Full wwPDB X-ray Structure Validation Report ⓘ

Apr 11, 2016 – 05:17 AM EDT

PDB ID	:	5DGQ
Title	:	Crystal structure of GH9 exo-beta-D-glucosaminidase PBPRA0520
Authors	:	Suzuki, K.; Honda, Y.; Fushinobu, S.
Deposited on	:	2015-08-28
Resolution	:	1.90 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see references ①) were used in the production of this report:

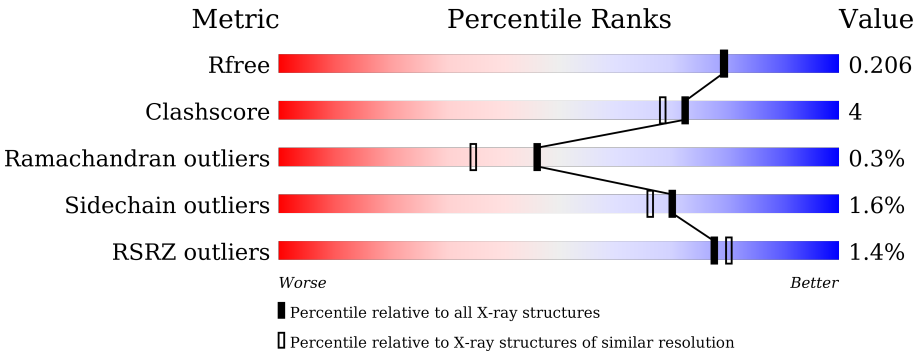
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027107
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027107

1 Overall quality at a glance (i)

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

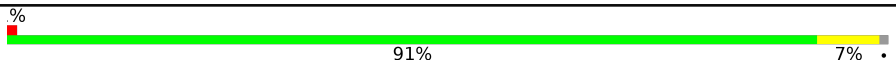
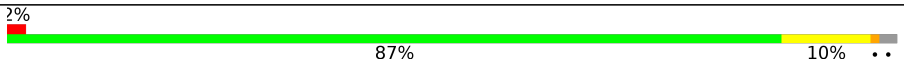
The reported resolution of this entry is 1.90 Å.

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	4755 (1.90-1.90)
Clashscore	102246	5398 (1.90-1.90)
Ramachandran outliers	100387	5338 (1.90-1.90)
Sidechain outliers	100360	5339 (1.90-1.90)
RSRZ outliers	91569	4766 (1.90-1.90)

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	586	
1	B	586	

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 10003 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 Putative endoglucanase-related protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	578	Total	C	N	O	S	0	0	0
			4608	2912	793	870	33			
1	B	573	Total	C	N	O	S	0	0	0
			4573	2888	788	864	33			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	579	LEU	-	expression tag	UNP Q6LUT2
A	580	GLU	-	expression tag	UNP Q6LUT2
A	581	HIS	-	expression tag	UNP Q6LUT2
A	582	HIS	-	expression tag	UNP Q6LUT2
A	583	HIS	-	expression tag	UNP Q6LUT2
A	584	HIS	-	expression tag	UNP Q6LUT2
A	585	HIS	-	expression tag	UNP Q6LUT2
A	586	HIS	-	expression tag	UNP Q6LUT2
B	579	LEU	-	expression tag	UNP Q6LUT2
B	580	GLU	-	expression tag	UNP Q6LUT2
B	581	HIS	-	expression tag	UNP Q6LUT2
B	582	HIS	-	expression tag	UNP Q6LUT2
B	583	HIS	-	expression tag	UNP Q6LUT2
B	584	HIS	-	expression tag	UNP Q6LUT2
B	585	HIS	-	expression tag	UNP Q6LUT2
B	586	HIS	-	expression tag	UNP Q6LUT2

- Molecule 2 is SODIUM ION (three-letter code: NA) (formula: Na).

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

- Molecule 3 is water.

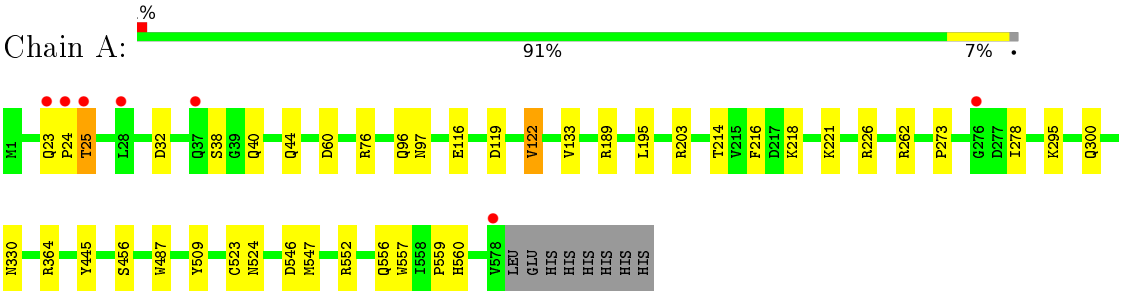
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	446	Total O 446 446	0	0
3	B	374	Total O 374 374	0	0

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.

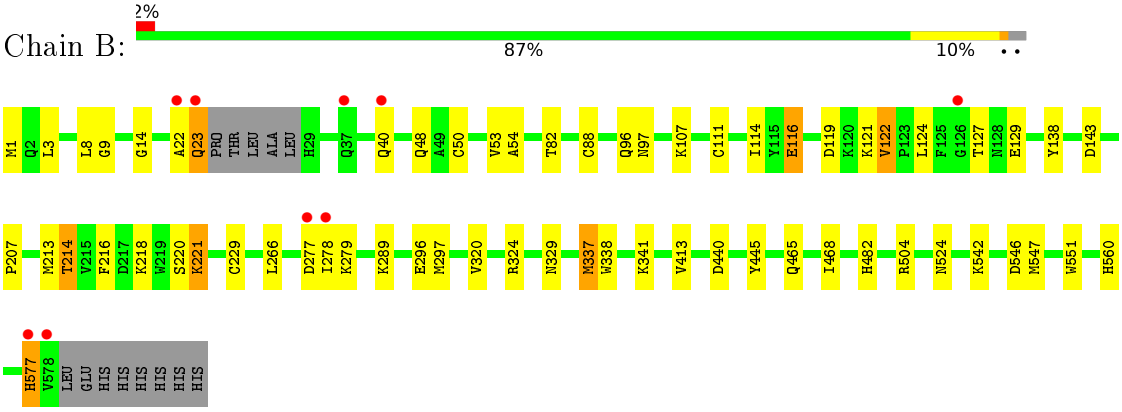
- Molecule 1: Putative endoglucanase-related protein

Chain A:



- Molecule 1: Putative endoglucanase-related protein

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	61.93Å 102.32Å 90.28Å 90.00° 97.72° 90.00°	Depositor
Resolution (Å)	31.87 – 1.90 31.87 – 1.89	Depositor EDS
% Data completeness (in resolution range)	99.0 (31.87-1.90) 99.0 (31.87-1.89)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.30 (at 1.89Å)	Xtriage
Refinement program	REFMAC 5.8.0107	Depositor
R, R_{free}	0.157 , 0.206 0.158 , 0.206	Depositor DCC
R_{free} test set	4408 reflections (5.29%)	DCC
Wilson B-factor (Å^2)	17.7	Xtriage
Anisotropy	0.055	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 47.1	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Outliers	0 of 87798 reflections	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	10003	wwPDB-VP
Average B, all atoms (Å^2)	19.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.24% 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, 0.375 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

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	1.20	3/4727 (0.1%)	1.02	10/6411 (0.2%)
1	B	1.14	4/4690 (0.1%)	0.99	8/6357 (0.1%)
All	All	1.17	7/9417 (0.1%)	1.00	18/12768 (0.1%)

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	A	0	1

All (7) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	456	SER	CA-CB	6.61	1.62	1.52
1	B	551	TRP	CG-CD1	5.46	1.44	1.36
1	B	338	TRP	CE3-CZ3	5.43	1.47	1.38
1	B	551	TRP	CZ3-CH2	5.43	1.48	1.40
1	A	509	TYR	CG-CD1	5.34	1.46	1.39
1	B	116	GLU	CD-OE2	5.27	1.31	1.25
1	A	487	TRP	CG-CD1	5.10	1.43	1.36

All (18) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	203	ARG	NE-CZ-NH2	-10.08	115.26	120.30
1	A	203	ARG	NE-CZ-NH1	9.77	125.18	120.30
1	B	504	ARG	NE-CZ-NH2	8.72	124.66	120.30
1	B	107	LYS	CD-CE-NZ	-7.85	93.65	111.70

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	504	ARG	NE-CZ-NH1	-7.55	116.53	120.30
1	B	143	ASP	CB-CG-OD2	-6.77	112.20	118.30
1	A	60	ASP	CB-CG-OD2	-6.71	112.26	118.30
1	A	364	ARG	NE-CZ-NH1	6.31	123.45	120.30
1	B	440	ASP	CB-CG-OD1	-6.30	112.63	118.30
1	A	221	LYS	CD-CE-NZ	-5.96	97.99	111.70
1	A	189	ARG	NE-CZ-NH2	-5.88	117.36	120.30
1	B	546	ASP	CB-CG-OD1	5.86	123.58	118.30
1	A	552	ARG	NE-CZ-NH2	-5.82	117.39	120.30
1	A	546	ASP	CB-CG-OD1	5.70	123.43	118.30
1	A	226	ARG	NE-CZ-NH1	5.32	122.96	120.30
1	B	122	VAL	CB-CA-C	-5.30	101.33	111.40
1	B	221	LYS	CD-CE-NZ	-5.26	99.60	111.70
1	A	547	MET	CG-SD-CE	-5.07	92.08	100.20

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	24	PRO	Peptide

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	4608	0	4364	24	0
1	B	4573	0	4322	52	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	446	0	0	6	0
3	B	374	0	0	14	0
All	All	10003	0	8686	75	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 4.

All (75) 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:B:337:MET:HE1	1:B:341:LYS:CE	1.60	1.29
1:B:337:MET:CE	1:B:341:LYS:CE	2.13	1.26
1:B:337:MET:HE1	1:B:341:LYS:HE3	1.28	1.12
1:B:337:MET:CE	1:B:341:LYS:HE2	1.76	1.12
1:A:38:SER:OG	1:A:40:GLN:HG2	1.65	0.96
1:B:337:MET:HE1	1:B:341:LYS:CD	1.97	0.95
1:A:116:GLU:HG3	3:A:1060:HOH:O	1.67	0.94
1:B:116:GLU:HG3	3:B:894:HOH:O	1.68	0.92
1:B:50:CYS:HB3	3:B:826:HOH:O	1.71	0.90
1:B:337:MET:HE3	1:B:341:LYS:HE2	1.56	0.86
1:B:413:VAL:HG12	3:B:751:HOH:O	1.76	0.85
1:B:48:GLN:HB3	3:B:1031:HOH:O	1.77	0.85
1:B:127:THR:HG23	1:B:129:GLU:H	1.47	0.79
1:B:337:MET:CE	1:B:341:LYS:CD	2.59	0.79
1:B:337:MET:HE1	1:B:341:LYS:HD2	1.64	0.78
1:B:337:MET:CE	1:B:341:LYS:HE3	1.98	0.75
1:B:214:THR:HG23	1:B:229:CYS:O	1.89	0.73
1:B:14:GLY:HA3	3:B:808:HOH:O	1.88	0.72
1:B:278:ILE:HD11	3:B:1068:HOH:O	1.89	0.72
1:A:40:GLN:OE1	3:A:701:HOH:O	2.07	0.71
1:B:337:MET:CE	1:B:341:LYS:HD2	2.21	0.70
1:A:32:ASP:OD1	1:A:44:GLN:HG2	2.02	0.60
1:B:1:MET:HG2	1:B:22:ALA:HB2	1.83	0.59
1:B:207:PRO:O	1:B:289:LYS:NZ	2.25	0.57
1:B:8:LEU:HD12	1:B:8:LEU:N	2.20	0.56
1:B:124:LEU:O	1:B:127:THR:HG22	2.07	0.55
1:B:337:MET:HE3	1:B:341:LYS:CE	2.15	0.55
1:A:119:ASP:HA	1:A:122:VAL:HG13	1.89	0.54
1:A:122:VAL:CG2	1:A:133:VAL:HG22	2.38	0.54
1:B:214:THR:HG21	3:B:1027:HOH:O	2.06	0.54
1:A:216:PHE:CZ	1:A:218:LYS:HA	2.45	0.52
1:B:23:GLN:C	3:B:937:HOH:O	2.47	0.52
1:B:524:ASN:HA	1:B:560:HIS:NE2	2.25	0.52
1:B:220:SER:O	1:B:221:LYS:HB2	2.09	0.51
1:B:14:GLY:HA2	1:B:482:HIS:CE1	2.46	0.51
1:B:14:GLY:CA	3:B:808:HOH:O	2.55	0.50
1:A:330:ASN:HB2	3:A:1023:HOH:O	2.11	0.50
1:B:577:HIS:ND1	1:B:577:HIS:O	2.43	0.50
1:B:127:THR:HG23	1:B:129:GLU:N	2.22	0.49
1:B:1:MET:HG2	1:B:22:ALA:CB	2.42	0.49
1:B:296:GLU:HG2	1:B:297:MET:HG2	1.95	0.48
1:B:278:ILE:HG13	3:B:999:HOH:O	2.13	0.48

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:96:GLN:HG2	1:A:97:ASN:OD1	2.14	0.48
1:A:295:LYS:HD2	1:A:295:LYS:HA	1.70	0.47
1:B:337:MET:HE2	1:B:341:LYS:CD	2.43	0.47
1:B:96:GLN:HG2	1:B:97:ASN:OD1	2.13	0.47
1:A:524:ASN:HA	1:A:560:HIS:NE2	2.30	0.47
1:B:468:ILE:HG22	3:B:701:HOH:O	2.15	0.46
1:B:465:GLN:OE1	1:B:577:HIS:CE1	2.69	0.46
1:B:337:MET:HE2	1:B:341:LYS:HE2	1.82	0.46
1:B:413:VAL:HG11	3:B:1049:HOH:O	2.15	0.45
1:A:116:GLU:CG	3:A:1060:HOH:O	2.39	0.45
1:B:111:CYS:HA	1:B:138:TYR:CZ	2.52	0.45
1:A:23:GLN:C	1:A:25:THR:H	2.19	0.45
1:A:330:ASN:ND2	3:A:714:HOH:O	2.49	0.44
1:A:557:TRP:HB3	1:A:560:HIS:CD2	2.53	0.44
1:B:9:GLY:HA3	1:B:88:CYS:O	2.18	0.44
1:A:262:ARG:HD3	1:A:262:ARG:HH21	1.63	0.44
1:B:116:GLU:CG	3:B:894:HOH:O	2.47	0.43
1:A:122:VAL:HG21	1:A:133:VAL:CG2	2.49	0.43
1:A:122:VAL:HG21	1:A:133:VAL:HG22	2.01	0.43
1:A:273:PRO:HB2	1:B:114:ILE:HD12	2.00	0.43
1:A:278:ILE:HD13	1:A:278:ILE:HG21	1.83	0.43
1:B:119:ASP:HA	1:B:122:VAL:CG2	2.48	0.42
1:A:300:GLN:HG3	3:A:1035:HOH:O	2.20	0.42
1:A:557:TRP:CD2	1:A:559:PRO:HD2	2.56	0.41
1:B:116:GLU:CB	3:B:894:HOH:O	2.69	0.41
1:A:195:LEU:HD23	1:A:195:LEU:HA	1.83	0.41
1:B:213:MET:C	1:B:214:THR:HG22	2.41	0.41
1:B:216:PHE:CZ	1:B:218:LYS:HA	2.56	0.41
1:B:320:VAL:O	1:B:324:ARG:HG3	2.20	0.41
1:B:3:LEU:HD12	1:B:82:THR:HG22	2.02	0.41
1:B:53:VAL:O	1:B:54:ALA:C	2.58	0.41
1:B:216:PHE:HB3	1:B:229:CYS:SG	2.62	0.40
1:A:523:CYS:HA	1:A:556:GLN:HA	2.02	0.40

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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	576/586 (98%)	556 (96%)	19 (3%)	1 (0%)	52	42
1	B	569/586 (97%)	551 (97%)	16 (3%)	2 (0%)	39	27
All	All	1145/1172 (98%)	1107 (97%)	35 (3%)	3 (0%)	46	35

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	279	LYS
1	A	445	TYR
1	B	445	TYR

5.3.2 Protein sidechains

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	480/488 (98%)	476 (99%)	4 (1%)	86	86
1	B	476/488 (98%)	465 (98%)	11 (2%)	58	51
All	All	956/976 (98%)	941 (98%)	15 (2%)	70	66

All (15) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	25	THR
1	A	76	ARG
1	A	122	VAL

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	214	THR
1	B	23	GLN
1	B	40	GLN
1	B	121	LYS
1	B	214	THR
1	B	266	LEU
1	B	277	ASP
1	B	329	ASN
1	B	337	MET
1	B	542	LYS
1	B	547	MET
1	B	577	HIS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	40	GLN

5.3.3 RNA

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 carbohydrates 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	578/586 (98%)	-0.48	7 (1%)	81	83	8, 15, 34, 66	0
1	B	573/586 (97%)	-0.33	9 (1%)	74	78	9, 19, 36, 70	0
All	All	1151/1172 (98%)	-0.41	16 (1%)	78	80	8, 17, 36, 70	0

All (16) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	277	ASP	8.0
1	B	278	ILE	6.3
1	A	24	PRO	4.9
1	B	577	HIS	3.8
1	A	276	GLY	3.8
1	B	23	GLN	3.1
1	A	25	THR	2.5
1	A	23	GLN	2.5
1	B	37	GLN	2.4
1	B	578	VAL	2.4
1	A	28	LEU	2.3
1	B	126	GLY	2.3
1	B	40	GLN	2.2
1	A	578	VAL	2.2
1	B	22	ALA	2.1
1	A	37	GLN	2.0

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 (i)

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
2	NA	B	601	1/1	0.98	0.04	-1.24	19,19,19,19	0
2	NA	A	601	1/1	1.00	0.03	-3.47	12,12,12,12	0

6.5 Other polymers (i)

There are no such residues in this entry.