



Full wwPDB X-ray Structure Validation Report ⓘ

Mar 8, 2026 – 05:42 PM UTC

PDB ID : 4CGK / pdb_00004cgc
Title : Crystal structure of the essential protein PcsB from *Streptococcus pneumoniae*
Authors : Bartual, S.G.; Straume, D.; Stamsas, G.A.; Alfonso, C.; Martinez-Ripoll, M.; Havarstein, L.S.; Hermoso, J.A.
Deposited on : 2013-11-25
Resolution : 2.55 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	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

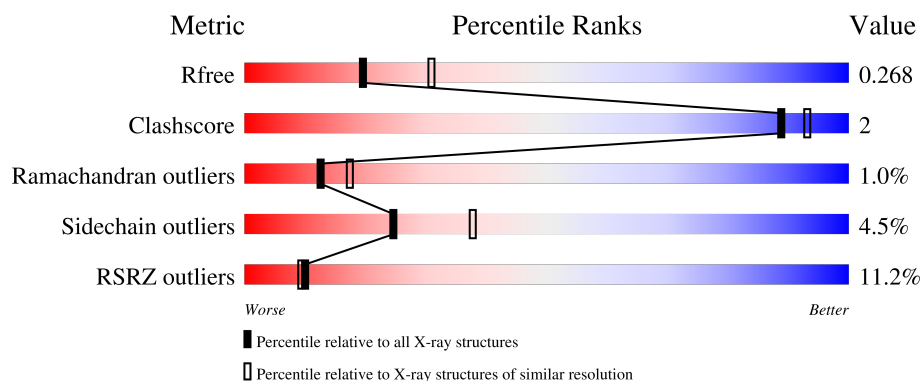
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.55 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}	180053	1091 (2.54-2.54)
Clashscore	190562	1120 (2.54-2.54)
Ramachandran outliers	187476	1106 (2.54-2.54)
Sidechain outliers	187428	1106 (2.54-2.54)
RSRZ outliers	180081	1091 (2.54-2.54)

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	392	8% 84% 5% 10%
1	B	392	12% 80% 8% 11%

2 Entry composition [i](#)

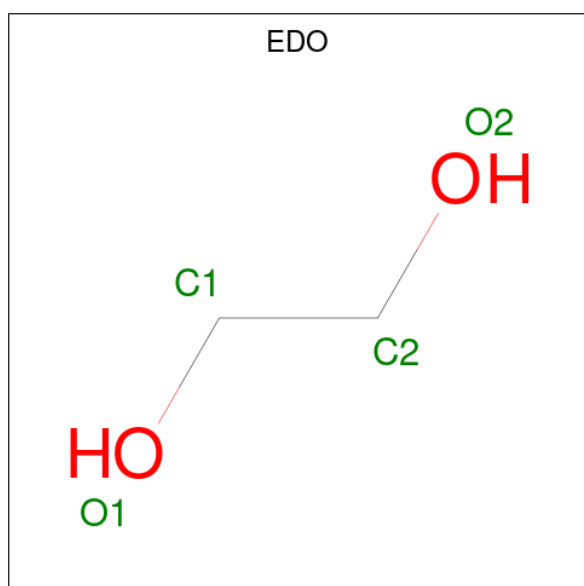
There are 6 unique types of molecules in this entry. The entry contains 5709 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 SECRETED 45 KDA PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	351	Total	C	N	O	S	0	0	0
			2624	1601	472	547	4			
1	B	348	Total	C	N	O	S	0	1	0
			2607	1591	469	543	4			

- Molecule 2 is 1,2-ETHANEDIOL (CCD ID: EDO) (formula: C₂H₆O₂).



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

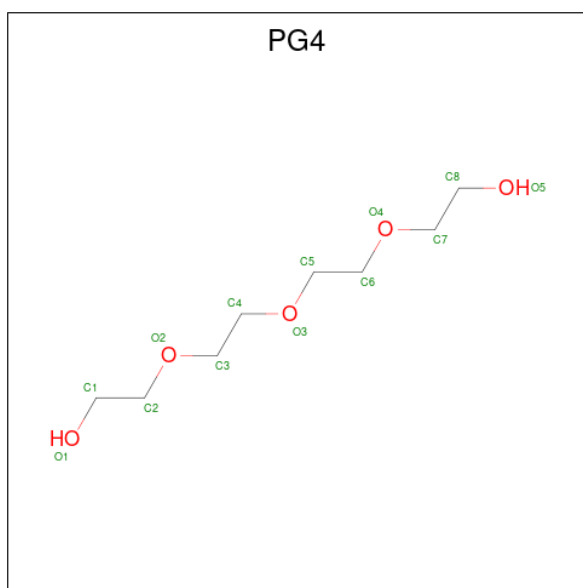
- Molecule 3 is CHLORIDE ION (CCD ID: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	7	Total	Cl	0	0
			7	7		
3	B	1	Total	Cl	0	0
			1	1		

- Molecule 4 is MAGNESIUM ION (CCD ID: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Mg	0	0
			1	1		

- Molecule 5 is TETRAETHYLENE GLYCOL (CCD ID: PG4) (formula: C₈H₁₈O₅).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	B	1	Total	C	O	0	0
			13	8	5		

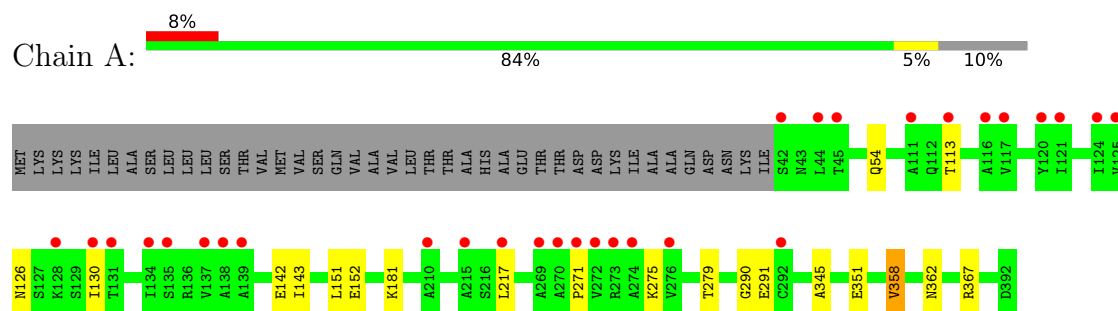
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	256	Total	O	0	0
			256	256		
6	B	192	Total	O	0	0
			192	192		

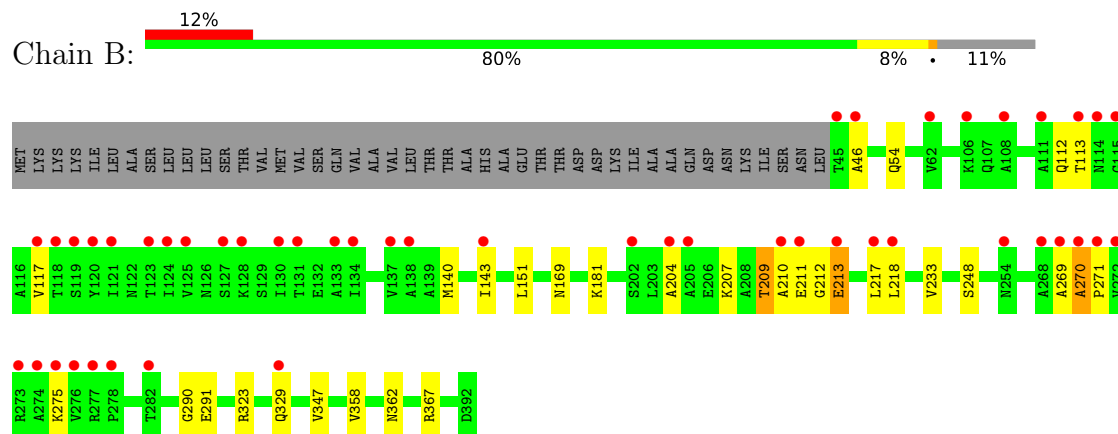
3 Residue-property plots

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: SECRETED 45 KDA PROTEIN



• Molecule 1: SECRETED 45 KDA PROTEIN



4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α , β , γ	125.81Å 125.81Å 126.64Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	44.63 – 2.55 44.63 – 2.55	Depositor EDS
% Data completeness (in resolution range)	99.8 (44.63-2.55) 99.9 (44.63-2.55)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.60 (at 2.54Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, R_{free}	0.238 , 0.272 0.238 , 0.268	Depositor DCC
R_{free} test set	1907 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	43.1	Xtriage
Anisotropy	1.009	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 36.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.027 for -h,-k,l	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	5709	wwPDB-VP
Average B, all atoms (Å ²)	76.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.66% of the height of the origin peak. No significant pseudotranslation is detected.*

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: EDO, CL, MG, PG4

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	0/2655	0.71	0/3602
1	B	0.41	0/2638	0.72	0/3579
All	All	0.42	0/5293	0.71	0/7181

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	2624	0	2569	3	0
1	B	2607	0	2548	17	0
2	A	8	0	12	0	0
3	A	7	0	0	0	0
3	B	1	0	0	0	0
4	B	1	0	0	0	0
5	B	13	0	18	0	0
6	A	256	0	0	2	0
6	B	192	0	0	1	0
All	All	5709	0	5147	20	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.

All (20) 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:269:ALA:HA	1:B:270:ALA:HB3	1.65	0.78
1:B:211:GLU:HB3	1:B:213:GLU:N	2.16	0.61
6:A:2207:HOH:O	1:B:323:ARG:NH1	2.42	0.53
6:A:2205:HOH:O	1:B:329[A]:GLN:NE2	2.41	0.52
1:B:211:GLU:CG	1:B:212:GLY:HA3	2.39	0.52
1:B:112:GLN:HA	1:B:117:VAL:HG22	1.92	0.52
1:B:211:GLU:HB3	1:B:212:GLY:CA	2.40	0.51
1:A:345:ALA:HB1	1:A:358:VAL:HG13	1.94	0.50
1:B:211:GLU:CB	1:B:212:GLY:HA3	2.42	0.49
1:B:112:GLN:HA	1:B:117:VAL:CG2	2.43	0.48
1:B:347:VAL:HA	1:B:358:VAL:HG12	1.95	0.48
1:B:211:GLU:HB3	1:B:212:GLY:C	2.39	0.46
1:A:291:GLU:OE2	1:A:367:ARG:NH2	2.50	0.43
1:B:290:GLY:O	1:B:362:ASN:HB2	2.19	0.43
1:A:290:GLY:O	1:A:362:ASN:HB2	2.19	0.42
1:B:169:ASN:ND2	6:B:2059:HOH:O	2.52	0.42
1:B:291:GLU:OE2	1:B:367:ARG:NH2	2.51	0.42
1:B:209:THR:N	1:B:210:ALA:HA	2.35	0.42
1:B:204:ALA:O	1:B:207:LYS:HG2	2.19	0.42
1:B:211:GLU:CB	1:B:212:GLY:CA	2.98	0.41

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	349/392 (89%)	334 (96%)	12 (3%)	3 (1%)	14	20
1	B	347/392 (88%)	331 (95%)	12 (4%)	4 (1%)	10	13

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
All	All	696/784 (89%)	665 (96%)	24 (3%)	7 (1%)	12	17

All (7) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	209	THR
1	A	126	ASN
1	B	270	ALA
1	B	46	ALA
1	A	130	ILE
1	A	271	PRO
1	B	271	PRO

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	269/304 (88%)	257 (96%)	12 (4%)	24	38
1	B	266/304 (88%)	254 (96%)	12 (4%)	24	38
All	All	535/608 (88%)	511 (96%)	24 (4%)	24	38

All (24) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	54	GLN
1	A	113	THR
1	A	142	GLU
1	A	143	ILE
1	A	151	LEU
1	A	152	GLU
1	A	181	LYS
1	A	217	LEU
1	A	275	LYS
1	A	279	THR
1	A	351	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	358	VAL
1	B	54	GLN
1	B	113	THR
1	B	140	MET
1	B	143	ILE
1	B	151	LEU
1	B	181	LYS
1	B	213	GLU
1	B	217	LEU
1	B	218	LEU
1	B	233	VAL
1	B	248	SER
1	B	275	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (12) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	52	GLN
1	A	54	GLN
1	A	59	GLN
1	A	61	GLN
1	A	147	ASN
1	A	372	HIS
1	B	52	GLN
1	B	54	GLN
1	B	59	GLN
1	B	147	ASN
1	B	165	GLN
1	B	372	HIS

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 oligosaccharides in this entry.

5.6 Ligand geometry

Of 12 ligands modelled in this entry, 9 are monoatomic - leaving 3 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	$\# Z > 2$
5	PG4	B	1394	-	12,12,12	0.48	0	11,11,11	0.20	0
2	EDO	A	1393	-	3,3,3	0.45	0	2,2,2	0.28	0
2	EDO	A	1394	-	3,3,3	0.42	0	2,2,2	0.32	0

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
5	PG4	B	1394	-	-	7/10/10/10	-
2	EDO	A	1393	-	-	1/1/1/1	-
2	EDO	A	1394	-	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (9) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	B	1394	PG4	O1-C1-C2-O2
5	B	1394	PG4	O4-C7-C8-O5
5	B	1394	PG4	O3-C5-C6-O4
5	B	1394	PG4	O2-C3-C4-O3
2	A	1393	EDO	O1-C1-C2-O2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
2	A	1394	EDO	O1-C1-C2-O2
5	B	1394	PG4	C1-C2-O2-C3
5	B	1394	PG4	C8-C7-O4-C6
5	B	1394	PG4	C4-C3-O2-C2

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

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, 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	351/392 (89%)	0.47	30 (8%) 16 16	35, 58, 186, 212	1 (0%)
1	B	348/392 (88%)	0.63	48 (13%) 6 5	24, 63, 193, 225	1 (0%)
All	All	699/784 (89%)	0.55	78 (11%) 10 9	24, 60, 190, 225	2 (0%)

All (78) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	271	PRO	5.2
1	B	210	ALA	4.8
1	A	269	ALA	4.6
1	B	272	VAL	4.5
1	B	273	ARG	4.4
1	A	121	ILE	4.2
1	B	276	VAL	4.2
1	B	274	ALA	4.1
1	B	45	THR	4.1
1	A	292	CYS	4.1
1	A	272	VAL	4.0
1	B	121	ILE	4.0
1	A	125	VAL	3.7
1	B	111	ALA	3.6
1	B	115	GLY	3.6
1	B	205	ALA	3.5
1	B	270	ALA	3.4
1	B	128	LYS	3.4
1	A	271	PRO	3.4
1	A	134	ILE	3.3
1	A	210	ALA	3.3
1	A	116	ALA	3.3
1	B	120	TYR	3.3
1	A	273	ARG	3.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	130	ILE	3.3
1	B	269	ALA	3.3
1	A	44	LEU	3.2
1	A	124	ILE	3.2
1	B	125	VAL	3.1
1	A	130	ILE	3.1
1	A	111	ALA	3.1
1	B	268	ALA	3.1
1	B	217	LEU	3.0
1	A	276	VAL	2.9
1	B	134	ILE	2.9
1	A	45	THR	2.9
1	B	114	ASN	2.8
1	B	137	VAL	2.8
1	A	270	ALA	2.8
1	B	143	ILE	2.8
1	A	42	SER	2.8
1	B	282	THR	2.8
1	B	124	ILE	2.8
1	B	119	SER	2.8
1	B	118	THR	2.8
1	B	254	ASN	2.7
1	B	46	ALA	2.7
1	B	278	PRO	2.7
1	B	117	VAL	2.7
1	B	277	ARG	2.7
1	A	113	THR	2.6
1	B	131	THR	2.6
1	A	117	VAL	2.6
1	A	137	VAL	2.5
1	B	113	THR	2.5
1	B	329[A]	GLN	2.4
1	A	139	ALA	2.4
1	B	275	LYS	2.4
1	A	120	TYR	2.3
1	A	128	LYS	2.3
1	B	211	GLU	2.3
1	A	131	THR	2.3
1	A	217	LEU	2.3
1	A	274	ALA	2.3
1	A	135	SER	2.3
1	B	127	SER	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	106	LYS	2.3
1	B	213	GLU	2.3
1	A	138	ALA	2.3
1	A	215	ALA	2.2
1	B	123	THR	2.2
1	B	218	LEU	2.2
1	B	138	ALA	2.1
1	B	204	ALA	2.1
1	B	202	SER	2.1
1	B	62	VAL	2.0
1	B	133	ALA	2.0
1	B	108	ALA	2.0

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 oligosaccharides 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
5	PG4	B	1394	13/13	0.77	0.19	76,82,85,85	0
3	CL	A	1401	1/1	0.79	0.15	105,105,105,105	0
2	EDO	A	1394	4/4	0.79	0.31	82,83,83,86	0
3	CL	A	1399	1/1	0.81	0.31	128,128,128,128	0
2	EDO	A	1393	4/4	0.83	0.21	82,83,83,84	0
3	CL	B	1395	1/1	0.88	0.15	89,89,89,89	0
3	CL	A	1395	1/1	0.89	0.13	85,85,85,85	0
3	CL	A	1397	1/1	0.89	0.14	81,81,81,81	0
4	MG	B	1393	1/1	0.90	0.10	75,75,75,75	0
3	CL	A	1396	1/1	0.94	0.17	85,85,85,85	0
3	CL	A	1398	1/1	0.95	0.08	80,80,80,80	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($\text{\AA}^2$)	Q<0.9
3	CL	A	1400	1/1	0.96	0.08	76,76,76,76	0

6.5 Other polymers [i](#)

There are no such residues in this entry.