wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 12:26 AM GMT

PDB ID	:	2AF7
Title	:	Crystal structure of the gamma-carboxymuconolactone decarboxylase from Methanobacterium thermoautotrophicum. Northeast Structural Genomics Consortium target TT747.
Authors	:	Vorobiev, S.M.; Kuzin, A.; Skarina, T.; Savchenko, A.; Semesi, A.; Arrowsmith, C.; Edwards, A.; Montelione, G.T.; Tong, L.; Northeast Structural Genomics Consortium (NESG)
Deposited on	:	2005-07-25
Resolution	:	2.81 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:

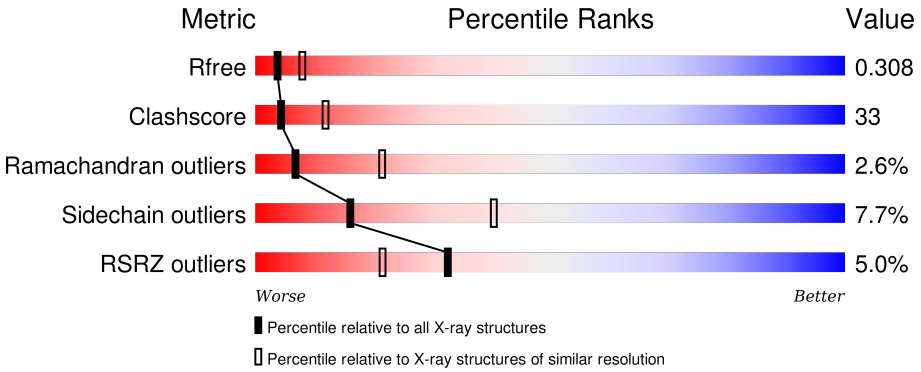
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 2.81 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	2676 (2.84-2.80)
Clashscore	102246	3124 (2.84-2.80)
Ramachandran outliers	100387	3072 (2.84-2.80)
Sidechain outliers	100360	3074 (2.84-2.80)
RSRZ outliers	91569	2690 (2.84-2.80)

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	125	2% 54% 35% 5% • 5%
1	B	125	3% 51% 41% • 6%
1	C	125	3% 48% 41% • • 6%
1	D	125	6% 46% 42% 9% • •
1	E	125	5% 50% 39% 7% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	125	6% 48% 44%
1	G	125	9% 43% 44% 7% 6%
1	H	125	3% 53% 39% 5%
1	I	125	5% 47% 43% 6%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 8293 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 gamma-carboxymuconolactone decarboxylase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	119	Total	C	N	O	S	Se	0	0	0
			909	576	162	165	1	5			
1	B	118	Total	C	N	O	S	Se	0	0	0
			900	571	158	166	1	4			
1	C	118	Total	C	N	O	S	Se	0	0	0
			910	576	161	168	1	4			
1	D	122	Total	C	N	O	S	Se	0	0	0
			935	590	166	173	1	5			
1	E	120	Total	C	N	O	S	Se	0	0	0
			908	575	160	167	1	5			
1	F	120	Total	C	N	O	S	Se	0	0	0
			888	566	153	164	1	4			
1	G	118	Total	C	N	O	S	Se	0	0	0
			914	578	161	170	1	4			
1	H	119	Total	C	N	O	S	Se	0	0	0
			922	582	163	172	1	4			
1	I	120	Total	C	N	O	S	Se	0	0	0
			913	578	161	169	1	4			

There are 45 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	MODIFIED RESIDUE	UNP O26336
A	8	MSE	MET	MODIFIED RESIDUE	UNP O26336
A	14	MSE	MET	MODIFIED RESIDUE	UNP O26336
A	94	MSE	MET	MODIFIED RESIDUE	UNP O26336
A	97	MSE	MET	MODIFIED RESIDUE	UNP O26336
B	1	MSE	MET	MODIFIED RESIDUE	UNP O26336
B	8	MSE	MET	MODIFIED RESIDUE	UNP O26336
B	14	MSE	MET	MODIFIED RESIDUE	UNP O26336
B	94	MSE	MET	MODIFIED RESIDUE	UNP O26336
B	97	MSE	MET	MODIFIED RESIDUE	UNP O26336
C	1	MSE	MET	MODIFIED RESIDUE	UNP O26336

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	8	MSE	MET	MODIFIED RESIDUE	UNP O26336
C	14	MSE	MET	MODIFIED RESIDUE	UNP O26336
C	94	MSE	MET	MODIFIED RESIDUE	UNP O26336
C	97	MSE	MET	MODIFIED RESIDUE	UNP O26336
D	1	MSE	MET	MODIFIED RESIDUE	UNP O26336
D	8	MSE	MET	MODIFIED RESIDUE	UNP O26336
D	14	MSE	MET	MODIFIED RESIDUE	UNP O26336
D	94	MSE	MET	MODIFIED RESIDUE	UNP O26336
D	97	MSE	MET	MODIFIED RESIDUE	UNP O26336
E	1	MSE	MET	MODIFIED RESIDUE	UNP O26336
E	8	MSE	MET	MODIFIED RESIDUE	UNP O26336
E	14	MSE	MET	MODIFIED RESIDUE	UNP O26336
E	94	MSE	MET	MODIFIED RESIDUE	UNP O26336
E	97	MSE	MET	MODIFIED RESIDUE	UNP O26336
F	1	MSE	MET	MODIFIED RESIDUE	UNP O26336
F	8	MSE	MET	MODIFIED RESIDUE	UNP O26336
F	14	MSE	MET	MODIFIED RESIDUE	UNP O26336
F	94	MSE	MET	MODIFIED RESIDUE	UNP O26336
F	97	MSE	MET	MODIFIED RESIDUE	UNP O26336
G	1	MSE	MET	MODIFIED RESIDUE	UNP O26336
G	8	MSE	MET	MODIFIED RESIDUE	UNP O26336
G	14	MSE	MET	MODIFIED RESIDUE	UNP O26336
G	94	MSE	MET	MODIFIED RESIDUE	UNP O26336
G	97	MSE	MET	MODIFIED RESIDUE	UNP O26336
H	1	MSE	MET	MODIFIED RESIDUE	UNP O26336
H	8	MSE	MET	MODIFIED RESIDUE	UNP O26336
H	14	MSE	MET	MODIFIED RESIDUE	UNP O26336
H	94	MSE	MET	MODIFIED RESIDUE	UNP O26336
H	97	MSE	MET	MODIFIED RESIDUE	UNP O26336
I	1	MSE	MET	MODIFIED RESIDUE	UNP O26336
I	8	MSE	MET	MODIFIED RESIDUE	UNP O26336
I	14	MSE	MET	MODIFIED RESIDUE	UNP O26336
I	94	MSE	MET	MODIFIED RESIDUE	UNP O26336
I	97	MSE	MET	MODIFIED RESIDUE	UNP O26336

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	13	Total O 13 13	0	0
2	B	8	Total O 8 8	0	0

Continued on next page...

Continued from previous page...

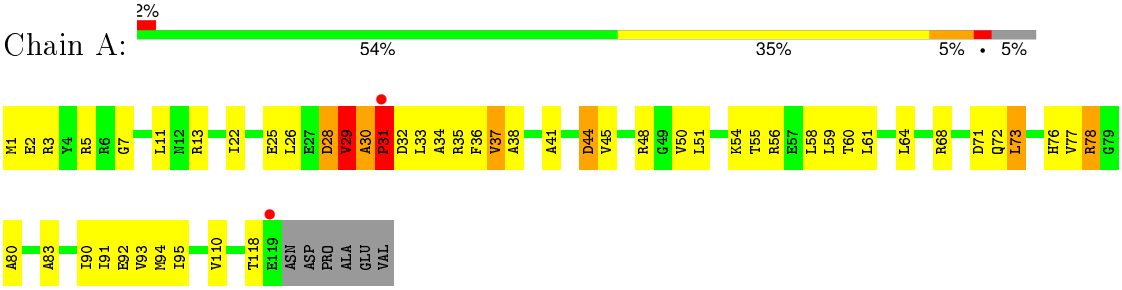
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	C	12	Total O 12 12	0	0
2	D	11	Total O 11 11	0	0
2	E	5	Total O 5 5	0	0
2	F	9	Total O 9 9	0	0
2	G	10	Total O 10 10	0	0
2	H	13	Total O 13 13	0	0
2	I	13	Total O 13 13	0	0

3 Residue-property plots

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.

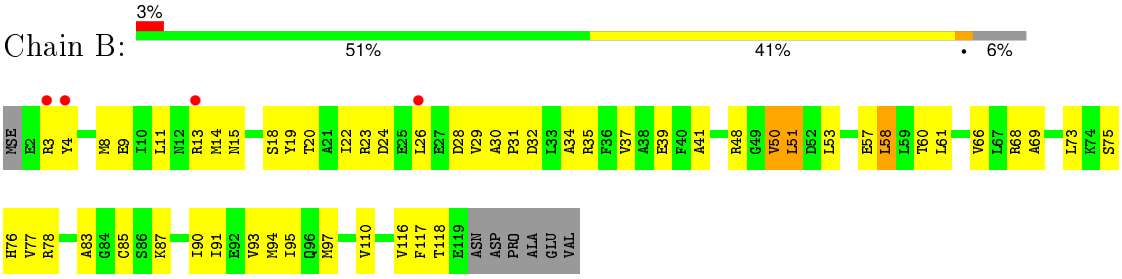
- Molecule 1: gamma-carboxymuconolactone decarboxylase

Chain A:



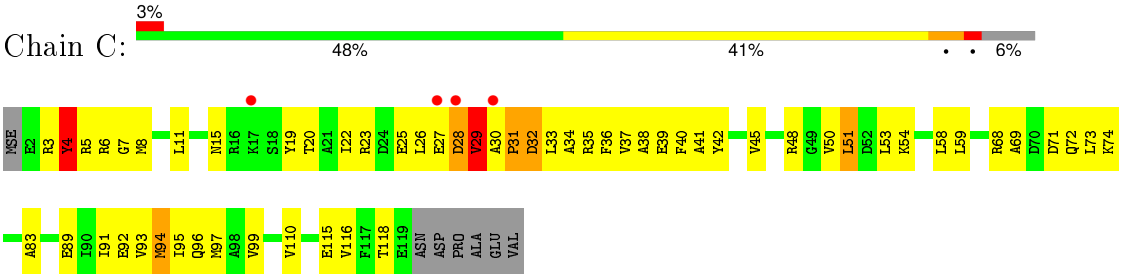
- Molecule 1: gamma-carboxymuconolactone decarboxylase

Chain B:



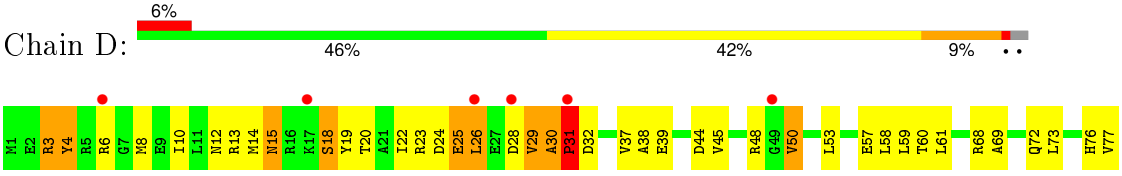
- Molecule 1: gamma-carboxymuconolactone decarboxylase

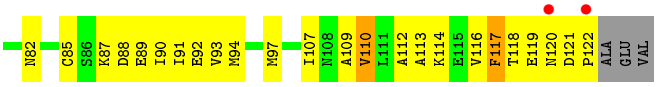
Chain C:



- Molecule 1: gamma-carboxymuconolactone decarboxylase

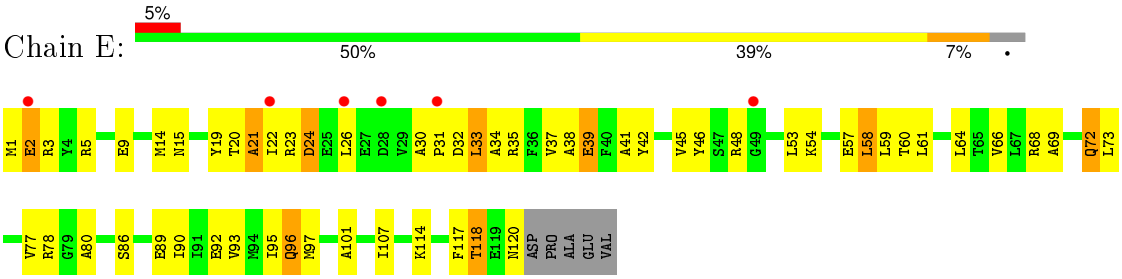
Chain D:





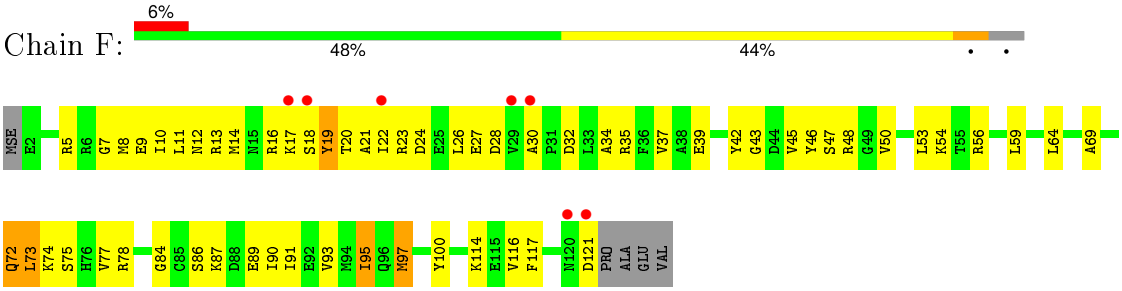
• Molecule 1: gamma-carboxymuconolactone decarboxylase

Chain E:



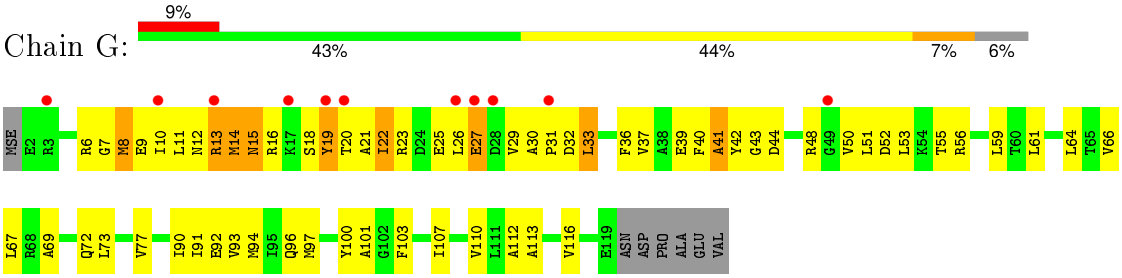
• Molecule 1: gamma-carboxymuconolactone decarboxylase

Chain F:



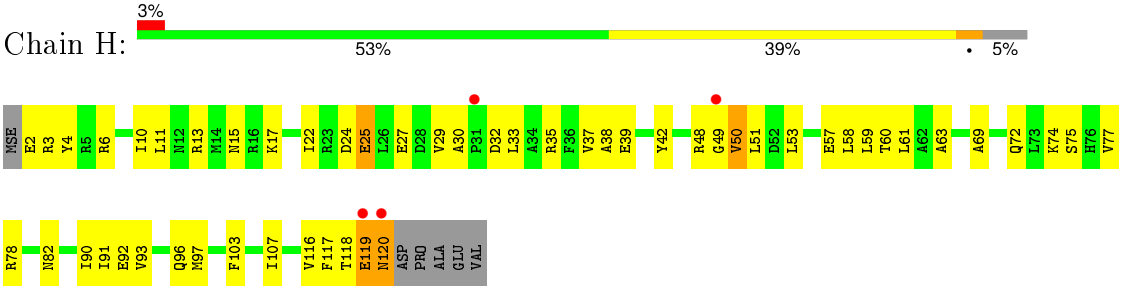
• Molecule 1: gamma-carboxymuconolactone decarboxylase

Chain G:



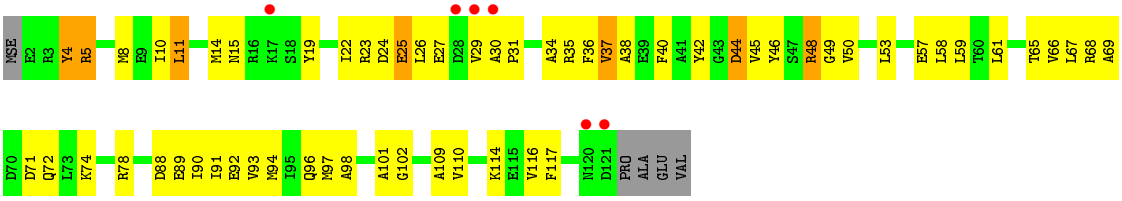
• Molecule 1: gamma-carboxymuconolactone decarboxylase

Chain H:



• Molecule 1: gamma-carboxymuconolactone decarboxylase

Chain I:


MSE E2 R3 Y4 R5 M8 E9 I10 L11 M14 N15 R16 K17 S18 Y19 I22 R23 D24 E25 L26 E27 D28 V29 A30 P31 A34 R35 F36 V37 A38 E39 F40 A41 Y42 G43 D44 V45 Y46 S47 R48 G49 V50 L53 E57 L58 L59 T60 L61 T65 V66 L67 R68 A69
D70 D71 Q72 L73 K74 R78 D88 E89 I90 I91 E92 V93 M94 I95 Q96 M97 A98 A101 G102 A109 V110 K114 E115 V116 F117 N120 D121 PRO ALA GLU VAL

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	183.31Å 119.11Å 73.62Å 90.00° 102.80° 90.00°	Depositor
Resolution (Å)	29.97 – 2.81 29.97 – 2.81	Depositor EDS
% Data completeness (in resolution range)	80.3 (29.97-2.81) 96.4 (29.97-2.81)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.20 (at 2.80Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.245 , 0.292 0.255 , 0.308	Depositor DCC
R_{free} test set	1393 reflections (3.98%)	DCC
Wilson B-factor (Å^2)	46.3	Xtriage
Anisotropy	0.287	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 57.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Outliers	1 of 73346 reflections (0.001%)	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	8293	wwPDB-VP
Average B, all atoms (Å^2)	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.65% 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, 0.375 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	0.64	2/916 (0.2%)	0.90	5/1229 (0.4%)
1	B	0.38	0/907	0.69	0/1219
1	C	0.42	0/917	0.80	1/1231 (0.1%)
1	D	0.51	0/943	0.85	5/1267 (0.4%)
1	E	0.45	0/915	0.77	2/1231 (0.2%)
1	F	0.45	0/895	0.75	1/1207 (0.1%)
1	G	0.40	0/921	0.68	0/1236
1	H	0.47	0/929	0.75	2/1247 (0.2%)
1	I	0.44	0/920	0.73	1/1237 (0.1%)
All	All	0.47	2/8263 (0.0%)	0.77	17/11104 (0.2%)

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	A	0	1
1	D	0	1
1	G	0	1
All	All	0	3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	29	VAL	CA-C	-8.35	1.31	1.52
1	A	30	ALA	CA-CB	-8.31	1.34	1.52

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	29	VAL	N-CA-C	9.74	137.30	111.00
1	A	28	ASP	N-CA-C	-7.94	89.56	111.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	F	32	ASP	N-CA-C	-7.02	92.05	111.00
1	A	29	VAL	CB-CA-C	-6.87	98.35	111.40
1	A	30	ALA	C-N-CA	-6.51	94.65	122.00

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	29	VAL	Mainchain
1	D	28	ASP	Peptide
1	G	100	TYR	Sidechain

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	909	0	919	84	0
1	B	900	0	905	58	0
1	C	910	0	920	72	0
1	D	935	0	938	71	0
1	E	908	0	908	71	0
1	F	888	0	872	67	0
1	G	914	0	924	82	0
1	H	922	0	930	62	0
1	I	913	0	913	66	0
2	A	13	0	0	1	0
2	B	8	0	0	0	0
2	C	12	0	0	1	0
2	D	11	0	0	0	0
2	E	5	0	0	1	0
2	F	9	0	0	1	0
2	G	10	0	0	2	0
2	H	13	0	0	0	0
2	I	13	0	0	2	0
All	All	8293	0	8229	540	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 33.

The worst 5 of 540 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:30:ALA:HB1	1:A:33:LEU:H	1.01	1.10
1:B:29:VAL:HG22	1:B:30:ALA:H	1.16	1.05
1:F:91:ILE:HD11	1:F:117:PHE:HE2	1.22	1.03
1:A:30:ALA:HB1	1:A:33:LEU:N	1.76	1.01
1:A:56:ARG:NH2	1:C:96:GLN:HB2	1.73	1.00

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	117/125 (94%)	108 (92%)	7 (6%)	2 (2%)	11 34
1	B	116/125 (93%)	99 (85%)	15 (13%)	2 (2%)	11 34
1	C	116/125 (93%)	95 (82%)	18 (16%)	3 (3%)	7 21
1	D	120/125 (96%)	96 (80%)	20 (17%)	4 (3%)	5 16
1	E	118/125 (94%)	96 (81%)	20 (17%)	2 (2%)	11 34
1	F	118/125 (94%)	92 (78%)	23 (20%)	3 (2%)	7 23
1	G	116/125 (93%)	95 (82%)	14 (12%)	7 (6%)	2 5
1	H	117/125 (94%)	105 (90%)	11 (9%)	1 (1%)	21 53
1	I	118/125 (94%)	101 (86%)	14 (12%)	3 (2%)	7 23
All	All	1056/1125 (94%)	887 (84%)	142 (13%)	27 (3%)	7 21

5 of 27 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	28	ASP
1	B	50	VAL
1	D	29	VAL
1	F	27	GLU
1	G	41	ALA

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	89/96 (93%)	83 (93%)	6 (7%)	20	49
1	B	89/96 (93%)	84 (94%)	5 (6%)	26	58
1	C	91/96 (95%)	81 (89%)	10 (11%)	8	22
1	D	93/96 (97%)	87 (94%)	6 (6%)	21	50
1	E	89/96 (93%)	82 (92%)	7 (8%)	15	39
1	F	84/96 (88%)	77 (92%)	7 (8%)	14	37
1	G	92/96 (96%)	86 (94%)	6 (6%)	21	50
1	H	93/96 (97%)	86 (92%)	7 (8%)	17	42
1	I	90/96 (94%)	82 (91%)	8 (9%)	12	33
All	All	810/864 (94%)	748 (92%)	62 (8%)	16	40

5 of 62 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	15	ASN
1	F	19	TYR
1	I	15	ASN
1	E	58	LEU
1	F	72	GLN

Some 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	72	GLN
1	F	72	GLN
1	H	120	ASN
1	E	15	ASN
1	B	15	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 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	114/125 (91%)	-0.22	2 (1%)	71	61	11, 39, 75, 102	0
1	B	114/125 (91%)	0.04	4 (3%)	48	35	20, 55, 95, 102	0
1	C	114/125 (91%)	-0.14	4 (3%)	48	35	11, 42, 91, 101	0
1	D	117/125 (93%)	0.17	8 (6%)	20	12	17, 48, 102, 102	0
1	E	115/125 (92%)	0.11	6 (5%)	31	20	13, 52, 102, 102	0
1	F	116/125 (92%)	-0.05	7 (6%)	25	15	18, 48, 95, 102	0
1	G	114/125 (91%)	0.23	11 (9%)	10	5	15, 49, 102, 102	0
1	H	115/125 (92%)	-0.14	4 (3%)	48	35	14, 39, 81, 102	0
1	I	116/125 (92%)	0.12	6 (5%)	31	20	21, 50, 96, 102	0
All	All	1035/1125 (92%)	0.01	52 (5%)	32	21	11, 46, 97, 102	0

The worst 5 of 52 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	I	28	ASP	5.2
1	G	19	TYR	4.1
1	G	49	GLY	4.1
1	D	120	ASN	4.0
1	H	31	PRO	4.0

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

There are no such residues in this entry.