wwPDB X-ray Structure Validation Summary Report (i)

Oct 23, 2024 – 08:49 AM EDT

PDB ID	:	4LRV
Title	:	Crystal structure of DndE from Escherichia coli B7A involved in DNA phosphorothioation modification
Authors	:	Hu, W.; Wang, C.K.; Liang, J.D.; Zhang, T.L.; Yang, M.; Hu, Z.P.; Wang, Z.J.; Lan, W.X.; Wu, H.M.; Ding, J.P.; Wu, G.; Deng, Z.X.; Cao, C.
Deposited on	:	2013-07-21
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 (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 (1)) were used in the production of this report:

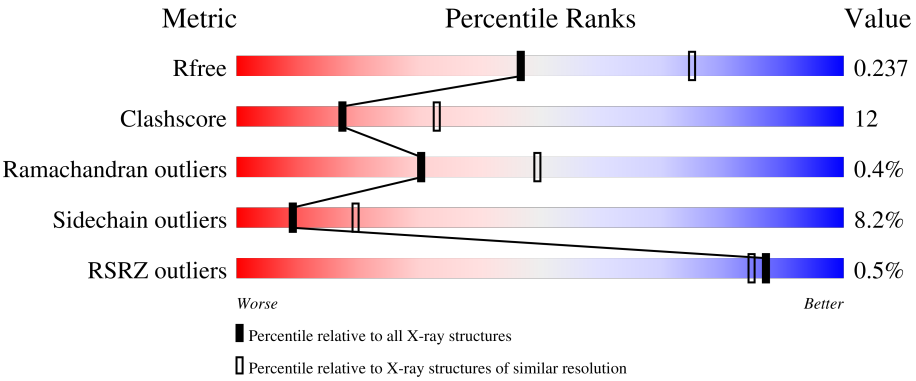
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (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 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.



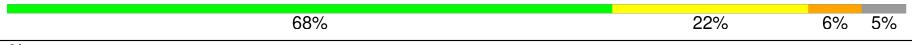
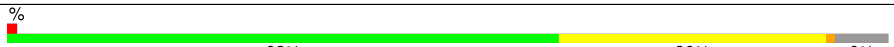
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	5504 (2.50-2.50)
Clashscore	180529	6282 (2.50-2.50)
Ramachandran outliers	177936	6191 (2.50-2.50)
Sidechain outliers	177891	6193 (2.50-2.50)
RSRZ outliers	164620	5504 (2.50-2.50)

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 <=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	109	68% 28%
1	B	109	76% 19%
1	C	109	65% 27% 5%
1	D	109	75% 21%
1	E	109	76% 16% 6%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	109	
1	G	109	
1	H	109	
1	I	109	
1	J	109	
1	K	109	
1	L	109	
1	M	109	
1	N	109	
1	O	109	
1	P	109	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 13759 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 DNA sulfur modification protein DndE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	107	Total	C	N	O	Se	0	0	0
			854	545	149	157	3			
1	B	106	Total	C	N	O	Se	0	0	0
			848	542	148	155	3			
1	C	105	Total	C	N	O	Se	0	0	0
			840	538	146	153	3			
1	D	107	Total	C	N	O	Se	0	0	0
			854	545	149	157	3			
1	E	102	Total	C	N	O	Se	0	0	0
			815	525	140	147	3			
1	F	104	Total	C	N	O	Se	0	0	0
			829	532	142	152	3			
1	G	106	Total	C	N	O	Se	0	0	0
			846	541	147	155	3			
1	H	106	Total	C	N	O	Se	0	0	0
			846	541	147	155	3			
1	I	102	Total	C	N	O	Se	0	0	0
			812	522	139	148	3			
1	J	106	Total	C	N	O	Se	0	0	0
			848	542	148	155	3			
1	K	104	Total	C	N	O	Se	0	0	0
			833	535	145	150	3			
1	L	100	Total	C	N	O	Se	0	0	0
			797	514	137	143	3			
1	M	106	Total	C	N	O	Se	0	0	0
			848	544	148	153	3			
1	N	104	Total	C	N	O	Se	0	0	0
			828	531	143	151	3			
1	O	106	Total	C	N	O	Se	0	0	0
			846	541	147	155	3			
1	P	104	Total	C	N	O	Se	0	0	0
			831	532	145	151	3			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2	SER	-	expression tag	UNP B3HI59
B	2	SER	-	expression tag	UNP B3HI59
C	2	SER	-	expression tag	UNP B3HI59
D	2	SER	-	expression tag	UNP B3HI59
E	2	SER	-	expression tag	UNP B3HI59
F	2	SER	-	expression tag	UNP B3HI59
G	2	SER	-	expression tag	UNP B3HI59
H	2	SER	-	expression tag	UNP B3HI59
I	2	SER	-	expression tag	UNP B3HI59
J	2	SER	-	expression tag	UNP B3HI59
K	2	SER	-	expression tag	UNP B3HI59
L	2	SER	-	expression tag	UNP B3HI59
M	2	SER	-	expression tag	UNP B3HI59
N	2	SER	-	expression tag	UNP B3HI59
O	2	SER	-	expression tag	UNP B3HI59
P	2	SER	-	expression tag	UNP B3HI59

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	32	Total O 32 32	0	0
2	B	34	Total O 34 34	0	0
2	C	36	Total O 36 36	0	0
2	D	13	Total O 13 13	0	0
2	E	13	Total O 13 13	0	0
2	F	27	Total O 27 27	0	0
2	G	35	Total O 35 35	0	0
2	H	23	Total O 23 23	0	0
2	I	16	Total O 16 16	0	0
2	J	30	Total O 30 30	0	0
2	K	12	Total O 12 12	0	0

Continued on next page...

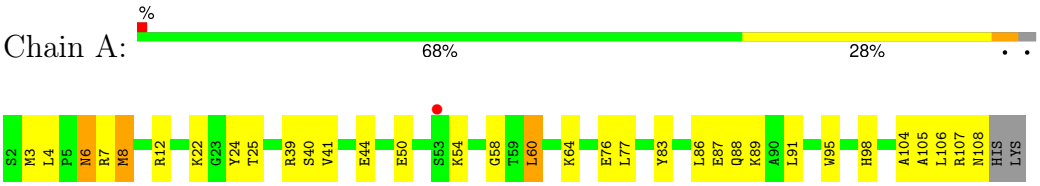
Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	L	11	Total O 11 11	0	0
2	M	27	Total O 27 27	0	0
2	N	29	Total O 29 29	0	0
2	O	33	Total O 33 33	0	0
2	P	13	Total O 13 13	0	0

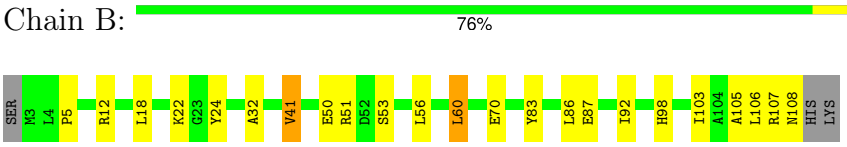
3 Residue-property plots (i)

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.

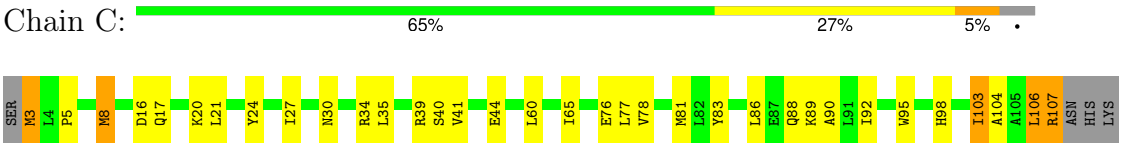
• Molecule 1: DNA sulfur modification protein DndE



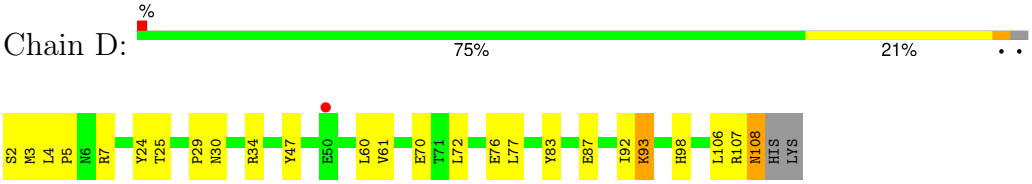
• Molecule 1: DNA sulfur modification protein DndE



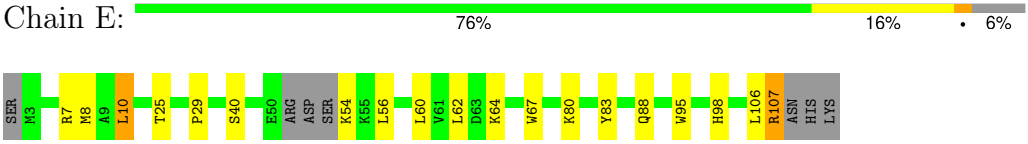
• Molecule 1: DNA sulfur modification protein DndE



• Molecule 1: DNA sulfur modification protein DndE

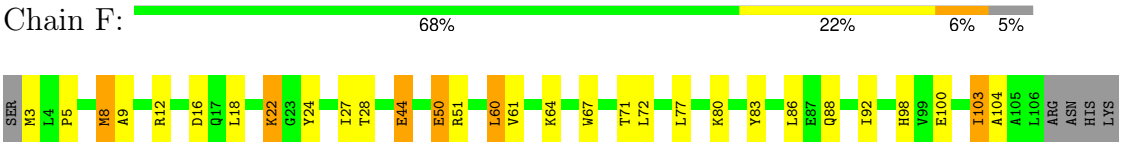


• Molecule 1: DNA sulfur modification protein DndE



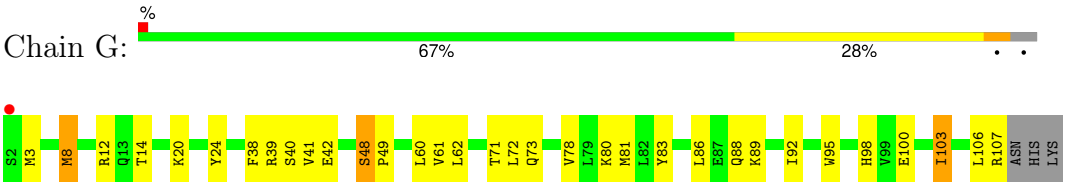
• Molecule 1: DNA sulfur modification protein DndE

Chain F:



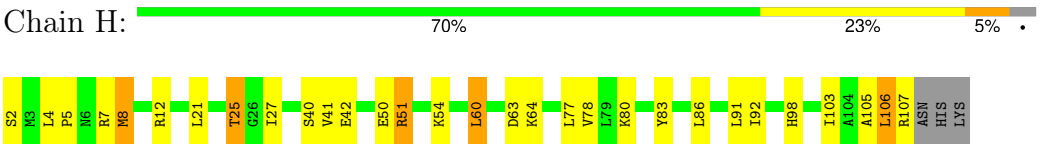
• Molecule 1: DNA sulfur modification protein DndE

Chain G:



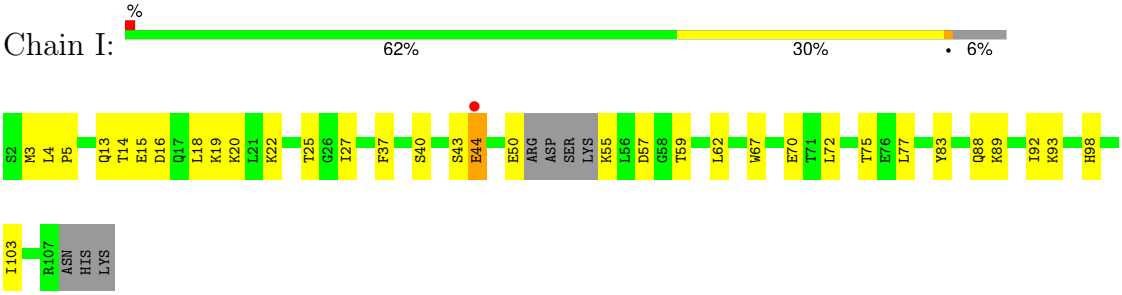
• Molecule 1: DNA sulfur modification protein DndE

Chain H:



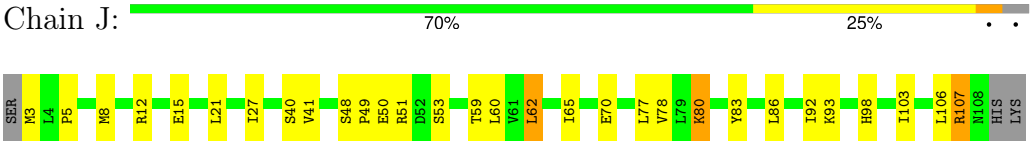
• Molecule 1: DNA sulfur modification protein DndE

Chain I:



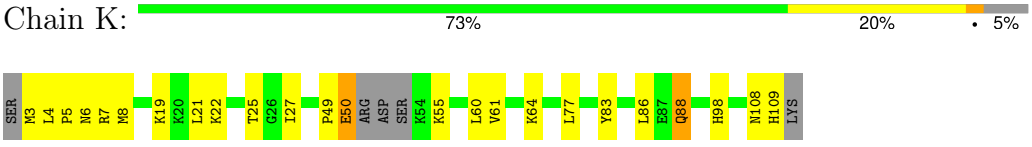
• Molecule 1: DNA sulfur modification protein DndE

Chain J:



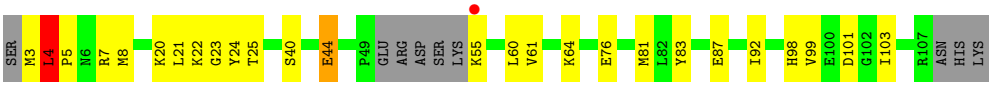
• Molecule 1: DNA sulfur modification protein DndE

Chain K:



• Molecule 1: DNA sulfur modification protein DndE

Chain L:



• Molecule 1: DNA sulfur modification protein DndE

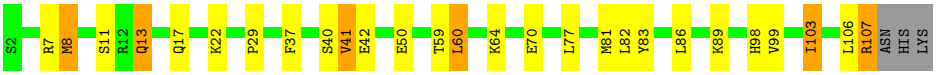
Chain M:

• Molecule 1: DNA sulfur modification protein DndE

Chain N:

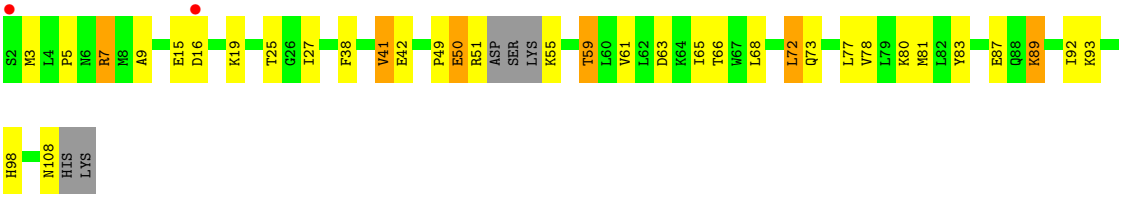
• Molecule 1: DNA sulfur modification protein DndE

Chain O:



• Molecule 1: DNA sulfur modification protein DndE

Chain P:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	97.84Å 97.31Å 218.53Å 90.00° 95.81° 90.00°	Depositor
Resolution (Å)	50.00 – 2.50 50.00 – 2.50	Depositor EDS
% Data completeness (in resolution range)	88.3 (50.00-2.50) 88.3 (50.00-2.50)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	8.54 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.201 , 0.242 0.198 , 0.237	Depositor DCC
R_{free} test set	3139 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	24.9	Xtriage
Anisotropy	0.147	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 37.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	13759	wwPDB-VP
Average B, all atoms (Å^2)	24.0	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 30.81 % 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 1.2296e-03. 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

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	0.44	0/865	0.59	0/1161
1	B	0.42	0/860	0.59	0/1156
1	C	0.43	0/852	0.59	0/1145
1	D	0.40	0/865	0.56	0/1161
1	E	0.42	0/826	0.58	0/1109
1	F	0.44	0/841	0.57	0/1131
1	G	0.43	0/857	0.57	0/1150
1	H	0.41	0/857	0.56	0/1150
1	I	0.39	0/822	0.55	0/1103
1	J	0.40	0/860	0.54	0/1156
1	K	0.41	0/845	0.58	0/1135
1	L	0.39	0/808	0.58	1/1086 (0.1%)
1	M	0.41	0/860	0.55	0/1154
1	N	0.41	0/839	0.55	0/1127
1	O	0.40	0/857	0.55	0/1150
1	P	0.40	0/841	0.58	0/1128
All	All	0.41	0/13555	0.57	1/18202 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	L	4	LEU	CA-CB-CG	5.54	128.04	115.30

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	854	0	885	32	0
1	B	848	0	880	14	0
1	C	840	0	874	25	0
1	D	854	0	885	23	0
1	E	815	0	851	14	0
1	F	829	0	861	29	0
1	G	846	0	879	28	0
1	H	846	0	879	28	0
1	I	812	0	843	27	0
1	J	848	0	880	23	0
1	K	833	0	864	18	0
1	L	797	0	832	27	0
1	M	848	0	882	28	0
1	N	828	0	860	12	0
1	O	846	0	879	23	0
1	P	831	0	862	30	0
2	A	32	0	0	3	0
2	B	34	0	0	0	0
2	C	36	0	0	0	0
2	D	13	0	0	0	0
2	E	13	0	0	0	0
2	F	27	0	0	0	0
2	G	35	0	0	1	0
2	H	23	0	0	1	0
2	I	16	0	0	1	0
2	J	30	0	0	2	0
2	K	12	0	0	0	0
2	L	11	0	0	2	0
2	M	27	0	0	1	0
2	N	29	0	0	0	0
2	O	33	0	0	1	0
2	P	13	0	0	0	0
All	All	13759	0	13896	336	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 12.

The worst 5 of 336 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:I:5:PRO:CD	1:I:92:ILE:HD11	1.71	1.21
1:M:83:TYR:HB3	1:M:86:LEU:HD13	1.32	1.11
1:H:5:PRO:HD2	1:H:92:ILE:HD11	1.29	1.08
1:I:5:PRO:HD2	1:I:92:ILE:HD11	1.08	1.07
1:I:5:PRO:HD2	1:I:92:ILE:CD1	1.90	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	105/109 (96%)	100 (95%)	4 (4%)	1 (1%)	13 25
1	B	104/109 (95%)	99 (95%)	4 (4%)	1 (1%)	13 25
1	C	103/109 (94%)	100 (97%)	3 (3%)	0	100 100
1	D	105/109 (96%)	98 (93%)	7 (7%)	0	100 100
1	E	98/109 (90%)	95 (97%)	3 (3%)	0	100 100
1	F	102/109 (94%)	97 (95%)	4 (4%)	1 (1%)	13 25
1	G	104/109 (95%)	100 (96%)	4 (4%)	0	100 100
1	H	104/109 (95%)	99 (95%)	4 (4%)	1 (1%)	13 25
1	I	98/109 (90%)	94 (96%)	3 (3%)	1 (1%)	13 25
1	J	104/109 (95%)	100 (96%)	4 (4%)	0	100 100
1	K	100/109 (92%)	96 (96%)	3 (3%)	1 (1%)	13 25
1	L	96/109 (88%)	91 (95%)	5 (5%)	0	100 100
1	M	102/109 (94%)	98 (96%)	4 (4%)	0	100 100
1	N	100/109 (92%)	98 (98%)	2 (2%)	0	100 100
1	O	104/109 (95%)	101 (97%)	3 (3%)	0	100 100
1	P	100/109 (92%)	94 (94%)	5 (5%)	1 (1%)	13 25

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1629/1744 (93%)	1560 (96%)	62 (4%)	7 (0%)	30 49

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	K	88	GLN
1	A	105	ALA
1	B	105	ALA
1	F	104	ALA
1	H	106	LEU

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	91/90 (101%)	84 (92%)	7 (8%)	10 22
1	B	90/90 (100%)	80 (89%)	10 (11%)	5 10
1	C	89/90 (99%)	81 (91%)	8 (9%)	8 16
1	D	91/90 (101%)	82 (90%)	9 (10%)	6 13
1	E	86/90 (96%)	80 (93%)	6 (7%)	12 26
1	F	88/90 (98%)	80 (91%)	8 (9%)	7 16
1	G	90/90 (100%)	86 (96%)	4 (4%)	24 47
1	H	90/90 (100%)	82 (91%)	8 (9%)	8 17
1	I	86/90 (96%)	80 (93%)	6 (7%)	12 26
1	J	90/90 (100%)	82 (91%)	8 (9%)	8 17
1	K	88/90 (98%)	82 (93%)	6 (7%)	13 27
1	L	84/90 (93%)	80 (95%)	4 (5%)	21 43
1	M	90/90 (100%)	84 (93%)	6 (7%)	13 28
1	N	88/90 (98%)	79 (90%)	9 (10%)	6 12
1	O	90/90 (100%)	80 (89%)	10 (11%)	5 10
1	P	88/90 (98%)	81 (92%)	7 (8%)	10 20

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1419/1440 (98%)	1303 (92%)	116 (8%)	9 19

5 of 116 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	H	63	ASP
1	P	7	ARG
1	J	80	LYS
1	P	3	MSE
1	O	8	MSE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 33 such sidechains are listed below:

Mol	Chain	Res	Type
1	O	85	GLN
1	O	98	HIS
1	P	108	ASN
1	G	17	GLN
1	F	98	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 (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

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	104/109 (95%)	-0.43	1 (0%) 79 76	11, 17, 29, 38	0
1	B	103/109 (94%)	-0.46	0 100 100	14, 19, 32, 44	0
1	C	102/109 (93%)	-0.41	0 100 100	12, 19, 30, 39	0
1	D	104/109 (95%)	-0.10	1 (0%) 79 76	16, 26, 45, 58	0
1	E	99/109 (90%)	-0.33	0 100 100	13, 22, 39, 49	0
1	F	101/109 (92%)	-0.36	0 100 100	16, 22, 33, 41	0
1	G	103/109 (94%)	-0.34	1 (0%) 79 76	15, 22, 34, 44	0
1	H	103/109 (94%)	-0.32	0 100 100	14, 22, 36, 47	0
1	I	99/109 (90%)	-0.14	1 (1%) 79 76	17, 26, 38, 50	0
1	J	103/109 (94%)	-0.35	0 100 100	16, 22, 34, 40	0
1	K	101/109 (92%)	-0.24	0 100 100	15, 22, 39, 46	0
1	L	97/109 (88%)	-0.05	1 (1%) 79 76	18, 29, 43, 47	0
1	M	103/109 (94%)	-0.41	0 100 100	12, 20, 37, 45	0
1	N	101/109 (92%)	-0.44	1 (0%) 79 76	12, 21, 35, 40	0
1	O	103/109 (94%)	-0.41	0 100 100	14, 20, 35, 43	0
1	P	101/109 (92%)	-0.16	2 (1%) 64 62	16, 27, 44, 53	0
All	All	1627/1744 (93%)	-0.31	8 (0%) 87 85	11, 22, 38, 58	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	L	55	LYS	2.6
1	A	53	SER	2.3
1	P	16	ASP	2.2
1	D	50	GLU	2.1
1	I	44	GLU	2.1

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.