W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Aug 29, 2020 – 06:34 PM BST

PDB ID	:	6U08
Title	:	Double-stranded DNA-specific cytidine deaminase type VI secretion system effector and cognate immunity complex from Burkholderia cenocepacia
Authors	:	Bosch, D.E.; de Moraes, M.M.H.; Mougous, J.D.
Deposited on	:	2019-08-13
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 (i)) 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	:	2.13
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)	:	2.13

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
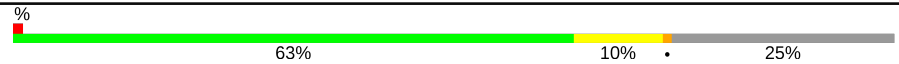*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.49 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.
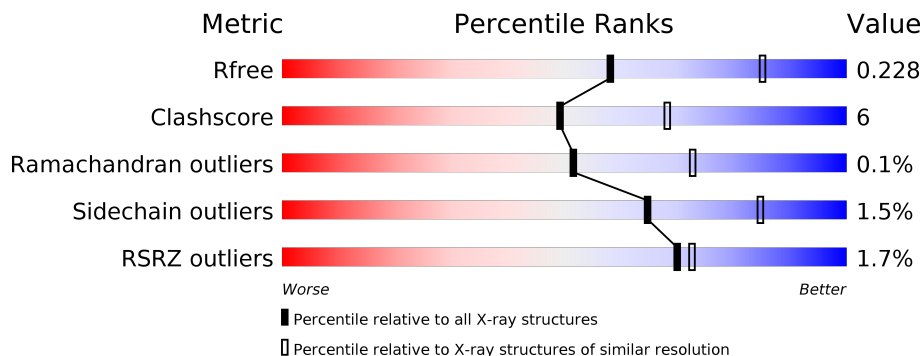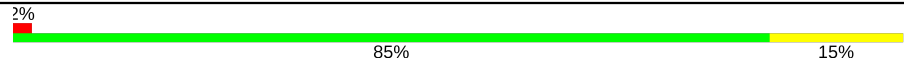

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	178	% 63% 10% · 25%
1	C	178	65% 10% 25%
1	E	178	2% 65% 8% · 26%
1	G	178	3% 67% 7% 26%
2	B	123	% 85% 12% ··
2	D	123	89% 10% ·

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	123	2% 85% 15%
2	H	123	2% 84% 15%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 7889 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 Double-stranded DNA-specific cytidine deaminase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	133	Total	C	N	O	S	Se	0	0	0
			978	615	163	195	2	3			
1	C	133	Total	C	N	O	S	Se	0	0	0
			978	615	163	195	2	3			
1	E	132	Total	C	N	O	S	Se	0	0	0
			974	613	162	194	2	3			
1	G	132	Total	C	N	O	S	Se	0	0	0
			974	613	162	194	2	3			

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1250	MSE	-	expression tag	UNP A0A1V2VU04
A	1251	GLY	-	expression tag	UNP A0A1V2VU04
A	1252	SER	-	expression tag	UNP A0A1V2VU04
A	1253	SER	-	expression tag	UNP A0A1V2VU04
A	1254	HIS	-	expression tag	UNP A0A1V2VU04
A	1255	HIS	-	expression tag	UNP A0A1V2VU04
A	1256	HIS	-	expression tag	UNP A0A1V2VU04
A	1257	HIS	-	expression tag	UNP A0A1V2VU04
A	1258	HIS	-	expression tag	UNP A0A1V2VU04
A	1259	HIS	-	expression tag	UNP A0A1V2VU04
A	1260	SER	-	expression tag	UNP A0A1V2VU04
C	1250	MSE	-	expression tag	UNP A0A1V2VU04
C	1251	GLY	-	expression tag	UNP A0A1V2VU04
C	1252	SER	-	expression tag	UNP A0A1V2VU04
C	1253	SER	-	expression tag	UNP A0A1V2VU04
C	1254	HIS	-	expression tag	UNP A0A1V2VU04
C	1255	HIS	-	expression tag	UNP A0A1V2VU04
C	1256	HIS	-	expression tag	UNP A0A1V2VU04
C	1257	HIS	-	expression tag	UNP A0A1V2VU04
C	1258	HIS	-	expression tag	UNP A0A1V2VU04
C	1259	HIS	-	expression tag	UNP A0A1V2VU04

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	1260	SER	-	expression tag	UNP A0A1V2VU04
E	1250	MSE	-	expression tag	UNP A0A1V2VU04
E	1251	GLY	-	expression tag	UNP A0A1V2VU04
E	1252	SER	-	expression tag	UNP A0A1V2VU04
E	1253	SER	-	expression tag	UNP A0A1V2VU04
E	1254	HIS	-	expression tag	UNP A0A1V2VU04
E	1255	HIS	-	expression tag	UNP A0A1V2VU04
E	1256	HIS	-	expression tag	UNP A0A1V2VU04
E	1257	HIS	-	expression tag	UNP A0A1V2VU04
E	1258	HIS	-	expression tag	UNP A0A1V2VU04
E	1259	HIS	-	expression tag	UNP A0A1V2VU04
E	1260	SER	-	expression tag	UNP A0A1V2VU04
G	1250	MSE	-	expression tag	UNP A0A1V2VU04
G	1251	GLY	-	expression tag	UNP A0A1V2VU04
G	1252	SER	-	expression tag	UNP A0A1V2VU04
G	1253	SER	-	expression tag	UNP A0A1V2VU04
G	1254	HIS	-	expression tag	UNP A0A1V2VU04
G	1255	HIS	-	expression tag	UNP A0A1V2VU04
G	1256	HIS	-	expression tag	UNP A0A1V2VU04
G	1257	HIS	-	expression tag	UNP A0A1V2VU04
G	1258	HIS	-	expression tag	UNP A0A1V2VU04
G	1259	HIS	-	expression tag	UNP A0A1V2VU04
G	1260	SER	-	expression tag	UNP A0A1V2VU04

- Molecule 2 is a protein called DddI.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	120	Total 963	C 614	N 150	O 195	S 1	Se 3	0	0	0
2	D	121	Total 969	C 617	N 151	O 197	S 1	Se 3	0	0	0
2	F	123	Total 980	C 623	N 153	O 200	S 1	Se 3	0	0	0
2	H	123	Total 980	C 623	N 153	O 200	S 1	Se 3	0	0	0

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	G	1	Total 1	Zn 1	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total 1	Zn 1	0	0
3	C	1	Total 1	Zn 1	0	0
3	E	1	Total 1	Zn 1	0	0

- Molecule 4 is water.

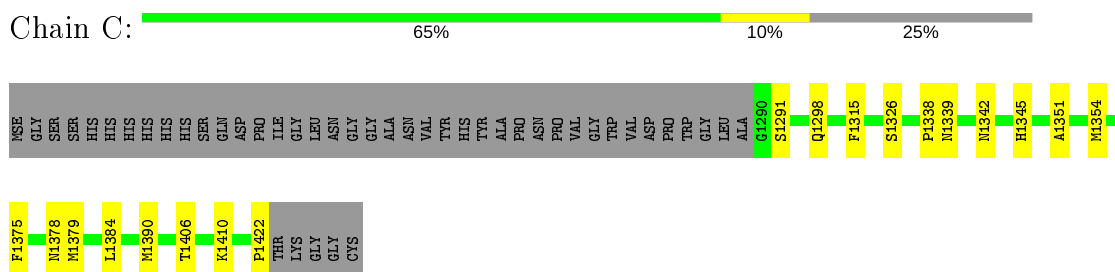| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	10	Total 10	O 10	0	0
4	B	9	Total 9	O 9	0	0
4	C	18	Total 18	O 18	0	0
4	D	3	Total 3	O 3	0	0
4	E	14	Total 14	O 14	0	0
4	F	10	Total 10	O 10	0	0
4	G	17	Total 17	O 17	0	0
4	H	8	Total 8	O 8	0	0

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.
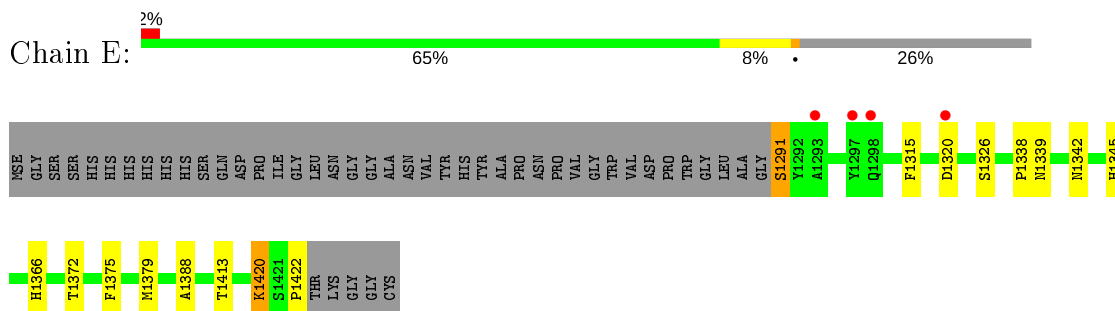
- Molecule 1: Double-stranded DNA-specific cytidine deaminase

Chain A:
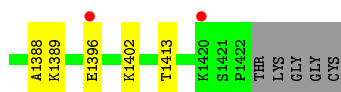

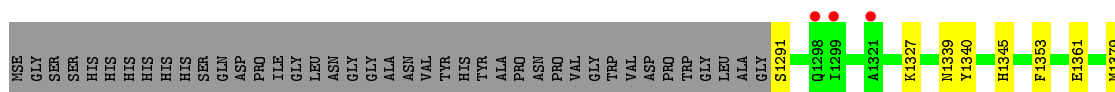- Molecule 1: Double-stranded DNA-specific cytidine deaminase

Chain C:
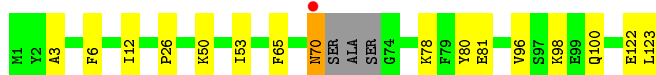

- Molecule 1: Double-stranded DNA-specific cytidine deaminase

Chain E:

- Molecule 1: Double-stranded DNA-specific cytidine deaminase

Chain G:

MSE GLY SER SER HIS HIS HIS HIS HIS HIS SER GLN ASP PRO ILE GLY LEU ASN GLY GLY ALA ASN VAL TYR HIS TYR ALA PRO ASN PRO VAL GLY TRP VAL ASP PRO TRP GLY LEU ALA GLY S1291 Q1298 I1299 A1321 K1327 N1339 Y1340 H1345 F1353 E1361 M1379

A1388 K1389 E1396 K1402 T1413 K1420 S1421 P1422 THR LYS GLY GLY CYS

● Molecule 2: DddI

Chain B:

M1 Y2 A3 F6 I12 P26 K50 I53 F65 N70 SER ALA SER G74 K78 F79 Y80 E81 V96 S97 K98 E99 Q100 E122 L123

● Molecule 2: DddI

Chain D:

M1 F6 P26 D29 F33 I47 N51 D52 E63 N64 F65 S71 ALA SER G74 E81 L118 L123

● Molecule 2: DddI

Chain F:

M1 D4 F33 V34 L35 T36 F37 F48 A54 Y58 N70 S71 A72 S73 G74 G75 F79 Y80 E81 L90 S91 K92 S97 Q100 K112 Q113 R114 L123

● Molecule 2: DddI

Chain H:

M1 D4 D5 I12 D13 L18 V19 E20 R24 F33 V34 P43 D52 M59 E63 N64 F65 A72 S73 K78 F79 Y80 E81 N82 K83 Q107 T111 K112 Q113 L123

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	126.76Å 144.95Å 64.19Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.60 – 2.49 38.60 – 2.49	Depositor EDS
% Data completeness (in resolution range)	98.9 (38.60-2.49) 94.8 (38.60-2.49)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.37 (at 2.48Å)	Xtriage
Refinement program	PHENIX 1.14_3260	Depositor
R, R_{free}	0.172 , 0.228 0.172 , 0.228	Depositor DCC
R_{free} test set	1999 reflections (4.80%)	wwPDB-VP
Wilson B-factor (Å^2)	46.2	Xtriage
Anisotropy	0.368	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 36.9	EDS
L-test for twinning[2]	$<|L|> = 0.49$, $<L^2> = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	7889	wwPDB-VP
Average B, all atoms (Å^2)	59.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.81% 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 ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	0.59	0/1002	0.61	0/1362
1	C	0.60	0/1002	0.66	0/1362
1	E	0.56	0/998	0.60	0/1357
1	G	0.55	0/998	0.60	0/1357
2	B	0.60	0/979	0.66	0/1314
2	D	0.54	0/985	0.62	0/1322
2	F	0.56	0/997	0.63	0/1340
2	H	0.58	0/997	0.66	0/1340
All	All	0.57	0/7958	0.63	0/10754

There are no bond length outliers.

There are no bond angle outliers.

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	978	0	926	14	0
1	C	978	0	925	20	0
1	E	974	0	922	12	0
1	G	974	0	922	8	0
2	B	963	0	910	11	0

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	D	969	0	915	8	0
2	F	980	0	926	10	0
2	H	980	0	926	14	0
3	A	1	0	0	0	0
3	C	1	0	0	0	0
3	E	1	0	0	0	0
3	G	1	0	0	0	0
4	A	10	0	0	0	0
4	B	9	0	0	0	0
4	C	18	0	0	0	0
4	D	3	0	0	0	0
4	E	14	0	0	1	0
4	F	10	0	0	0	0
4	G	17	0	0	1	0
4	H	8	0	0	1	0
All	All	7889	0	7372	90	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 6.

The worst 5 of 90 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:E:1366:HIS:HE2	1:E:1372:THR:HG22	1.22	1.03
1:C:1354:MSE:SE	1:C:1390:MSE:HE2	2.18	0.93
2:H:111:THR:HG23	2:H:113:GLN:H	1.44	0.82
1:C:1384:LEU:HD21	1:C:1390:MSE:HE3	1.61	0.81
1:C:1351:ALA:HA	1:C:1390:MSE:HE1	1.61	0.81

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	131/178 (74%)	131 (100%)	0	0	100	100
1	C	131/178 (74%)	130 (99%)	1 (1%)	0	100	100
1	E	130/178 (73%)	130 (100%)	0	0	100	100
1	G	130/178 (73%)	130 (100%)	0	0	100	100
2	B	116/123 (94%)	115 (99%)	1 (1%)	0	100	100
2	D	117/123 (95%)	116 (99%)	1 (1%)	0	100	100
2	F	121/123 (98%)	116 (96%)	4 (3%)	1 (1%)	19	35
2	H	121/123 (98%)	119 (98%)	2 (2%)	0	100	100
All	All	997/1204 (83%)	987 (99%)	9 (1%)	1 (0%)	51	73

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	F	73	SER

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	105/135 (78%)	103 (98%)	2 (2%)	57	80
1	C	105/135 (78%)	105 (100%)	0	100	100
1	E	105/135 (78%)	102 (97%)	3 (3%)	42	69
1	G	105/135 (78%)	103 (98%)	2 (2%)	57	80
2	B	104/103 (101%)	102 (98%)	2 (2%)	57	80
2	D	105/103 (102%)	104 (99%)	1 (1%)	76	90
2	F	106/103 (103%)	106 (100%)	0	100	100
2	H	106/103 (103%)	103 (97%)	3 (3%)	43	70
All	All	841/952 (88%)	828 (98%)	13 (2%)	65	85

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	1291	SER
1	E	1320	ASP
2	H	5	ASP
2	D	52	ASP
1	G	1402	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	1357	ASN
2	H	117	GLN
1	G	1310	GLN
2	D	64	ASN
1	G	1298	GLN

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)

Of 4 ligands modelled in this entry, 4 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 (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	130/178 (73%)	-0.15	1 (0%)	86	87	33, 51, 74, 88	0
1	C	130/178 (73%)	-0.18	0	100	100	33, 49, 71, 90	0
1	E	129/178 (72%)	0.11	4 (3%)	49	52	32, 57, 87, 96	0
1	G	129/178 (72%)	0.08	5 (3%)	39	42	32, 57, 92, 103	0
2	B	117/123 (95%)	-0.21	1 (0%)	84	86	37, 59, 81, 91	0
2	D	118/123 (95%)	-0.32	0	100	100	41, 62, 85, 104	0
2	F	120/123 (97%)	-0.08	3 (2%)	57	61	35, 60, 91, 115	0
2	H	120/123 (97%)	-0.14	3 (2%)	57	61	38, 63, 95, 105	0
All	All	993/1204 (82%)	-0.11	17 (1%)	70	72	32, 57, 88, 115	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	72	ALA	5.1
2	B	70	ASN	3.7
1	E	1298	GLN	3.6
1	G	1321	ALA	3.5
2	H	72	ALA	3.2

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)

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. The B-factors column lists the minimum, median, 95th 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	B-factors($Å^2$)	Q<0.9
3	ZN	C	1501	1/1	0.70	0.26	171,171,171,171	0
3	ZN	A	1501	1/1	0.77	0.17	156,156,156,156	0
3	ZN	E	1501	1/1	0.86	0.26	189,189,189,189	0
3	ZN	G	1501	1/1	0.91	0.28	171,171,171,171	0

6.5 Other polymers (i)

There are no such residues in this entry.