wwPDB X-ray Structure Validation Summary Report (i)

May 13, 2020 – 09:27 pm BST

PDB ID	:	5D50
Title	:	Crystal structure of Rep-Ant complex from Salmonella-temperate phage
Authors	:	Son, S.H.; Yoon, H.J.; Ryu, S.; Lee, H.H.
Deposited on	:	2015-08-10
Resolution	:	2.49 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:

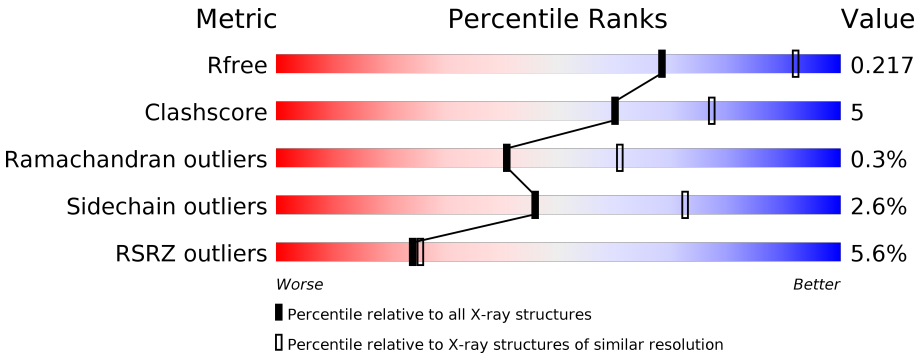
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.49 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.



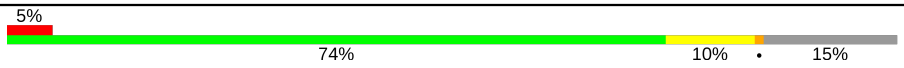
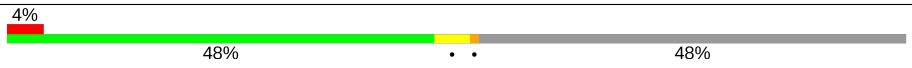
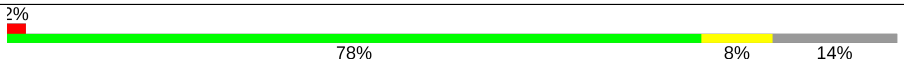
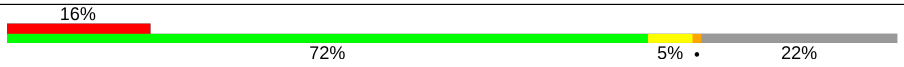
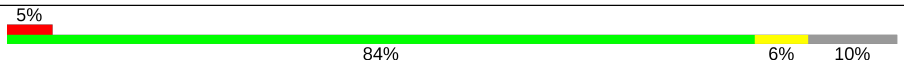
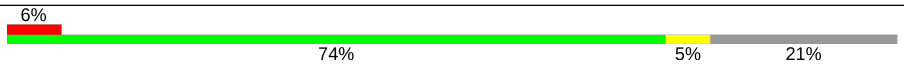
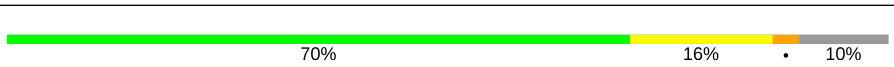
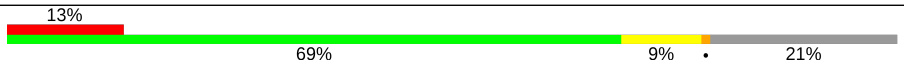
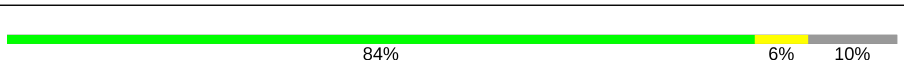
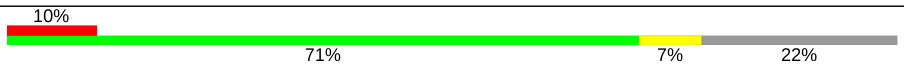
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 on 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	199	4% 72% 13% • 14%
1	B	199	2% 42% 10% • 47%
1	C	199	% 73% 11% • 14%
1	D	199	2% 45% 7% 49%
1	I	199	5% 76% 10% 15%
1	J	199	4% 48% • 48%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	K	199	
1	L	199	
2	E	86	
2	F	86	
2	G	86	
2	H	86	
2	M	86	
2	N	86	
2	O	86	
2	P	86	

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 13325 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 Repressor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	171	Total	C	N	O	S	0	0	0
			1375	872	248	249	6			
1	B	105	Total	C	N	O	S	0	0	0
			816	521	139	152	4			
1	C	171	Total	C	N	O	S	0	0	0
			1375	872	248	249	6			
1	D	102	Total	C	N	O	S	0	0	0
			791	506	132	149	4			
1	I	170	Total	C	N	O	S	0	0	0
			1370	869	247	248	6			
1	J	103	Total	C	N	O	S	0	0	0
			806	515	137	150	4			
1	K	169	Total	C	N	O	S	0	0	0
			1361	863	245	247	6			
1	L	104	Total	C	N	O	S	0	0	0
			811	518	138	151	4			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	ALA	-	expression tag	UNP T1S9Z0
B	0	ALA	-	expression tag	UNP T1S9Z0
C	0	ALA	-	expression tag	UNP T1S9Z0
D	0	ALA	-	expression tag	UNP T1S9Z0
I	0	ALA	-	expression tag	UNP T1S9Z0
J	0	ALA	-	expression tag	UNP T1S9Z0
K	0	ALA	-	expression tag	UNP T1S9Z0
L	0	ALA	-	expression tag	UNP T1S9Z0

- Molecule 2 is a protein called Anti-repressor protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	74	Total	C	N	O	S	0	0	0
			592	369	102	117	4			
2	G	77	Total	C	N	O	S	0	0	0
			614	381	105	124	4			
2	M	77	Total	C	N	O	S	0	0	0
			614	381	105	124	4			
2	O	77	Total	C	N	O	S	0	0	0
			614	381	105	124	4			
2	F	67	Total	C	N	O	S	0	0	0
			527	328	90	105	4			
2	H	68	Total	C	N	O	S	0	0	0
			534	332	91	107	4			
2	N	68	Total	C	N	O	S	0	0	0
			534	332	91	107	4			
2	P	67	Total	C	N	O	S	0	0	0
			527	328	90	105	4			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	4	Total	O	0	0
			4	4		
3	B	5	Total	O	0	0
			5	5		
3	C	7	Total	O	0	0
			7	7		
3	D	4	Total	O	0	0
			4	4		
3	E	4	Total	O	0	0
			4	4		
3	G	3	Total	O	0	0
			3	3		
3	I	6	Total	O	0	0
			6	6		
3	J	4	Total	O	0	0
			4	4		
3	K	6	Total	O	0	0
			6	6		
3	L	4	Total	O	0	0
			4	4		
3	M	2	Total	O	0	0
			2	2		
3	O	2	Total	O	0	0
			2	2		

Continued on next page...

Continued from previous page...

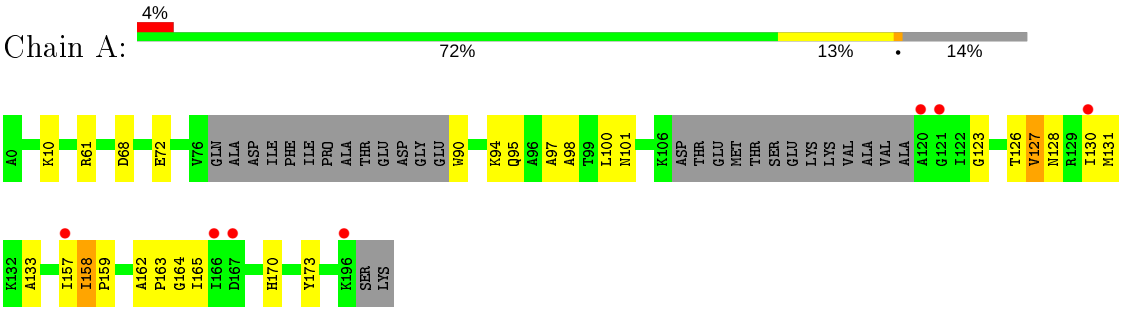
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	F	2	Total O 2 2	0	0
3	H	1	Total O 1 1	0	0
3	N	5	Total O 5 5	0	0
3	P	5	Total O 5 5	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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.

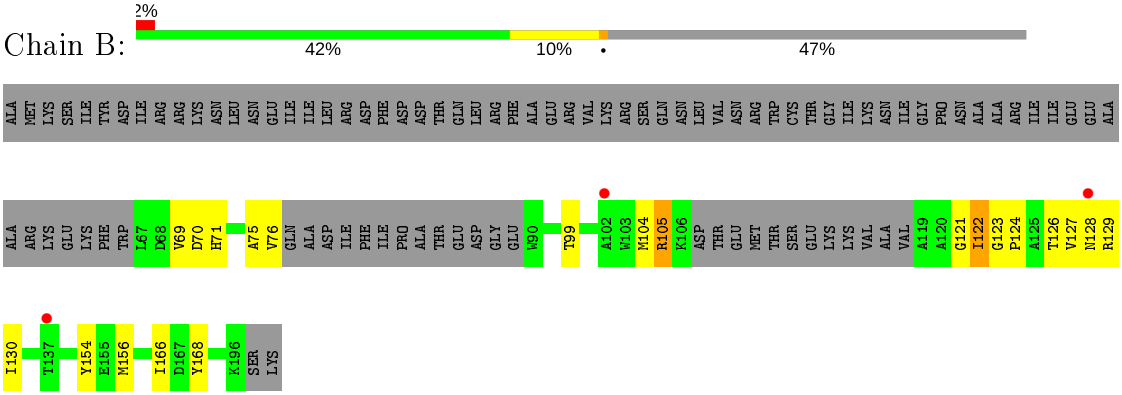
- Molecule 1: Repressor

Chain A:



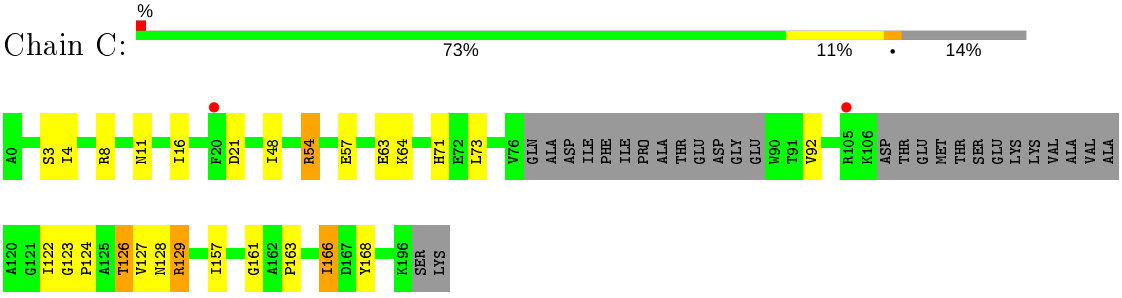
- Molecule 1: Repressor

Chain B:



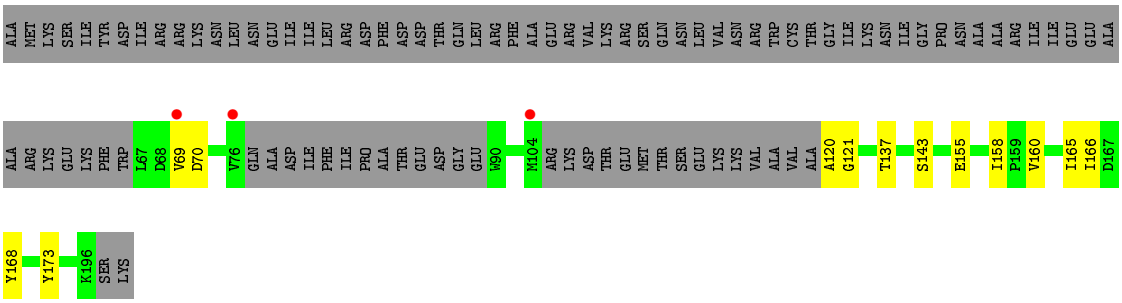
- Molecule 1: Repressor

Chain C:



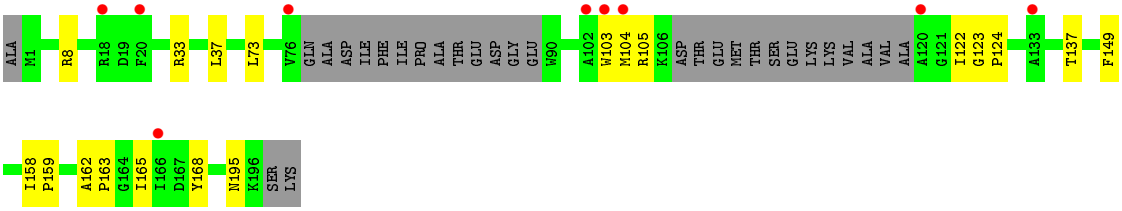
- Molecule 1: Repressor

Chain D:



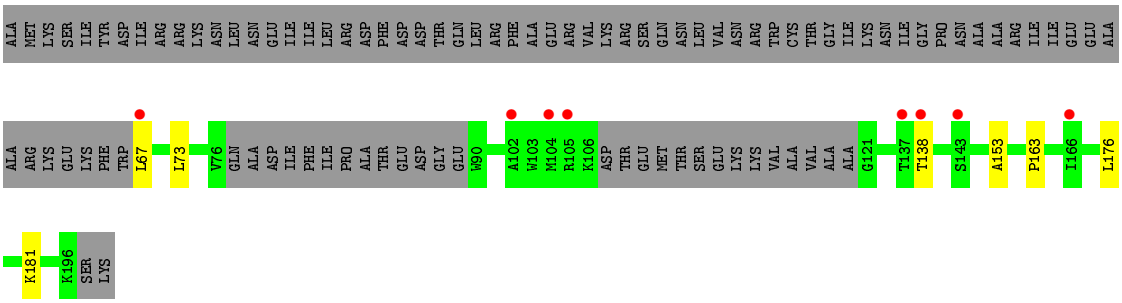
• Molecule 1: Repressor

Chain I:



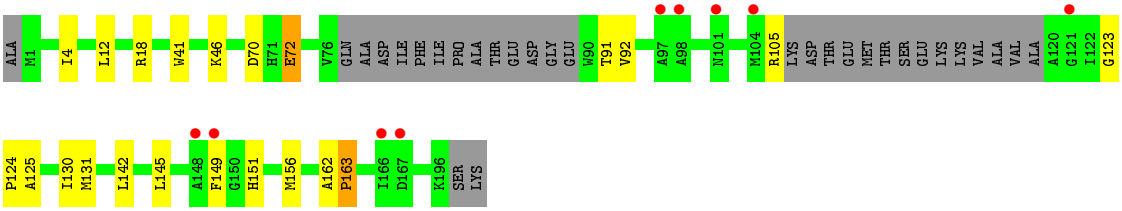
• Molecule 1: Repressor

Chain J:



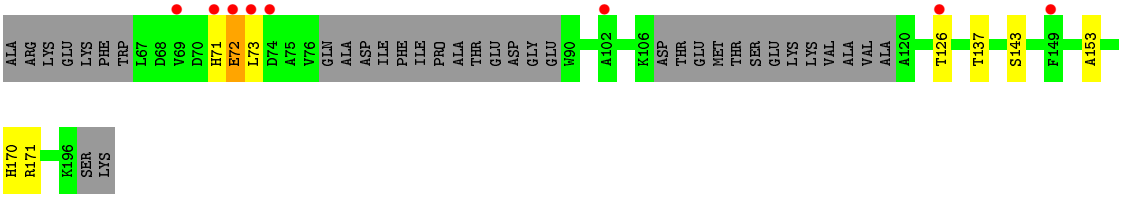
• Molecule 1: Repressor

Chain K:



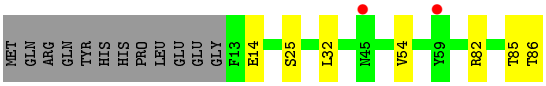
• Molecule 1: Repressor

Chain L:



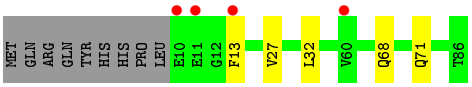
- Molecule 2: Anti-repressor protein

Chain E:



- Molecule 2: Anti-repressor protein

Chain G:



- Molecule 2: Anti-repressor protein

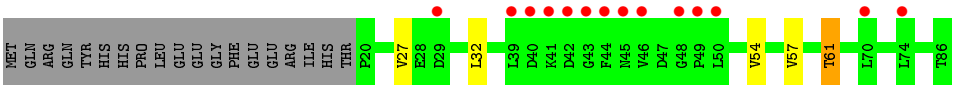
Chain M:

- Molecule 2: Anti-repressor protein

Chain O:

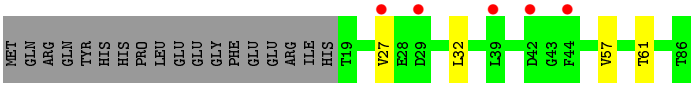
- Molecule 2: Anti-repressor protein

Chain F:



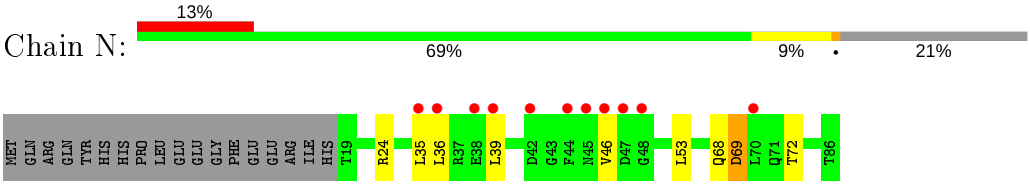
- Molecule 2: Anti-repressor protein

Chain H:



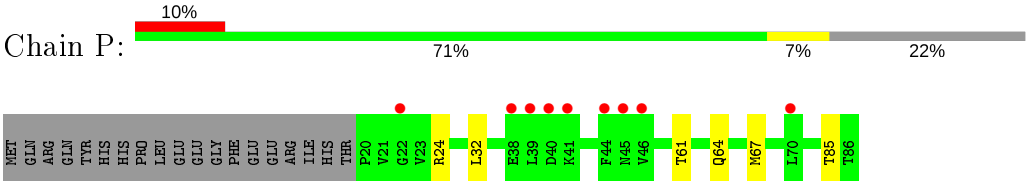
• Molecule 2: Anti-repressor protein

Chain N:



• Molecule 2: Anti-repressor protein

Chain P:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32	Depositor
Cell constants a, b, c, α, β, γ	86.61Å 86.61Å 337.03Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	36.61 – 2.49 36.61 – 2.49	Depositor EDS
% Data completeness (in resolution range)	98.2 (36.61-2.49) 98.2 (36.61-2.49)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	54.36 (at 2.48Å)	Xtriage
Refinement program	REFMAC 5.8.0073	Depositor
R, R_{free}	0.201 , 0.210 0.205 , 0.217	Depositor DCC
R_{free} test set	4785 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å^2)	58.3	Xtriage
Anisotropy	0.005	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 63.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.103 for -h,-k,l 0.119 for h,-h-k,-l 0.117 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	13325	wwPDB-VP
Average B, all atoms (Å^2)	65.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.42% 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

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	0.31	0/1399	0.54	0/1885
1	B	0.36	0/831	0.55	0/1124
1	C	0.31	0/1399	0.53	0/1885
1	D	0.33	0/806	0.50	0/1092
1	I	0.32	0/1394	0.55	0/1878
1	J	0.33	0/821	0.55	0/1110
1	K	0.32	0/1385	0.55	0/1867
1	L	0.34	0/826	0.53	0/1117
2	E	0.32	0/600	0.49	0/811
2	F	0.28	0/533	0.47	0/720
2	G	0.31	0/622	0.49	0/840
2	H	0.30	0/540	0.49	0/731
2	M	0.31	0/622	0.53	0/840
2	N	0.30	0/540	0.50	0/731
2	O	0.32	0/622	0.48	0/840
2	P	0.28	0/533	0.46	0/720
All	All	0.32	0/13473	0.52	0/18191

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	L	0	1
2	M	0	1
All	All	0	2

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	L	73	LEU	Peptide
2	M	69	ASP	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	1375	0	1387	26	0
1	B	816	0	809	23	0
1	C	1375	0	1387	17	0
1	D	791	0	778	9	0
1	I	1370	0	1382	10	0
1	J	806	0	799	3	0
1	K	1361	0	1369	13	0
1	L	811	0	804	6	0
2	E	592	0	586	5	0
2	F	527	0	528	9	0
2	G	614	0	601	3	0
2	H	534	0	534	4	0
2	M	614	0	601	13	0
2	N	534	0	534	5	0
2	O	614	0	601	3	0
2	P	527	0	528	6	0
3	A	4	0	0	0	0
3	B	5	0	0	0	0
3	C	7	0	0	0	0
3	D	4	0	0	0	0
3	E	4	0	0	0	0
3	F	2	0	0	0	0
3	G	3	0	0	0	0
3	H	1	0	0	0	0
3	I	6	0	0	0	0
3	J	4	0	0	0	0
3	K	6	0	0	0	0
3	L	4	0	0	0	0
3	M	2	0	0	0	0
3	N	5	0	0	0	0
3	O	2	0	0	0	0
3	P	5	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	13325	0	13228	122	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 122 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:123:GLY:HA2	1:A:127:VAL:HG11	1.50	0.93
1:B:71:HIS:ND1	2:F:61:THR:OG1	2.16	0.75
1:A:127:VAL:HG13	1:A:128:ASN:H	1.55	0.69
1:C:92:VAL:HG21	1:D:143:SER:HB2	1.75	0.69
1:C:73:LEU:HD12	2:G:27:VAL:HG21	1.76	0.68

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	165/199 (83%)	158 (96%)	6 (4%)	1 (1%)	25 43
1	B	99/199 (50%)	92 (93%)	5 (5%)	2 (2%)	7 12
1	C	165/199 (83%)	155 (94%)	9 (6%)	1 (1%)	25 43
1	D	96/199 (48%)	85 (88%)	11 (12%)	0	100 100
1	I	164/199 (82%)	154 (94%)	10 (6%)	0	100 100
1	J	97/199 (49%)	91 (94%)	5 (5%)	1 (1%)	15 28
1	K	163/199 (82%)	149 (91%)	14 (9%)	0	100 100
1	L	98/199 (49%)	89 (91%)	9 (9%)	0	100 100
2	E	72/86 (84%)	69 (96%)	3 (4%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	F	65/86 (76%)	63 (97%)	2 (3%)	0	100	100
2	G	75/86 (87%)	73 (97%)	2 (3%)	0	100	100
2	H	66/86 (77%)	64 (97%)	2 (3%)	0	100	100
2	M	75/86 (87%)	71 (95%)	4 (5%)	0	100	100
2	N	66/86 (77%)	66 (100%)	0	0	100	100
2	O	75/86 (87%)	73 (97%)	2 (3%)	0	100	100
2	P	65/86 (76%)	61 (94%)	4 (6%)	0	100	100
All	All	1606/2280 (70%)	1513 (94%)	88 (6%)	5 (0%)	41	61

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	128	ASN
1	A	127	VAL
1	J	153	ALA
1	B	124	PRO
1	C	166	ILE

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	143/166 (86%)	139 (97%)	4 (3%)	43	70
1	B	84/166 (51%)	82 (98%)	2 (2%)	49	74
1	C	143/166 (86%)	139 (97%)	4 (3%)	43	70
1	D	82/166 (49%)	81 (99%)	1 (1%)	71	88
1	I	143/166 (86%)	139 (97%)	4 (3%)	43	70
1	J	84/166 (51%)	82 (98%)	2 (2%)	49	74
1	K	142/166 (86%)	137 (96%)	5 (4%)	36	62
1	L	84/166 (51%)	80 (95%)	4 (5%)	25	48

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	E	68/79 (86%)	67 (98%)	1 (2%)	65 85
2	F	61/79 (77%)	60 (98%)	1 (2%)	62 84
2	G	70/79 (89%)	69 (99%)	1 (1%)	67 86
2	H	62/79 (78%)	62 (100%)	0	100 100
2	M	70/79 (89%)	67 (96%)	3 (4%)	29 53
2	N	62/79 (78%)	60 (97%)	2 (3%)	39 65
2	O	70/79 (89%)	68 (97%)	2 (3%)	42 69
2	P	61/79 (77%)	60 (98%)	1 (2%)	62 84
All	All	1429/1960 (73%)	1392 (97%)	37 (3%)	46 72

5 of 37 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	I	195	ASN
1	K	46	LYS
2	N	46	VAL
1	J	67	LEU
1	J	138	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	K	151	HIS
2	P	58	ASN
1	K	195	ASN
2	G	71	GLN
2	M	18	HIS

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

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	171/199 (85%)	0.18	7 (4%) 37 40	40, 63, 100, 112	0
1	B	105/199 (52%)	0.17	3 (2%) 51 55	35, 65, 102, 110	0
1	C	171/199 (85%)	0.02	2 (1%) 79 80	45, 61, 89, 108	0
1	D	102/199 (51%)	0.07	3 (2%) 51 55	39, 67, 89, 108	0
1	I	170/199 (85%)	0.23	9 (5%) 26 28	41, 62, 94, 124	0
1	J	103/199 (51%)	0.13	8 (7%) 13 13	40, 67, 90, 98	0
1	K	169/199 (84%)	0.24	9 (5%) 26 28	42, 63, 101, 117	0
1	L	104/199 (52%)	0.32	8 (7%) 13 13	37, 69, 111, 140	0
2	E	74/86 (86%)	0.33	2 (2%) 54 58	36, 57, 79, 93	0
2	F	67/86 (77%)	0.92	14 (20%) 1 0	38, 66, 103, 111	0
2	G	77/86 (89%)	0.45	4 (5%) 27 29	36, 61, 99, 119	0
2	H	68/86 (79%)	0.49	5 (7%) 14 15	40, 66, 94, 110	0
2	M	77/86 (89%)	0.28	0 100 100	40, 57, 88, 109	0
2	N	68/86 (79%)	0.82	11 (16%) 1 1	43, 66, 93, 97	0
2	O	77/86 (89%)	0.23	0 100 100	40, 56, 82, 98	0
2	P	67/86 (77%)	0.69	9 (13%) 3 2	43, 63, 88, 92	0
All	All	1670/2280 (73%)	0.29	94 (5%) 24 25	35, 63, 97, 140	0

The worst 5 of 94 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	K	149	PHE	8.3
1	A	166	ILE	6.8
2	F	46	VAL	6.4
1	B	102	ALA	6.3
1	I	102	ALA	5.8

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.