wwPDB X-ray Structure Validation Summary Report (i)

Jun 20, 2021 – 04:02 AM BST

PDB ID	:	6YES
Title	:	Crystal structure of type I-D CRISPR-Cas nuclease Cas10d
Authors	:	Brodersen, D.E.; Van, L.B.; Manav, M.C.
Deposited on	:	2020-03-25
Resolution	:	4.10 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:

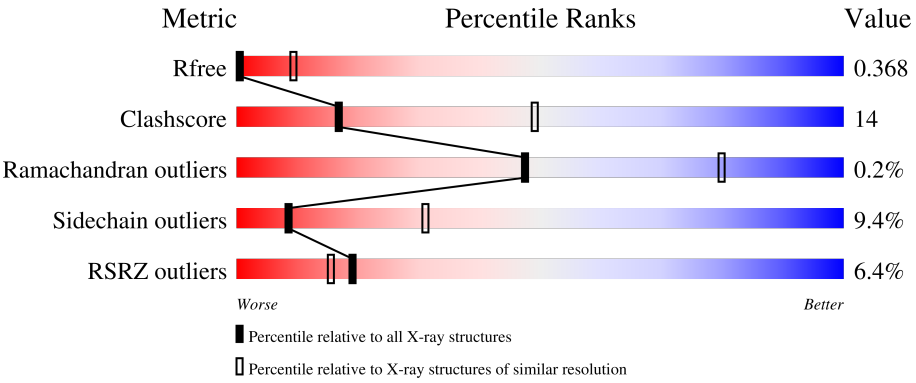
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.20
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.20

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 4.10 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	1193 (4.50-3.70)
Clashscore	141614	1003 (4.44-3.76)
Ramachandran outliers	138981	1005 (4.48-3.72)
Sidechain outliers	138945	1199 (4.50-3.70)
RSRZ outliers	127900	1034 (4.50-3.70)

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	855	2% 63% 27% 7%
1	B	855	10% 62% 26% 9%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 12534 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 CRISPR-associated protein, CscA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	791	Total 6325	C 4088	N 1009	O 1203	S 25	0	0	0
1	B	776	Total 6207	C 4013	N 992	O 1180	S 22	0	0	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	848	LEU	-	expression tag	UNP M9U4Y8
A	849	GLU	-	expression tag	UNP M9U4Y8
A	850	HIS	-	expression tag	UNP M9U4Y8
A	851	HIS	-	expression tag	UNP M9U4Y8
A	852	HIS	-	expression tag	UNP M9U4Y8
A	853	HIS	-	expression tag	UNP M9U4Y8
A	854	HIS	-	expression tag	UNP M9U4Y8
A	855	HIS	-	expression tag	UNP M9U4Y8
B	848	LEU	-	expression tag	UNP M9U4Y8
B	849	GLU	-	expression tag	UNP M9U4Y8
B	850	HIS	-	expression tag	UNP M9U4Y8
B	851	HIS	-	expression tag	UNP M9U4Y8
B	852	HIS	-	expression tag	UNP M9U4Y8
B	853	HIS	-	expression tag	UNP M9U4Y8
B	854	HIS	-	expression tag	UNP M9U4Y8
B	855	HIS	-	expression tag	UNP M9U4Y8

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

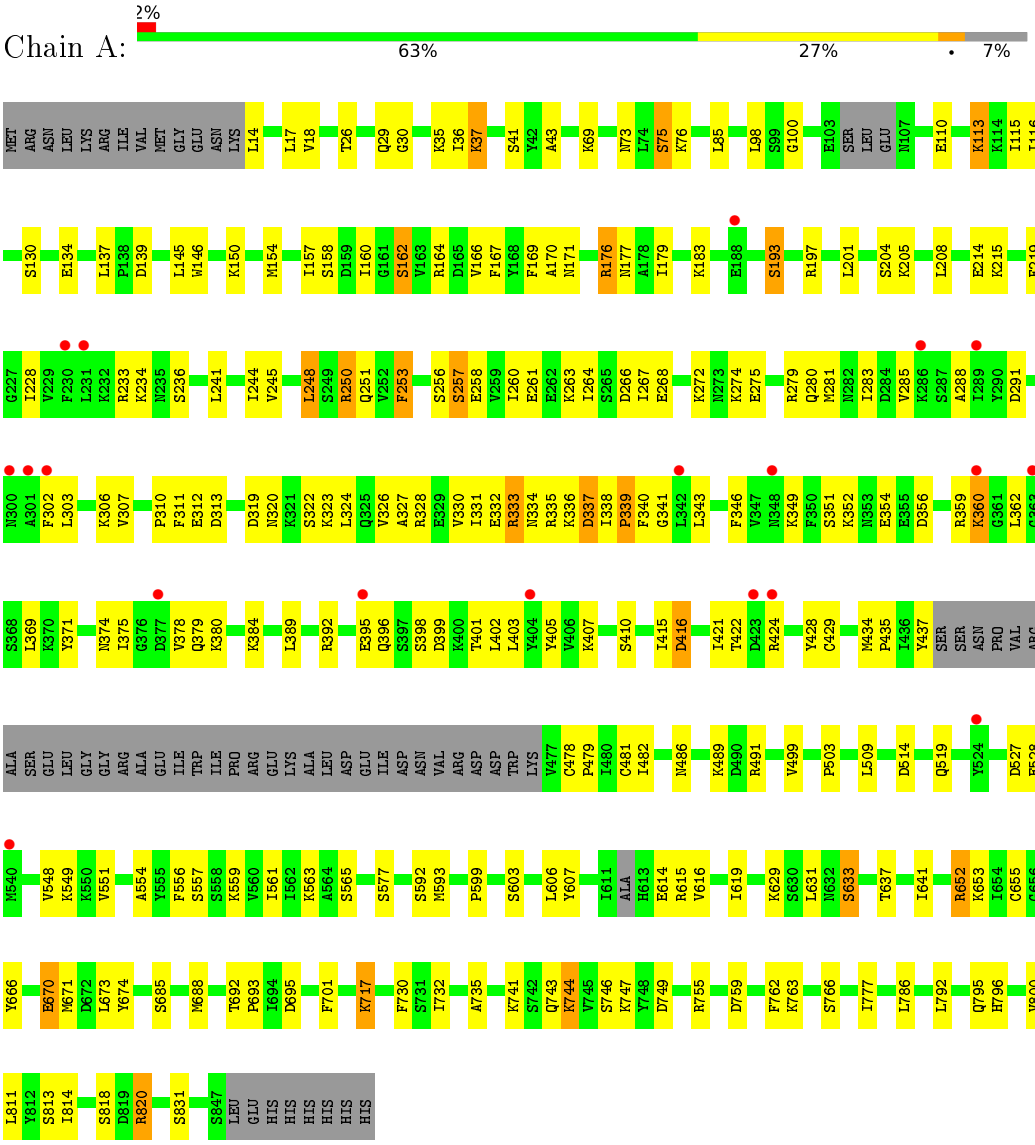
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total 1	Zn 1	0	0
2	B	1	Total 1	Zn 1	0	0

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.

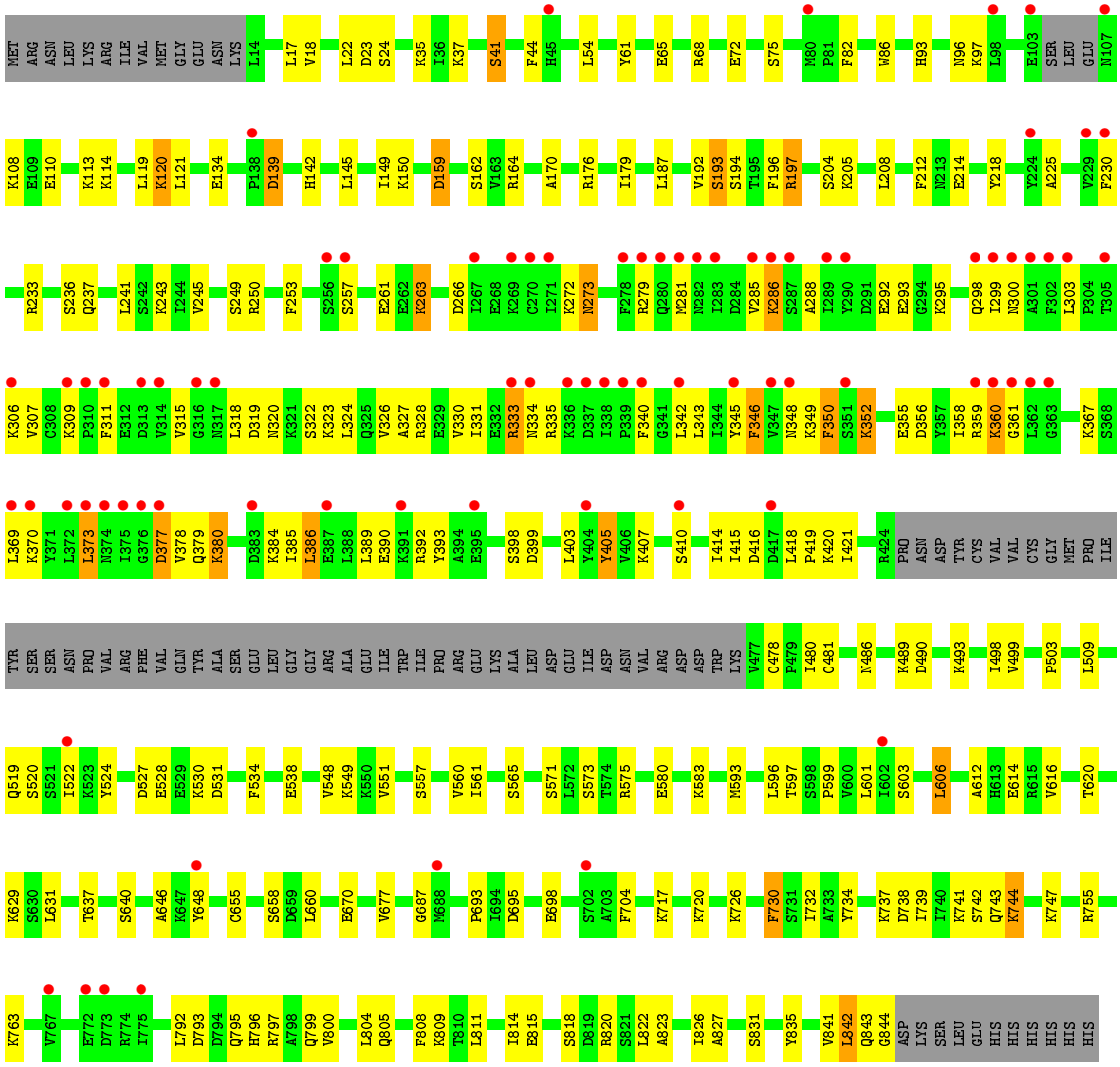
- Molecule 1: CRISPR-associated protein, CscA

Chain A:



- Molecule 1: CRISPR-associated protein, CscA

Chain B:


MET ARG ASN LEU LYS ARG ILE VAL MET GLY GLU ASN LYS L14 L17 V18 L22 D23 S24 K35 I36 K37 S41 F44 H45 L54 Y61 E65 R68 E72 S75 M80 P81 F82 W86 H93 N96 K97 L98 E103 SER LEU GLU N107
K108 E109 E110 K113 K114 L119 K120 L121 E134 P138 D139 H142 L145 I149 K150 D159 S162 V163 R164 A170 R176 I179 L187 V192 S193 S194 T195 F196 R197 S204 K205 L208 F212 M213 E214 Y218 Y224 A225 V229 F230
R233 S236 Q237 L241 S242 K243 I244 V245 S249 R250 F253 S256 S257 E261 E262 K263 D266 I267 E268 K269 C270 I271 K272 N273 F278 R279 Q280 M281 N282 I283 D284 V285 K286 S287 A288 I289 Y290 D291 E292 E293 G294 K295 Q298 I299 N300 A301 F302 L303 P304 T305
K306 V307 C308 K309 P310 F311 E312 D313 V314 V315 G316 N317 L318 D319 N320 K321 S322 K323 L324 Q325 V326 A327 R328 E329 V330 I331 E332 R333 N334 R335 K336 D337 I338 P339 F340 G341 L342 L343 I344 Y345 F346 V347 N348 K349 F350 S351 K352 E355 D356 Y357 I358 R359 K360 G361 L362 G363 K367 S368
L369 K370 Y371 L372 L373 N374 I375 G376 D377 V378 Q379 K380 D383 K384 I385 L386 E387 L388 L389 E390 K391 R392 Y393 A394 E395 S398 D399 L403 Y404 Y405 V406 K407 S410 I414 I415 D416 D417 L418 P419 K420 I421 R424 PRO ASN ASP TYR CYS VAL VAL CYS GLY MET PRO ILE
TYR SER SER ASN PRO VAL ARG PHE VAL GLN TYR ALA SER GLU LEU GLY GLY ARG ALA GLU ILE TRP PRO ARG GLU LYS ALA LEU ASP GLU ILE ASP ASN VAL ARG ASP ASP TRP LYS V477 C478 P479 I480 C481 N486 K489 D490 K493 I498 V499 P503 L509
Q519 S520 S521 I522 K523 Y524 D527 E528 E529 K530 D531 F534 E538 V548 K549 K550 V551 S557 V560 I561 S565 S571 L572 S573 T574 R575 E580 K583 H593 L596 T597 S598 P599 V600 L601 I602 S603 L606 A612 H613 E614 R615 V616 T620
K629 S630 L631 T637 S640 A646 K647 Y648 G655 S658 D659 L660 E670 V677 G687 W688 P693 I694 D695 E698 S702 A703 F704 K717 K720 K726 F730 S731 I732 A733 Y734 K737 D738 I739 T740 K741 S742 Q743 K744 K747 R755
K763 V767 E772 D773 R774 I775 L792 D793 D794 Q795 H796 R797 A798 Q799 V800 L804 Q805 F808 K809 I810 L811 I814 E815 S818 D819 R820 S821 L822 A823 I826 A827 S831 Y835 V841 L842 Q843 G844 ASP LYS SER LEU GLU HIS HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	141.43Å 141.43Å 223.87Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	91.31 – 4.10 119.57 – 4.00	Depositor EDS
% Data completeness (in resolution range)	97.9 (91.31-4.10) 80.6 (119.57-4.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.32 (at 4.01Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.343 , 0.370 0.341 , 0.368	Depositor DCC
R_{free} test set	953 reflections (4.82%)	wwPDB-VP
Wilson B-factor (Å^2)	163.5	Xtriage
Anisotropy	0.325	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 208.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	12534	wwPDB-VP
Average B, all atoms (Å^2)	263.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.10% 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	0.61	0/6446	0.75	2/8697 (0.0%)
1	B	0.62	0/6325	0.72	0/8533
All	All	0.62	0/12771	0.73	2/17230 (0.0%)

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	B	0	1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	339	PRO	N-CA-C	-7.30	93.12	112.10
1	A	337	ASP	CB-CG-OD2	5.26	123.04	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	356	ASP	Mainchain

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	6325	0	6422	183	7
1	B	6207	0	6315	172	31
2	A	1	0	0	0	0
2	B	1	0	0	0	0
All	All	12534	0	12737	355	38

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 14.

The worst 5 of 355 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:359:ARG:NH1	1:A:384:LYS:HB2	1.52	1.22
1:A:735:ALA:HB1	1:A:796:HIS:CE1	1.76	1.19
1:A:334:ASN:O	1:A:337:ASP:HB2	1.42	1.16
1:A:214:GLU:HG3	1:A:219:PHE:CE1	1.83	1.13
1:B:379:GLN:N	1:B:379:GLN:OE1	1.81	1.11

The worst 5 of 38 symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:292:GLU:CD	1:B:841:VAL:C[6_555]	0.69	1.51
1:B:292:GLU:OE2	1:B:841:VAL:CA[6_555]	0.80	1.40
1:B:292:GLU:OE2	1:B:841:VAL:C[6_555]	0.98	1.22
1:B:293:GLU:OE2	1:B:843:GLN:C[6_555]	1.11	1.09
1:B:293:GLU:CD	1:B:844:GLY:N[6_555]	1.18	1.02

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	783/855 (92%)	698 (89%)	85 (11%)	0	100	100
1	B	770/855 (90%)	691 (90%)	76 (10%)	3 (0%)	34	71
All	All	1553/1710 (91%)	1389 (89%)	161 (10%)	3 (0%)	47	80

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	261	GLU
1	B	360	LYS
1	B	386	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	718/775 (93%)	653 (91%)	65 (9%)	9	33
1	B	703/775 (91%)	635 (90%)	68 (10%)	8	29
All	All	1421/1550 (92%)	1288 (91%)	133 (9%)	8	31

5 of 133 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	571	SER
1	B	603	SER
1	B	815	GLU
1	A	657	ARG
1	A	652	ARG

Sometimes 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

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no monosaccharides in this entry.

5.6 Ligand geometry

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

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	791/855 (92%)	0.08	19 (2%) 59 49	138, 225, 308, 329	0
1	B	776/855 (90%)	0.53	82 (10%) 6 6	176, 264, 467, 602	0
All	All	1567/1710 (91%)	0.30	101 (6%) 19 15	138, 241, 441, 602	0

The worst 5 of 101 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	282	ASN	10.2
1	B	374	ASN	8.2
1	B	301	ALA	8.2
1	B	376	GLY	7.8
1	B	310	PRO	7.7

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 monosaccharides in this entry.

6.4 Ligands

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, 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	B-factors(Å^2)	Q<0.9
2	ZN	A	901	1/1	0.51	0.14	284,284,284,284	0
2	ZN	B	901	1/1	0.89	0.27	257,257,257,257	0

6.5 Other polymers (i)

There are no such residues in this entry.