



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 06:31 pm GMT

PDB ID : 4KAF
Title : Crystal Structure of Haloalkane dehalogenase HaloTag7 at the resolution 1.5Å, Northeast Structural Genomics Consortium (NESG) Target OR151
Authors : Kuzin, A.P.; Lew, S.; Neklesa, T.K.; Noblin, D.; Seetharaman, J.; Maglaqui, M.; Xiao, R.; Kohan, E.; Wang, H.; Everett, J.K.; Acton, T.B.; Kornhaber, G.; Montelione, G.T.; Crews, C.; Tong, L.; Northeast Structural Genomics Consortium (NESG)
Deposited on : 2013-04-22
Resolution : 1.50 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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	recalc28949

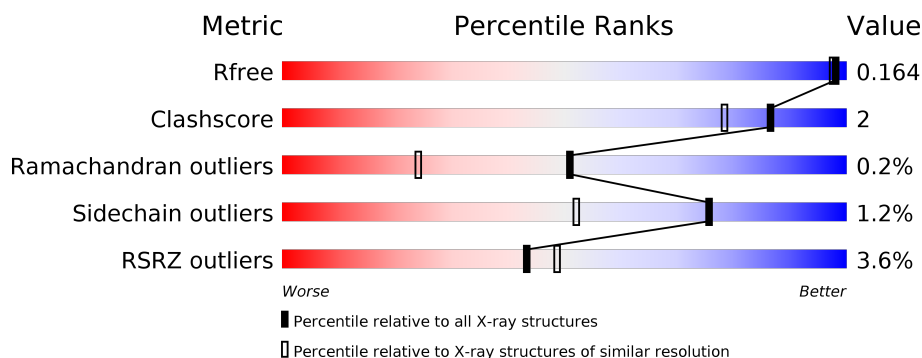
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.50 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}	100719	2279 (1.50-1.50)
Clashscore	112137	2503 (1.50-1.50)
Ramachandran outliers	110173	2445 (1.50-1.50)
Sidechain outliers	110143	2443 (1.50-1.50)
RSRZ outliers	101464	2305 (1.50-1.50)

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. 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	307	<div> <div>4%</div> <div> <div></div> <div>92%</div> <div>5%</div> <div>.</div> </div> </div>
1	B	307	<div> <div>3%</div> <div> <div></div> <div>94%</div> <div>.</div> <div>.</div> </div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	SCN	A	401	-	X	-	-
2	SCN	B	401	-	X	-	-
3	NA	A	405	-	-	-	X
4	EDO	B	404	-	-	-	X
4	EDO	B	405	-	-	X	X
5	PEG	B	406[A]	-	-	X	X
5	PEG	B	406[B]	-	-	X	X

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 10306 atoms, of which 4742 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 Haloalkane dehalogenase.

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
1	A	296	Total	C	H	N	O	S	Se	0	9	0
			4785	1568	2370	404	432	2	9			
1	B	299	Total	C	H	N	O	S	Se	0	8	0
			4800	1574	2372	408	435	2	9			

There are 74 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	EXPRESSION TAG	UNP P0A3G3
A	2	GLY	-	EXPRESSION TAG	UNP P0A3G3
A	3	HIS	-	EXPRESSION TAG	UNP P0A3G3
A	4	HIS	-	EXPRESSION TAG	UNP P0A3G3
A	5	HIS	-	EXPRESSION TAG	UNP P0A3G3
A	6	HIS	-	EXPRESSION TAG	UNP P0A3G3
A	7	HIS	-	EXPRESSION TAG	UNP P0A3G3
A	8	HIS	-	EXPRESSION TAG	UNP P0A3G3
A	10	HIS	-	EXPRESSION TAG	UNP P0A3G3
A	11	MSE	-	EXPRESSION TAG	UNP P0A3G3
A	12	ALA	-	EXPRESSION TAG	UNP P0A3G3
A	57	VAL	LEU	engineered mutation	UNP P0A3G3
A	68	THR	SER	engineered mutation	UNP P0A3G3
A	88	GLY	ASP	engineered mutation	UNP P0A3G3
A	97	PHE	TYR	engineered mutation	UNP P0A3G3
A	98	MSE	LEU	engineered mutation	UNP P0A3G3
A	133	ILE	VAL	engineered mutation	UNP P0A3G3
A	138	PHE	CYS	engineered mutation	UNP P0A3G3
A	165	THR	ALA	engineered mutation	UNP P0A3G3
A	170	LYS	GLU	engineered mutation	UNP P0A3G3
A	177	VAL	ALA	engineered mutation	UNP P0A3G3
A	182	THR	ALA	engineered mutation	UNP P0A3G3
A	185	MSE	LYS	engineered mutation	UNP P0A3G3
A	186	GLY	CYS	engineered mutation	UNP P0A3G3
A	205	ASN	LYS	engineered mutation	UNP P0A3G3

Continued on next page...

Continued from previous page...

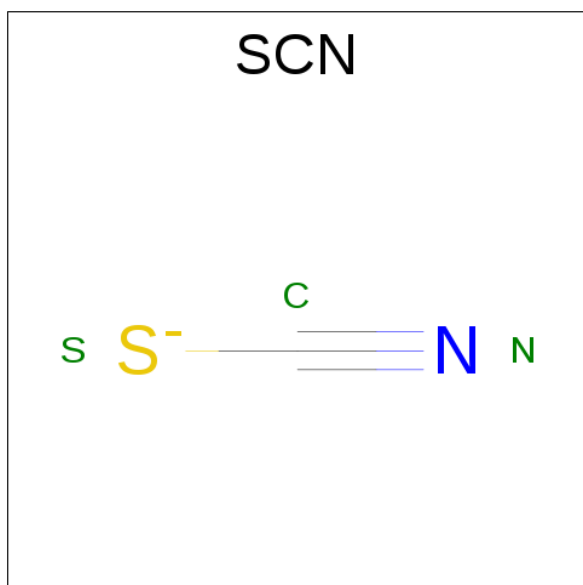
Chain	Residue	Modelled	Actual	Comment	Reference
A	234	GLU	ALA	engineered mutation	UNP P0A3G3
A	237	ASP	ASN	engineered mutation	UNP P0A3G3
A	267	LYS	GLU	engineered mutation	UNP P0A3G3
A	274	ALA	THR	engineered mutation	UNP P0A3G3
A	282	ASN	HIS	engineered mutation	UNP P0A3G3
A	283	LEU	TYR	engineered mutation	UNP P0A3G3
A	301	SER	PRO	engineered mutation	UNP P0A3G3
A	302	THR	ALA	engineered mutation	UNP P0A3G3
A	304	GLU	-	EXPRESSION TAG	UNP P0A3G3
A	305	ILE	-	EXPRESSION TAG	UNP P0A3G3
A	306	SER	-	EXPRESSION TAG	UNP P0A3G3
A	307	GLY	-	EXPRESSION TAG	UNP P0A3G3
B	-9	MSE	-	EXPRESSION TAG	UNP P0A3G3
B	-8	GLY	-	EXPRESSION TAG	UNP P0A3G3
B	-7	HIS	-	EXPRESSION TAG	UNP P0A3G3
B	-6	HIS	-	EXPRESSION TAG	UNP P0A3G3
B	-5	HIS	-	EXPRESSION TAG	UNP P0A3G3
B	-4	HIS	-	EXPRESSION TAG	UNP P0A3G3
B	-3	HIS	-	EXPRESSION TAG	UNP P0A3G3
B	-2	HIS	-	EXPRESSION TAG	UNP P0A3G3
B	0	HIS	-	EXPRESSION TAG	UNP P0A3G3
B	1	MSE	-	EXPRESSION TAG	UNP P0A3G3
B	12	ALA	-	EXPRESSION TAG	UNP P0A3G3
B	57	VAL	LEU	engineered mutation	UNP P0A3G3
B	68	THR	SER	engineered mutation	UNP P0A3G3
B	88	GLY	ASP	engineered mutation	UNP P0A3G3
B	97	PHE	TYR	engineered mutation	UNP P0A3G3
B	98	MSE	LEU	engineered mutation	UNP P0A3G3
B	133	ILE	VAL	engineered mutation	UNP P0A3G3
B	138	PHE	CYS	engineered mutation	UNP P0A3G3
B	165	THR	ALA	engineered mutation	UNP P0A3G3
B	170	LYS	GLU	engineered mutation	UNP P0A3G3
B	177	VAL	ALA	engineered mutation	UNP P0A3G3
B	182	THR	ALA	engineered mutation	UNP P0A3G3
B	185	MSE	LYS	engineered mutation	UNP P0A3G3
B	186	GLY	CYS	engineered mutation	UNP P0A3G3
B	205	ASN	LYS	engineered mutation	UNP P0A3G3
B	234	GLU	ALA	engineered mutation	UNP P0A3G3
B	237	ASP	ASN	engineered mutation	UNP P0A3G3
B	267	LYS	GLU	engineered mutation	UNP P0A3G3
B	274	ALA	THR	engineered mutation	UNP P0A3G3
B	282	ASN	HIS	engineered mutation	UNP P0A3G3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	283	LEU	TYR	engineered mutation	UNP P0A3G3
B	301	SER	PRO	engineered mutation	UNP P0A3G3
B	302	THR	ALA	engineered mutation	UNP P0A3G3
B	304	GLU	-	EXPRESSION TAG	UNP P0A3G3
B	305	ILE	-	EXPRESSION TAG	UNP P0A3G3
B	306	SER	-	EXPRESSION TAG	UNP P0A3G3
B	307	GLY	-	EXPRESSION TAG	UNP P0A3G3

- Molecule 2 is THIOCYANATE ION (three-letter code: SCN) (formula: CNS).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	S	0	0
			3	1	1	1		
2	B	1	Total	C	N	S	0	0
			3	1	1	1		

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

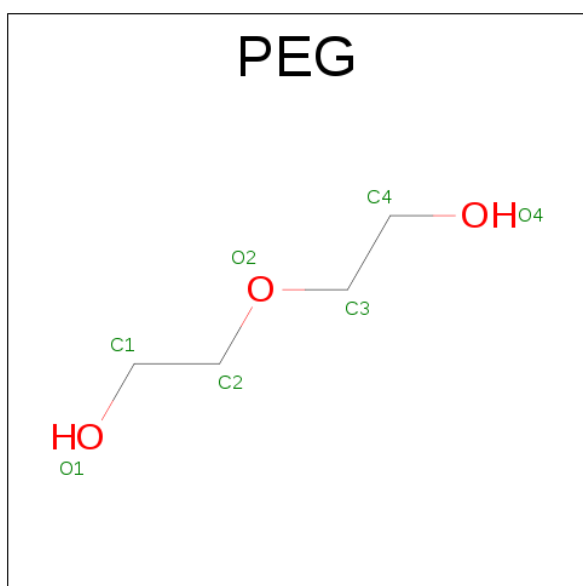
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	2	Total	Na	0	0
			2	2		
3	A	4	Total	Na	0	0
			4	4		

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C₂H₆O₂).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	B	1	Total	C	O	0	0
			4	2	2		
4	B	1	Total	C	O	0	0
			4	2	2		

- Molecule 5 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	B	1	Total	C	O	0	1
			14	8	6		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	329	Total 330	O 330	0	5
6	B	357	Total 357	O 357	0	5

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 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: Haloalkane dehalogenase



- Molecule 1: Haloalkane dehalogenase



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	68.93Å 94.80Å 100.50Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	44.40 – 1.50 44.40 – 1.50	Depositor EDS
% Data completeness (in resolution range)	98.9 (44.40-1.50) 96.3 (44.40-1.50)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	5.56 (at 1.50Å)	Xtriage
Refinement program	PHENIX (phenix.refine: dev_1269)	Depositor
R, R_{free}	0.148 , 0.165 0.147 , 0.164	Depositor DCC
R_{free} test set	3860 reflections (1.90%)	DCC
Wilson B-factor (Å ²)	9.7	Xtriage
Anisotropy	0.044	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.42 , 46.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	10306	wwPDB-VP
Average B, all atoms (Å ²)	13.0	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 55.99 % 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 2.9182e-05. 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: NA, PEG, SCN, EDO

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	0.43	0/2509	0.64	0/3416
1	B	0.42	0/2527	0.64	0/3437
All	All	0.42	0/5036	0.64	0/6853

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	2415	2370	2339	6	0
1	B	2428	2372	2346	15	0
2	A	3	0	0	0	0
2	B	3	0	0	0	0
3	A	4	0	0	0	0
3	B	2	0	0	0	0
4	B	8	0	12	7	0
5	B	14	0	20	10	0
6	A	330	0	0	1	0
6	B	357	0	0	2	0
All	All	5564	4742	4717	24	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 24 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:228:ILE:HD13	5:B:406[B]:PEG:H41	1.52	0.92
1:B:92:ASP:OD1	5:B:406[B]:PEG:H21	1.82	0.78
1:B:228:ILE:CD1	5:B:406[B]:PEG:H41	2.19	0.72
4:B:405:EDO:H11	6:B:644:HOH:O	1.94	0.67
1:B:227:ASN:HB2	5:B:406[A]:PEG:H31	1.77	0.67

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	302/307 (98%)	294 (97%)	8 (3%)	0	100	100
1	B	305/307 (99%)	295 (97%)	9 (3%)	1 (0%)	44	19
All	All	607/614 (99%)	589 (97%)	17 (3%)	1 (0%)	51	24

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	306	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	262/255 (103%)	258 (98%)	4 (2%)	70	42
1	B	264/255 (104%)	262 (99%)	2 (1%)	85	68
All	All	526/510 (103%)	520 (99%)	6 (1%)	75	56

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	178	PHE
1	B	235	TYR
1	A	235	TYR
1	A	153	GLU
1	B	178	PHE

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 [i](#)

Of 12 ligands modelled in this entry, 6 are monoatomic - leaving 6 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
2	SCN	A	401	-	1,2,2	7.32	1 (100%)	0,1,1	0.00	-
2	SCN	B	401	-	1,2,2	7.48	1 (100%)	0,1,1	0.00	-
4	EDO	B	404	-	3,3,3	0.46	0	2,2,2	0.39	0
4	EDO	B	405	-	3,3,3	0.77	0	2,2,2	0.23	0
5	PEG	B	406[A]	-	6,6,6	0.58	0	5,5,5	0.71	0
5	PEG	B	406[B]	-	6,6,6	0.47	0	5,5,5	1.07	1 (20%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	SCN	A	401	-	-	0/0/0/0	0/0/0/0
2	SCN	B	401	-	-	0/0/0/0	0/0/0/0
4	EDO	B	404	-	-	0/1/1/1	0/0/0/0
4	EDO	B	405	-	-	0/1/1/1	0/0/0/0
5	PEG	B	406[A]	-	-	0/4/4/4	0/0/0/0
5	PEG	B	406[B]	-	-	0/4/4/4	0/0/0/0

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	401	SCN	C-N	7.32	1.40	1.15
2	B	401	SCN	C-N	7.48	1.41	1.15

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	B	406[B]	PEG	O2-C3-C4	2.12	119.95	110.15

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 17 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	405	EDO	7	0
5	B	406[A]	PEG	4	0
5	B	406[B]	PEG	6	0

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	289/307 (94%)	-0.16	13 (4%) 34 38	6, 10, 24, 46	0
1	B	291/307 (94%)	-0.29	8 (2%) 55 61	6, 9, 21, 50	0
All	All	580/614 (94%)	-0.22	21 (3%) 43 49	6, 10, 24, 50	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	307	GLY	8.6
1	B	306	SER	7.7
1	A	305	ILE	7.6
1	A	306	SER	6.9
1	A	154	PHE	5.4

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 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, 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	LLDF	B-factors(\AA^2)	Q<0.9
5	PEG	B	406[A]	7/7	0.88	0.35	47.85	25,27,28,31	7
5	PEG	B	406[B]	7/7	0.88	0.35	45.44	21,29,33,34	7
3	NA	A	405	1/1	0.41	2.84	40.27	60,60,60,60	0
4	EDO	B	405	4/4	0.86	0.24	12.29	18,24,29,35	0
4	EDO	B	404	4/4	0.93	0.13	8.55	16,19,20,28	0
2	SCN	B	401	3/3	0.99	0.08	1.55	9,9,15,24	0
2	SCN	A	401	3/3	1.00	0.07	0.09	9,9,21,22	0
3	NA	A	403	1/1	0.98	0.08	-0.22	19,19,19,19	0
3	NA	B	403	1/1	0.94	0.11	-	33,33,33,33	0
3	NA	A	402	1/1	0.98	0.09	-	24,24,24,24	0
3	NA	B	402	1/1	0.94	0.10	-	28,28,28,28	0
3	NA	A	404	1/1	0.87	0.15	-	28,28,28,28	0

6.5 Other polymers [i](#)

There are no such residues in this entry.