



Full wwPDB X-ray Structure Validation Report ⓘ

Sep 25, 2023 – 10:29 AM EDT

PDB ID : 5W0Z
Title : Crystal structure of MBP fused activation-induced cytidine deaminase (AID)
Authors : Qiao, Q.; Wang, L.; Wu, H.
Deposited on : 2017-06-01
Resolution : 3.61 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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
Xtriage (Phenix) : 1.13
EDS : 2.35.1
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

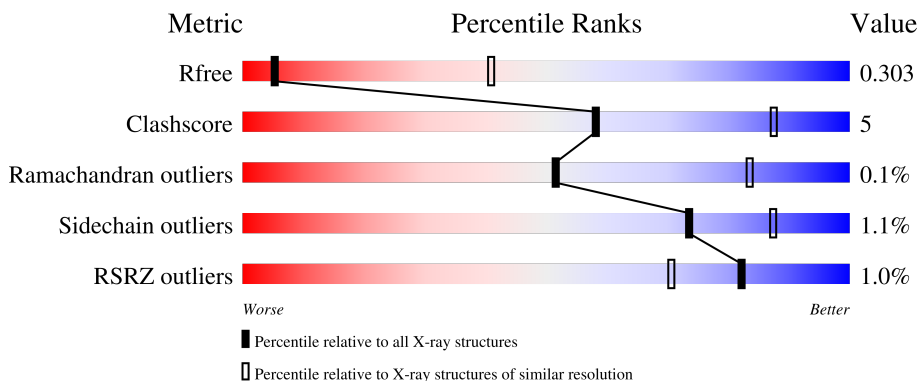
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.61 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	1290 (3.74-3.50)
Clashscore	141614	1387 (3.74-3.50)
Ramachandran outliers	138981	1339 (3.74-3.50)
Sidechain outliers	138945	1339 (3.74-3.50)
RSRZ outliers	127900	1191 (3.74-3.50)

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	549	
1	B	549	

2 Entry composition i

There are 2 unique types of molecules in this entry. The entry contains 8536 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 MBP fused activation-induced cytidine deaminase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	B	537	4260	2740	717	788	15	0	0	0
1	A	539	4274	2749	720	790	15	0	0	0

There are 56 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	MET	-	initiating methionine	UNP P0AEY0
B	83	ALA	ASP	engineered mutation	UNP P0AEY0
B	84	ALA	LYS	engineered mutation	UNP P0AEY0
B	173	ALA	GLU	engineered mutation	UNP P0AEY0
B	174	ALA	ASN	engineered mutation	UNP P0AEY0
B	240	ALA	LYS	engineered mutation	UNP P0AEY0
B	360	ALA	GLU	engineered mutation	UNP P0AEY0
B	363	ALA	LYS	engineered mutation	UNP P0AEY0
B	364	ALA	ASP	engineered mutation	UNP P0AEY0
B	368	ASN	-	linker	UNP P0AEY0
B	369	ALA	-	linker	UNP P0AEY0
B	370	ALA	-	linker	UNP P0AEY0
B	371	ALA	-	linker	UNP P0AEY0
B	372	GLU	-	linker	UNP P0AEY0
B	373	PHE	-	linker	UNP P0AEY0
B	1006	MET	-	linker	UNP P0AEY0
B	1007	ASP	-	linker	UNP P0AEY0
B	1008	PRO	-	linker	UNP P0AEY0
B	1009	ALA	-	linker	UNP P0AEY0
B	1010	THR	-	linker	UNP P0AEY0
B	1011	PHE	-	linker	UNP P0AEY0
B	1012	THR	-	linker	UNP P0AEY0
B	1042	GLU	PHE	engineered mutation	UNP Q9GZX7
B	1130	ALA	HIS	engineered mutation	UNP Q9GZX7
B	1131	GLU	ARG	engineered mutation	UNP Q9GZX7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	1141	TYR	PHE	engineered mutation	UNP Q9GZX7
B	1145	GLU	PHE	engineered mutation	UNP Q9GZX7
B	1181	GLN	LEU	engineered mutation	UNP Q9GZX7
A	1	MET	-	initiating methionine	UNP P0AEY0
A	83	ALA	ASP	engineered mutation	UNP P0AEY0
A	84	ALA	LYS	engineered mutation	UNP P0AEY0
A	173	ALA	GLU	engineered mutation	UNP P0AEY0
A	174	ALA	ASN	engineered mutation	UNP P0AEY0
A	240	ALA	LYS	engineered mutation	UNP P0AEY0
A	360	ALA	GLU	engineered mutation	UNP P0AEY0
A	363	ALA	LYS	engineered mutation	UNP P0AEY0
A	364	ALA	ASP	engineered mutation	UNP P0AEY0
A	368	ASN	-	linker	UNP P0AEY0
A	369	ALA	-	linker	UNP P0AEY0
A	370	ALA	-	linker	UNP P0AEY0
A	371	ALA	-	linker	UNP P0AEY0
A	372	GLU	-	linker	UNP P0AEY0
A	373	PHE	-	linker	UNP P0AEY0
A	1006	MET	-	linker	UNP P0AEY0
A	1007	ASP	-	linker	UNP P0AEY0
A	1008	PRO	-	linker	UNP P0AEY0
A	1009	ALA	-	linker	UNP P0AEY0
A	1010	THR	-	linker	UNP P0AEY0
A	1011	PHE	-	linker	UNP P0AEY0
A	1012	THR	-	linker	UNP P0AEY0
A	1042	GLU	PHE	engineered mutation	UNP Q9GZX7
A	1130	ALA	HIS	engineered mutation	UNP Q9GZX7
A	1131	GLU	ARG	engineered mutation	UNP Q9GZX7
A	1141	TYR	PHE	engineered mutation	UNP Q9GZX7
A	1145	GLU	PHE	engineered mutation	UNP Q9GZX7
A	1181	GLN	LEU	engineered mutation	UNP Q9GZX7

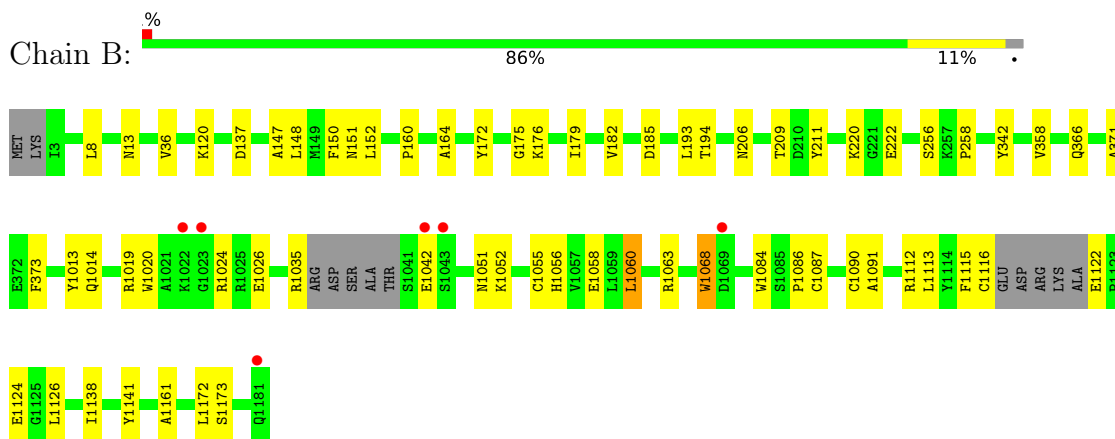
- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total Zn 1 1	0	0
2	A	1	Total Zn 1 1	0	0

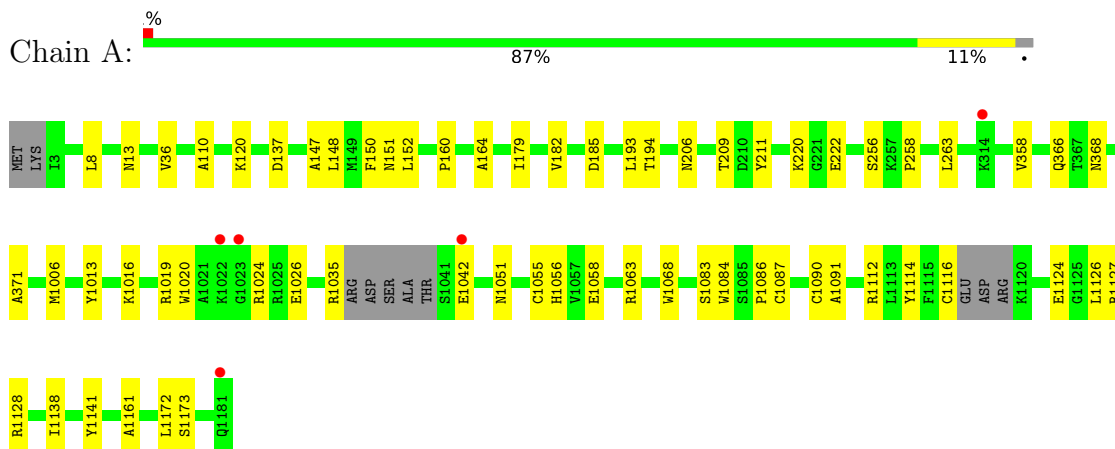
3 Residue-property plots [i](#)

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: MBP fused activation-induced cytidine deaminase



- Molecule 1: MBP fused activation-induced cytidine deaminase



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	38.68Å 167.65Å 188.30Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	25.00 – 3.61 24.47 – 3.61	Depositor EDS
% Data completeness (in resolution range)	88.7 (25.00-3.61) 89.1 (24.47-3.61)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.32 (at 3.64Å)	Xtrriage
Refinement program	REFMAC 5.8.0103	Depositor
R, R_{free}	0.274 , 0.307 0.275 , 0.303	Depositor DCC
R_{free} test set	672 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å ²)	59.3	Xtrriage
Anisotropy	0.122	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.27 , 7.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.41$, $\langle L^2 \rangle = 0.24$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.85	EDS
Total number of atoms	8536	wwPDB-VP
Average B, all atoms (Å ²)	65.0	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 25.34 % 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 3.2324e-03. 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](#)

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	0.45	0/4384	0.63	0/5948
1	B	0.45	0/4370	0.63	0/5930
All	All	0.45	0/8754	0.63	0/11878

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 [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	4274	0	4188	38	0
1	B	4260	0	4170	43	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
All	All	8536	0	8358	81	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.

All (81) 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:1013:TYR:O	1:A:1024:ARG:NH1	1.91	1.04
1:B:1014:GLN:O	1:B:1112:ARG:NH1	2.09	0.85
1:B:172:TYR:HE1	1:B:176:LYS:N	1.80	0.80
1:B:172:TYR:CE1	1:B:176:LYS:N	2.55	0.74
1:A:185:ASP:N	1:A:366:GLN:OE1	2.20	0.72
1:B:185:ASP:N	1:B:366:GLN:OE1	2.20	0.71
1:B:1013:TYR:CD1	1:B:1024:ARG:NH1	2.60	0.70
1:B:206:ASN:OD1	1:B:209:THR:OG1	2.08	0.70
1:B:1013:TYR:CE1	1:B:1024:ARG:NH1	2.61	0.69
1:A:206:ASN:OD1	1:A:209:THR:OG1	2.08	0.69
1:B:1113:LEU:HB2	1:B:1115:PHE:HE2	1.59	0.67
1:B:1115:PHE:HA	1:B:1122:GLU:OE2	1.94	0.66
1:A:1083:SER:O	1:A:1112:ARG:HG2	1.98	0.63
1:B:371:ALA:HB1	1:B:373:PHE:CE2	2.34	0.62
1:B:1122:GLU:N	1:B:1122:GLU:OE1	2.36	0.58
1:A:1058:GLU:OE1	1:A:1058:GLU:N	2.32	0.58
1:B:1086:PRO:HD2	1:B:1126:LEU:HD21	1.87	0.57
1:A:1086:PRO:HD2	1:A:1126:LEU:HD21	1.87	0.57
1:B:8:LEU:HB2	1:B:36:VAL:HG12	1.88	0.54
1:A:8:LEU:HB2	1:A:36:VAL:HG12	1.88	0.54
1:B:1138:ILE:HD13	1:B:1172:LEU:HB3	1.89	0.54
1:A:1138:ILE:HD13	1:A:1172:LEU:HB3	1.89	0.54
1:A:1035:ARG:HB2	1:A:1042:GLU:HB3	1.91	0.53
1:A:1016:LYS:HB2	1:A:1024:ARG:HH12	1.73	0.53
1:B:1084:TRP:CE2	1:B:1112:ARG:NH2	2.77	0.53
1:B:1113:LEU:HB2	1:B:1115:PHE:CE2	2.43	0.52
1:A:1013:TYR:O	1:A:1024:ARG:CZ	2.56	0.52
1:B:1055:CYS:SG	1:B:1063:ARG:NH2	2.82	0.52
1:B:1115:PHE:CD1	1:B:1122:GLU:OE2	2.64	0.51
1:B:194:THR:HG22	1:B:358:VAL:HG21	1.93	0.51
1:A:1084:TRP:CE3	1:A:1112:ARG:HG3	2.46	0.51
1:A:194:THR:HG22	1:A:358:VAL:HG21	1.93	0.51
1:B:1035:ARG:HB2	1:B:1042:GLU:HB3	1.94	0.50
1:A:1019:ARG:O	1:A:1020:TRP:CD1	2.65	0.50
1:B:1019:ARG:O	1:B:1020:TRP:CD1	2.65	0.50
1:A:193:LEU:HD12	1:A:358:VAL:HG13	1.93	0.50
1:A:150:PHE:CE1	1:A:152:LEU:HD23	2.48	0.49
1:A:179:ILE:HG22	1:A:371:ALA:HB2	1.94	0.49
1:B:193:LEU:HD12	1:B:358:VAL:HG13	1.93	0.49
1:B:137:ASP:HA	1:B:147:ALA:HB2	1.95	0.49
1:A:137:ASP:HA	1:A:147:ALA:HB2	1.95	0.49
1:A:1055:CYS:SG	1:A:1063:ARG:NH2	2.86	0.49

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:150:PHE:CE1	1:B:152:LEU:HD23	2.47	0.49
1:B:160:PRO:HG3	1:B:258:PRO:HA	1.95	0.49
1:A:1124:GLU:O	1:A:1127:ARG:N	2.47	0.48
1:A:160:PRO:HG3	1:A:258:PRO:HA	1.95	0.48
1:A:1016:LYS:HB2	1:A:1024:ARG:NH1	2.29	0.48
1:B:342:TYR:HE2	1:B:1068:TRP:CZ2	2.31	0.48
1:B:179:ILE:HG22	1:B:371:ALA:HB2	1.95	0.48
1:B:220:LYS:NZ	1:B:222:GLU:OE2	2.48	0.47
1:A:220:LYS:NZ	1:A:222:GLU:OE2	2.47	0.47
1:A:1114:TYR:HH	1:A:1116:CYS:HG	1.61	0.47
1:B:1019:ARG:O	1:B:1020:TRP:CG	2.67	0.47
1:B:1138:ILE:HD12	1:B:1173:SER:HA	1.96	0.47
1:A:1138:ILE:HD12	1:A:1173:SER:HA	1.97	0.47
1:B:1026:GLU:HG3	1:B:1051:ASN:HB2	1.95	0.47
1:A:1019:ARG:O	1:A:1020:TRP:CG	2.67	0.47
1:A:1026:GLU:HB2	1:A:1051:ASN:HB2	1.95	0.47
1:B:148:LEU:HD21	1:B:150:PHE:CE2	2.51	0.46
1:A:1128:ARG:HD2	1:A:1128:ARG:HA	1.77	0.46
1:A:182:VAL:O	1:A:366:GLN:NE2	2.42	0.46
1:A:148:LEU:HD21	1:A:150:PHE:CE2	2.51	0.46
1:B:1113:LEU:CB	1:B:1115:PHE:CE2	2.99	0.46
1:A:1056:HIS:HB2	1:A:1090:CYS:SG	2.56	0.46
1:B:1058:GLU:OE1	1:B:1058:GLU:N	2.41	0.45
1:A:1114:TYR:OH	1:A:1116:CYS:SG	2.74	0.44
1:B:1052:LYS:HD2	1:B:1060:LEU:HD11	1.98	0.44
1:B:1087:CYS:O	1:B:1091:ALA:N	2.32	0.44
1:B:182:VAL:O	1:B:366:GLN:NE2	2.43	0.44
1:B:1056:HIS:CB	1:B:1090:CYS:SG	3.07	0.43
1:A:164:ALA:HB2	1:A:256:SER:HA	1.99	0.43
1:A:1141:TYR:HE1	1:A:1161:ALA:HB2	1.84	0.43
1:B:164:ALA:HB2	1:B:256:SER:HA	1.99	0.43
1:B:1056:HIS:HB2	1:B:1090:CYS:SG	2.59	0.43
1:B:1141:TYR:HE1	1:B:1161:ALA:HB2	1.84	0.43
1:A:1087:CYS:O	1:A:1091:ALA:N	2.32	0.43
1:B:151:ASN:OD1	1:B:211:TYR:N	2.53	0.42
1:B:172:TYR:OH	1:B:175:GLY:HA2	2.20	0.42
1:A:151:ASN:OD1	1:A:211:TYR:N	2.53	0.42
1:A:110:ALA:HB3	1:A:263:LEU:HB3	2.04	0.40
1:A:1056:HIS:CB	1:A:1090:CYS:SG	3.10	0.40

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	533/549 (97%)	511 (96%)	21 (4%)	1 (0%)	47	79
1	B	531/549 (97%)	508 (96%)	23 (4%)	0	100	100
All	All	1064/1098 (97%)	1019 (96%)	44 (4%)	1 (0%)	51	83

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1006	MET

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	437/446 (98%)	433 (99%)	4 (1%)	78	89
1	B	436/446 (98%)	430 (99%)	6 (1%)	67	84
All	All	873/892 (98%)	863 (99%)	10 (1%)	73	87

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	13	ASN
1	B	120	LYS
1	B	1060	LEU
1	B	1068	TRP
1	B	1116	CYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	1124	GLU
1	A	13	ASN
1	A	120	LYS
1	A	368	ASN
1	A	1068	TRP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	13	ASN
1	A	13	ASN
1	A	368	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](#)

Of 2 ligands modelled in this entry, 2 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	539/549 (98%)	-0.19	5 (0%) 84 73	27, 61, 98, 162	0
1	B	537/549 (97%)	-0.17	6 (1%) 80 68	28, 62, 114, 145	0
All	All	1076/1098 (97%)	-0.18	11 (1%) 82 70	27, 62, 108, 162	0

All (11) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1022	LYS	3.7
1	A	1042	GLU	3.7
1	B	1181	GLN	2.9
1	B	1023	GLY	2.8
1	A	1181	GLN	2.7
1	B	1022	LYS	2.5
1	B	1069	ASP	2.4
1	A	1023	GLY	2.4
1	B	1043	SER	2.3
1	B	1042	GLU	2.3
1	A	314	LYS	2.2

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
2	ZN	B	2001	1/1	0.95	0.09	80,80,80,80	0
2	ZN	A	2001	1/1	0.97	0.10	43,43,43,43	0

6.5 Other polymers [i](#)

There are no such residues in this entry.