



wwPDB X-ray Structure Validation Summary Report

Aug 15, 2023 – 07:42 AM EDT

PDB ID : 1TO0
Title : X-ray structure of Northeast Structural Genomics target protein sr145 from *Bacillus subtilis*
Authors : Kuzin, A.P.; Edstrom, W.; Vorobiev, S.M.; Shastry, R.; Ma, L.-C.; Xiao, R.; Acton, T.; Montelione, G.T.; Tong, L.; Hunt, J.F.; Northeast Structural Genomics Consortium (NESG)
Deposited on : 2004-06-11
Resolution : 2.50 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  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](#) ) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.35
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) : 2.35

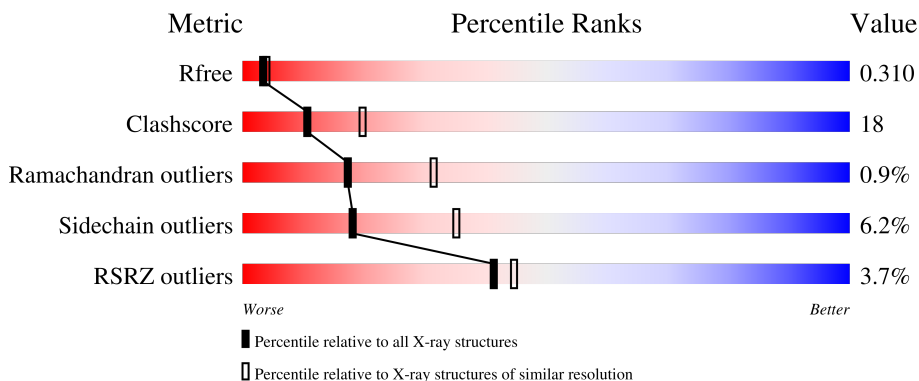
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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.





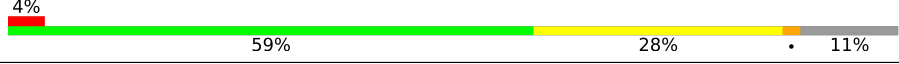

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 $\leq 5\%$. 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	167	
1	B	167	
1	C	167	
1	D	167	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	E	167	 <p>%</p> <p>53% 26% 17%</p>
1	F	167	 <p>%</p> <p>56% 28% 15%</p>
1	G	167	 <p>4%</p> <p>%</p> <p>59% 28% 11%</p>
1	H	167	 <p>4%</p> <p>%</p> <p>55% 30% 5% 10%</p>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 9527 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 UPF0247 protein yda.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
1	A	148	1180	757	198	219	6	0	0	0
1	B	148	1179	755	197	221	6	0	0	0
1	C	148	1178	752	199	221	6	0	0	0
1	D	151	1201	768	201	226	6	0	0	0
1	E	139	1105	708	187	205	5	0	0	0
1	F	142	1126	722	191	208	5	0	0	0
1	G	148	1179	755	197	221	6	0	0	0
1	H	150	1197	766	201	224	6	0	0	0

There are 112 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	modified residue	UNP Q45601
A	52	MSE	MET	modified residue	UNP Q45601
A	82	MSE	MET	modified residue	UNP Q45601
A	118	MSE	MET	modified residue	UNP Q45601
A	130	MSE	MET	modified residue	UNP Q45601
A	137	MSE	MET	modified residue	UNP Q45601
A	160	LEU	-	cloning artifact	UNP Q45601
A	161	GLU	-	cloning artifact	UNP Q45601
A	162	HIS	-	expression tag	UNP Q45601
A	163	HIS	-	expression tag	UNP Q45601
A	164	HIS	-	expression tag	UNP Q45601
A	165	HIS	-	expression tag	UNP Q45601
A	166	HIS	-	expression tag	UNP Q45601

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	167	HIS	-	expression tag	UNP Q45601
B	1	MSE	MET	modified residue	UNP Q45601
B	52	MSE	MET	modified residue	UNP Q45601
B	82	MSE	MET	modified residue	UNP Q45601
B	118	MSE	MET	modified residue	UNP Q45601
B	130	MSE	MET	modified residue	UNP Q45601
B	137	MSE	MET	modified residue	UNP Q45601
B	160	LEU	-	cloning artifact	UNP Q45601
B	161	GLU	-	cloning artifact	UNP Q45601
B	162	HIS	-	expression tag	UNP Q45601
B	163	HIS	-	expression tag	UNP Q45601
B	164	HIS	-	expression tag	UNP Q45601
B	165	HIS	-	expression tag	UNP Q45601
B	166	HIS	-	expression tag	UNP Q45601
B	167	HIS	-	expression tag	UNP Q45601
C	1	MSE	MET	modified residue	UNP Q45601
C	52	MSE	MET	modified residue	UNP Q45601
C	82	MSE	MET	modified residue	UNP Q45601
C	118	MSE	MET	modified residue	UNP Q45601
C	130	MSE	MET	modified residue	UNP Q45601
C	137	MSE	MET	modified residue	UNP Q45601
C	160	LEU	-	cloning artifact	UNP Q45601
C	161	GLU	-	cloning artifact	UNP Q45601
C	162	HIS	-	expression tag	UNP Q45601
C	163	HIS	-	expression tag	UNP Q45601
C	164	HIS	-	expression tag	UNP Q45601
C	165	HIS	-	expression tag	UNP Q45601
C	166	HIS	-	expression tag	UNP Q45601
C	167	HIS	-	expression tag	UNP Q45601
D	1	MSE	MET	modified residue	UNP Q45601
D	52	MSE	MET	modified residue	UNP Q45601
D	82	MSE	MET	modified residue	UNP Q45601
D	118	MSE	MET	modified residue	UNP Q45601
D	130	MSE	MET	modified residue	UNP Q45601
D	137	MSE	MET	modified residue	UNP Q45601
D	160	LEU	-	cloning artifact	UNP Q45601
D	161	GLU	-	cloning artifact	UNP Q45601
D	162	HIS	-	expression tag	UNP Q45601
D	163	HIS	-	expression tag	UNP Q45601
D	164	HIS	-	expression tag	UNP Q45601
D	165	HIS	-	expression tag	UNP Q45601
D	166	HIS	-	expression tag	UNP Q45601

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	167	HIS	-	expression tag	UNP Q45601
E	1	MSE	MET	modified residue	UNP Q45601
E	52	MSE	MET	modified residue	UNP Q45601
E	82	MSE	MET	modified residue	UNP Q45601
E	118	MSE	MET	modified residue	UNP Q45601
E	130	MSE	MET	modified residue	UNP Q45601
E	137	MSE	MET	modified residue	UNP Q45601
E	160	LEU	-	cloning artifact	UNP Q45601
E	161	GLU	-	cloning artifact	UNP Q45601
E	162	HIS	-	expression tag	UNP Q45601
E	163	HIS	-	expression tag	UNP Q45601
E	164	HIS	-	expression tag	UNP Q45601
E	165	HIS	-	expression tag	UNP Q45601
E	166	HIS	-	expression tag	UNP Q45601
E	167	HIS	-	expression tag	UNP Q45601
F	1	MSE	MET	modified residue	UNP Q45601
F	52	MSE	MET	modified residue	UNP Q45601
F	82	MSE	MET	modified residue	UNP Q45601
F	118	MSE	MET	modified residue	UNP Q45601
F	130	MSE	MET	modified residue	UNP Q45601
F	137	MSE	MET	modified residue	UNP Q45601
F	160	LEU	-	cloning artifact	UNP Q45601
F	161	GLU	-	cloning artifact	UNP Q45601
F	162	HIS	-	expression tag	UNP Q45601
F	163	HIS	-	expression tag	UNP Q45601
F	164	HIS	-	expression tag	UNP Q45601
F	165	HIS	-	expression tag	UNP Q45601
F	166	HIS	-	expression tag	UNP Q45601
F	167	HIS	-	expression tag	UNP Q45601
G	1	MSE	MET	modified residue	UNP Q45601
G	52	MSE	MET	modified residue	UNP Q45601
G	82	MSE	MET	modified residue	UNP Q45601
G	118	MSE	MET	modified residue	UNP Q45601
G	130	MSE	MET	modified residue	UNP Q45601
G	137	MSE	MET	modified residue	UNP Q45601
G	160	LEU	-	cloning artifact	UNP Q45601
G	161	GLU	-	cloning artifact	UNP Q45601
G	162	HIS	-	expression tag	UNP Q45601
G	163	HIS	-	expression tag	UNP Q45601
G	164	HIS	-	expression tag	UNP Q45601
G	165	HIS	-	expression tag	UNP Q45601
G	166	HIS	-	expression tag	UNP Q45601

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	167	HIS	-	expression tag	UNP Q45601
H	1	MSE	MET	modified residue	UNP Q45601
H	52	MSE	MET	modified residue	UNP Q45601
H	82	MSE	MET	modified residue	UNP Q45601
H	118	MSE	MET	modified residue	UNP Q45601
H	130	MSE	MET	modified residue	UNP Q45601
H	137	MSE	MET	modified residue	UNP Q45601
H	160	LEU	-	cloning artifact	UNP Q45601
H	161	GLU	-	cloning artifact	UNP Q45601
H	162	HIS	-	expression tag	UNP Q45601
H	163	HIS	-	expression tag	UNP Q45601
H	164	HIS	-	expression tag	UNP Q45601
H	165	HIS	-	expression tag	UNP Q45601
H	166	HIS	-	expression tag	UNP Q45601
H	167	HIS	-	expression tag	UNP Q45601

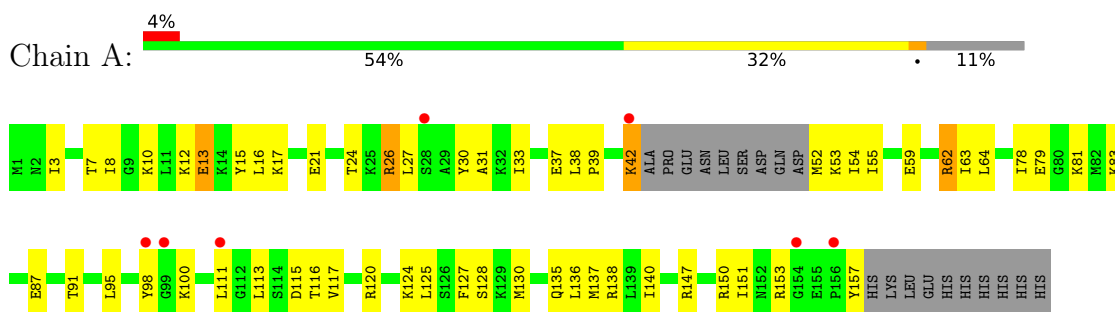
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	13	Total O 13 13	0	0
2	B	28	Total O 28 28	0	0
2	C	34	Total O 34 34	0	0
2	D	9	Total O 9 9	0	0
2	E	21	Total O 21 21	0	0
2	F	27	Total O 27 27	0	0
2	G	25	Total O 25 25	0	0
2	H	25	Total O 25 25	0	0

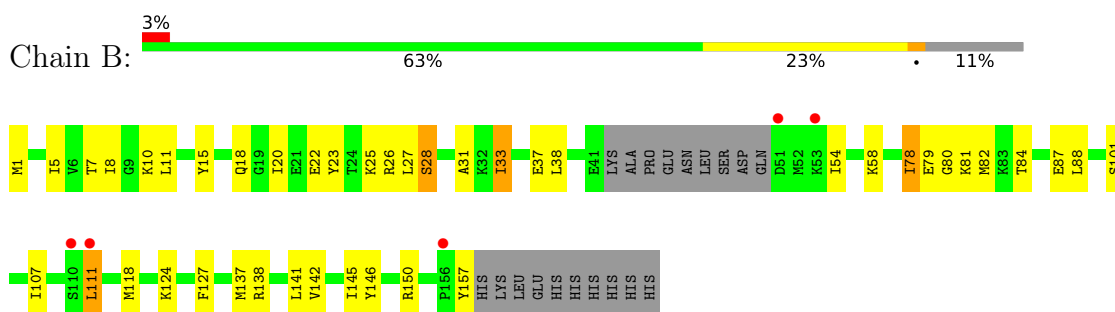
3 Residue-property plots [i](#)

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.

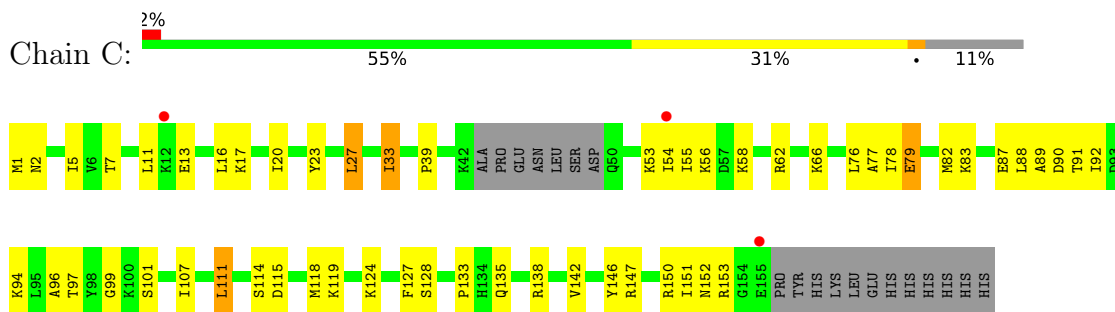
- Molecule 1: Hypothetical UPF0247 protein yda



- Molecule 1: Hypothetical UPF0247 protein yda

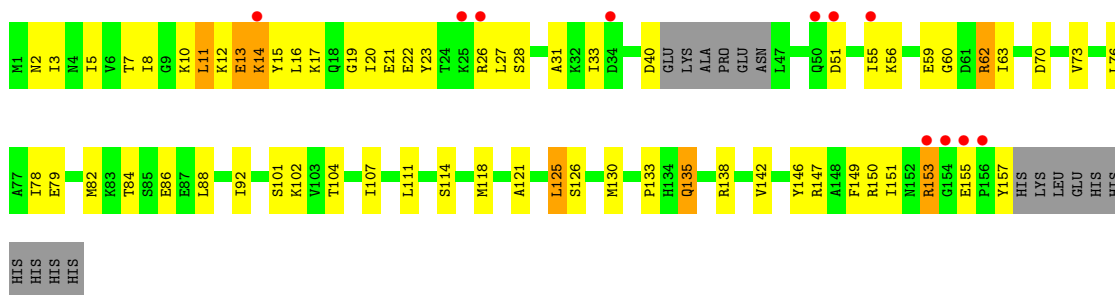


- Molecule 1: Hypothetical UPF0247 protein yda

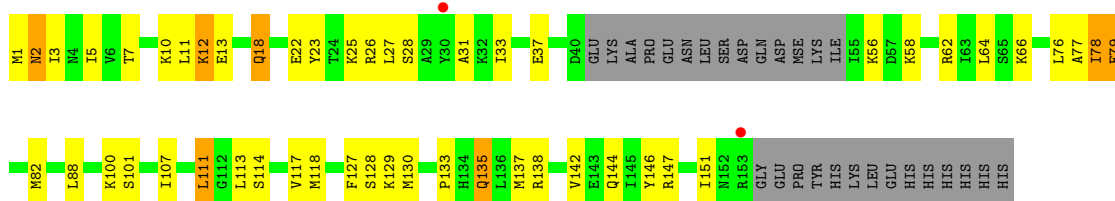


- Molecule 1: Hypothetical UPF0247 protein yda

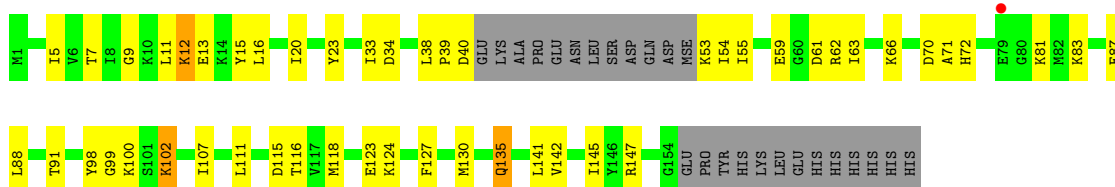




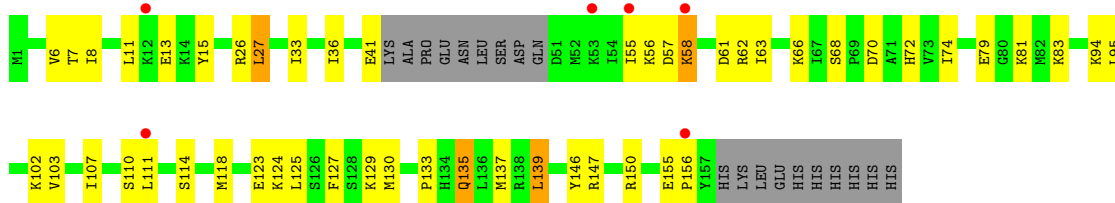
• Molecule 1: Hypothetical UPF0247 protein yda



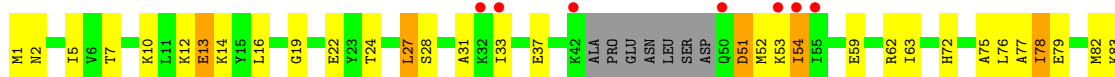
• Molecule 1: Hypothetical UPF0247 protein yda



• Molecule 1: Hypothetical UPF0247 protein yda



• Molecule 1: Hypothetical UPF0247 protein yda





4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	98.69Å 70.55Å 100.49Å 90.00° 95.16° 90.00°	Depositor
Resolution (Å)	20.00 – 2.50 29.72 – 2.31	Depositor EDS
% Data completeness (in resolution range)	84.6 (20.00-2.50) 94.6 (29.72-2.31)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	0.10	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.31 (at 2.31Å)	Xtrriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.237 , 0.290 0.255 , 0.310	Depositor DCC
R_{free} test set	5508 reflections (4.85%)	wwPDB-VP
Wilson B-factor (Å ²)	30.5	Xtrriage
Anisotropy	0.709	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 48.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.000 for l,-k,h	Xtrriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	9527	wwPDB-VP
Average B, all atoms (Å ²)	42.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 41.74 % 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 2.2580e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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	0.35	0/1190	0.58	0/1583
1	B	0.40	0/1188	0.63	0/1580
1	C	0.39	0/1185	0.65	0/1573
1	D	0.35	0/1210	0.57	0/1610
1	E	0.37	0/1113	0.61	0/1481
1	F	0.40	0/1134	0.64	0/1508
1	G	0.39	0/1188	0.60	0/1580
1	H	0.35	0/1206	0.61	0/1603
All	All	0.38	0/9414	0.61	0/12518

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	F	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	F	98	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	1180	0	1254	49	0
1	B	1179	0	1245	49	0
1	C	1178	0	1250	48	0
1	D	1201	0	1267	51	0
1	E	1105	0	1177	57	0
1	F	1126	0	1204	42	0
1	G	1179	0	1245	34	0
1	H	1197	0	1266	46	0
2	A	13	0	0	0	0
2	B	28	0	0	1	0
2	C	34	0	0	4	0
2	D	9	0	0	0	0
2	E	21	0	0	2	0
2	F	27	0	0	0	0
2	G	25	0	0	0	0
2	H	25	0	0	2	0
All	All	9527	0	9908	354	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 18.

The worst 5 of 354 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:B:111:LEU:H	1:B:111:LEU:HD12	1.24	1.03
1:B:78:ILE:HD13	1:B:78:ILE:H	1.26	1.00
1:C:111:LEU:HD23	1:C:111:LEU:H	1.29	0.98
1:H:118:MSE:HE3	1:H:124:LYS:HE3	1.46	0.97
1:F:83:LYS:NZ	1:F:91:THR:HG21	1.86	0.89

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	144/167 (86%)	137 (95%)	5 (4%)	2 (1%)	11	20
1	B	144/167 (86%)	138 (96%)	6 (4%)	0	100	100
1	C	144/167 (86%)	135 (94%)	8 (6%)	1 (1%)	22	39
1	D	147/167 (88%)	131 (89%)	13 (9%)	3 (2%)	7	12
1	E	135/167 (81%)	129 (96%)	4 (3%)	2 (2%)	10	18
1	F	138/167 (83%)	133 (96%)	5 (4%)	0	100	100
1	G	144/167 (86%)	139 (96%)	5 (4%)	0	100	100
1	H	146/167 (87%)	136 (93%)	8 (6%)	2 (1%)	11	20
All	All	1142/1336 (86%)	1078 (94%)	54 (5%)	10 (1%)	17	31

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	11	LEU
1	D	12	LYS
1	D	14	LYS
1	H	53	LYS
1	E	2	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	130/142 (92%)	125 (96%)	5 (4%)	33	58
1	B	130/142 (92%)	123 (95%)	7 (5%)	22	42
1	C	130/142 (92%)	122 (94%)	8 (6%)	18	35
1	D	133/142 (94%)	128 (96%)	5 (4%)	33	58
1	E	122/142 (86%)	115 (94%)	7 (6%)	20	39
1	F	124/142 (87%)	114 (92%)	10 (8%)	11	23

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	G	130/142 (92%)	119 (92%)	11 (8%)	10	21
1	H	132/142 (93%)	121 (92%)	11 (8%)	11	22
All	All	1031/1136 (91%)	967 (94%)	64 (6%)	18	35

5 of 64 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	H	51	ASP
1	H	88	LEU
1	D	153	ARG
1	D	135	GLN
1	H	93	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	72	HIS
1	F	2	ASN
1	E	135	GLN
1	F	72	HIS
1	D	4	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 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	142/167 (85%)	0.28	7 (4%) 29 31	16, 45, 81, 92	0
1	B	142/167 (85%)	0.15	5 (3%) 44 47	9, 32, 71, 102	0
1	C	142/167 (85%)	0.02	3 (2%) 63 66	11, 31, 75, 89	0
1	D	145/167 (86%)	0.53	11 (7%) 13 14	20, 51, 88, 100	0
1	E	134/167 (80%)	0.07	2 (1%) 73 75	16, 35, 73, 80	0
1	F	137/167 (82%)	0.03	1 (0%) 87 89	9, 31, 71, 84	0
1	G	142/167 (85%)	-0.10	6 (4%) 36 39	12, 32, 69, 100	0
1	H	144/167 (86%)	0.20	7 (4%) 29 31	14, 40, 77, 106	0
All	All	1128/1336 (84%)	0.15	42 (3%) 41 45	9, 36, 78, 106	0

The worst 5 of 42 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	55	ILE	4.9
1	D	156	PRO	4.5
1	B	53	LYS	4.5
1	A	98	TYR	4.0
1	G	12	LYS	3.7

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

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.