Full wwPDB X-ray Structure Validation Report ⓘ

May 29, 2020 – 07:54 am BST

PDB ID	:	5NZ7
Title	:	Clostridium thermocellum cellodextrin phosphorylase ligand free form
Authors	:	O'Neill, E.C.; Pergolizzi, G.; Stevenson, C.E.M.; Lawson, D.M.; Nepogodiev, S.A.; Field, R.A.
Deposited on	:	2017-05-12
Resolution	:	2.30 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 following versions of software and data (see references ⓘ) were used in the production of this report:

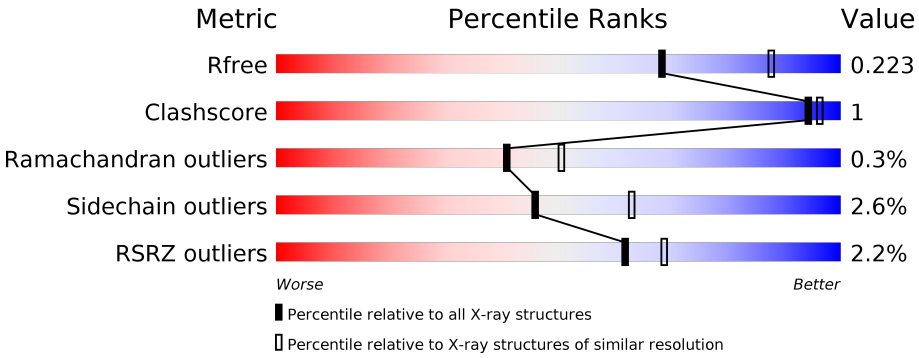
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance

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

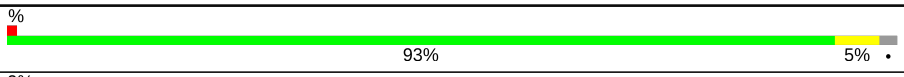
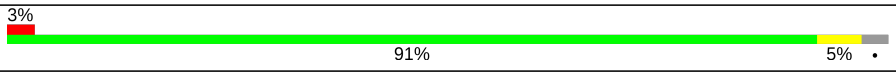
The reported resolution of this entry is 2.30 Å.

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	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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	1009	
1	B	1009	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 15347 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 Cellodextrin phosphorylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	984	Total	C	N	O	S	0	0	0
			7596	4878	1259	1427	32			
1	B	978	Total	C	N	O	S	0	0	0
			7512	4819	1252	1409	32			

There are 50 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-24	MET	-	initiating methionine	UNP Q93HT8
A	-23	GLY	-	expression tag	UNP Q93HT8
A	-22	SER	-	expression tag	UNP Q93HT8
A	-21	SER	-	expression tag	UNP Q93HT8
A	-20	HIS	-	expression tag	UNP Q93HT8
A	-19	HIS	-	expression tag	UNP Q93HT8
A	-18	HIS	-	expression tag	UNP Q93HT8
A	-17	HIS	-	expression tag	UNP Q93HT8
A	-16	HIS	-	expression tag	UNP Q93HT8
A	-15	HIS	-	expression tag	UNP Q93HT8
A	-14	SER	-	expression tag	UNP Q93HT8
A	-13	SER	-	expression tag	UNP Q93HT8
A	-12	GLY	-	expression tag	UNP Q93HT8
A	-11	LEU	-	expression tag	UNP Q93HT8
A	-10	VAL	-	expression tag	UNP Q93HT8
A	-9	PRO	-	expression tag	UNP Q93HT8
A	-8	ARG	-	expression tag	UNP Q93HT8
A	-7	GLY	-	expression tag	UNP Q93HT8
A	-6	SER	-	expression tag	UNP Q93HT8
A	-5	HIS	-	expression tag	UNP Q93HT8
A	-4	MET	-	expression tag	UNP Q93HT8
A	-3	LEU	-	expression tag	UNP Q93HT8
A	-2	GLU	-	expression tag	UNP Q93HT8
A	-1	ASP	-	expression tag	UNP Q93HT8
A	0	PRO	-	expression tag	UNP Q93HT8

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-24	MET	-	initiating methionine	UNP Q93HT8
B	-23	GLY	-	expression tag	UNP Q93HT8
B	-22	SER	-	expression tag	UNP Q93HT8
B	-21	SER	-	expression tag	UNP Q93HT8
B	-20	HIS	-	expression tag	UNP Q93HT8
B	-19	HIS	-	expression tag	UNP Q93HT8
B	-18	HIS	-	expression tag	UNP Q93HT8
B	-17	HIS	-	expression tag	UNP Q93HT8
B	-16	HIS	-	expression tag	UNP Q93HT8
B	-15	HIS	-	expression tag	UNP Q93HT8
B	-14	SER	-	expression tag	UNP Q93HT8
B	-13	SER	-	expression tag	UNP Q93HT8
B	-12	GLY	-	expression tag	UNP Q93HT8
B	-11	LEU	-	expression tag	UNP Q93HT8
B	-10	VAL	-	expression tag	UNP Q93HT8
B	-9	PRO	-	expression tag	UNP Q93HT8
B	-8	ARG	-	expression tag	UNP Q93HT8
B	-7	GLY	-	expression tag	UNP Q93HT8
B	-6	SER	-	expression tag	UNP Q93HT8
B	-5	HIS	-	expression tag	UNP Q93HT8
B	-4	MET	-	expression tag	UNP Q93HT8
B	-3	LEU	-	expression tag	UNP Q93HT8
B	-2	GLU	-	expression tag	UNP Q93HT8
B	-1	ASP	-	expression tag	UNP Q93HT8
B	0	PRO	-	expression tag	UNP Q93HT8

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

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

- Molecule 3 is water.

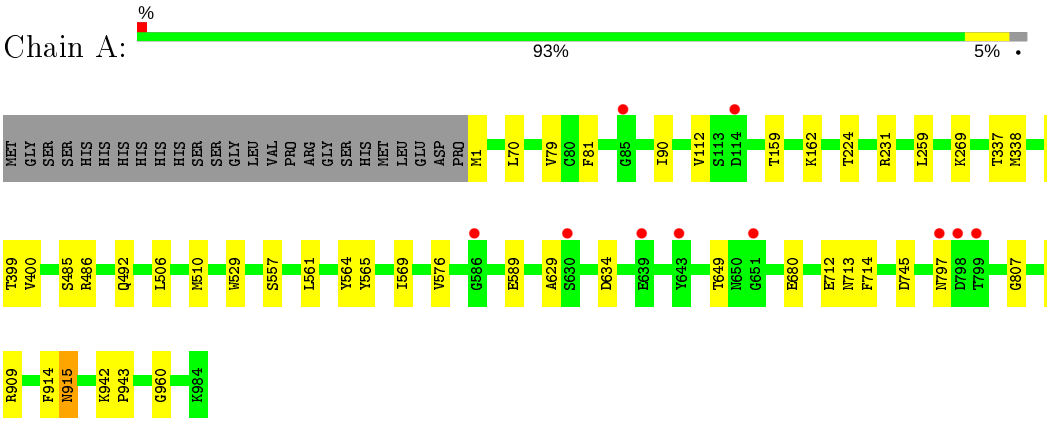
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	119	Total O 119 119	0	0
3	B	114	Total O 114 114	0	0

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.

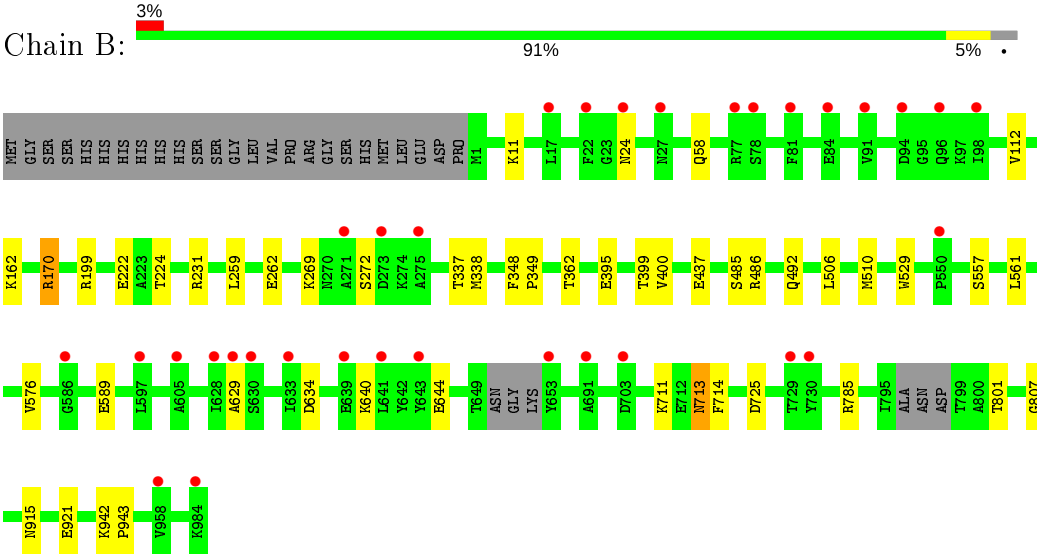
- Molecule 1: Cellodextrin phosphorylase

Chain A:



- Molecule 1: Cellodextrin phosphorylase

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	84.64Å 151.81Å 91.97Å 90.00° 114.60° 90.00°	Depositor
Resolution (Å)	47.65 – 2.30 47.61 – 2.30	Depositor EDS
% Data completeness (in resolution range)	99.5 (47.65-2.30) 99.6 (47.61-2.30)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.49 (at 2.29Å)	Xtriage
Refinement program	REFMAC 5	Depositor
R, R_{free}	0.191 , 0.223 0.191 , 0.223	Depositor DCC
R_{free} test set	4668 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	55.4	Xtriage
Anisotropy	0.303	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 38.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.025 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	15347	wwPDB-VP
Average B, all atoms (Å^2)	66.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.16% 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: 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	0.44	0/7770	0.59	0/10569
1	B	0.43	1/7683 (0.0%)	0.60	1/10449 (0.0%)
All	All	0.43	1/15453 (0.0%)	0.59	1/21018 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	272	SER	CB-OG	5.60	1.49	1.42

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	170	ARG	NE-CZ-NH1	5.24	122.92	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	7596	0	7189	21	0
1	B	7512	0	7064	16	0
2	A	3	0	0	0	0
2	B	3	0	0	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	119	0	0	1	0
3	B	114	0	0	0	0
All	All	15347	0	14253	37	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 1.

All (37) 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:199:ARG:NE	1:B:262:GLU:OE2	2.21	0.69
1:A:680:GLU:HG2	1:A:909:ARG:HD3	1.77	0.66
1:A:269:LYS:CB	1:A:399:THR:HG22	2.35	0.56
1:B:269:LYS:CB	1:B:399:THR:HG22	2.37	0.55
1:A:112:VAL:HG12	1:A:112:VAL:O	2.07	0.55
1:A:79:VAL:CG1	1:A:90:ILE:HD11	2.37	0.54
1:A:506:LEU:O	1:A:510:MET:HB2	2.08	0.53
1:B:170:ARG:HG2	1:B:170:ARG:HH11	1.73	0.53
1:B:112:VAL:HG12	1:B:112:VAL:O	2.09	0.53
1:B:506:LEU:O	1:B:510:MET:HB2	2.08	0.52
1:B:486:ARG:O	1:B:492:GLN:HG2	2.10	0.51
1:A:486:ARG:O	1:A:492:GLN:HG2	2.14	0.48
1:A:392:PHE:CD2	1:A:400:VAL:HG21	2.49	0.48
1:A:79:VAL:HG11	1:A:90:ILE:HD11	1.95	0.47
1:A:564:TYR:HE1	1:A:576:VAL:CG1	2.29	0.46
1:A:81:PHE:CE1	1:A:90:ILE:HD13	2.51	0.46
1:B:564:TYR:HE1	1:B:576:VAL:CG1	2.29	0.45
1:B:565:TYR:CZ	1:B:569:ILE:HG13	2.52	0.45
1:A:337:THR:HG22	1:A:338:MET:N	2.32	0.45
1:A:81:PHE:CE1	1:A:90:ILE:CD1	3.00	0.45
1:B:337:THR:HG22	1:B:338:MET:N	2.32	0.45
1:B:889:SER:HB2	2:B:1003:CL:CL	2.54	0.45
1:A:942:LYS:HG3	1:A:943:PRO:O	2.18	0.44
1:B:942:LYS:HG3	1:B:943:PRO:O	2.18	0.44
1:A:70:LEU:HD21	1:A:81:PHE:CE2	2.53	0.43
1:A:159:THR:HB	1:A:259:LEU:HD11	2.00	0.43
1:B:159:THR:HB	1:B:259:LEU:HD11	2.01	0.43
1:B:711:LYS:O	1:B:713:ASN:N	2.44	0.43
1:A:565:TYR:CZ	1:A:569:ILE:HG13	2.54	0.42
1:A:915:ASN:ND2	3:A:1105:HOH:O	2.51	0.42
1:A:81:PHE:CD1	1:A:90:ILE:HD12	2.54	0.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:348:PHE:HB3	1:B:349:PRO:CD	2.50	0.41
1:A:745:ASP:C	1:A:745:ASP:OD1	2.59	0.41
1:B:807:GLY:O	1:B:877:ASN:HA	2.21	0.41
1:A:807:GLY:O	1:A:877:ASN:HA	2.21	0.41
1:B:640:LYS:O	1:B:644:GLU:HG3	2.21	0.41
1:A:904:GLY:O	1:A:914:PHE:HA	2.22	0.40

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	982/1009 (97%)	940 (96%)	39 (4%)	3 (0%)	41	50
1	B	972/1009 (96%)	932 (96%)	37 (4%)	3 (0%)	41	50
All	All	1954/2018 (97%)	1872 (96%)	76 (4%)	6 (0%)	41	50

All (6) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	629	ALA
1	A	649	THR
1	B	11	LYS
1	B	629	ALA
1	A	960	GLY
1	B	725	ASP

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	768/872 (88%)	752 (98%)	16 (2%)	53 70
1	B	753/872 (86%)	730 (97%)	23 (3%)	40 55
All	All	1521/1744 (87%)	1482 (97%)	39 (3%)	46 63

All (39) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1	MET
1	A	162	LYS
1	A	224	THR
1	A	231	ARG
1	A	362	THR
1	A	485	SER
1	A	529	TRP
1	A	557	SER
1	A	561	LEU
1	A	589	GLU
1	A	634	ASP
1	A	712	GLU
1	A	713	ASN
1	A	714	PHE
1	A	797	ASN
1	A	915	ASN
1	B	24	ASN
1	B	58	GLN
1	B	155	ARG
1	B	162	LYS
1	B	222	GLU
1	B	224	THR
1	B	231	ARG
1	B	362	THR
1	B	395	GLU
1	B	400	VAL
1	B	437	GLU
1	B	485	SER
1	B	529	TRP
1	B	557	SER
1	B	561	LEU
1	B	589	GLU
1	B	634	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	713	ASN
1	B	714	PHE
1	B	785	ARG
1	B	801	THR
1	B	915	ASN
1	B	921	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (7) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	797	ASN
1	A	915	ASN
1	B	58	GLN
1	B	373	ASN
1	B	452	ASN
1	B	793	ASN
1	B	915	ASN

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 6 ligands modelled in this entry, 6 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	984/1009 (97%)	-0.15	10 (1%)	82	86	39, 61, 92, 128	0
1	B	978/1009 (96%)	0.13	33 (3%)	45	52	38, 68, 107, 131	0
All	All	1962/2018 (97%)	-0.01	43 (2%)	62	69	38, 64, 101, 131	0

All (43) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	84	GLU	4.1
1	B	81	PHE	3.9
1	B	629	ALA	3.8
1	B	630	SER	3.7
1	B	22	PHE	3.7
1	A	630	SER	3.4
1	B	984	LYS	3.3
1	B	605	ALA	3.2
1	B	550	PRO	3.1
1	B	633	ILE	3.0
1	B	643	TYR	2.9
1	B	273	ASP	2.9
1	B	628	ILE	2.8
1	B	17	LEU	2.8
1	A	586	GLY	2.7
1	B	98	ILE	2.6
1	A	85	GLY	2.6
1	B	653	TYR	2.6
1	A	797	ASN	2.6
1	B	91	VAL	2.6
1	B	958	VAL	2.6
1	A	643	TYR	2.6
1	B	78	SER	2.5
1	B	729	THR	2.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	94	ASP	2.5
1	A	639	GLU	2.4
1	A	799	THR	2.4
1	B	586	GLY	2.3
1	B	730	TYR	2.3
1	A	651	GLY	2.3
1	A	114	ASP	2.3
1	B	96	GLN	2.3
1	B	639	GLU	2.3
1	B	24	ASN	2.2
1	B	597	LEU	2.2
1	B	275	ALA	2.2
1	B	271	ALA	2.2
1	B	691	ALA	2.1
1	B	77	ARG	2.1
1	B	703	ASP	2.1
1	A	798	ASP	2.1
1	B	27	ASN	2.0
1	B	641	LEU	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 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. 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	CL	A	1003	1/1	0.88	0.16	75,75,75,75	0
2	CL	B	1002	1/1	0.95	0.09	67,67,67,67	0
2	CL	B	1003	1/1	0.98	0.28	72,72,72,72	0
2	CL	A	1002	1/1	0.99	0.07	52,52,52,52	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	CL	A	1001	1/1	0.99	0.12	46,46,46,46	0
2	CL	B	1001	1/1	0.99	0.20	43,43,43,43	0

6.5 Other polymers (i)

There are no such residues in this entry.