wwPDB X-ray Structure Validation Summary Report (i)

Sep 28, 2024 – 08:06 pm BST

PDB ID	:	1GU9
Title	:	Crystal Structure of Mycobacterium tuberculosis Alkylperoxidase AhpD
Authors	:	Nunn, C.M.; Djordjevic, S.; Hillas, P.J.; Nishida, C.; Ortiz de Montellano, P.R.
Deposited on	:	2002-01-24
Resolution	:	1.90 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
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:

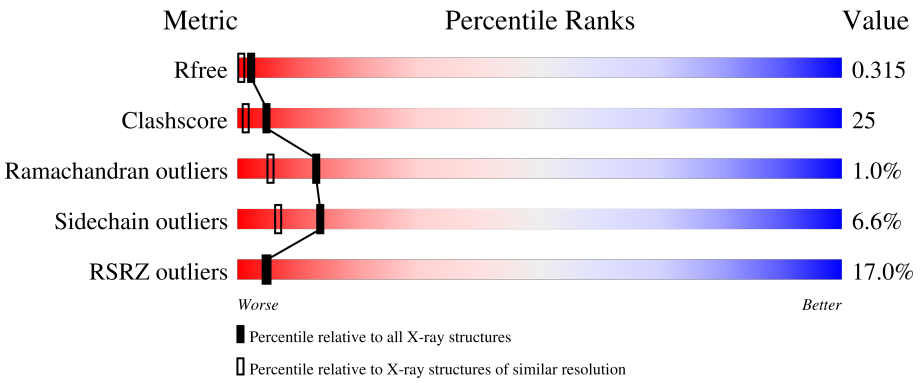
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

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 1.90 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}	164625	7293 (1.90-1.90)
Clashscore	180529	8090 (1.90-1.90)
Ramachandran outliers	177936	8022 (1.90-1.90)
Sidechain outliers	177891	8022 (1.90-1.90)
RSRZ outliers	164620	7292 (1.90-1.90)

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 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	177	10% 67% 25% 5% •
1	B	177	14% 67% 27% • •
1	C	177	10% 65% 25% 5% 5%
1	D	177	9% 63% 29% 5% •
1	E	177	27% 54% 37% 6% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	177	13% 67% 25% 5%
1	G	177	15% 63% 28% 5%
1	H	177	14% 63% 31%
1	I	177	26% 61% 32% 5%
1	J	177	16% 62% 32%
1	K	177	16% 58% 36%
1	L	177	25% 57% 33% 5% 5%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 16174 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 ALKYLHYDROPEROXIDASE D.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	172	1284	808	232	239	3	2	0	0	0
1	B	173	1292	814	233	240	3	2	8	0	0
1	C	168	1256	791	228	232	3	2	0	0	0
1	D	172	1284	808	232	239	3	2	0	0	0
1	E	173	1292	814	233	240	3	2	8	0	0
1	F	169	1264	797	229	233	3	2	0	0	0
1	G	168	1256	791	228	232	3	2	11	0	0
1	H	171	1275	803	231	236	3	2	0	0	0
1	I	173	1292	814	233	240	3	2	0	0	0
1	J	172	1284	808	232	239	3	2	0	0	0
1	K	173	1292	814	233	240	3	2	0	0	0
1	L	168	1256	791	228	232	3	2	11	0	0

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	61	61	61	0	0
2	B	78	78	78	0	0

Continued on next page...

Continued from previous page...

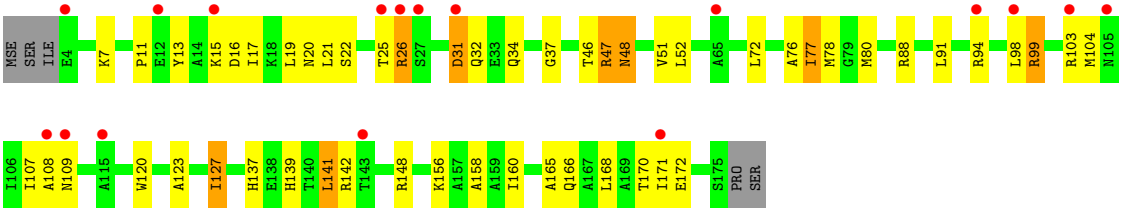
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	C	88	Total O 88 88	0	0
2	D	70	Total O 70 70	0	0
2	E	53	Total O 53 53	0	0
2	F	108	Total O 108 108	0	0
2	G	66	Total O 66 66	0	0
2	H	65	Total O 65 65	0	0
2	I	57	Total O 57 57	0	0
2	J	76	Total O 76 76	0	0
2	K	58	Total O 58 58	0	0
2	L	67	Total O 67 67	0	0

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.

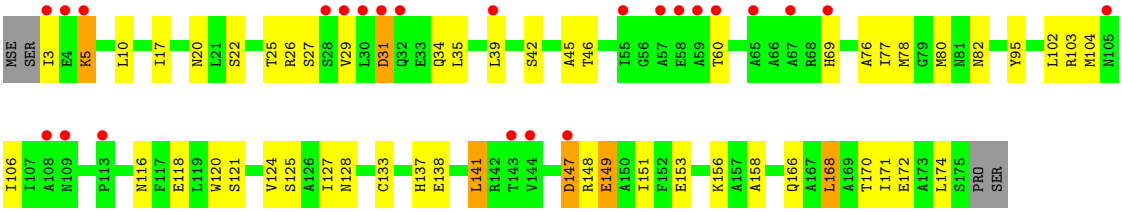
• Molecule 1: ALKYLHYDROPEROXIDASE D

Chain A:



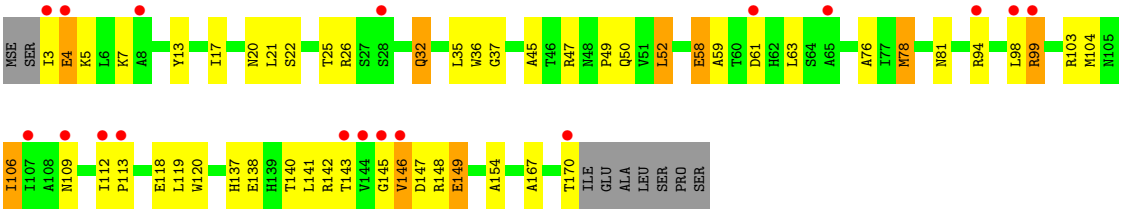
• Molecule 1: ALKYLHYDROPEROXIDASE D

Chain B:



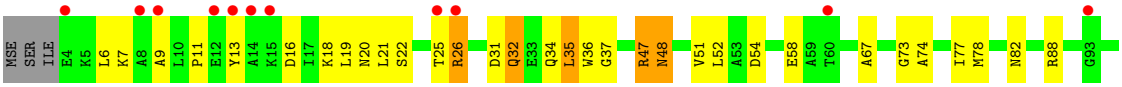
• Molecule 1: ALKYLHYDROPEROXIDASE D

Chain C:



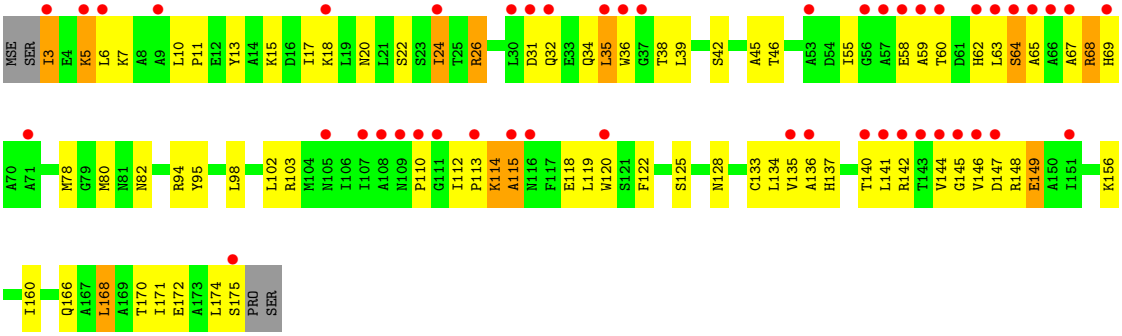
• Molecule 1: ALKYLHYDROPEROXIDASE D

Chain D:



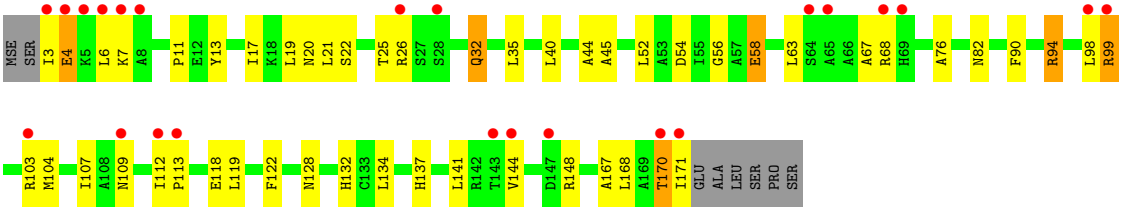
• Molecule 1: ALKYLHYDROPEROXIDASE D

Chain E:



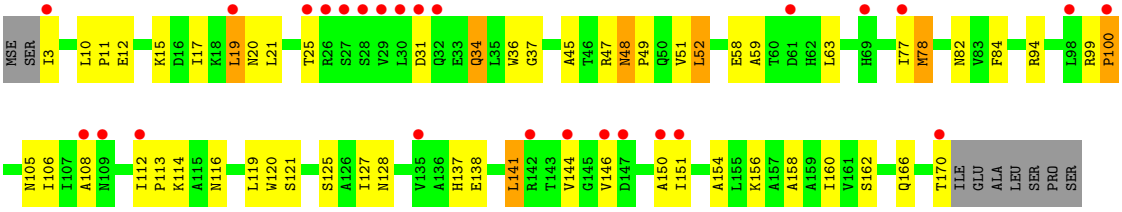
• Molecule 1: ALKYLHYDROPEROXIDASE D

Chain F:



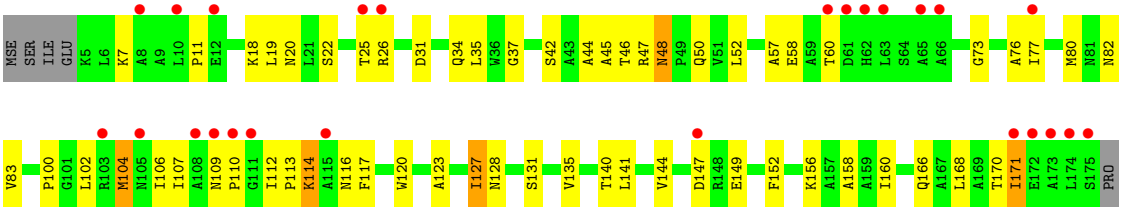
• Molecule 1: ALKYLHYDROPEROXIDASE D

Chain G:



• Molecule 1: ALKYLHYDROPEROXIDASE D

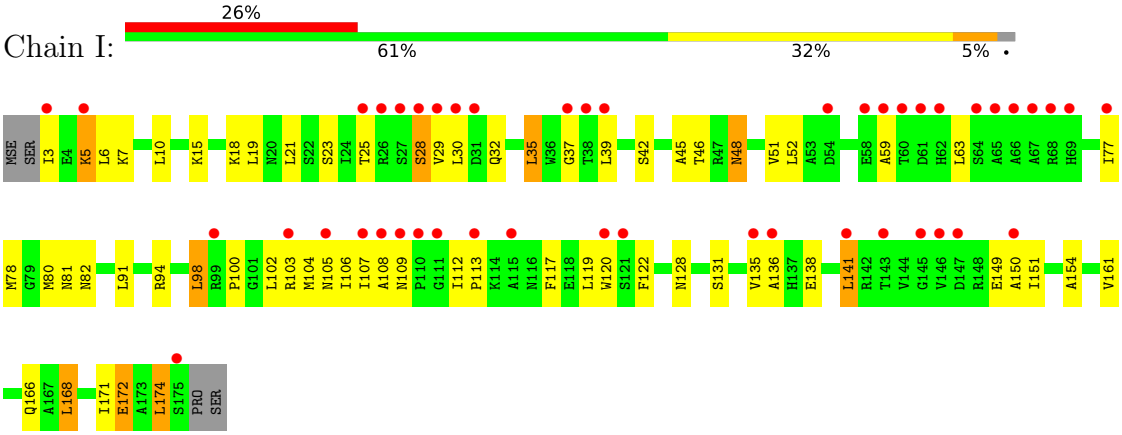
Chain H:



SER

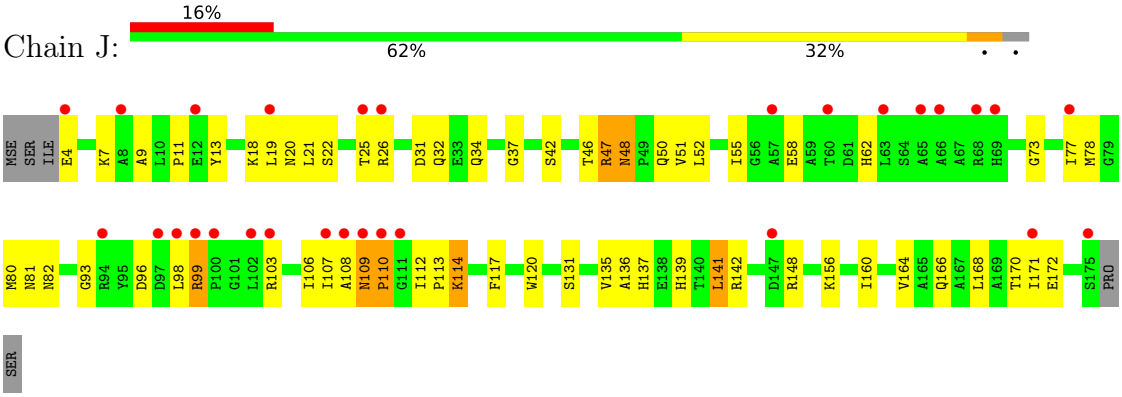
• Molecule 1: ALKYLHYDROPEROXIDASE D

Chain I:



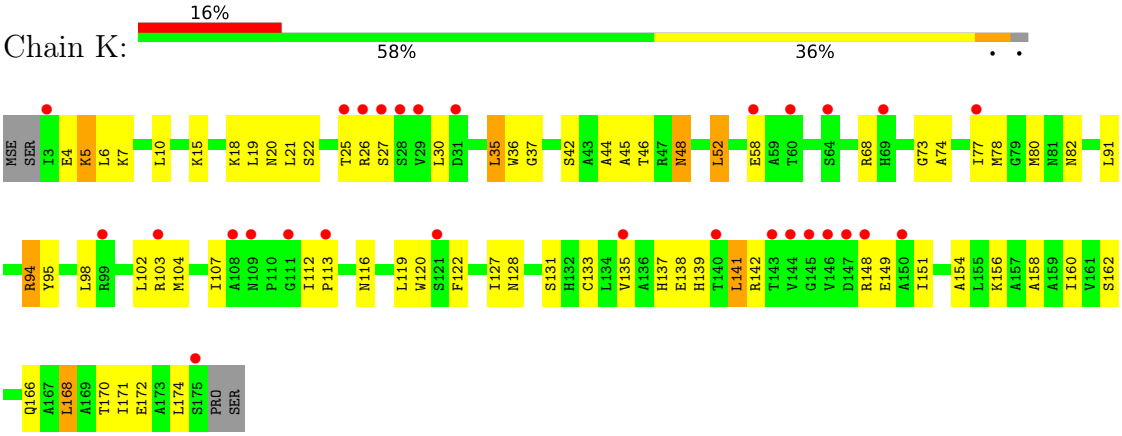
• Molecule 1: ALKYLHYDROPEROXIDASE D

Chain J:



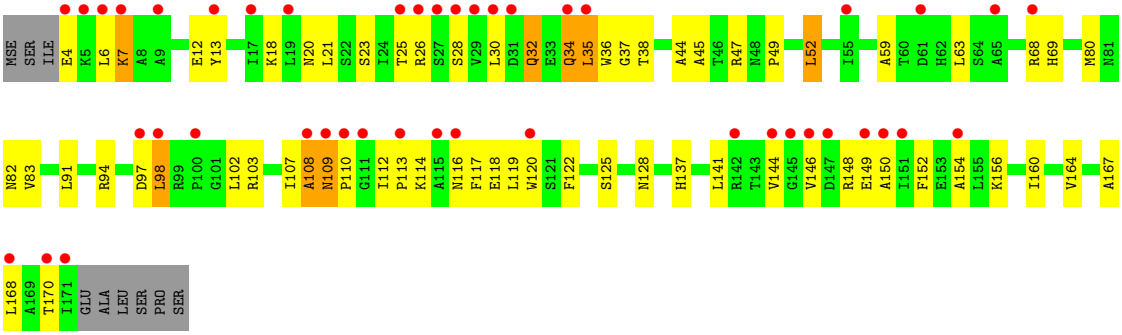
• Molecule 1: ALKYLHYDROPEROXIDASE D

Chain K:



• Molecule 1: ALKYLHYDROPEROXIDASE D

Chain L:


MSE
SER
ILE
E4
K5
L6
K7
A8
A9
E12
Y13
I17
K18
L19
N20
L21
S22
S23
I24
T25
R26
S27
S28
V29
L30
D31
Q32
E33
Q34
L35
W36
G37
T38
A44
A45
T46
R47
N48
P49
L52
I55
A59
T60
D61
H62
L63
S64
A65
R68
H69
M80
N81
N82
V83
L91
R94
D97
L98
R99
P100
G101
L102
R103
I107
A108
N109
P110
G111
I112
P113
K114
A115
N116
F117
E118
L119
W120
S121
F122
S125
N128
H137
L141
R142
T143
V144
G145
V146
D147
R148
E149
A150
I151
F152
E153
A154
L155
K156
I160
V164
A167
L168
A169
T170
I171
GLU
ALA
LEU
SER
PRO
SER

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	186.38Å 117.28Å 88.99Å 90.00° 113.97° 90.00°	Depositor
Resolution (Å)	29.60 – 1.90 29.60 – 1.90	Depositor EDS
% Data completeness (in resolution range)	99.5 (29.60-1.90) 99.4 (29.60-1.90)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.11 (at 1.91Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.243 , 0.313 0.245 , 0.315	Depositor DCC
R_{free} test set	13616 reflections (9.98%)	wwPDB-VP
Wilson B-factor (Å^2)	21.9	Xtriage
Anisotropy	0.227	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 56.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	16174	wwPDB-VP
Average B, all atoms (Å^2)	32.0	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 88.34 % 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 5.0701e-08. 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, 0.333 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	0.45	0/1303	0.65	0/1764
1	B	0.44	0/1311	0.64	0/1775
1	C	0.46	0/1275	0.66	0/1726
1	D	0.47	0/1303	0.64	0/1764
1	E	0.44	0/1311	0.64	0/1775
1	F	0.45	0/1283	0.65	0/1737
1	G	0.47	0/1275	0.66	1/1726 (0.1%)
1	H	0.45	0/1294	0.65	1/1752 (0.1%)
1	I	0.41	0/1311	0.66	0/1775
1	J	0.44	0/1303	0.63	0/1764
1	K	0.43	0/1311	0.64	0/1775
1	L	0.43	0/1275	0.64	0/1726
All	All	0.45	0/15555	0.65	2/21059 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	H	104	MET	CG-SD-CE	-5.99	90.62	100.20
1	G	78	MSE	CG-SE-CE	-5.94	85.84	98.90

There are no chirality outliers.

There are no planarity outliers.

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	1284	0	1296	75	0
1	B	1292	0	1307	59	0
1	C	1256	0	1269	66	0
1	D	1284	0	1296	75	0
1	E	1292	0	1307	81	0
1	F	1264	0	1280	56	0
1	G	1256	0	1269	59	0
1	H	1275	0	1290	74	0
1	I	1292	0	1307	79	0
1	J	1284	0	1296	84	0
1	K	1292	0	1307	72	0
1	L	1256	0	1269	79	0
2	A	61	0	0	4	0
2	B	78	0	0	6	0
2	C	88	0	0	6	0
2	D	70	0	0	3	0
2	E	53	0	0	4	0
2	F	108	0	0	2	0
2	G	66	0	0	2	0
2	H	65	0	0	6	0
2	I	57	0	0	7	0
2	J	76	0	0	3	0
2	K	58	0	0	5	0
2	L	67	0	0	5	0
All	All	16174	0	15493	769	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 25.

The worst 5 of 769 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:J:78:MSE:HE2	1:J:82:ASN:ND2	1.58	1.19
1:G:78:MSE:HE2	1:G:121:SER:HB3	1.15	1.14
1:K:5:LYS:HA	1:K:5:LYS:HE3	1.32	1.10
1:C:3:ILE:HG22	1:C:4:GLU:H	1.16	1.10
1:F:99:ARG:H	1:F:99:ARG:HD2	1.07	1.10

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	170/177 (96%)	162 (95%)	6 (4%)	2 (1%)	11 4
1	B	171/177 (97%)	165 (96%)	6 (4%)	0	100 100
1	C	166/177 (94%)	157 (95%)	7 (4%)	2 (1%)	11 4
1	D	170/177 (96%)	164 (96%)	5 (3%)	1 (1%)	22 13
1	E	171/177 (97%)	157 (92%)	10 (6%)	4 (2%)	5 1
1	F	167/177 (94%)	165 (99%)	1 (1%)	1 (1%)	22 13
1	G	166/177 (94%)	155 (93%)	9 (5%)	2 (1%)	11 4
1	H	169/177 (96%)	161 (95%)	7 (4%)	1 (1%)	22 13
1	I	171/177 (97%)	160 (94%)	10 (6%)	1 (1%)	22 13
1	J	170/177 (96%)	158 (93%)	9 (5%)	3 (2%)	7 2
1	K	171/177 (97%)	163 (95%)	7 (4%)	1 (1%)	22 13
1	L	166/177 (94%)	155 (93%)	8 (5%)	3 (2%)	7 2
All	All	2028/2124 (96%)	1922 (95%)	85 (4%)	21 (1%)	13 5

5 of 21 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	99	ARG
1	E	64	SER
1	E	115	ALA
1	G	100	PRO
1	J	109	ASN

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	128/130 (98%)	119 (93%)	9 (7%)	12	6
1	B	129/130 (99%)	123 (95%)	6 (5%)	22	14
1	C	125/130 (96%)	115 (92%)	10 (8%)	10	4
1	D	128/130 (98%)	118 (92%)	10 (8%)	10	4
1	E	129/130 (99%)	121 (94%)	8 (6%)	15	7
1	F	126/130 (97%)	116 (92%)	10 (8%)	10	4
1	G	125/130 (96%)	117 (94%)	8 (6%)	14	7
1	H	127/130 (98%)	119 (94%)	8 (6%)	15	7
1	I	129/130 (99%)	118 (92%)	11 (8%)	8	3
1	J	128/130 (98%)	123 (96%)	5 (4%)	27	20
1	K	129/130 (99%)	121 (94%)	8 (6%)	15	7
1	L	125/130 (96%)	117 (94%)	8 (6%)	14	7
All	All	1528/1560 (98%)	1427 (93%)	101 (7%)	14	7

5 of 101 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	47	ARG
1	I	35	LEU
1	L	98	LEU
1	G	52	LEU
1	H	114	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 84 such sidechains are listed below:

Mol	Chain	Res	Type
1	I	81	ASN
1	K	32	GLN
1	I	105	ASN
1	J	48	ASN
1	K	105	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 oligosaccharides 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	170/177 (96%)	0.64	17 (10%)	14	14	9, 26, 59, 73	0
1	B	170/177 (96%)	0.80	24 (14%)	7	7	10, 30, 56, 71	0
1	C	166/177 (93%)	0.64	18 (10%)	12	13	9, 25, 50, 69	0
1	D	170/177 (96%)	0.48	16 (9%)	15	16	7, 24, 54, 67	0
1	E	170/177 (96%)	1.25	48 (28%)	1	1	10, 35, 63, 73	0
1	F	167/177 (94%)	0.51	23 (13%)	8	8	7, 24, 53, 69	0
1	G	165/177 (93%)	0.83	26 (15%)	6	5	11, 29, 55, 72	0
1	H	169/177 (95%)	0.71	25 (14%)	7	7	9, 27, 56, 75	0
1	I	171/177 (96%)	1.16	46 (26%)	2	1	11, 34, 59, 74	0
1	J	170/177 (96%)	0.88	29 (17%)	5	4	10, 31, 61, 73	0
1	K	171/177 (96%)	0.86	29 (16%)	5	5	11, 33, 59, 76	0
1	L	165/177 (93%)	1.35	44 (26%)	2	1	14, 33, 59, 77	0
All	All	2024/2124 (95%)	0.84	345 (17%)	5	5	7, 30, 59, 77	0

The worst 5 of 345 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	J	108	ALA	10.3
1	J	25	THR	6.1
1	F	3	ILE	6.0
1	L	109	ASN	5.8
1	H	108	ALA	5.6

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.