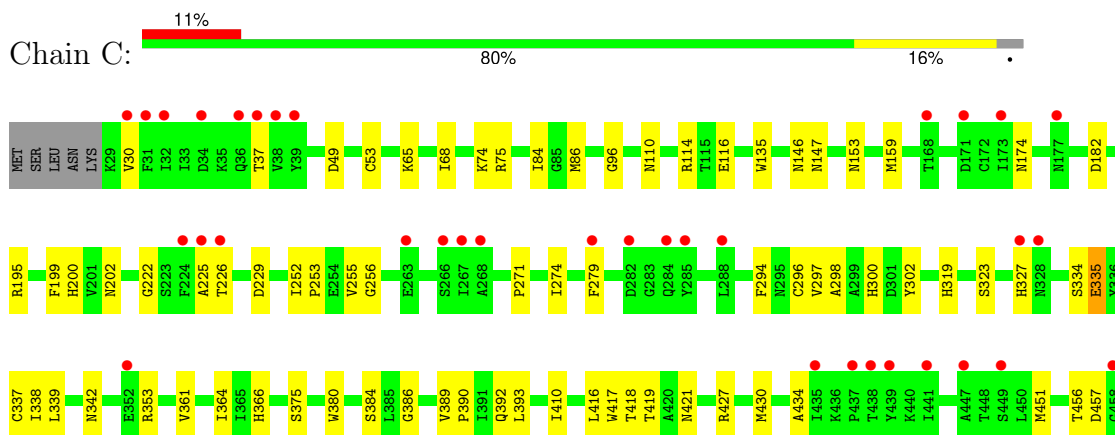
- Molecule 1: Conserved exported protein

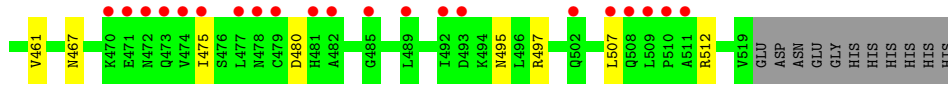


- Molecule 1: Conserved exported protein

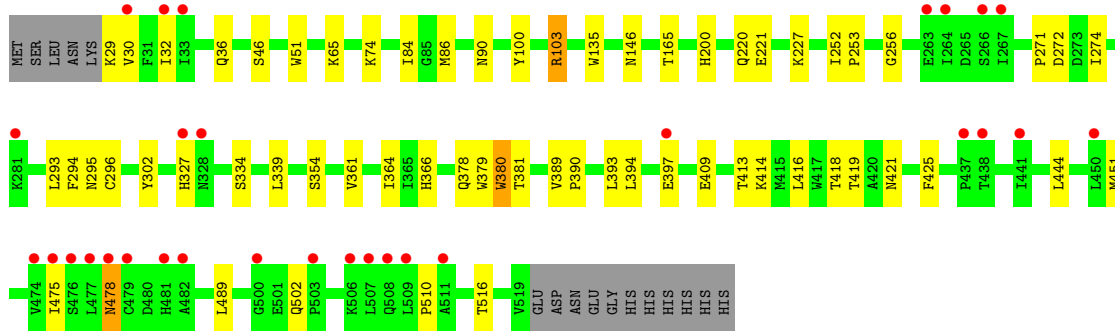
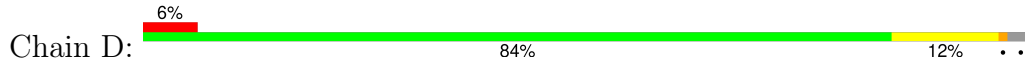


- Molecule 1: Conserved exported protein

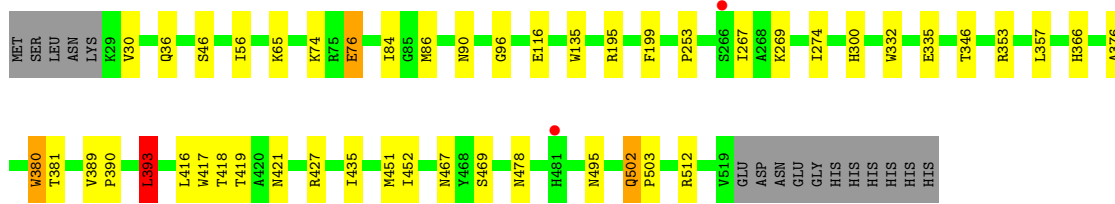
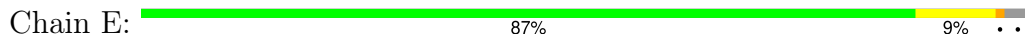




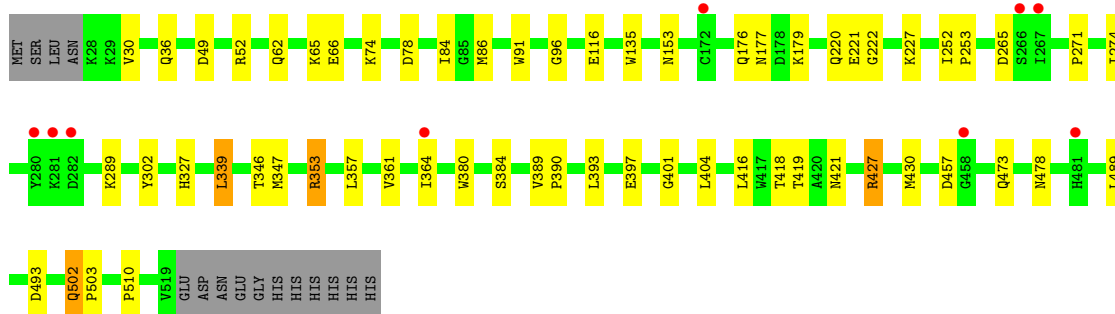
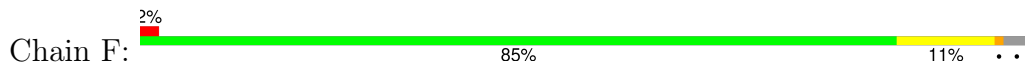
● Molecule 1: Conserved exported protein



● Molecule 1: Conserved exported protein



● Molecule 1: Conserved exported protein



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	87.53Å 116.26Å 393.35Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.20 48.42 – 2.20	Depositor EDS
% Data completeness (in resolution range)	84.5 (20.00-2.20) 84.4 (48.42-2.20)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	0.10	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.53 (at 2.20Å)	Xtrriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.199 , 0.243 0.202 , 0.245	Depositor DCC
R_{free} test set	8682 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å ²)	31.4	Xtrriage
Anisotropy	0.146	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 43.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	24381	wwPDB-VP
Average B, all atoms (Å ²)	46.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.38% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

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	0.70	0/4015	0.71	3/5447 (0.1%)
1	B	0.70	0/3988	0.69	0/5413
1	C	0.67	0/3994	0.69	0/5423
1	D	0.66	0/3981	0.69	0/5404
1	E	0.74	0/4001	0.72	2/5431 (0.0%)
1	F	0.70	0/4034	0.72	5/5467 (0.1%)
All	All	0.69	0/24013	0.70	10/32585 (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	E	0	1
1	F	0	1
All	All	0	2

There are no bond length outliers.

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	F	427	ARG	NE-CZ-NH2	-7.08	116.76	120.30
1	F	353	ARG	NE-CZ-NH1	-7.06	116.77	120.30
1	A	427	ARG	NE-CZ-NH2	-6.45	117.08	120.30
1	F	427	ARG	NE-CZ-NH1	6.14	123.37	120.30
1	E	427	ARG	NE-CZ-NH2	-5.66	117.47	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	E	346	THR	Peptide
1	F	346	THR	Peptide

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	3918	0	3767	30	0
1	B	3895	0	3734	41	1
1	C	3897	0	3725	48	0
1	D	3888	0	3726	44	0
1	E	3901	0	3738	30	0
1	F	3938	0	3815	39	0
2	A	220	0	0	1	0
2	B	132	0	0	1	1
2	C	56	0	0	1	0
2	D	130	0	0	1	0
2	E	219	0	0	0	0
2	F	187	0	0	0	0
All	All	24381	0	22505	229	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 5.

The worst 5 of 229 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:86:MET:CE	1:F:416:LEU:HA	1.92	0.98
1:C:74:LYS:HB3	1:C:84:ILE:HD11	1.43	0.97
1:C:86:MET:CE	1:C:416:LEU:HA	1.94	0.97
1:B:86:MET:HE3	1:B:390:PRO:HB3	1.49	0.94
1:F:86:MET:HE3	1:F:416:LEU:HA	1.53	0.88

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:324:ALA:O	2:B:589:HOH:O[4_575]	2.18	0.02

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	491/507 (97%)	473 (96%)	17 (4%)	1 (0%)	47	55
1	B	490/507 (97%)	465 (95%)	24 (5%)	1 (0%)	47	55
1	C	490/507 (97%)	457 (93%)	28 (6%)	5 (1%)	15	14
1	D	489/507 (96%)	457 (94%)	29 (6%)	3 (1%)	25	26
1	E	491/507 (97%)	469 (96%)	21 (4%)	1 (0%)	47	55
1	F	491/507 (97%)	464 (94%)	26 (5%)	1 (0%)	47	55
All	All	2942/3042 (97%)	2785 (95%)	145 (5%)	12 (0%)	34	37

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	380	TRP
1	A	380	TRP
1	C	174	ASN
1	C	480	ASP
1	B	380	TRP

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	416/443 (94%)	409 (98%)	7 (2%)	60	74
1	B	411/443 (93%)	402 (98%)	9 (2%)	52	65
1	C	411/443 (93%)	403 (98%)	8 (2%)	57	71
1	D	410/443 (93%)	404 (98%)	6 (2%)	65	78
1	E	412/443 (93%)	406 (98%)	6 (2%)	65	78
1	F	421/443 (95%)	410 (97%)	11 (3%)	46	58
All	All	2481/2658 (93%)	2434 (98%)	47 (2%)	57	71

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

Mol	Chain	Res	Type
1	D	478	ASN
1	E	502	GLN
1	D	502	GLN
1	E	76	GLU
1	F	36	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 30 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	366	HIS
1	F	421	ASN
1	D	366	HIS
1	F	508	GLN
1	F	90	ASN

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 monosaccharides 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	492/507 (97%)	-0.03	3 (0%) 89 88	28, 39, 56, 68	0
1	B	492/507 (97%)	0.07	8 (1%) 72 70	33, 48, 64, 71	0
1	C	491/507 (96%)	0.56	56 (11%) 5 4	35, 55, 75, 85	0
1	D	491/507 (96%)	0.29	30 (6%) 21 20	33, 48, 71, 76	0
1	E	491/507 (96%)	-0.03	2 (0%) 92 91	28, 40, 57, 69	0
1	F	492/507 (97%)	0.16	9 (1%) 68 66	31, 42, 56, 69	0
All	All	2949/3042 (96%)	0.17	108 (3%) 41 39	28, 45, 66, 85	0

The worst 5 of 108 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	511	ALA	6.4
1	C	474	VAL	6.1
1	D	30	VAL	5.9
1	D	511	ALA	5.8
1	C	282	ASP	5.7

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.