wwPDB X-ray Structure Validation Summary Report (i)

May 14, 2020 – 03:53 pm BST

PDB ID	:	2VSY
Title	:	Xanthomonas campestris putative OGT (XCC0866), apostructure
Authors	:	Schuettelkopf, A.W.; Clarke, A.J.; van Aalten, D.M.F.
Deposited on	:	2008-05-01
Resolution	:	2.10 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:

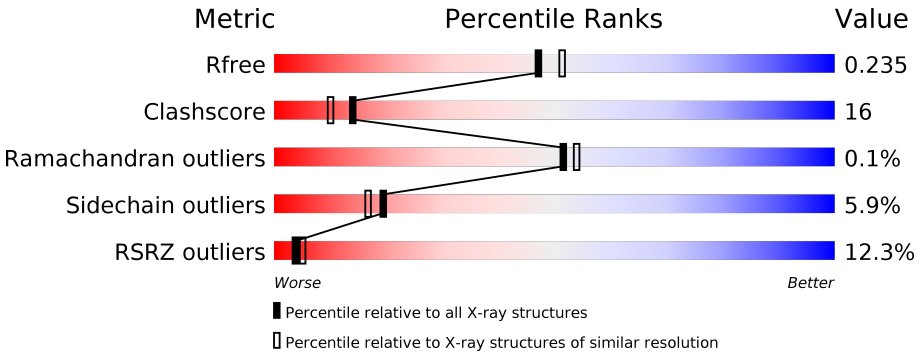
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.10 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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	568	11% 74% 20%
1	B	568	13% 74% 20%

2 Entry composition

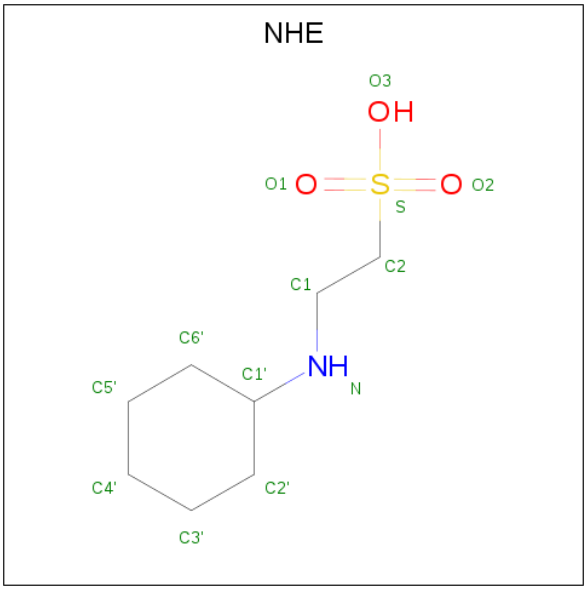
There are 6 unique types of molecules in this entry. The entry contains 8733 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 XCC0866.

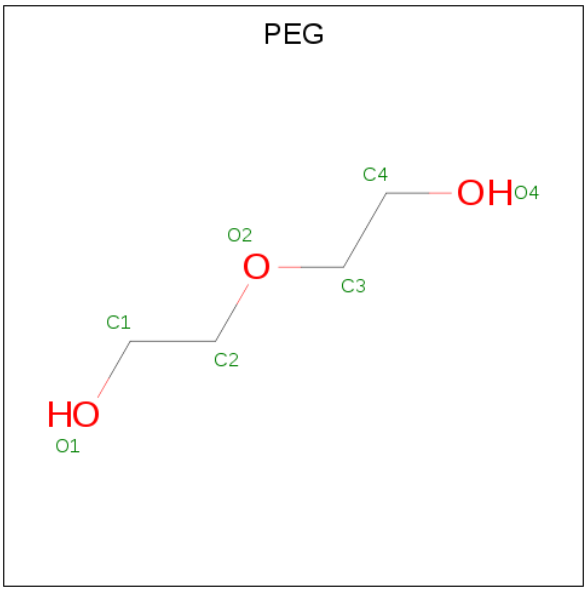
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	547	Total	C	N	O	S	0	1	0
			4166	2632	770	748	16			
1	B	547	Total	C	N	O	S	0	4	0
			4189	2648	778	747	16			

- Molecule 2 is 2-[N-CYCLOHEXYLAMINO]ETHANE SULFONIC ACID (three-letter code: NHE) (formula: $C_8H_{17}NO_3S$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	S	0	0
			13	8	1	3	1		
2	B	1	Total	C	N	O	S	0	0
			13	8	1	3	1		
2	B	1	Total	C	N	O	S	0	0
			13	8	1	3	1		

- Molecule 3 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 7 4 3	0	0
3	A	1	Total C O 7 4 3	0	0

- Molecule 4 is PRASEODYMIUM ION (three-letter code: PR) (formula: Pr).

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

- Molecule 5 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	B	2	Total Cl 2 2	0	0

- Molecule 6 is water.

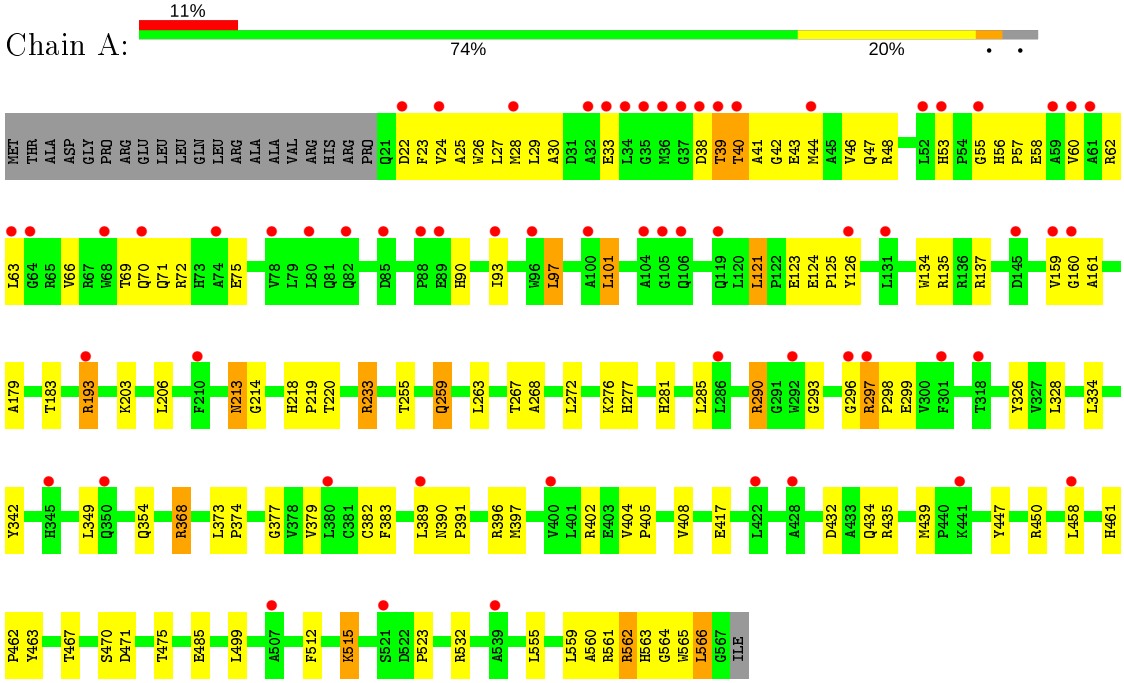
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	153	Total O 153 153	0	0
6	B	167	Total O 167 167	0	0

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 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.

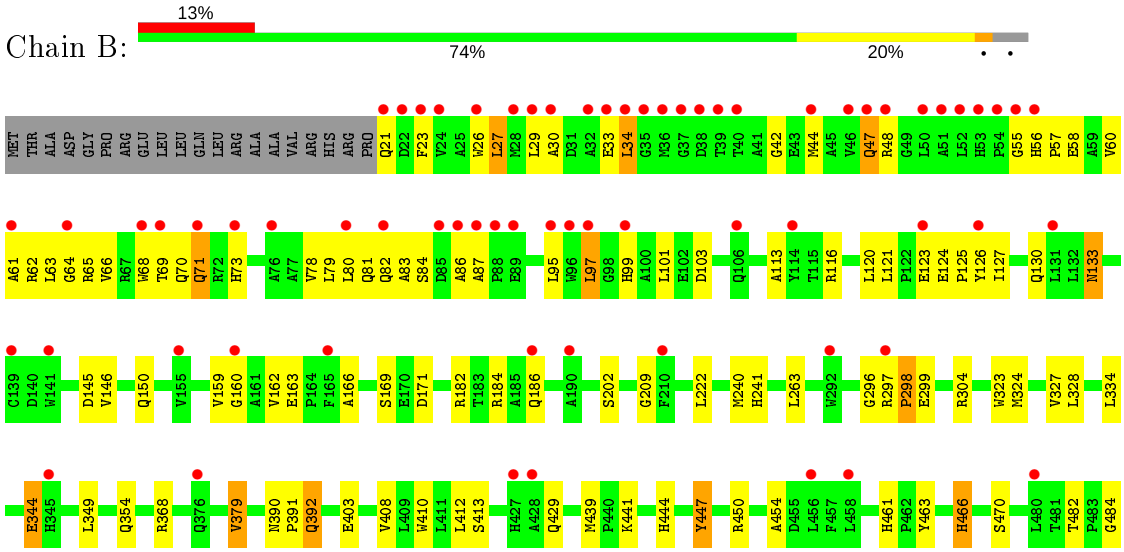
- Molecule 1: XCC0866

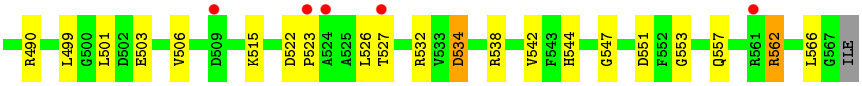
Chain A:



- Molecule 1: XCC0866

Chain B:




R490
L499
G500
L501
D502
E503
V506
D509
K515
D522
P523
A524
A525
L526
T527
R532
V533
D534
R538
V542
F543
H544
G547
D551
F552
G553
Q557
R561
R562
L566
G567
ILE

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	81.52Å 100.10Å 156.27Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.10 19.97 – 2.10	Depositor EDS
% Data completeness (in resolution range)	93.4 (20.00-2.10) 93.4 (19.97-2.10)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.93 (at 2.11Å)	Xtriage
Refinement program	REFMAC 5.4.0069	Depositor
R, R_{free}	0.193 , 0.236 0.194 , 0.235	Depositor DCC
R_{free} test set	1399 reflections (2.00%)	wwPDB-VP
Wilson B-factor (Å^2)	30.4	Xtriage
Anisotropy	0.819	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 71.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	8733	wwPDB-VP
Average B, all atoms (Å^2)	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.98% 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: PR, NHE, PEG, CL

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	0.69	0/4278	0.80	7/5839 (0.1%)
1	B	0.76	0/4311	0.80	2/5882 (0.0%)
All	All	0.73	0/8589	0.80	9/11721 (0.1%)

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	135	ARG	NE-CZ-NH1	-6.85	116.87	120.30
1	A	290	ARG	NE-CZ-NH1	6.71	123.65	120.30
1	A	290	ARG	NE-CZ-NH2	-6.71	116.95	120.30
1	A	368	ARG	NE-CZ-NH1	-6.38	117.11	120.30
1	A	368	ARG	NE-CZ-NH2	6.02	123.31	120.30

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	4166	0	4089	132	0
1	B	4189	0	4124	128	0
2	A	13	0	17	1	0
2	B	26	0	34	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	14	0	20	1	0
4	A	2	0	0	0	0
4	B	1	0	0	0	0
5	B	2	0	0	0	0
6	A	153	0	0	3	0
6	B	167	0	0	5	0
All	All	8733	0	8284	261	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 16.

The worst 5 of 261 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:97:LEU:C	1:B:97:LEU:HD23	1.65	1.17
1:B:34:LEU:HD23	1:B:34:LEU:N	1.55	1.16
1:B:78:VAL:O	1:B:82:GLN:HG3	1.47	1.15
1:A:555:LEU:C	1:A:555:LEU:HD23	1.68	1.14
1:B:562:ARG:CG	1:B:562:ARG:HH11	1.63	1.10

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	546/568 (96%)	533 (98%)	12 (2%)	1 (0%)	47 49
1	B	549/568 (97%)	530 (96%)	19 (4%)	0	100 100
All	All	1095/1136 (96%)	1063 (97%)	31 (3%)	1 (0%)	51 54

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	39	THR

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	413/429 (96%)	392 (95%)	21 (5%)	24 22
1	B	416/429 (97%)	388 (93%)	28 (7%)	16 13
All	All	829/858 (97%)	780 (94%)	49 (6%)	19 17

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	27	LEU
1	B	97	LEU
1	B	466	HIS
1	B	47	GLN
1	B	116	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 21 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	281	HIS
1	A	392	GLN
1	B	281	HIS
1	A	277	HIS
1	B	385	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 10 ligands modelled in this entry, 5 are monoatomic - leaving 5 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	$\#\|Z\| > 2$
3	PEG	A	1569	-	6,6,6	0.71	0	5,5,5	0.71	0
3	PEG	A	1570	-	6,6,6	0.62	0	5,5,5	1.52	1 (20%)
2	NHE	A	1568	-	13,13,13	1.87	3 (23%)	16,17,17	1.70	5 (31%)
2	NHE	B	1568	-	13,13,13	1.72	3 (23%)	16,17,17	1.34	2 (12%)
2	NHE	B	1569	-	13,13,13	2.22	4 (30%)	16,17,17	2.73	4 (25%)

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
3	PEG	A	1569	-	-	3/4/4/4	-
3	PEG	A	1570	-	-	2/4/4/4	-
2	NHE	A	1568	-	-	2/7/15/15	0/1/1/1
2	NHE	B	1568	-	-	1/7/15/15	0/1/1/1
2	NHE	B	1569	-	-	1/7/15/15	0/1/1/1

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1568	NHE	C2-S	5.03	1.84	1.77
2	B	1569	NHE	C2-S	4.77	1.84	1.77
2	B	1569	NHE	O1-S	4.38	1.57	1.45
2	B	1569	NHE	O2-S	3.93	1.56	1.45
2	B	1568	NHE	O1-S	3.71	1.56	1.45

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	1569	NHE	O1-S-C2	8.86	117.58	106.92
2	B	1568	NHE	O2-S-C2	3.41	111.02	106.92
2	B	1569	NHE	C6'-C1'-C2'	-3.37	104.98	110.82
3	A	1570	PEG	C3-O2-C2	-3.33	98.88	113.29
2	A	1568	NHE	C5'-C6'-C1'	-3.30	104.89	111.11

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	1569	NHE	N-C1-C2-S
2	A	1568	NHE	C2'-C1'-N-C1
3	A	1569	PEG	O1-C1-C2-O2
3	A	1570	PEG	C1-C2-O2-C3
3	A	1570	PEG	O1-C1-C2-O2

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1570	PEG	1	0
2	A	1568	NHE	1	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

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	547/568 (96%)	0.74	62 (11%)	5 6	29, 38, 48, 58	2 (0%)
1	B	547/568 (96%)	0.82	73 (13%)	3 4	28, 38, 49, 60	2 (0%)
All	All	1094/1136 (96%)	0.78	135 (12%)	4 5	28, 38, 48, 60	4 (0%)

The worst 5 of 135 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	37	GLY	7.8
1	B	29	LEU	7.5
1	B	35	GLY	7.4
1	B	38	ASP	7.3
1	B	55	GLY	7.0

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

There are no carbohydrates 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	NHE	B	1569	13/13	0.86	0.28	34,46,48,48	0
3	PEG	A	1570	7/7	0.87	0.18	51,52,53,55	0
4	PR	A	1572	1/1	0.92	0.07	54,54,54,54	1
5	CL	B	1570	1/1	0.94	0.06	66,66,66,66	0
3	PEG	A	1569	7/7	0.94	0.23	48,51,58,58	0
5	CL	B	1571	1/1	0.94	0.10	72,72,72,72	0
2	NHE	B	1568	13/13	0.97	0.19	37,38,42,42	0
2	NHE	A	1568	13/13	0.97	0.13	36,41,43,44	0
4	PR	A	1571	1/1	0.99	0.11	36,36,36,36	1
4	PR	B	1572	1/1	0.99	0.13	57,57,57,57	1

6.5 Other polymers (i)

There are no such residues in this entry.