



Full wwPDB X-ray Structure Validation Report ⓘ

May 15, 2025 – 12:03 PM EDT

PDB ID : 4OEM / pdb_00004oem
Title : Crystal structure of Cathepsin C in complex with dipeptide substrates
Authors : Zhao, B.; Smallwood, A.; Concha, N.
Deposited on : 2014-01-13
Resolution : 1.52 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 Phenix2.0rc1
Mogul : 2022.3.0, CSD as543be (2022)
Xtriage (Phenix) : 2.0rc1
EDS : 3.0
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4 : 9.0.006 (Gargrove)
Density-Fitness : 1.0.12
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.43.1

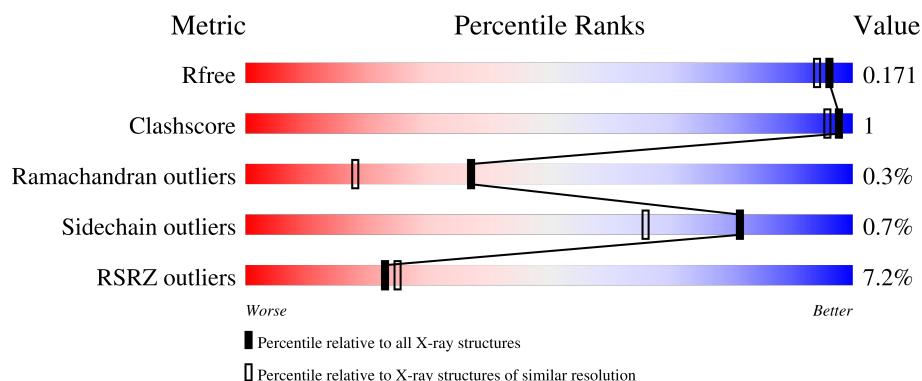
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.52 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}	164625	5293 (1.54-1.50)
Clashscore	180529	5759 (1.54-1.50)
Ramachandran outliers	177936	5653 (1.54-1.50)
Sidechain outliers	177891	5650 (1.54-1.50)
RSRZ outliers	164620	5293 (1.54-1.50)

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 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	370	<div> <div>6%</div> <div> <div></div> <div>74%</div> <div>24%</div> </div> </div>
2	B	70	<div> <div>6%</div> <div> <div></div> <div>94%</div> </div> </div>

2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 5690 atoms, of which 2550 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 Dipeptidyl peptidase 1 Heavy chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	280	Total	C	H	N	O	S	7	2	0
			4234	1418	2025	362	410	19			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	129	THR	ILE	engineered mutation	UNP P53634

- Molecule 2 is a protein called Dipeptidyl peptidase 1 Light chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	68	Total	C	H	N	O	S	0	0	0
			1030	339	501	88	100	2			

There is a discrepancy between the modelled and reference sequences:

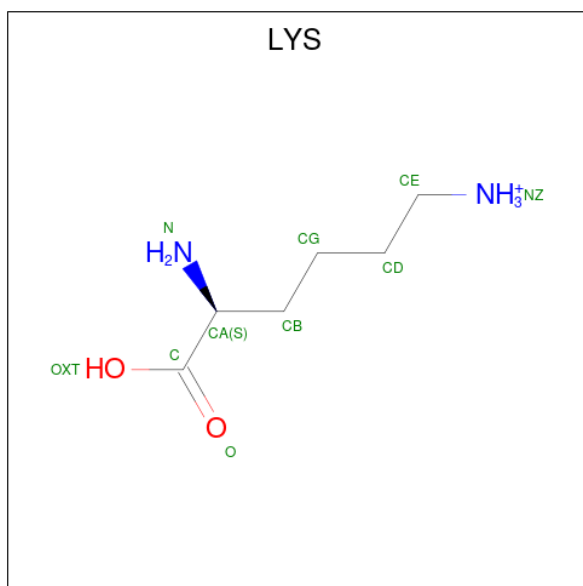
Chain	Residue	Modelled	Actual	Comment	Reference
B	440	GLU	-	expression tag	UNP P53634

- Molecule 3 is GLYCINE (CCD ID: GLY) (formula: C₂H₅NO₂).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	1
			5	2	2	1		

- Molecule 4 is LYSINE (CCD ID: LYS) (formula: $C_6H_{15}N_2O_2$).



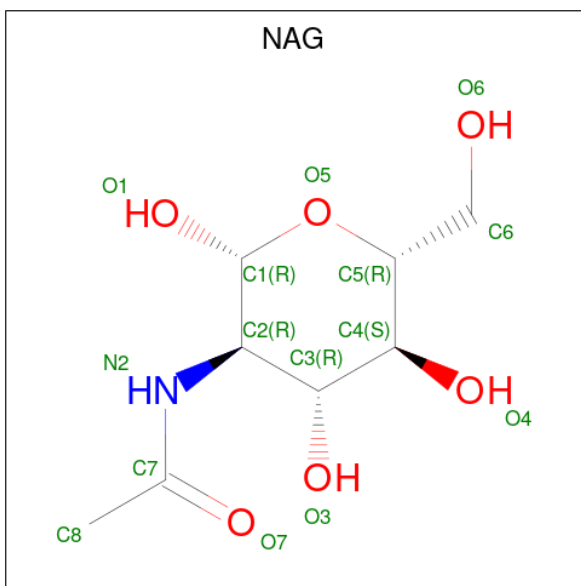
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	N	O	0	0
			9	6	2	1		

- Molecule 5 is SULFATE ION (CCD ID: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	O	S	0	0
			5	4	1		

- Molecule 6 is 2-acetamido-2-deoxy-beta-D-glucopyranose (CCD ID: NAG) (formula: $C_8H_{15}NO_6$).



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

Continued on next page...

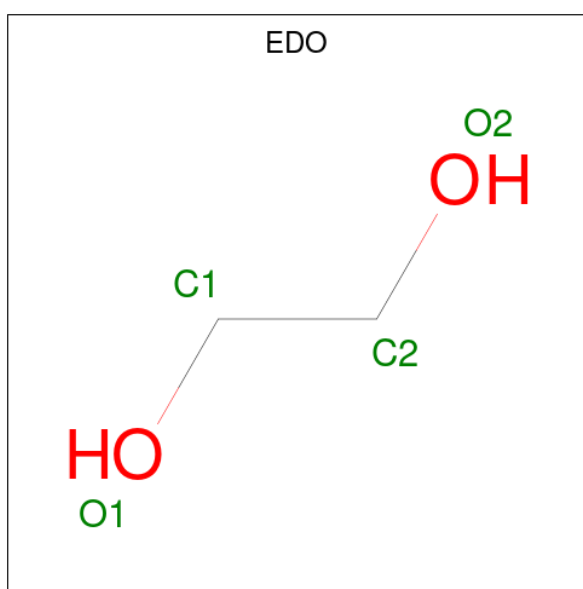
Continued from previous page...

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

- Molecule 7 is CHLORIDE ION (CCD ID: CL) (formula: Cl).

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

- Molecule 8 is 1,2-ETHANEDIOL (CCD ID: EDO) (formula: C₂H₆O₂).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
8	A	1	Total	C	H	O	0	0
			10	2	6	2		
8	A	1	Total	C	H	O	0	0
			10	2	6	2		
8	A	1	Total	C	H	O	0	0
			10	2	6	2		
8	A	1	Total	C	H	O	0	0
			10	2	6	2		

- Molecule 9 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	A	262	Total	O	0	0
			262	262		

Continued on next page...

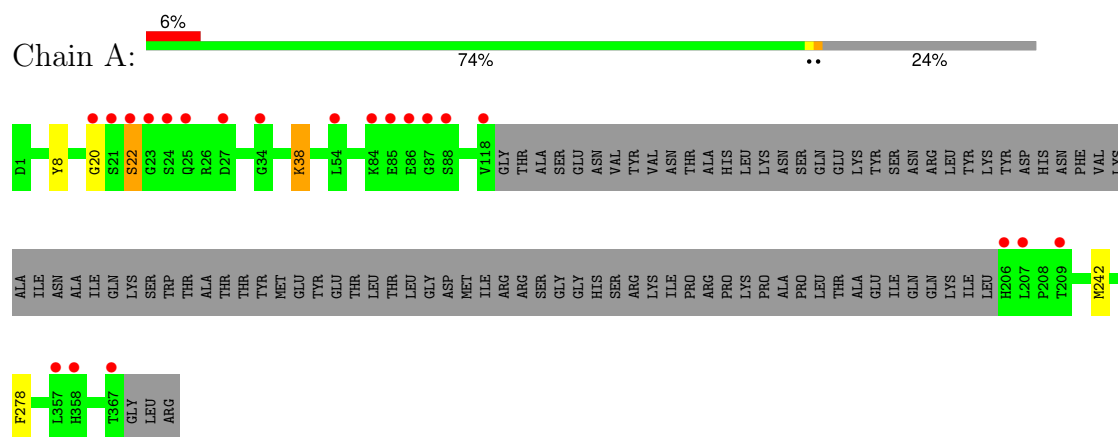
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	B	62	Total	O	0	0
			62	62		

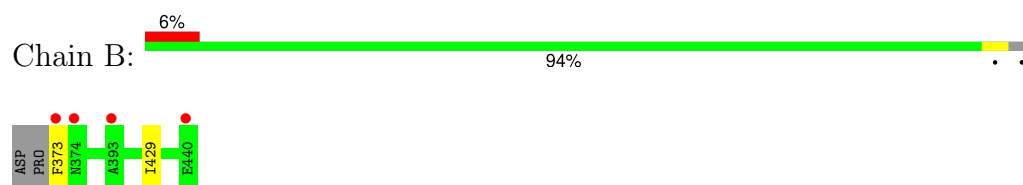
3 Residue-property plots

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: Dipeptidyl peptidase 1 Heavy chain



• Molecule 2: Dipeptidyl peptidase 1 Light chain



4 Data and refinement statistics

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α , β , γ	84.92Å 89.02Å 115.10Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	27.49 – 1.52 27.49 – 1.52	Depositor EDS
% Data completeness (in resolution range)	94.5 (27.49-1.52) 94.5 (27.49-1.52)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.51 (at 1.52Å)	Xtriage
Refinement program	BUSTER-TNT, BUSTER 2.11.5	Depositor
R, R_{free}	0.164 , 0.178 0.157 , 0.171	Depositor DCC
R_{free} test set	3759 reflections (5.92%)	wwPDB-VP
Wilson B-factor (Å ²)	19.3	Xtriage
Anisotropy	0.140	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.42 , 54.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.014 for -k,-h,-l	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	5690	wwPDB-VP
Average B, all atoms (Å ²)	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.29% 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 [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: CL, SO4, NAG, 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	0.79	1/2281 (0.0%)	0.93	1/3098 (0.0%)
2	B	0.84	0/544	0.95	0/741
All	All	0.80	1/2825 (0.0%)	0.94	1/3839 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	242	MET	SD-CE	-7.44	1.60	1.79

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	278	PHE	CA-CB-CG	-5.35	108.45	113.80

There are no chirality outliers.

There are no planarity outliers.

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	2209	2025	2033	4	0
2	B	529	501	500	1	0
3	A	5	0	4	0	0
4	A	9	0	13	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	A	5	0	0	0	0
6	A	42	0	39	0	0
7	A	1	0	0	0	0
8	A	16	24	24	0	0
9	A	262	0	0	1	0
9	B	62	0	0	0	0
All	All	3140	2550	2613	4	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 (4) 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:A:20:GLY:C	9:A:839:HOH:O	2.51	0.54
1:A:22:SER:OG	1:A:38:LYS:HE2	2.08	0.53
1:A:38:LYS:HB2	1:A:38:LYS:HE3	1.69	0.42
1:A:8:TYR:CE2	2:B:373:PHE:HB3	2.55	0.41

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	278/370 (75%)	269 (97%)	9 (3%)	0	100	100
2	B	66/70 (94%)	64 (97%)	1 (2%)	1 (2%)	8	1
All	All	344/440 (78%)	333 (97%)	10 (3%)	1 (0%)	37	18

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	429	ILE

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	231/318 (73%)	229 (99%)	2 (1%)	75	56
2	B	54/56 (96%)	54 (100%)	0	100	100
All	All	285/374 (76%)	283 (99%)	2 (1%)	81	66

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	22	SER
1	A	38	LYS

Sometimes 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 oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 12 ligands modelled in this entry, 1 is monoatomic - leaving 11 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	$\# Z > 2$
8	EDO	A	509	-	3,3,3	0.40	0	2,2,2	0.53	0
6	NAG	A	505	-	14,14,15	0.26	0	17,19,21	1.18	2 (11%)
3	GLY	A	501[B]	-	3,3,4	0.36	0	1,2,4	1.79	0
3	GLY	A	501[A]	-	3,3,4	0.52	0	1,2,4	1.19	0
6	NAG	A	506	-	14,14,15	0.28	0	17,19,21	1.72	3 (17%)
8	EDO	A	510	-	3,3,3	0.27	0	2,2,2	0.46	0
4	LYS	A	502	3,1	7,8,9	0.59	0	3,8,10	0.54	0
8	EDO	A	508	-	3,3,3	0.51	0	2,2,2	0.20	0
6	NAG	A	504	1	14,14,15	0.38	0	17,19,21	0.58	0
5	SO4	A	503	-	4,4,4	1.07	0	6,6,6	1.16	1 (16%)
8	EDO	A	511	-	3,3,3	0.54	0	2,2,2	0.78	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
8	EDO	A	509	-	-	0/1/1/1	-
6	NAG	A	505	-	-	0/6/23/26	0/1/1/1
3	GLY	A	501[B]	-	-	0/0/1/2	-
3	GLY	A	501[A]	-	-	0/0/1/2	-
6	NAG	A	506	-	-	2/6/23/26	0/1/1/1
8	EDO	A	510	-	-	0/1/1/1	-
4	LYS	A	502	3,1	-	2/6/7/9	-
8	EDO	A	508	-	-	0/1/1/1	-
6	NAG	A	504	1	-	0/6/23/26	0/1/1/1
8	EDO	A	511	-	-	0/1/1/1	-

There are no bond length outliers.

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	A	506	NAG	C1-C2-N2	4.72	117.87	110.43
6	A	506	NAG	O5-C1-C2	-4.45	104.41	111.29
6	A	505	NAG	O5-C1-C2	-3.95	105.18	111.29
5	A	503	SO4	O4-S-O2	2.50	122.64	109.56
6	A	506	NAG	C2-N2-C7	2.15	125.78	122.90
6	A	505	NAG	C1-C2-N2	2.06	113.68	110.43

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	502	LYS	O-C-CA-CB
6	A	506	NAG	C1-C2-N2-C7
6	A	506	NAG	O5-C5-C6-O6
4	A	502	LYS	CA-CB-CG-CD

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	280/370 (75%)	-0.14	21 (7%) 22 24	11, 22, 42, 67	3 (1%)
2	B	68/70 (97%)	-0.20	4 (5%) 29 33	16, 23, 42, 66	0
All	All	348/440 (79%)	-0.15	25 (7%) 23 25	11, 22, 42, 67	3 (0%)

All (25) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	367	THR	10.0
1	A	118	VAL	3.8
1	A	34	GLY	3.6
2	B	440	GLU	3.3
1	A	23	GLY	3.2
1	A	21[A]	SER	2.9
1	A	358	HIS	2.7
1	A	24	SER	2.7
1	A	86	GLU	2.7
1	A	207	LEU	2.7
1	A	88	SER	2.7
1	A	22	SER	2.6
1	A	27	ASP	2.5
1	A	85	GLU	2.4
1	A	357	LEU	2.4
2	B	373	PHE	2.4
1	A	87	GLY	2.3
2	B	393	ALA	2.3
1	A	20	GLY	2.3
1	A	54	LEU	2.2
1	A	206	HIS	2.2
1	A	209	THR	2.2
1	A	25	GLN	2.2
2	B	374	ASN	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	84	LYS	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 monosaccharides 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
6	NAG	A	506	14/15	0.56	0.22	37,64,72,75	0
4	LYS	A	502	9/10	0.67	0.24	35,40,59,61	0
6	NAG	A	505	14/15	0.68	0.19	72,73,75,75	0
5	SO4	A	503	5/5	0.73	0.18	28,28,30,37	5
6	NAG	A	504	14/15	0.81	0.13	30,39,46,46	0
3	GLY	A	501[B]	4/5	0.83	0.19	31,33,35,35	1
3	GLY	A	501[A]	4/5	0.83	0.19	25,31,33,35	1
8	EDO	A	508	4/4	0.89	0.10	37,39,41,41	0
8	EDO	A	511	4/4	0.90	0.17	27,29,35,35	0
8	EDO	A	510	4/4	0.91	0.10	39,40,42,42	0
8	EDO	A	509	4/4	0.95	0.14	25,30,36,37	0
7	CL	A	507	1/1	0.99	0.03	16,16,16,16	0

6.5 Other polymers [i](#)

There are no such residues in this entry.