wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 09:47 PM GMT

PDB ID	:	4W78
Title	:	Crystal structure of the ChsH1-ChsH2 complex from Mycobacterium tuberculosis
Authors	:	Guja, K.E.; Yang, M.; Sampson, N.; Garcia-Diaz, M.
Deposited on	:	2014-08-21
Resolution	:	1.54 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:

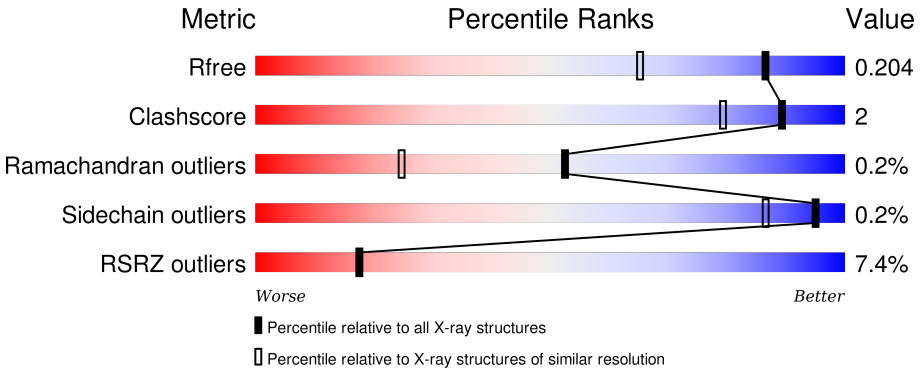
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 1.54 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.



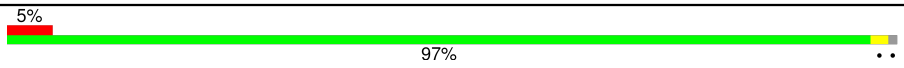
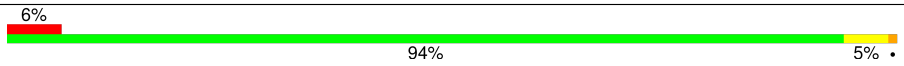
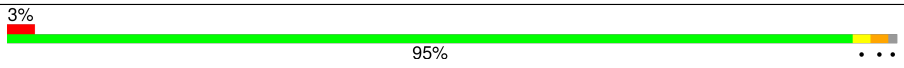
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1555 (1.56-1.52)
Clashscore	102246	1627 (1.56-1.52)
Ramachandran outliers	100387	1594 (1.56-1.52)
Sidechain outliers	100360	1592 (1.56-1.52)
RSRZ outliers	91569	1555 (1.56-1.52)

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	178	7%, 89%, 6%, 5%
1	C	178	10%, 97%, ·
1	E	178	3%, 89%, ·, 7%
1	G	178	18%, 84%, 8%, 8%
2	B	127	2%, 97%, ··

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	D	127	
2	F	127	
2	H	127	

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 10103 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 Hydratase ChsH1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	169	Total	C	N	O	S	0	1	0
			1287	821	223	237	6			
1	C	178	Total	C	N	O	S	0	1	0
			1322	843	225	247	7			
1	E	165	Total	C	N	O	S	0	1	0
			1259	804	219	230	6			
1	G	163	Total	C	N	O	S	0	2	0
			1204	769	210	218	7			

- Molecule 2 is a protein called Hydratase ChsH2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	126	Total	C	N	O	S	0	3	0
			974	625	164	184	1			
2	D	126	Total	C	N	O	S	0	1	0
			957	612	159	185	1			
2	F	127	Total	C	N	O	S	0	0	0
			961	613	164	182	2			
2	H	126	Total	C	N	O	S	0	1	0
			956	613	159	182	2			

- Molecule 3 is CADMIUM ION (three-letter code: CD) (formula: Cd).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	G	1	Total	Cd	0	0
			1	1		
3	D	3	Total	Cd	0	0
			3	3		
3	E	4	Total	Cd	0	0
			4	4		
3	H	1	Total	Cd	0	0
			1	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total Cd 1 1	0	0
3	C	2	Total Cd 2 2	0	0
3	F	1	Total Cd 1 1	0	0

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	H	1	Total Cl 1 1	0	0
4	B	1	Total Cl 1 1	0	0
4	D	1	Total Cl 1 1	0	0
4	F	1	Total Cl 1 1	0	0

- Molecule 5 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	D	1	Total Na 1 1	0	0

- Molecule 6 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	G	1	Total Ca 1 1	0	0
6	E	1	Total Ca 1 1	0	0

- Molecule 7 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	A	190	Total O 190 190	0	0
7	B	143	Total O 143 143	0	0
7	C	150	Total O 150 150	0	0

Continued on next page...

Continued from previous page...

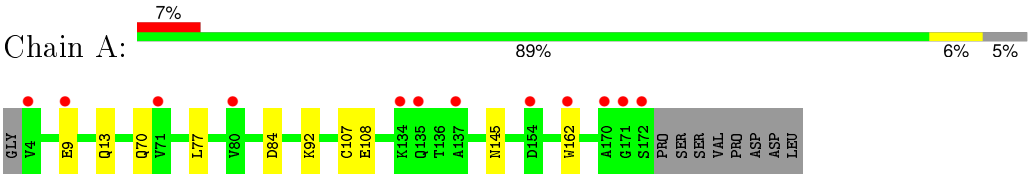
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	D	132	Total O 132 132	0	0
7	E	180	Total O 180 180	0	0
7	F	133	Total O 133 133	0	0
7	G	126	Total O 126 126	0	0
7	H	109	Total O 109 109	0	0

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 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.

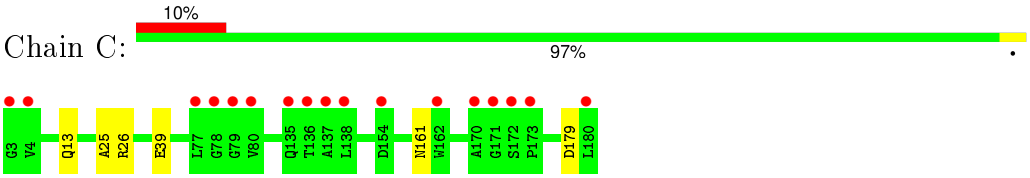
- Molecule 1: Hydratase ChsH1

Chain A:



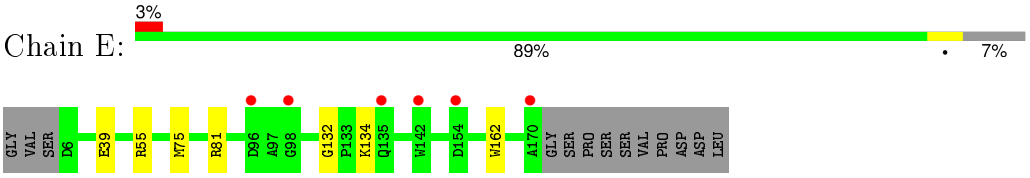
- Molecule 1: Hydratase ChsH1

Chain C:



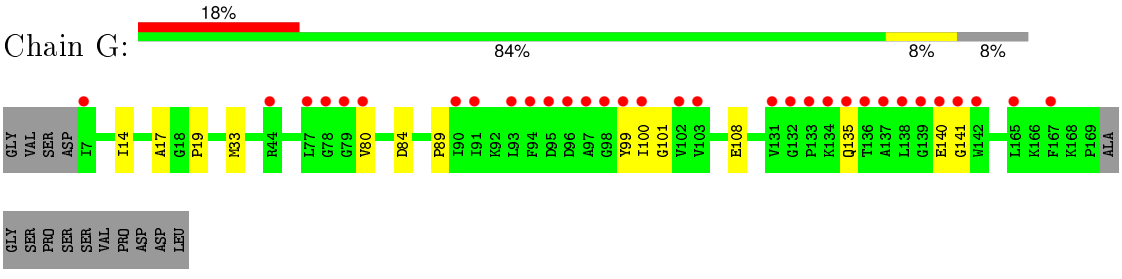
- Molecule 1: Hydratase ChsH1

Chain E:



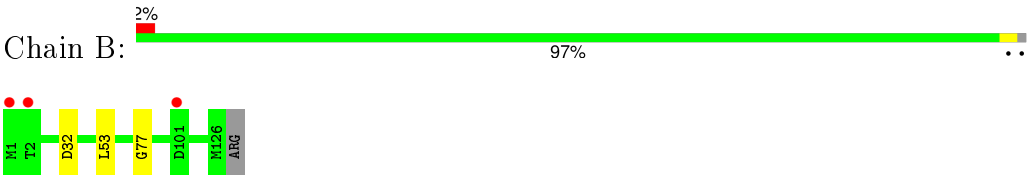
- Molecule 1: Hydratase ChsH1

Chain G:



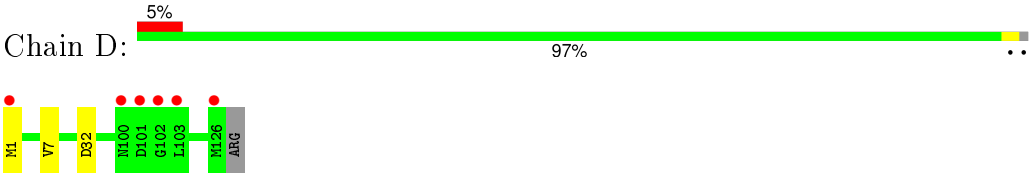
- Molecule 2: Hydratase ChsH2

Chain B:



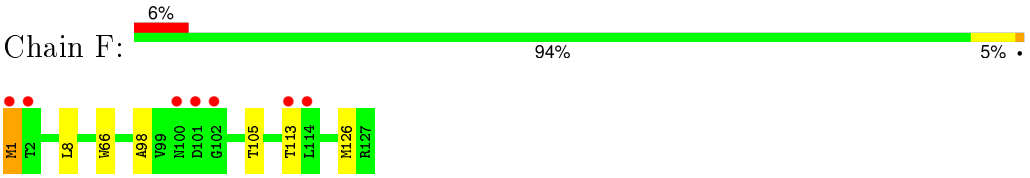
• Molecule 2: Hydratase ChsH2

Chain D:



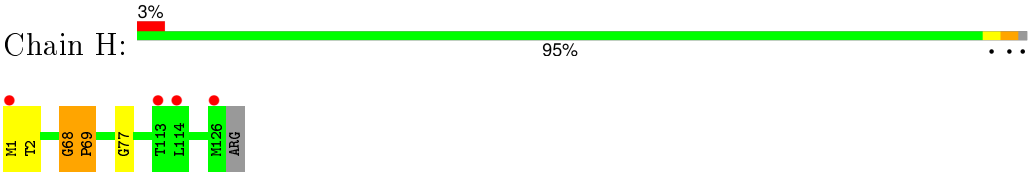
• Molecule 2: Hydratase ChsH2

Chain F:



• Molecule 2: Hydratase ChsH2

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	51.78Å 134.10Å 90.42Å 90.00° 91.58° 90.00°	Depositor
Resolution (Å)	37.60 – 1.54 37.63 – 1.54	Depositor EDS
% Data completeness (in resolution range)	99.6 (37.60-1.54) 96.3 (37.63-1.54)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.17 (at 1.54Å)	Xtriage
Refinement program	PHENIX (phenix.refine: dev_1391)	Depositor
R, R_{free}	0.177 , 0.205 0.176 , 0.204	Depositor DCC
R_{free} test set	1937 reflections (1.11%)	DCC
Wilson B-factor (Å^2)	16.1	Xtriage
Anisotropy	0.314	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 51.5	EDS
Estimated twinning fraction	0.000 for h,-k,-l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Outliers	1 of 180565 reflections (0.001%)	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	10103	wwPDB-VP
Average B, all atoms (Å^2)	25.0	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 29.10 % 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 1.6422e-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, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: NA, CA, CL, CD

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	0.36	0/1323	0.56	0/1808
1	C	0.48	2/1360 (0.1%)	0.57	2/1862 (0.1%)
1	E	0.35	0/1295	0.54	0/1770
1	G	0.32	0/1243	0.57	0/1704
2	B	0.37	0/1001	0.59	0/1367
2	D	0.35	0/978	0.58	0/1338
2	F	0.51	1/979 (0.1%)	0.68	2/1337 (0.1%)
2	H	0.73	3/977 (0.3%)	0.58	0/1336
All	All	0.45	6/9156 (0.1%)	0.58	4/12522 (0.0%)

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	H	69	PRO	N-CD	15.50	1.69	1.47
2	H	68	GLY	C-N	-12.62	1.10	1.34
1	C	25	ALA	C-N	12.08	1.61	1.34
2	F	1	MET	C-N	-11.25	1.08	1.34
1	C	26	ARG	C-N	5.96	1.47	1.34

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	F	1	MET	O-C-N	-8.01	109.89	122.70
1	C	26	ARG	O-C-N	6.65	133.34	122.70
2	F	1	MET	CA-C-N	5.40	129.08	117.20
1	C	26	ARG	CA-C-N	-5.05	106.10	117.20

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	1287	0	1266	11	0
1	C	1322	0	1273	4	0
1	E	1259	0	1236	4	0
1	G	1204	0	1132	14	0
2	B	974	0	996	3	0
2	D	957	0	959	4	0
2	F	961	0	962	10	0
2	H	956	0	966	5	0
3	B	1	0	0	0	0
3	C	2	0	0	0	0
3	D	3	0	0	0	0
3	E	4	0	0	0	0
3	F	1	0	0	0	0
3	G	1	0	0	0	0
3	H	1	0	0	0	0
4	B	1	0	0	1	0
4	D	1	0	0	0	0
4	F	1	0	0	0	0
4	H	1	0	0	0	0
5	D	1	0	0	0	0
6	E	1	0	0	0	0
6	G	1	0	0	0	0
7	A	190	0	0	1	0
7	B	143	0	0	2	0
7	C	150	0	0	3	0
7	D	132	0	0	1	0
7	E	180	0	0	1	0
7	F	133	0	0	0	0
7	G	126	0	0	1	0
7	H	109	0	0	0	0
All	All	10103	0	8790	42	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 2.

The worst 5 of 42 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)
2:H:69:PRO:N	2:H:69:PRO:CD	1.69	1.26
2:F:1:MET:SD	1:G:17:ALA:O	2.32	0.88
4:B:202:CL:CL	7:B:441:HOH:O	2.32	0.83
2:H:1:MET:SD	2:H:2:THR:N	2.62	0.73
2:H:68:GLY:C	2:H:69:PRO:CD	2.47	0.73

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	168/178 (94%)	166 (99%)	2 (1%)	0	100 100
1	C	177/178 (99%)	174 (98%)	3 (2%)	0	100 100
1	E	164/178 (92%)	162 (99%)	2 (1%)	0	100 100
1	G	163/178 (92%)	155 (95%)	6 (4%)	2 (1%)	16 2
2	B	127/127 (100%)	122 (96%)	5 (4%)	0	100 100
2	D	125/127 (98%)	121 (97%)	4 (3%)	0	100 100
2	F	125/127 (98%)	122 (98%)	3 (2%)	0	100 100
2	H	125/127 (98%)	122 (98%)	3 (2%)	0	100 100
All	All	1174/1220 (96%)	1144 (97%)	28 (2%)	2 (0%)	52 25

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	G	100	ILE
1	G	80	VAL

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	132/141 (94%)	132 (100%)	0	100	100
1	C	133/141 (94%)	133 (100%)	0	100	100
1	E	128/141 (91%)	127 (99%)	1 (1%)	86	70
1	G	114/141 (81%)	114 (100%)	0	100	100
2	B	107/106 (101%)	107 (100%)	0	100	100
2	D	104/106 (98%)	104 (100%)	0	100	100
2	F	103/106 (97%)	102 (99%)	1 (1%)	82	61
2	H	104/106 (98%)	104 (100%)	0	100	100
All	All	925/988 (94%)	923 (100%)	2 (0%)	95	86

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	162	TRP
2	F	113	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 carbohydrates in this entry.

5.6 Ligand geometry

Of 20 ligands modelled in this entry, 20 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	H	1
1	C	1
2	F	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	C	25:ALA	C	26:ARG	N	1.61
1	H	68:GLY	C	69:PRO	N	1.10
1	F	1:MET	C	2:THR	N	1.08

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	169/178 (94%)	0.26	12 (7%)	19	19	10, 22, 38, 52	0
1	C	178/178 (100%)	0.56	17 (9%)	10	11	11, 27, 51, 92	0
1	E	165/178 (92%)	0.09	6 (3%)	46	49	9, 21, 40, 54	0
1	G	163/178 (91%)	0.72	32 (19%)	1	1	11, 26, 60, 69	0
2	B	126/127 (99%)	0.02	3 (2%)	62	65	9, 21, 36, 49	0
2	D	126/127 (99%)	0.04	6 (4%)	34	36	11, 20, 33, 44	0
2	F	127/127 (100%)	0.07	7 (5%)	29	29	9, 19, 41, 61	0
2	H	126/127 (99%)	0.00	4 (3%)	51	54	11, 23, 38, 52	0
All	All	1180/1220 (96%)	0.25	87 (7%)	17	17	9, 22, 46, 92	0

The worst 5 of 87 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	171	GLY	10.6
1	C	173	PRO	10.3
1	C	172	SER	9.7
1	G	132	GLY	8.7
2	F	1	MET	7.6

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 (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å^2)	Q<0.9
3	CD	D	202	1/1	0.98	0.12	0.77	94,94,94,94	0
5	NA	D	205	1/1	0.84	0.10	0.16	20,20,20,20	0
3	CD	C	202	1/1	0.96	0.06	-1.33	34,34,34,34	0
4	CL	F	202	1/1	0.99	0.05	-2.05	14,14,14,14	0
4	CL	H	202	1/1	0.94	0.07	-2.23	27,27,27,27	0
4	CL	D	203	1/1	0.99	0.05	-	19,19,19,19	0
6	CA	G	202	1/1	0.86	0.12	-	55,55,55,55	0
3	CD	C	201	1/1	0.91	0.06	-	28,28,28,28	0
6	CA	E	205	1/1	0.84	0.12	-	35,35,35,35	0
3	CD	E	201	1/1	0.97	0.05	-	20,20,20,20	0
3	CD	E	202	1/1	0.95	0.07	-	33,33,33,33	0
3	CD	H	201	1/1	1.00	0.02	-	21,21,21,21	0
3	CD	G	201	1/1	0.84	0.06	-	79,79,79,79	0
3	CD	B	201	1/1	1.00	0.02	-	13,13,13,13	0
3	CD	D	201	1/1	1.00	0.05	-	16,16,16,16	0
3	CD	E	203	1/1	0.98	0.16	-	37,37,37,37	0
3	CD	F	201	1/1	1.00	0.04	-	12,12,12,12	0
3	CD	D	204	1/1	-0.15	0.25	-	187,187,187,187	0
3	CD	E	204	1/1	0.93	0.17	-	65,65,65,65	0
4	CL	B	202	1/1	0.98	0.08	-	18,18,18,18	0

6.5 Other polymers (i)

There are no such residues in this entry.