wwPDB X-ray Structure Validation Summary Report (i)

Aug 16, 2021 – 02:14 pm BST

PDB ID	:	7AE7
Title	:	Structure of Sedimentibacter hydroxybenzoicus vanillic acid decarboxylase (ShVdcCD) in open form, with truncated ShVdcD (V59X)
Authors	:	Marshall, S.A.; Leys, D.
Deposited on	:	2020-09-17
Resolution	:	2.66 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 following versions of software and data (see references (1)) were used in the production of this report:

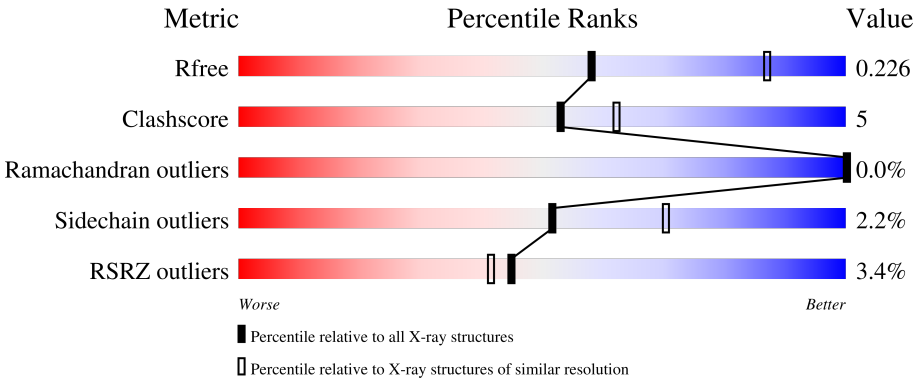
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.23.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.1

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 2.66 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	1332 (2.68-2.64)
Clashscore	141614	1374 (2.68-2.64)
Ramachandran outliers	138981	1349 (2.68-2.64)
Sidechain outliers	138945	1349 (2.68-2.64)
RSRZ outliers	127900	1318 (2.68-2.64)

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	480	86% 11%
1	B	480	5% 80% 17%
1	C	480	5% 81% 16%
1	D	480	85% 14%
1	E	480	4% 82% 14%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	480	2% 82% 16%
2	a	58	95%
2	b	58	19% 78% 5% 17%
2	c	58	10% 97%
2	d	58	95%
2	e	58	14% 98%
2	f	58	5% 91% 7%

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 24349 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 Phenolic acid decarboxylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	467	Total	C	N	O	S	0	0	0
			3608	2315	608	670	15			
1	B	468	Total	C	N	O	S	0	0	0
			3508	2252	590	652	14			
1	C	471	Total	C	N	O	S	0	0	0
			3534	2268	596	656	14			
1	D	475	Total	C	N	O	S	0	0	0
			3675	2352	617	691	15			
1	E	461	Total	C	N	O	S	0	0	0
			3491	2238	584	655	14			
1	F	471	Total	C	N	O	S	0	0	0
			3565	2286	601	663	15			

- Molecule 2 is a protein called Protein ShdD.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	a	57	Total	C	N	O	S	0	0	0
			454	280	79	86	9			
2	b	48	Total	C	N	O	S	0	0	0
			286	179	50	53	4			
2	c	57	Total	C	N	O	S	0	0	0
			415	259	74	74	8			
2	d	57	Total	C	N	O	S	0	0	0
			441	275	77	80	9			
2	e	57	Total	C	N	O	S	0	0	0
			406	257	72	68	9			
2	f	57	Total	C	N	O	S	0	0	0
			419	262	74	74	9			

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Na 1 1	0	0
3	B	1	Total Na 1 1	0	0
3	C	1	Total Na 1 1	0	0
3	E	1	Total Na 1 1	0	0
3	F	1	Total Na 1 1	0	0

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	a	1	Total Zn 1 1	0	0
4	b	1	Total Zn 1 1	0	0
4	c	1	Total Zn 1 1	0	0
4	d	1	Total Zn 1 1	0	0
4	e	1	Total Zn 1 1	0	0
4	f	1	Total Zn 1 1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	139	Total O 139 139	0	0
5	B	74	Total O 74 74	0	0
5	C	60	Total O 60 60	0	0
5	D	92	Total O 92 92	0	0
5	E	65	Total O 65 65	0	0
5	F	67	Total O 67 67	0	0
5	a	21	Total O 21 21	0	0

Continued on next page...

Continued from previous page...

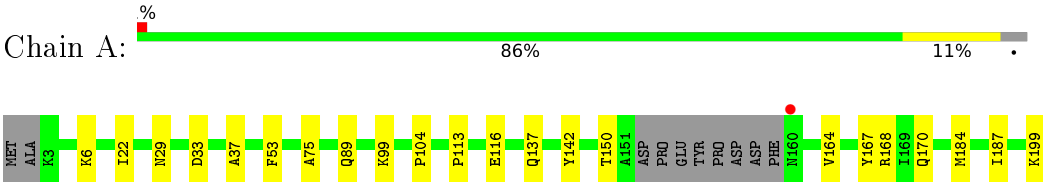
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	c	3	Total O 3 3	0	0
5	d	6	Total O 6 6	0	0
5	e	3	Total O 3 3	0	0
5	f	6	Total O 6 6	0	0

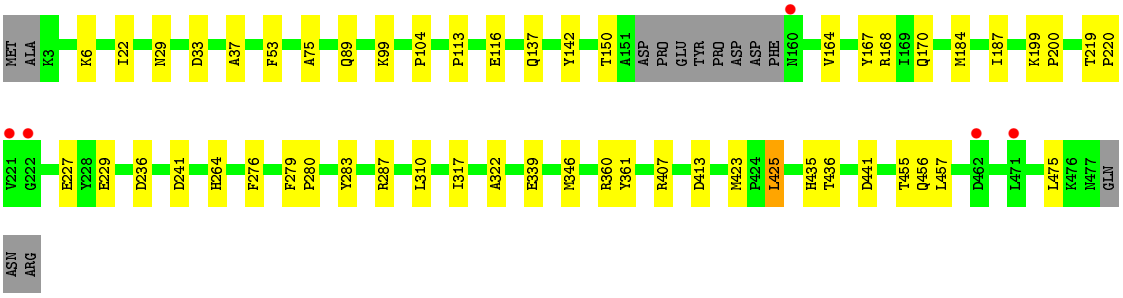
3 Residue-property plots (i)

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: Phenolic acid decarboxylase

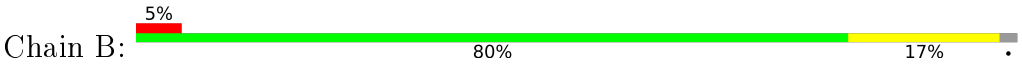
Chain A:

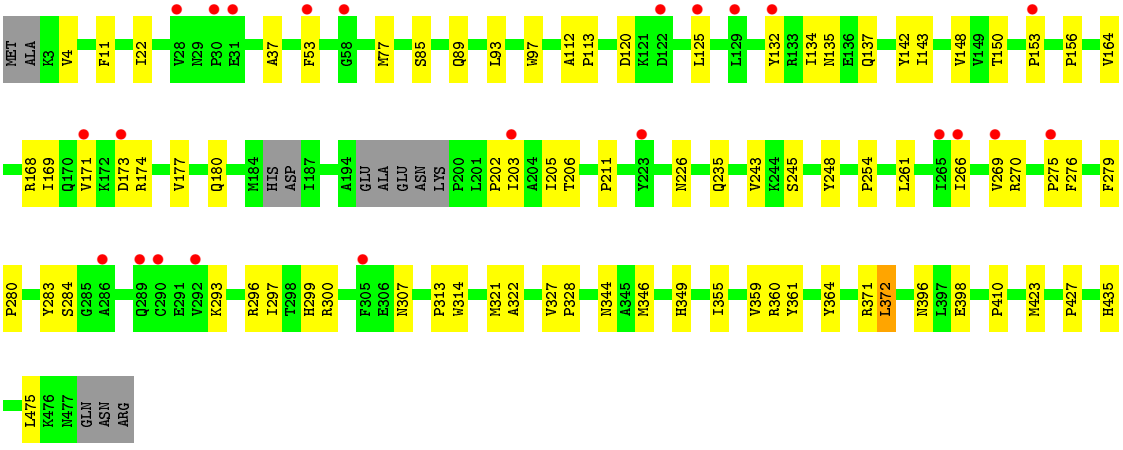




- Molecule 1: Phenolic acid decarboxylase

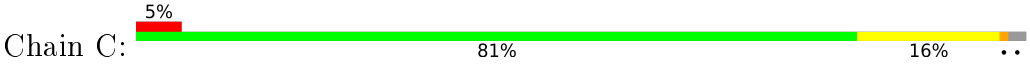
Chain B:

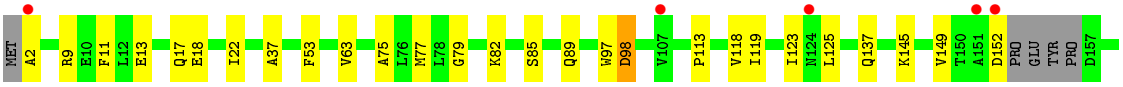


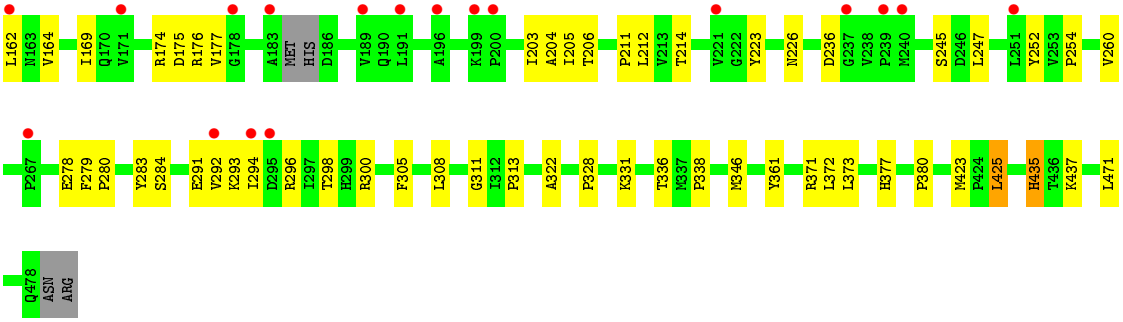


- Molecule 1: Phenolic acid decarboxylase

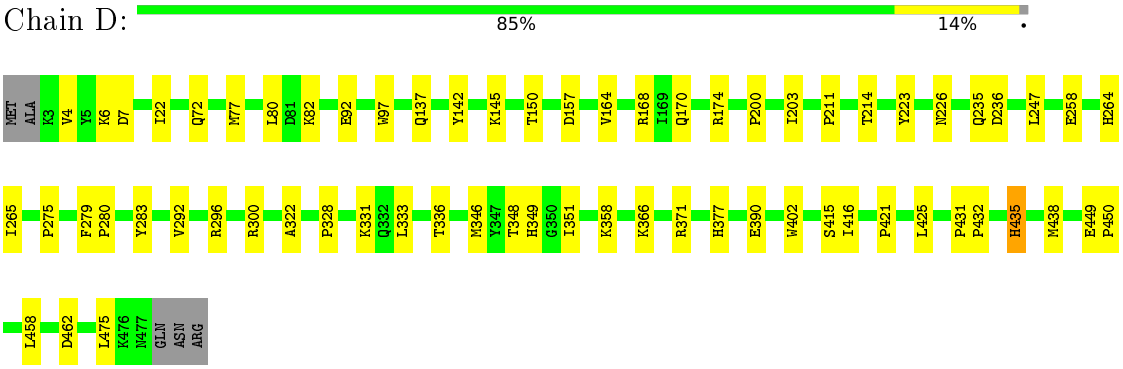
Chain C:



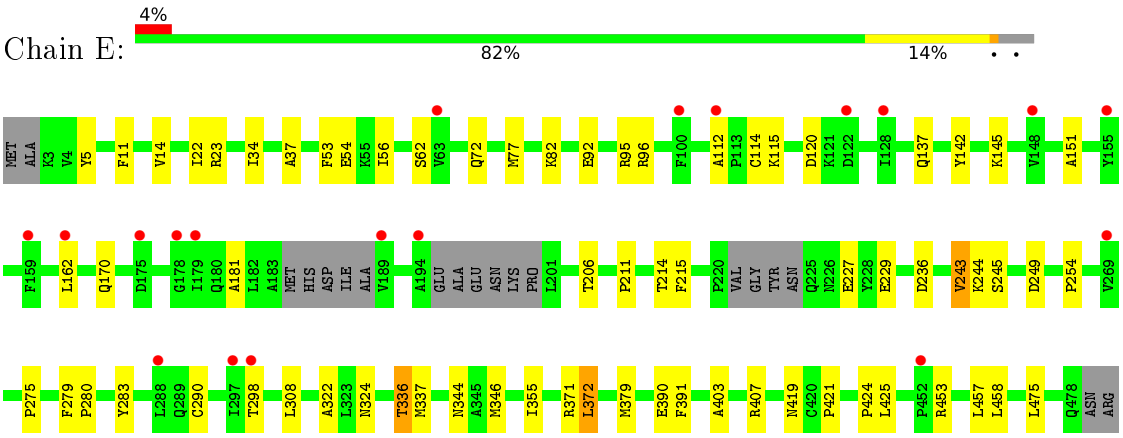




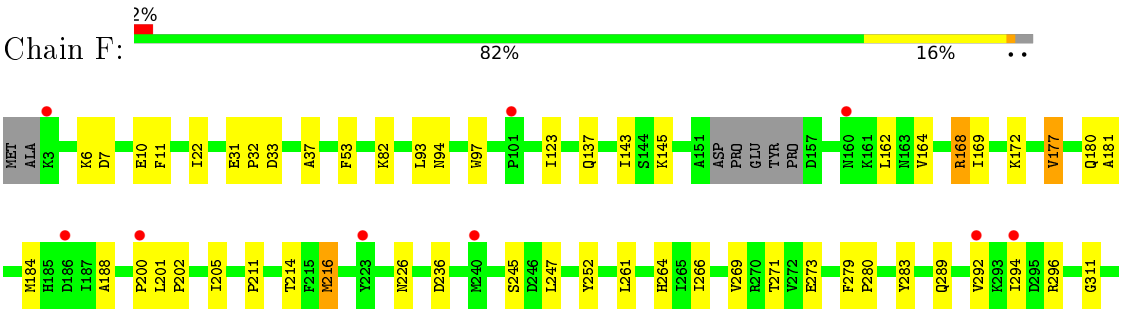
• Molecule 1: Phenolic acid decarboxylase



• Molecule 1: Phenolic acid decarboxylase

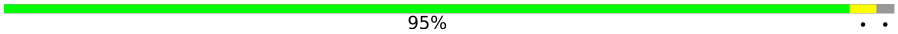


• Molecule 1: Phenolic acid decarboxylase



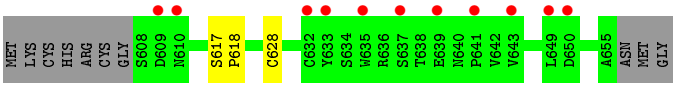
• Molecule 2: Protein ShdD

Chain a:



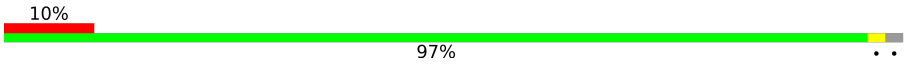
• Molecule 2: Protein ShdD

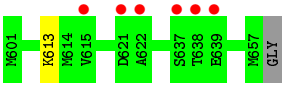
Chain b:



• Molecule 2: Protein ShdD

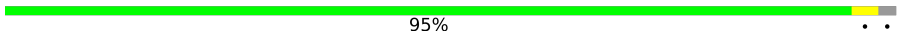
Chain c:





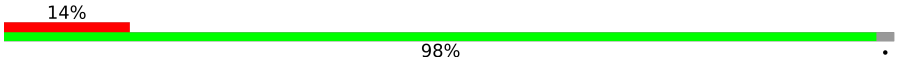
• Molecule 2: Protein ShdD

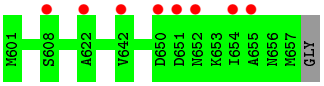
Chain d:



• Molecule 2: Protein ShdD

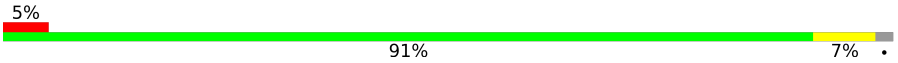
Chain e:

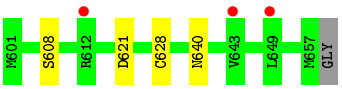




• Molecule 2: Protein ShdD

Chain f:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	101.55Å 200.94Å 201.99Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	71.42 – 2.66 71.42 – 2.66	Depositor EDS
% Data completeness (in resolution range)	99.5 (71.42-2.66) 99.5 (71.42-2.66)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.87 (at 2.65Å)	Xtriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.180 , 0.227 0.180 , 0.226	Depositor DCC
R_{free} test set	5824 reflections (4.89%)	wwPDB-VP
Wilson B-factor (Å^2)	37.4	Xtriage
Anisotropy	0.386	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 51.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.009 for -h,l,k	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	24349	wwPDB-VP
Average B, all atoms (Å^2)	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.32% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: NA, 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	0.27	0/3697	0.51	0/5048
1	B	0.28	0/3598	0.52	0/4930
1	C	0.26	0/3620	0.51	0/4961
1	D	0.26	0/3768	0.51	0/5153
1	E	0.26	0/3577	0.50	0/4898
1	F	0.26	0/3654	0.51	0/5003
2	a	0.25	0/463	0.51	0/623
2	b	0.24	0/292	0.48	0/406
2	c	0.26	0/424	0.52	0/577
2	d	0.26	0/450	0.53	0/606
2	e	0.25	0/415	0.50	0/564
2	f	0.24	0/428	0.52	0/581
All	All	0.26	0/24386	0.51	0/33350

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	3608	0	3544	29	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	3508	0	3326	50	0
1	C	3534	0	3379	47	0
1	D	3675	0	3573	39	0
1	E	3491	0	3338	39	0
1	F	3565	0	3416	43	0
2	a	454	0	426	0	0
2	b	286	0	186	0	0
2	c	415	0	362	0	0
2	d	441	0	412	0	0
2	e	406	0	357	0	0
2	f	419	0	371	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	E	1	0	0	0	0
3	F	1	0	0	0	0
4	a	1	0	0	0	0
4	b	1	0	0	0	0
4	c	1	0	0	0	0
4	d	1	0	0	0	0
4	e	1	0	0	0	0
4	f	1	0	0	0	0
5	A	139	0	0	1	0
5	B	74	0	0	2	0
5	C	60	0	0	1	0
5	D	92	0	0	1	0
5	E	65	0	0	1	0
5	F	67	0	0	2	0
5	a	21	0	0	0	0
5	c	3	0	0	0	0
5	d	6	0	0	0	0
5	e	3	0	0	0	0
5	f	6	0	0	0	0
All	All	24349	0	22690	235	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 5.

The worst 5 of 235 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:322:ALA:HB1	1:F:346:MET:HG2	1.62	0.81
1:E:77:MET:HG3	1:E:211:PRO:HB2	1.63	0.80
1:A:322:ALA:HB1	1:A:346:MET:HG2	1.64	0.80
1:B:112:ALA:HB2	1:B:243:VAL:HG11	1.69	0.73
1:F:289:GLN:NE2	5:F:601:HOH:O	2.24	0.69

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	463/480 (96%)	446 (96%)	17 (4%)	0	100	100
1	B	462/480 (96%)	453 (98%)	9 (2%)	0	100	100
1	C	465/480 (97%)	455 (98%)	10 (2%)	0	100	100
1	D	473/480 (98%)	464 (98%)	9 (2%)	0	100	100
1	E	453/480 (94%)	439 (97%)	14 (3%)	0	100	100
1	F	467/480 (97%)	456 (98%)	11 (2%)	0	100	100
2	a	55/58 (95%)	54 (98%)	1 (2%)	0	100	100
2	b	46/58 (79%)	43 (94%)	2 (4%)	1 (2%)	6	9
2	c	55/58 (95%)	53 (96%)	2 (4%)	0	100	100
2	d	55/58 (95%)	55 (100%)	0	0	100	100
2	e	55/58 (95%)	54 (98%)	1 (2%)	0	100	100
2	f	55/58 (95%)	53 (96%)	2 (4%)	0	100	100
All	All	3104/3228 (96%)	3025 (98%)	78 (2%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	b	618	PRO

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	386/419 (92%)	381 (99%)	5 (1%)	69 82
1	B	358/419 (85%)	354 (99%)	4 (1%)	73 85
1	C	362/419 (86%)	352 (97%)	10 (3%)	43 61
1	D	394/419 (94%)	389 (99%)	5 (1%)	69 82
1	E	364/419 (87%)	355 (98%)	9 (2%)	47 66
1	F	369/419 (88%)	358 (97%)	11 (3%)	41 59
2	a	52/53 (98%)	50 (96%)	2 (4%)	33 49
2	b	16/53 (30%)	14 (88%)	2 (12%)	4 6
2	c	41/53 (77%)	40 (98%)	1 (2%)	49 67
2	d	48/53 (91%)	46 (96%)	2 (4%)	30 45
2	e	38/53 (72%)	38 (100%)	0	100 100
2	f	42/53 (79%)	38 (90%)	4 (10%)	8 12
All	All	2470/2832 (87%)	2415 (98%)	55 (2%)	52 70

5 of 55 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	308	LEU
1	F	168	ARG
2	f	640	ASN
2	d	601	MET
1	E	336	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
2	c	610	ASN
1	F	289	GLN
1	D	377	HIS
1	F	17	GLN
1	D	160	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 monosaccharides in this entry.

5.6 Ligand geometry

Of 11 ligands modelled in this entry, 11 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

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	467/480 (97%)	-0.01	5 (1%) 80 79	16, 27, 76, 119	0
1	B	468/480 (97%)	0.26	23 (4%) 29 26	16, 46, 96, 128	0
1	C	471/480 (98%)	0.35	23 (4%) 29 26	19, 50, 90, 113	0
1	D	475/480 (98%)	-0.14	0 100 100	20, 37, 67, 110	0
1	E	461/480 (96%)	0.17	19 (4%) 37 33	20, 44, 82, 118	0
1	F	471/480 (98%)	0.06	9 (1%) 66 63	21, 44, 88, 113	0
2	a	57/58 (98%)	0.08	0 100 100	21, 30, 71, 85	0
2	b	48/58 (82%)	1.19	11 (22%) 0 0	58, 79, 94, 104	0
2	c	57/58 (98%)	0.86	6 (10%) 6 4	47, 60, 91, 101	0
2	d	57/58 (98%)	-0.06	0 100 100	30, 41, 60, 75	0
2	e	57/58 (98%)	0.85	8 (14%) 2 2	46, 60, 87, 102	0
2	f	57/58 (98%)	0.47	3 (5%) 26 23	43, 58, 83, 96	0
All	All	3146/3228 (97%)	0.16	107 (3%) 45 41	16, 43, 86, 128	0

The worst 5 of 107 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	b	643	VAL	3.5
1	B	153	PRO	3.4
1	F	294	ILE	3.4
1	B	58	GLY	3.3
1	B	171	VAL	3.3

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 (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^{th} 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^2)	Q<0.9
3	NA	A	501	1/1	0.82	0.23	49,49,49,49	0
4	ZN	b	701	1/1	0.85	0.09	84,84,84,84	0
3	NA	E	501	1/1	0.86	0.21	41,41,41,41	0
3	NA	C	501	1/1	0.88	0.26	31,31,31,31	0
3	NA	F	501	1/1	0.89	0.15	43,43,43,43	0
4	ZN	c	701	1/1	0.90	0.22	100,100,100,100	0
3	NA	B	501	1/1	0.94	0.24	28,28,28,28	0
4	ZN	d	701	1/1	0.95	0.19	56,56,56,56	0
4	ZN	e	701	1/1	0.98	0.07	69,69,69,69	0
4	ZN	f	701	1/1	0.98	0.09	56,56,56,56	0
4	ZN	a	701	1/1	0.99	0.15	28,28,28,28	0

6.5 Other polymers (i)

There are no such residues in this entry.