WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 15, 2020 – 02:57 am BST

PDB ID	:	5XHG
Title	:	Crystal structure of Trastuzumab Fab fragment bearing Ne-(o-azidobenzyloxycarbonyl)-L-lysine
Authors	:	Kuratani, M.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S.
Deposited on	:	2017-04-20
Resolution	:	1.76 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 (i)) were used in the production of this report:

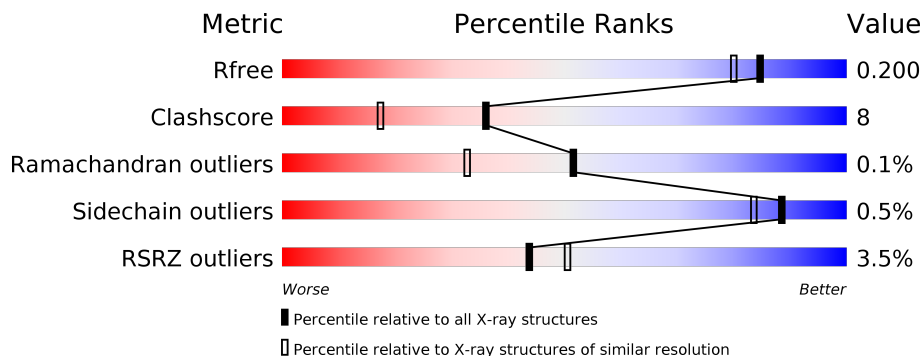|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (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 1.76 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	Percentile Ranks	Value
Rfree		0.200
Clashscore		8
Ramachandran outliers		0.1%
Sidechain outliers		0.5%
RSRZ outliers		3.5%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2340 (1.76-1.76)
Clashscore	141614	2466 (1.76-1.76)
Ramachandran outliers	138981	2437 (1.76-1.76)
Sidechain outliers	138945	2437 (1.76-1.76)
RSRZ outliers	127900	2298 (1.76-1.76)

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	214	5% / 85% / 14%
1	C	214	3% / 79% / 20%
2	B	220	3% / 90% / 7% ·
2	D	220	3% / 89% / 11%

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 7993 atoms, of which 74 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 polypeptide(L).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	213	Total	C	N	O	S	0	3	0
			1652	1036	276	335	5			
1	C	214	Total	C	N	O	S	0	3	0
			1660	1042	277	335	6			

- Molecule 2 is a protein called polypeptide(H).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	214	Total	C	N	O	S	0	0	0
			1611	1021	271	313	6			
2	D	220	Total	C	N	O	S	0	2	0
			1658	1048	279	324	7			

- Molecule 3 is (2-azidophenyl)methyl hydrogen carbonate (three-letter code: OAZ) (formula: $C_8H_7N_3O_2$).
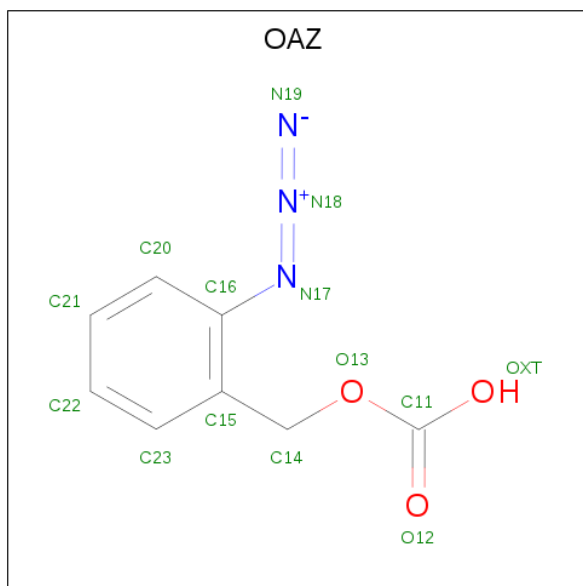
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total 13	C 8	N 3	O 2	0	0

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O_4S).

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

- Molecule 5 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).

EDO

O2

OH

C1

C2

HO

O1

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

- Molecule 6 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).

PEG

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	B	1	Total 17	C 4	H 10	O 3	0	0
6	C	1	Total 17	C 4	H 10	O 3	0	0

- Molecule 7 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	259	Total 259	O 259	0	0
7	B	317	Total 317	O 317	0	0
7	C	305	Total 305	O 305	0	0
7	D	379	Total 379	O 379	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.
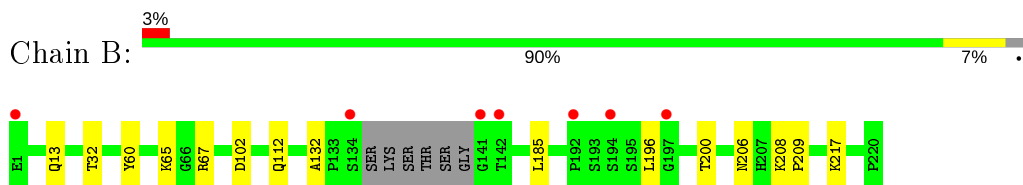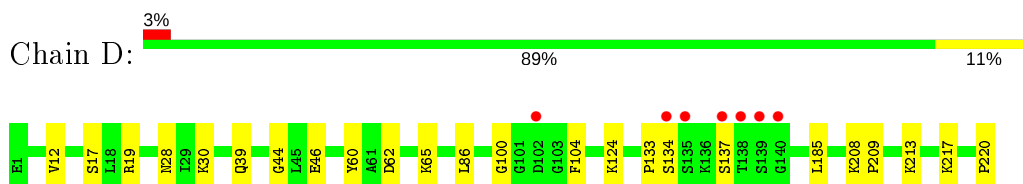
- Molecule 1: polypeptide(L)
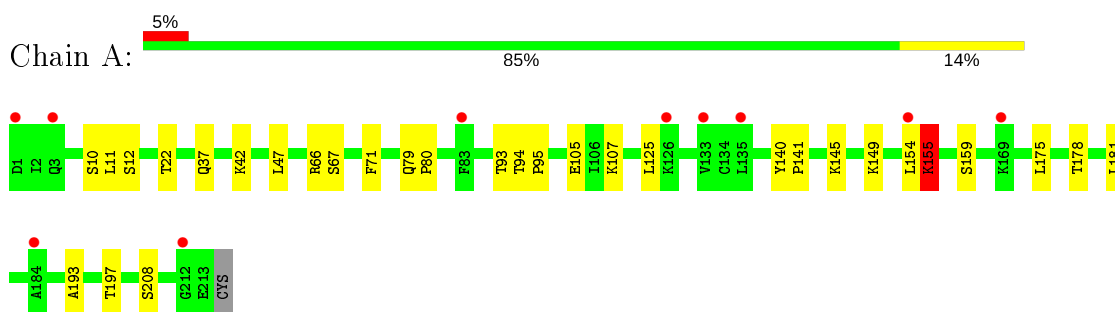
- Molecule 1: polypeptide(L)
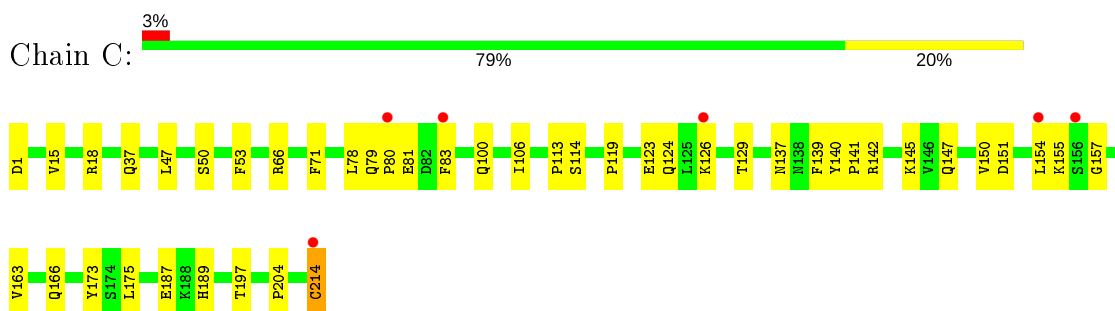
- Molecule 2: polypeptide(H)

- Molecule 2: polypeptide(H)

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	38.65Å　　79.89Å　　85.76Å 113.50°　　92.82°　　102.84°	Depositor
Resolution (Å)	25.44　　–　　1.76 25.44　　–　　1.60	Depositor EDS
% Data completeness (in resolution range)	92.6 (25.44-1.76) 87.9 (25.44-1.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.24 (at 1.60Å)	Xtriage
Refinement program	PHENIX dev_1839	Depositor
R, R_{free}	0.173　，　0.199 0.173　，　0.200	Depositor DCC
R_{free} test set	1993 reflections (1.72%)	wwPDB-VP
Wilson B-factor (Å^2)	15.1	Xtriage
Anisotropy	0.355	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 49.9	EDS
L-test for twinning[2]	$<|L|> = 0.49, <L^2> = 0.32$	Xtriage
Estimated twinning fraction	0.028 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	7993	wwPDB-VP
Average B, all atoms (Å^2)	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.24% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: PEG, EDO, OAZ, SO4

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	0.30	1/1698 (0.1%)	0.42	0/2307
1	C	0.25	0/1706	0.43	0/2319
2	B	0.28	0/1652	0.45	0/2253
2	D	0.29	0/1706	0.46	0/2326
All	All	0.28	1/6762 (0.0%)	0.44	0/9205

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	155	LYS	C-O	6.91	1.36	1.23

There are no bond angle outliers.

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	1652	0	1613	22	0
1	C	1660	0	1626	41	0
2	B	1611	0	1567	12	0
2	D	1658	0	1618	25	0
3	A	13	0	0	1	0
4	A	5	0	0	0	0

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	C	5	0	0	1	0
4	D	5	0	0	1	0
5	A	4	6	6	1	0
5	B	4	6	6	0	0
5	C	16	24	24	4	0
5	D	12	18	18	3	0
6	B	7	10	10	0	0
6	C	7	10	10	3	0
7	A	259	0	0	4	3
7	B	317	0	0	9	1
7	C	305	0	0	14	2
7	D	379	0	0	11	2
All	All	7919	74	6498	105	4

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 8.

The worst 5 of 105 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)
2:D:28:ASN:HB3	7:D:598:HOH:O	1.59	1.00
2:D:124:LYS:NZ	7:D:402:HOH:O	1.95	1.00
1:C:1:ASP:OD1	7:C:401:HOH:O	1.82	0.97
2:D:19:ARG:NH1	7:D:401:HOH:O	1.89	0.96
1:C:187:GLU:OE2	7:C:402:HOH:O	1.92	0.87

All (4) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
7:A:608:HOH:O	7:C:532:HOH:O[1_554]	1.83	0.37
7:A:659:HOH:O	7:D:762:HOH:O[1_545]	1.93	0.27
7:B:625:HOH:O	7:D:594:HOH:O[1_455]	2.15	0.05
7:A:645:HOH:O	7:C:602:HOH:O[1_444]	2.18	0.02

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	214/214 (100%)	211 (99%)	3 (1%)	0	100	100
1	C	215/214 (100%)	208 (97%)	6 (3%)	1 (0%)	29	12
2	B	210/220 (96%)	206 (98%)	4 (2%)	0	100	100
2	D	220/220 (100%)	219 (100%)	1 (0%)	0	100	100
All	All	859/868 (99%)	844 (98%)	14 (2%)	1 (0%)	51	33

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	157	GLY

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	191/189 (101%)	189 (99%)	2 (1%)	76	63
1	C	192/189 (102%)	191 (100%)	1 (0%)	88	83
2	B	176/181 (97%)	175 (99%)	1 (1%)	86	79
2	D	183/181 (101%)	183 (100%)	0	100	100
All	All	742/740 (100%)	738 (100%)	4 (0%)	88	83

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	125	LEU
1	A	155	LYS
2	B	196	LEU
1	C	214	CYS

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)

15 ligands are modelled in this entry.

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	#$|Z| > 2$
4	SO4	C	301	-	4,4,4	0.14	0	6,6,6	0.06	0
5	EDO	B	301	-	3,3,3	0.49	0	2,2,2	0.11	0
4	SO4	D	301	-	4,4,4	0.14	0	6,6,6	0.06	0
3	OAZ	A	301	1	11,13,14	1.08	1 (9%)	11,15,17	1.51	1 (9%)
5	EDO	A	303	-	3,3,3	0.46	0	2,2,2	0.27	0
6	PEG	C	306	-	6,6,6	0.45	0	5,5,5	0.56	0
5	EDO	C	302	-	3,3,3	0.47	0	2,2,2	0.27	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	#|Z| > 2	Counts	RMSZ	#|Z| > 2
4	SO4	A	302	-	4,4,4	0.14	0	6,6,6	0.06	0
5	EDO	D	302	-	3,3,3	0.47	0	2,2,2	0.22	0
5	EDO	D	303	-	3,3,3	0.40	0	2,2,2	0.34	0
5	EDO	C	303	-	3,3,3	0.46	0	2,2,2	0.35	0
6	PEG	B	302	-	6,6,6	0.51	0	5,5,5	0.63	0
5	EDO	C	305	-	3,3,3	0.47	0	2,2,2	0.27	0
5	EDO	D	304	-	3,3,3	0.47	0	2,2,2	0.28	0
5	EDO	C	304	-	3,3,3	0.44	0	2,2,2	0.35	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	EDO	B	301	-	-	0/1/1/1	-
5	EDO	C	303	-	-	0/1/1/1	-
3	OAZ	A	301	1	-	2/7/7/8	0/1/1/1
5	EDO	A	303	-	-	0/1/1/1	-
6	PEG	C	306	-	-	3/4/4/4	-
5	EDO	C	302	-	-	1/1/1/1	-
5	EDO	D	302	-	-	0/1/1/1	-
5	EDO	D	303	-	-	0/1/1/1	-
6	PEG	B	302	-	-	1/4/4/4	-
5	EDO	C	305	-	-	1/1/1/1	-
5	EDO	D	304	-	-	0/1/1/1	-
5	EDO	C	304	-	-	1/1/1/1	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	301	OAZ	N18-N17	-3.20	1.15	1.24

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	301	OAZ	C20-C16-N17	-4.62	117.13	123.26

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	301	OAZ	C15-C16-N17-N18
3	A	301	OAZ	C20-C16-N17-N18
6	B	302	PEG	O2-C3-C4-O4
5	C	305	EDO	O1-C1-C2-O2
6	C	306	PEG	C1-C2-O2-C3

There are no ring outliers.

9 monomers are involved in 14 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	C	301	SO4	1	0
4	D	301	SO4	1	0
3	A	301	OAZ	1	0
5	A	303	EDO	1	0
6	C	306	PEG	3	0
5	C	302	EDO	3	0
5	D	303	EDO	2	0
5	D	304	EDO	1	0
5	C	304	EDO	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 (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	213/214 (99%)	0.18	10 (4%)	31	37	11, 23, 39, 50	0
1	C	214/214 (100%)	0.15	6 (2%)	53	58	12, 22, 37, 57	0
2	B	214/220 (97%)	0.00	7 (3%)	46	53	8, 16, 39, 58	0
2	D	220/220 (100%)	-0.16	7 (3%)	47	54	8, 14, 38, 67	0
All	All	861/868 (99%)	0.04	30 (3%)	44	50	8, 18, 39, 67	0

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	141	GLY	6.4
1	C	214	CYS	6.1
2	D	139	SER	4.6
2	D	102	ASP	4.4
2	D	140	GLY	4.1

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, 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	EDO	C	305	4/4	0.56	0.22	57,69,70,71	0
6	PEG	C	306	7/7	0.64	0.26	31,50,63,63	0
3	OAZ	A	301	13/14	0.64	0.26	67,72,74,74	0
4	SO4	D	301	5/5	0.70	0.31	123,123,124,124	0
5	EDO	D	302	4/4	0.72	0.21	39,47,55,56	0
6	PEG	B	302	7/7	0.75	0.20	61,73,75,75	0
4	SO4	C	301	5/5	0.76	0.18	99,99,100,101	0
5	EDO	B	301	4/4	0.77	0.13	30,40,45,49	0
5	EDO	A	303	4/4	0.79	0.20	50,60,64,65	0
4	SO4	A	302	5/5	0.87	0.26	91,91,92,92	0
5	EDO	D	304	4/4	0.87	0.14	35,43,46,46	0
5	EDO	C	304	4/4	0.87	0.17	49,58,62,63	0
5	EDO	C	302	4/4	0.91	0.18	18,29,35,40	0
5	EDO	D	303	4/4	0.92	0.19	20,32,37,40	0
5	EDO	C	303	4/4	0.94	0.15	21,26,30,32	0

6.5 Other polymers (i)

There are no such residues in this entry.