wwPDB X-ray Structure Validation Summary Report (i)

Jun 15, 2024 – 10:15 AM EDT

PDB ID	:	1YOX
Title	:	Structure of the conserved Protein of Unknown Function PA3696 from Pseudomonas aeruginosa
Authors	:	Walker, J.R.; Xu, X.; Gu, J.; Joachimiak, A.; Edwards, A.; Savchenko, A.; Midwest Center for Structural Genomics (MCSG)
Deposited on	:	2005-01-28
Resolution	:	2.30 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:

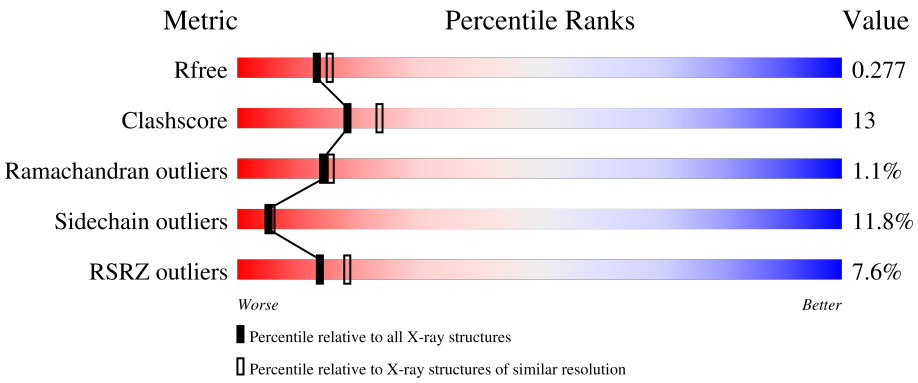
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.30 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	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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	250	14% 50% 22% • 24%
1	B	250	4% 64% 11% • 22%
1	C	250	3% 54% 20% • • 22%
1	D	250	3% 62% 18% • 16%
1	E	250	2% 62% 16% • 19%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	250	10% 47% 24% • 26%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 9193 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 hypothetical protein PA3696.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	190	1454	918	251	278	3	4	0	0	0
1	B	195	1474	933	253	281	3	4	0	0	0
1	C	196	1484	944	252	281	3	4	0	0	0
1	D	209	1576	1002	268	298	3	5	0	0	0
1	E	203	1541	980	261	292	3	5	0	0	0
1	F	186	1429	904	246	272	3	4	0	0	0

There are 54 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	CLONING ARTIFACT	UNP Q9HXU4
A	0	SER	-	CLONING ARTIFACT	UNP Q9HXU4
A	1	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
A	15	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
A	35	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
A	38	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
A	48	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
A	75	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
A	200	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
B	-1	GLY	-	CLONING ARTIFACT	UNP Q9HXU4
B	0	SER	-	CLONING ARTIFACT	UNP Q9HXU4
B	1	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
B	15	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
B	35	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
B	38	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
B	48	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
B	75	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	200	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
C	-1	GLY	-	CLONING ARTIFACT	UNP Q9HXU4
C	0	SER	-	CLONING ARTIFACT	UNP Q9HXU4
C	1	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
C	15	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
C	35	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
C	38	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
C	48	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
C	75	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
C	200	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
D	-1	GLY	-	CLONING ARTIFACT	UNP Q9HXU4
D	0	SER	-	CLONING ARTIFACT	UNP Q9HXU4
D	1	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
D	15	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
D	35	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
D	38	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
D	48	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
D	75	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
D	200	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
E	-1	GLY	-	CLONING ARTIFACT	UNP Q9HXU4
E	0	SER	-	CLONING ARTIFACT	UNP Q9HXU4
E	1	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
E	15	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
E	35	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
E	38	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
E	48	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
E	75	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
E	200	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
F	-1	GLY	-	CLONING ARTIFACT	UNP Q9HXU4
F	0	SER	-	CLONING ARTIFACT	UNP Q9HXU4
F	1	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
F	15	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
F	35	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
F	38	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
F	48	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
F	75	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4
F	200	MSE	MET	MODIFIED RESIDUE	UNP Q9HXU4

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	15	15	15	0	0

Continued on next page...

Continued from previous page...

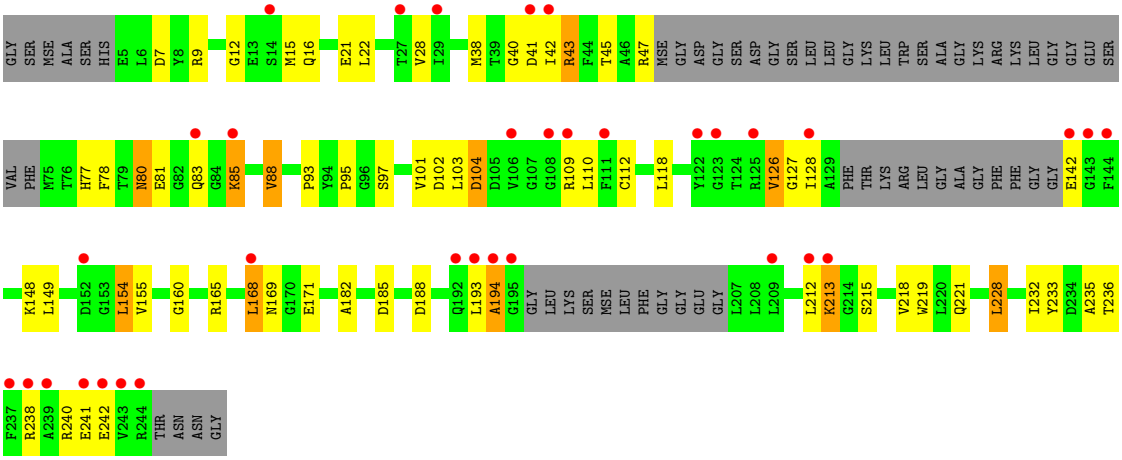
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	B	43	43	43	0	0
			Total	O		
2	C	35	35	35	0	0
			Total	O		
2	D	64	64	64	0	0
			Total	O		
2	E	52	52	52	0	0
			Total	O		
2	F	26	26	26	0	0

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.

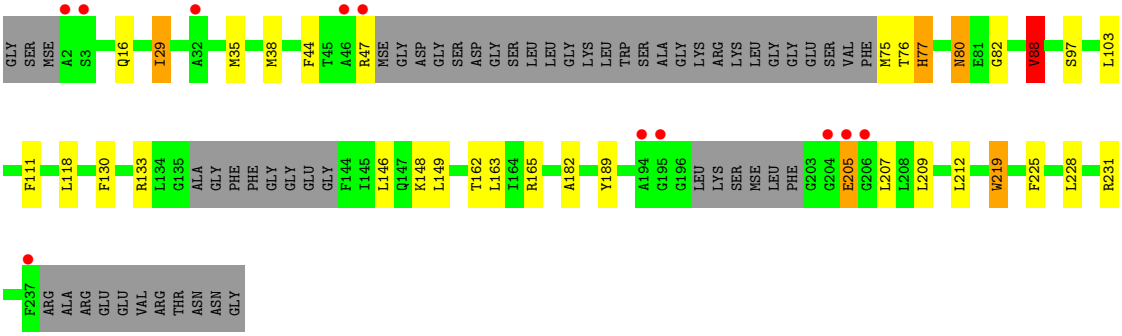
• Molecule 1: hypothetical protein PA3696

Chain A:



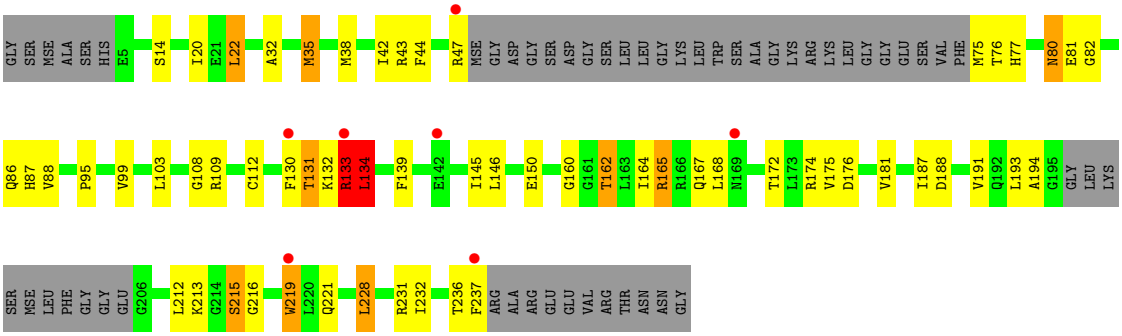
• Molecule 1: hypothetical protein PA3696

Chain B:



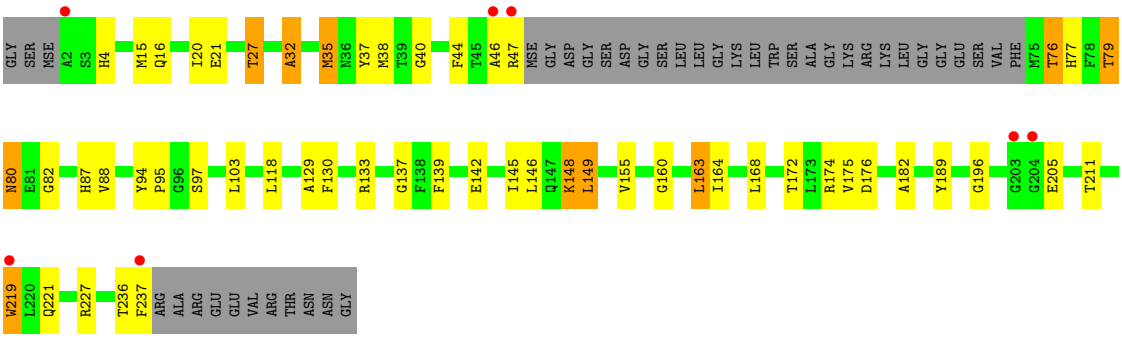
• Molecule 1: hypothetical protein PA3696

Chain C:



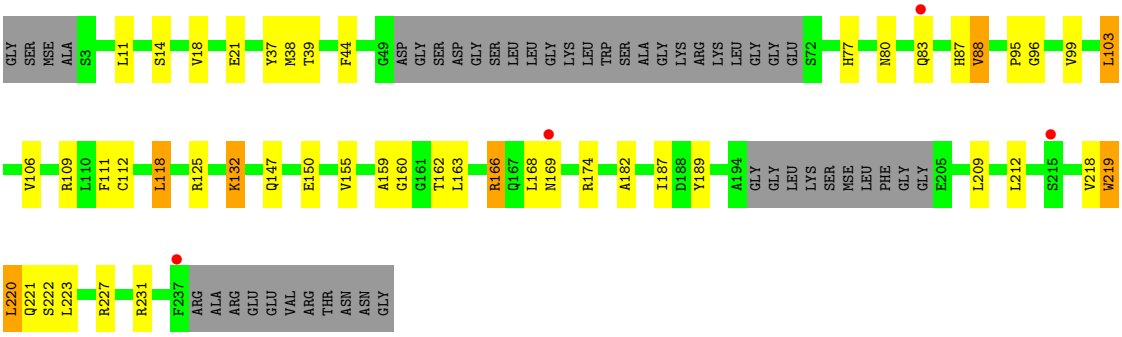
- Molecule 1: hypothetical protein PA3696

Chain D:



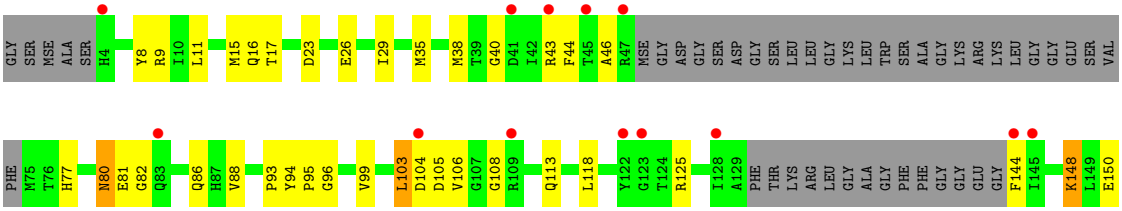
- Molecule 1: hypothetical protein PA3696

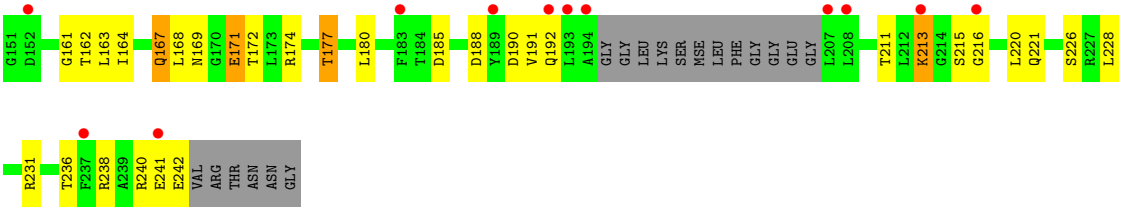
Chain E:



- Molecule 1: hypothetical protein PA3696

Chain F:




G151 D152 G161 T162 L163 I164 Q167 L168 N169 G170 E171 T172 L173 R174 T177 L180 F183 T184 D185 D188 Y189 D190 V191 Q192 L193 A194 GLY GLY LEU LYS SER MSE LEU PHE GLY GLY GLU GLY L207 L208 T211 L212 K213 G214 S215 G216 L220 Q221 S226 R227 L228
R231 T236 F237 R238 A239 R240 E241 E242 VAL ARG THR ASN ASN GLY

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	124.91Å 124.91Å 164.78Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.30 29.80 – 2.30	Depositor EDS
% Data completeness (in resolution range)	99.7 (30.00-2.30) 99.7 (29.80-2.30)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.99 (at 2.29Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.225 , 0.279 0.224 , 0.277	Depositor DCC
R_{free} test set	2952 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	44.2	Xtriage
Anisotropy	0.046	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 47.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	9193	wwPDB-VP
Average B, all atoms (Å^2)	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.48% 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 (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	0.59	0/1472	0.68	0/1981
1	B	0.70	0/1494	0.79	1/2010 (0.0%)
1	C	0.69	1/1506 (0.1%)	0.77	0/2027
1	D	0.80	0/1600	0.86	0/2151
1	E	0.78	1/1563 (0.1%)	0.83	0/2100
1	F	0.63	0/1448	0.69	0/1950
All	All	0.71	2/9083 (0.0%)	0.77	1/12219 (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

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	112	CYS	CB-SG	-7.27	1.69	1.82
1	C	112	CYS	CB-SG	-5.64	1.72	1.81

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	88	VAL	CB-CA-C	-5.05	101.81	111.40

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	215	SER	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	1454	0	1431	49	0
1	B	1474	0	1448	24	0
1	C	1484	0	1458	54	0
1	D	1576	0	1549	51	0
1	E	1541	0	1508	37	0
1	F	1429	0	1404	44	0
2	A	15	0	0	1	0
2	B	43	0	0	0	0
2	C	35	0	0	5	0
2	D	64	0	0	2	0
2	E	52	0	0	6	0
2	F	26	0	0	4	0
All	All	9193	0	8798	237	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 13.

The worst 5 of 237 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:A:15:MSE:HE3	1:A:95:PRO:HB3	1.32	1.08
1:A:101:VAL:HG23	1:A:155:VAL:HG12	1.46	0.96
1:C:131:THR:HG22	1:C:145:ILE:HG12	1.46	0.94
1:A:43:ARG:HH11	1:A:43:ARG:HG2	1.29	0.94
1:E:150:GLU:HG2	2:E:293:HOH:O	1.66	0.94

There are no symmetry-related clashes.

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	182/250 (73%)	167 (92%)	11 (6%)	4 (2%)	6	5
1	B	187/250 (75%)	179 (96%)	7 (4%)	1 (0%)	29	35
1	C	190/250 (76%)	174 (92%)	13 (7%)	3 (2%)	9	9
1	D	205/250 (82%)	197 (96%)	7 (3%)	1 (0%)	29	35
1	E	197/250 (79%)	186 (94%)	10 (5%)	1 (0%)	29	35
1	F	178/250 (71%)	165 (93%)	11 (6%)	2 (1%)	14	15
All	All	1139/1500 (76%)	1068 (94%)	59 (5%)	12 (1%)	14	15

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	194	ALA
1	F	105	ASP
1	A	40	GLY
1	A	126	VAL
1	C	133	ARG

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	150/183 (82%)	128 (85%)	22 (15%)	3	3
1	B	151/183 (82%)	134 (89%)	17 (11%)	6	6
1	C	151/183 (82%)	134 (89%)	17 (11%)	6	6
1	D	160/183 (87%)	146 (91%)	14 (9%)	10	12
1	E	158/183 (86%)	140 (89%)	18 (11%)	5	6
1	F	148/183 (81%)	128 (86%)	20 (14%)	4	4
All	All	918/1098 (84%)	810 (88%)	108 (12%)	5	5

5 of 108 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	35	MSE
1	E	14	SER
1	F	172	THR
1	D	79	THR
1	D	149	LEU

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	D	4	HIS
1	F	87	HIS
1	D	87	HIS
1	F	167	GLN
1	E	221	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 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	186/250 (74%)	1.09	34 (18%)	1	1	33, 61, 82, 90	0
1	B	191/250 (76%)	0.25	11 (5%)	23	29	28, 38, 54, 69	0
1	C	192/250 (76%)	0.42	7 (3%)	42	49	30, 45, 63, 69	0
1	D	204/250 (81%)	0.17	7 (3%)	45	52	25, 35, 51, 70	0
1	E	198/250 (79%)	0.26	4 (2%)	65	71	23, 37, 52, 65	0
1	F	182/250 (72%)	0.79	25 (13%)	3	4	31, 58, 76, 83	0
All	All	1153/1500 (76%)	0.49	88 (7%)	13	18	23, 43, 71, 90	0

The worst 5 of 88 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	142	GLU	8.3
1	A	194	ALA	7.9
1	D	47	ARG	7.9
1	B	204	GLY	7.1
1	E	237	PHE	6.4

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.