wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 10:20 PM GMT

PDB ID	:	1T72
Title	:	Crystal structure of phosphate transport system protein phoU from Aquifex aeolicus
Authors	:	Oganesyan, V.; Kim, S.-H.; Oganesyan, N.; Jancarik, J.; Adams, P.D.; Kim, R.; Berkeley Structural Genomics Center (BSGC)
Deposited on	:	2004-05-07
Resolution	:	2.90 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

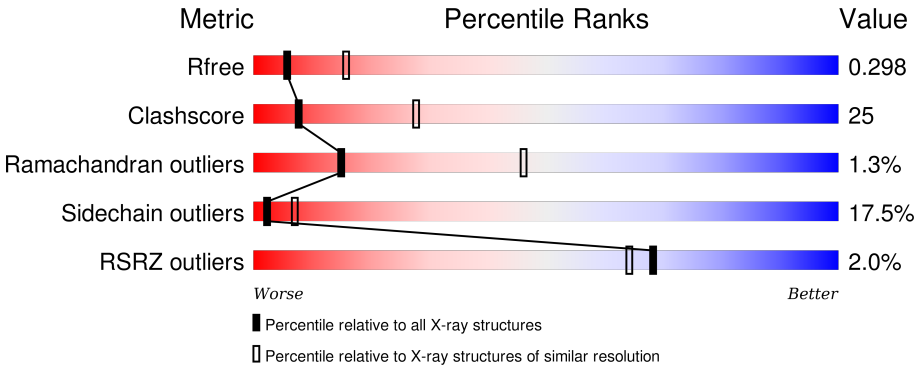
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)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

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.90 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	1451 (2.90-2.90)
Clashscore	102246	1668 (2.90-2.90)
Ramachandran outliers	100387	1630 (2.90-2.90)
Sidechain outliers	100360	1632 (2.90-2.90)
RSRZ outliers	91569	1456 (2.90-2.90)

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 <=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	227	3% 47% 37% 11% 5%
1	B	227	4% 51% 33% 11% 6%
1	D	227	2% 51% 37% 7% 5%
1	E	227	% 47% 37% 10% 6%
1	F	227	43% 46% 5% 6%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	227	

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 10391 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 Phosphate transport system protein phoU homolog.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	215	Total	C	N	O	S	Se	0	0	0
			1712	1076	286	339	1	10			
1	B	214	Total	C	N	O	S	Se	0	0	0
			1708	1074	285	338	1	10			
1	D	215	Total	C	N	O	S	Se	0	0	0
			1712	1076	286	339	1	10			
1	E	213	Total	C	N	O	S	Se	0	0	0
			1699	1069	284	335	1	10			
1	F	214	Total	C	N	O	S	Se	0	0	0
			1708	1074	285	338	1	10			
1	G	211	Total	C	N	O	S	Se	0	0	0
			1691	1065	282	333	1	10			

There are 96 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	CLONING ARTIFACT	UNP O67053
A	2	GLY	-	CLONING ARTIFACT	UNP O67053
A	3	GLY	-	CLONING ARTIFACT	UNP O67053
A	4	GLY	-	CLONING ARTIFACT	UNP O67053
A	5	GLY	-	CLONING ARTIFACT	UNP O67053
A	6	GLY	-	CLONING ARTIFACT	UNP O67053
A	7	MSE	MET	MODIFIED RESIDUE	UNP O67053
A	23	MSE	MET	MODIFIED RESIDUE	UNP O67053
A	70	MSE	MET	MODIFIED RESIDUE	UNP O67053
A	83	MSE	MET	MODIFIED RESIDUE	UNP O67053
A	85	MSE	MET	MODIFIED RESIDUE	UNP O67053
A	97	MSE	MET	MODIFIED RESIDUE	UNP O67053
A	126	MSE	MET	MODIFIED RESIDUE	UNP O67053
A	133	MSE	MET	MODIFIED RESIDUE	UNP O67053
A	171	MSE	MET	MODIFIED RESIDUE	UNP O67053
A	185	MSE	MET	MODIFIED RESIDUE	UNP O67053
B	1	GLY	-	CLONING ARTIFACT	UNP O67053

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	2	GLY	-	CLONING ARTIFACT	UNP O67053
B	3	GLY	-	CLONING ARTIFACT	UNP O67053
B	4	GLY	-	CLONING ARTIFACT	UNP O67053
B	5	GLY	-	CLONING ARTIFACT	UNP O67053
B	6	GLY	-	CLONING ARTIFACT	UNP O67053
B	7	MSE	MET	MODIFIED RESIDUE	UNP O67053
B	23	MSE	MET	MODIFIED RESIDUE	UNP O67053
B	70	MSE	MET	MODIFIED RESIDUE	UNP O67053
B	83	MSE	MET	MODIFIED RESIDUE	UNP O67053
B	85	MSE	MET	MODIFIED RESIDUE	UNP O67053
B	97	MSE	MET	MODIFIED RESIDUE	UNP O67053
B	126	MSE	MET	MODIFIED RESIDUE	UNP O67053
B	133	MSE	MET	MODIFIED RESIDUE	UNP O67053
B	171	MSE	MET	MODIFIED RESIDUE	UNP O67053
B	185	MSE	MET	MODIFIED RESIDUE	UNP O67053
D	1	GLY	-	CLONING ARTIFACT	UNP O67053
D	2	GLY	-	CLONING ARTIFACT	UNP O67053
D	3	GLY	-	CLONING ARTIFACT	UNP O67053
D	4	GLY	-	CLONING ARTIFACT	UNP O67053
D	5	GLY	-	CLONING ARTIFACT	UNP O67053
D	6	GLY	-	CLONING ARTIFACT	UNP O67053
D	7	MSE	MET	MODIFIED RESIDUE	UNP O67053
D	23	MSE	MET	MODIFIED RESIDUE	UNP O67053
D	70	MSE	MET	MODIFIED RESIDUE	UNP O67053
D	83	MSE	MET	MODIFIED RESIDUE	UNP O67053
D	85	MSE	MET	MODIFIED RESIDUE	UNP O67053
D	97	MSE	MET	MODIFIED RESIDUE	UNP O67053
D	126	MSE	MET	MODIFIED RESIDUE	UNP O67053
D	133	MSE	MET	MODIFIED RESIDUE	UNP O67053
D	171	MSE	MET	MODIFIED RESIDUE	UNP O67053
D	185	MSE	MET	MODIFIED RESIDUE	UNP O67053
E	1	GLY	-	CLONING ARTIFACT	UNP O67053
E	2	GLY	-	CLONING ARTIFACT	UNP O67053
E	3	GLY	-	CLONING ARTIFACT	UNP O67053
E	4	GLY	-	CLONING ARTIFACT	UNP O67053
E	5	GLY	-	CLONING ARTIFACT	UNP O67053
E	6	GLY	-	CLONING ARTIFACT	UNP O67053
E	7	MSE	MET	MODIFIED RESIDUE	UNP O67053
E	23	MSE	MET	MODIFIED RESIDUE	UNP O67053
E	70	MSE	MET	MODIFIED RESIDUE	UNP O67053
E	83	MSE	MET	MODIFIED RESIDUE	UNP O67053
E	85	MSE	MET	MODIFIED RESIDUE	UNP O67053

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	97	MSE	MET	MODIFIED RESIDUE	UNP O67053
E	126	MSE	MET	MODIFIED RESIDUE	UNP O67053
E	133	MSE	MET	MODIFIED RESIDUE	UNP O67053
E	171	MSE	MET	MODIFIED RESIDUE	UNP O67053
E	185	MSE	MET	MODIFIED RESIDUE	UNP O67053
F	1	GLY	-	CLONING ARTIFACT	UNP O67053
F	2	GLY	-	CLONING ARTIFACT	UNP O67053
F	3	GLY	-	CLONING ARTIFACT	UNP O67053
F	4	GLY	-	CLONING ARTIFACT	UNP O67053
F	5	GLY	-	CLONING ARTIFACT	UNP O67053
F	6	GLY	-	CLONING ARTIFACT	UNP O67053
F	7	MSE	MET	MODIFIED RESIDUE	UNP O67053
F	23	MSE	MET	MODIFIED RESIDUE	UNP O67053
F	70	MSE	MET	MODIFIED RESIDUE	UNP O67053
F	83	MSE	MET	MODIFIED RESIDUE	UNP O67053
F	85	MSE	MET	MODIFIED RESIDUE	UNP O67053
F	97	MSE	MET	MODIFIED RESIDUE	UNP O67053
F	126	MSE	MET	MODIFIED RESIDUE	UNP O67053
F	133	MSE	MET	MODIFIED RESIDUE	UNP O67053
F	171	MSE	MET	MODIFIED RESIDUE	UNP O67053
F	185	MSE	MET	MODIFIED RESIDUE	UNP O67053
G	1	GLY	-	CLONING ARTIFACT	UNP O67053
G	2	GLY	-	CLONING ARTIFACT	UNP O67053
G	3	GLY	-	CLONING ARTIFACT	UNP O67053
G	4	GLY	-	CLONING ARTIFACT	UNP O67053
G	5	GLY	-	CLONING ARTIFACT	UNP O67053
G	6	GLY	-	CLONING ARTIFACT	UNP O67053
G	7	MSE	MET	MODIFIED RESIDUE	UNP O67053
G	23	MSE	MET	MODIFIED RESIDUE	UNP O67053
G	70	MSE	MET	MODIFIED RESIDUE	UNP O67053
G	83	MSE	MET	MODIFIED RESIDUE	UNP O67053
G	85	MSE	MET	MODIFIED RESIDUE	UNP O67053
G	97	MSE	MET	MODIFIED RESIDUE	UNP O67053
G	126	MSE	MET	MODIFIED RESIDUE	UNP O67053
G	133	MSE	MET	MODIFIED RESIDUE	UNP O67053
G	171	MSE	MET	MODIFIED RESIDUE	UNP O67053
G	185	MSE	MET	MODIFIED RESIDUE	UNP O67053

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	33	Total O 33 33	0	0

Continued on next page...

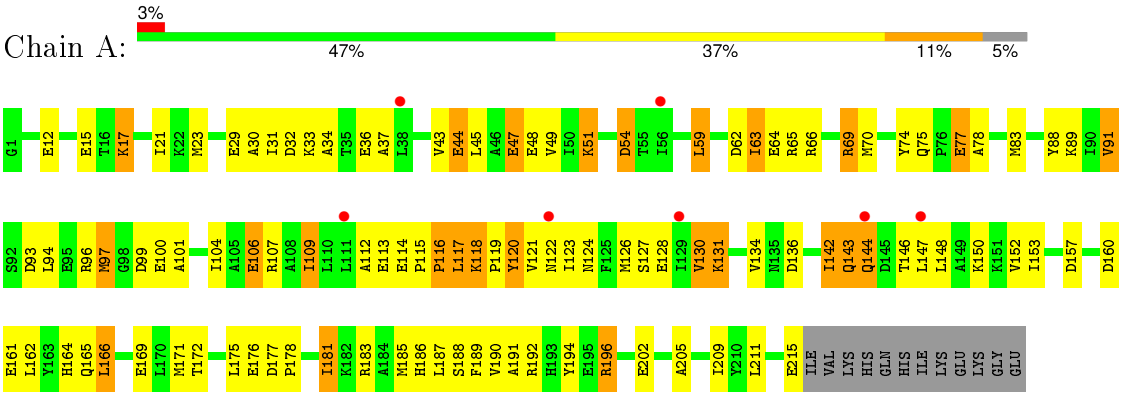
Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	26	Total O 26 26	0	0
2	D	30	Total O 30 30	0	0
2	E	18	Total O 18 18	0	0
2	F	27	Total O 27 27	0	0
2	G	27	Total O 27 27	0	0

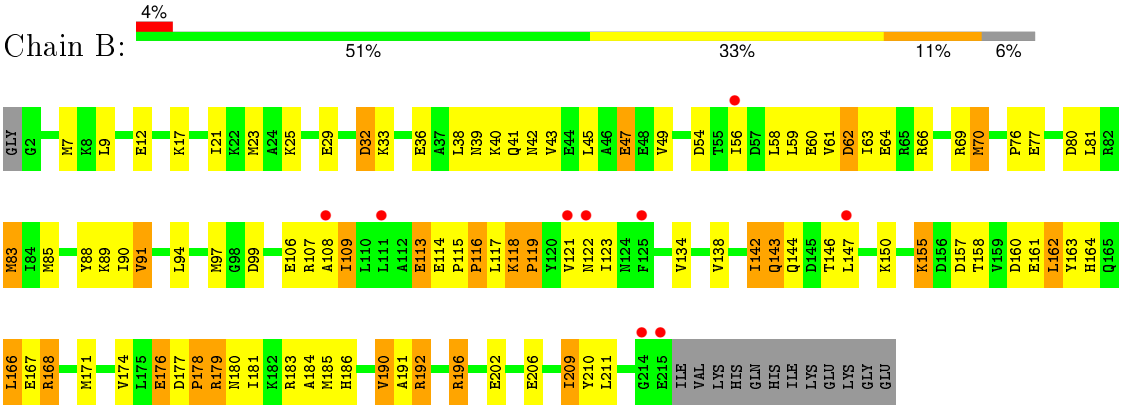
3 Residue-property plots (i)

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.

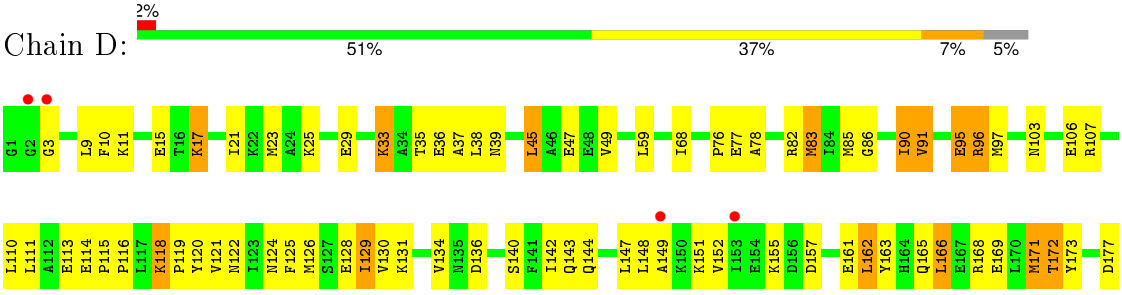
• Molecule 1: Phosphate transport system protein phoU homolog



• Molecule 1: Phosphate transport system protein phoU homolog

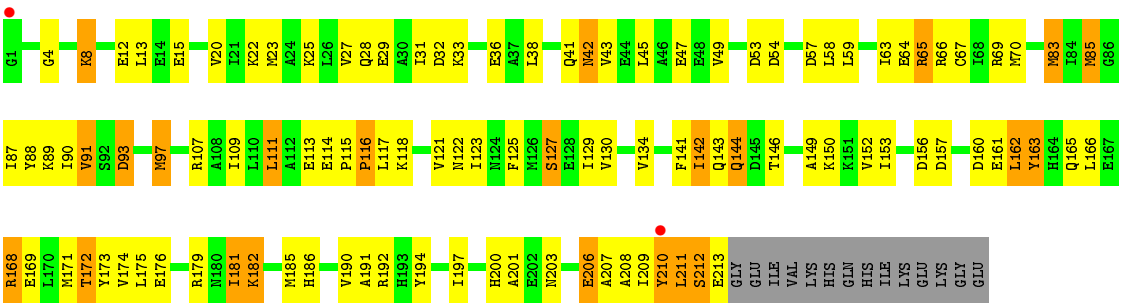


• Molecule 1: Phosphate transport system protein phoU homolog



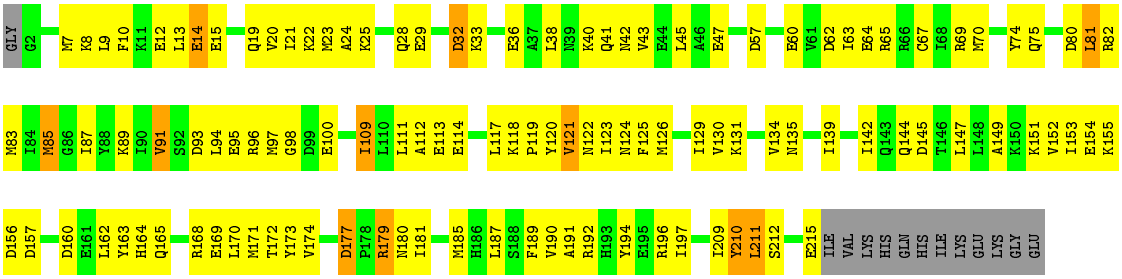
• Molecule 1: Phosphate transport system protein phoU homolog

Chain E:



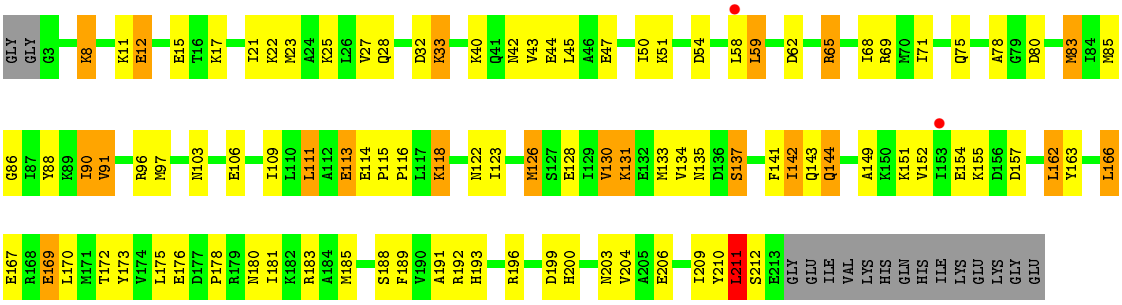
• Molecule 1: Phosphate transport system protein phoU homolog

Chain F:



• Molecule 1: Phosphate transport system protein phoU homolog

Chain G:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43	Depositor
Cell constants a, b, c, α, β, γ	113.50Å 113.50Å 155.00Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	15.00 – 2.90 27.91 – 2.80	Depositor EDS
% Data completeness (in resolution range)	84.5 (15.00-2.90) 84.5 (27.91-2.80)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	0.09	Depositor
$< I/\sigma(I) >$ [1]	2.74 (at 2.80Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, R_{free}	0.201 , 0.293 0.238 , 0.298	Depositor DCC
R_{free} test set	1930 reflections (5.42%)	DCC
Wilson B-factor (Å^2)	52.1	Xtriage
Anisotropy	0.132	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 1.0	EDS
Estimated twinning fraction	0.488 for h,-k,-l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Outliers	2 of 40646 reflections (0.005%)	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	10391	wwPDB-VP
Average B, all atoms (Å^2)	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 66.64 % of the origin peak, indicating pseudo translational symmetry.The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 5.8534e-06.The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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	0.65	1/1721 (0.1%)	0.98	11/2298 (0.5%)
1	B	0.63	0/1717	1.07	9/2293 (0.4%)
1	D	0.63	0/1721	0.92	6/2298 (0.3%)
1	E	0.67	1/1708 (0.1%)	1.09	11/2281 (0.5%)
1	F	0.67	0/1717	0.90	7/2293 (0.3%)
1	G	0.63	0/1700	0.96	7/2271 (0.3%)
All	All	0.65	2/10284 (0.0%)	0.99	51/13734 (0.4%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	97	MSE	SE-CE	-5.21	1.64	1.95
1	E	97	MSE	SE-CE	-5.17	1.65	1.95

The worst 5 of 51 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	115	PRO	CA-N-CD	-22.31	80.26	111.50
1	B	119	PRO	CA-N-CD	-19.37	84.38	111.50
1	B	116	PRO	CA-N-CD	-17.55	86.93	111.50
1	A	116	PRO	CA-N-CD	-17.50	87.00	111.50
1	D	76	PRO	CA-N-CD	-17.13	87.52	111.50

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	1712	0	1727	96	0
1	B	1708	0	1721	108	1
1	D	1712	0	1727	79	1
1	E	1699	0	1718	117	0
1	F	1708	0	1721	106	0
1	G	1691	0	1709	80	0
2	A	33	0	0	2	0
2	B	26	0	0	3	0
2	D	30	0	0	3	0
2	E	18	0	0	0	0
2	F	27	0	0	0	0
2	G	27	0	0	2	0
All	All	10391	0	10323	517	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 25.

The worst 5 of 517 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:85:MSE:SE	1:G:85:MSE:HE1	1.82	1.29
1:G:97:MSE:HE3	1:G:134:VAL:HG22	1.25	1.16
1:B:97:MSE:CE	1:B:134:VAL:HG22	1.75	1.16
1:D:97:MSE:CE	1:D:134:VAL:HG22	1.78	1.13
1:F:13:LEU:HA	1:F:70:MSE:HE1	1.34	1.10

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:179:ARG:NH1	1:D:179:ARG:NH1[4_565]	2.18	0.02

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	213/227 (94%)	189 (89%)	21 (10%)	3 (1%)	14	44
1	B	212/227 (93%)	185 (87%)	24 (11%)	3 (1%)	14	44
1	D	213/227 (94%)	186 (87%)	24 (11%)	3 (1%)	14	44
1	E	211/227 (93%)	191 (90%)	18 (8%)	2 (1%)	21	57
1	F	212/227 (93%)	191 (90%)	19 (9%)	2 (1%)	21	57
1	G	209/227 (92%)	184 (88%)	21 (10%)	4 (2%)	10	35
All	All	1270/1362 (93%)	1126 (89%)	127 (10%)	17 (1%)	15	46

5 of 17 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	78	ALA
1	A	117	LEU
1	B	41	GLN
1	D	180	ASN
1	E	176	GLU

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	185/186 (100%)	150 (81%)	35 (19%)	2	6
1	B	185/186 (100%)	152 (82%)	33 (18%)	2	6
1	D	185/186 (100%)	156 (84%)	29 (16%)	3	9
1	E	184/186 (99%)	150 (82%)	34 (18%)	2	6
1	F	185/186 (100%)	155 (84%)	30 (16%)	3	9
1	G	184/186 (99%)	151 (82%)	33 (18%)	2	6
All	All	1108/1116 (99%)	914 (82%)	194 (18%)	2	7

5 of 194 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	143	GLN
1	E	85	MSE
1	G	123	ILE
1	D	165	GLN
1	E	12	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	75	GLN
1	E	41	GLN
1	F	164	HIS
1	B	143	GLN
1	B	186	HIS

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 carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

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

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	205/227 (90%)	0.33	7 (3%) 49 41	9, 24, 38, 53	0
1	B	204/227 (89%)	0.30	9 (4%) 38 32	11, 24, 38, 61	0
1	D	205/227 (90%)	0.27	4 (1%) 68 64	6, 24, 41, 63	0
1	E	203/227 (89%)	0.26	2 (0%) 84 82	10, 23, 36, 56	0
1	F	204/227 (89%)	0.29	0 100 100	10, 23, 37, 55	0
1	G	201/227 (88%)	0.23	2 (0%) 84 82	5, 25, 35, 49	0
All	All	1222/1362 (89%)	0.28	24 (1%) 68 64	5, 24, 38, 63	0

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	122	ASN	4.3
1	B	121	VAL	3.3
1	B	111	LEU	3.1
1	A	147	LEU	3.0
1	G	58	LEU	3.0

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 carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.