wwPDB X-ray Structure Validation Summary Report (i)

Jun 18, 2024 – 05:42 PM EDT

PDB ID	:	4AEJ
Title	:	Crystal structure of Human fibrillar procollagen type III C- propeptide trimer
Authors	:	Bourhis, J.M.; Mariano, N.; Zhao, Y.; Harlos, K.; Jones, E.Y.; Moali, C.; Aghajari, N.; Hulmes, D.J.S.
Deposited on	:	2012-01-11
Resolution	:	2.21 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

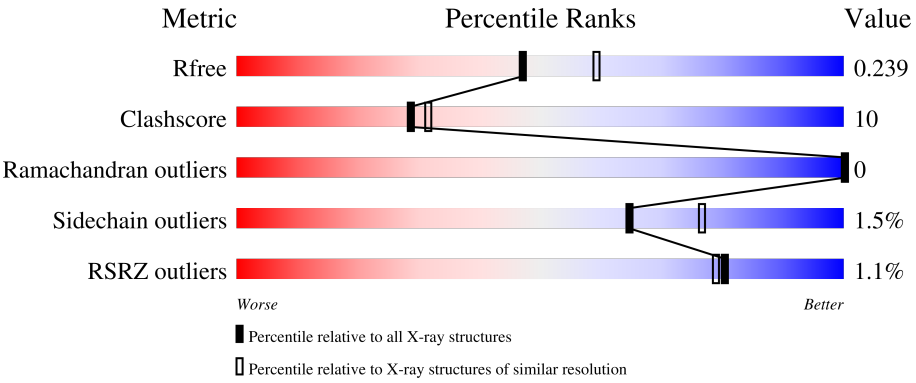
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.21 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	5912 (2.24-2.20)
Clashscore	141614	6646 (2.24-2.20)
Ramachandran outliers	138981	6543 (2.24-2.20)
Sidechain outliers	138945	6544 (2.24-2.20)
RSRZ outliers	127900	5797 (2.24-2.20)

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 <=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	256	69% 14% 17%
1	B	256	73% 12% 15%
1	C	256	68% 14% 17%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 5277 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 COLLAGEN ALPHA-1(III) CHAIN.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	213	Total	C	N	O	S	Se	0	1	0
			1677	1059	287	319	8	4			
1	B	218	Total	C	N	O	S	Se	0	2	0
			1730	1090	297	330	8	5			
1	C	212	Total	C	N	O	S	Se	0	0	0
			1646	1043	283	308	8	4			

There are 39 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-10	GLU	-	expression tag	UNP P02461
A	-9	THR	-	expression tag	UNP P02461
A	-8	GLY	-	expression tag	UNP P02461
A	-7	HIS	-	expression tag	UNP P02461
A	-6	HIS	-	expression tag	UNP P02461
A	-5	HIS	-	expression tag	UNP P02461
A	-4	HIS	-	expression tag	UNP P02461
A	-3	HIS	-	expression tag	UNP P02461
A	-2	HIS	-	expression tag	UNP P02461
A	-1	SER	-	expression tag	UNP P02461
A	0	ALA	-	expression tag	UNP P02461
A	132	GLN	HIS	variant	UNP P02461
A	146	GLN	ASN	engineered mutation	UNP P02461
B	-10	GLU	-	expression tag	UNP P02461
B	-9	THR	-	expression tag	UNP P02461
B	-8	GLY	-	expression tag	UNP P02461
B	-7	HIS	-	expression tag	UNP P02461
B	-6	HIS	-	expression tag	UNP P02461
B	-5	HIS	-	expression tag	UNP P02461
B	-4	HIS	-	expression tag	UNP P02461
B	-3	HIS	-	expression tag	UNP P02461
B	-2	HIS	-	expression tag	UNP P02461
B	-1	SER	-	expression tag	UNP P02461

Continued on next page...

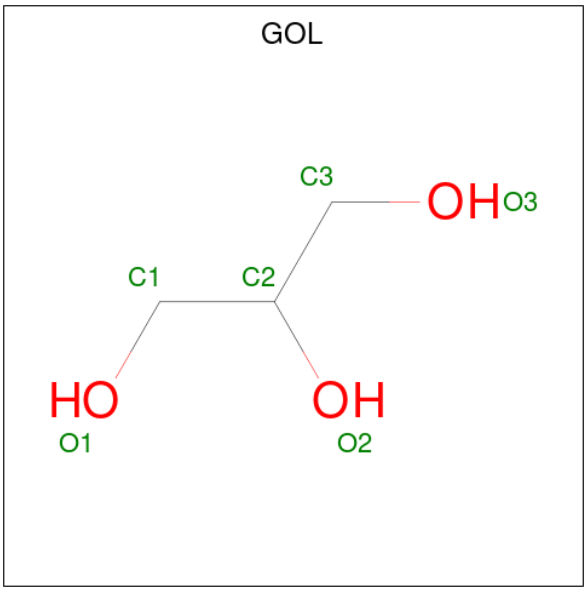
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	ALA	-	expression tag	UNP P02461
B	132	GLN	HIS	variant	UNP P02461
B	146	GLN	ASN	engineered mutation	UNP P02461
C	-10	GLU	-	expression tag	UNP P02461
C	-9	THR	-	expression tag	UNP P02461
C	-8	GLY	-	expression tag	UNP P02461
C	-7	HIS	-	expression tag	UNP P02461
C	-6	HIS	-	expression tag	UNP P02461
C	-5	HIS	-	expression tag	UNP P02461
C	-4	HIS	-	expression tag	UNP P02461
C	-3	HIS	-	expression tag	UNP P02461
C	-2	HIS	-	expression tag	UNP P02461
C	-1	SER	-	expression tag	UNP P02461
C	0	ALA	-	expression tag	UNP P02461
C	132	GLN	HIS	variant	UNP P02461
C	146	GLN	ASN	engineered mutation	UNP P02461

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total Ca 1 1	0	0
2	B	1	Total Ca 1 1	0	0
2	C	1	Total Ca 1 1	0	0

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 6 3 3	0	0
3	A	1	Total C O 6 3 3	0	0
3	A	1	Total C O 6 3 3	0	0
3	A	1	Total C O 6 3 3	0	0
3	A	1	Total C O 6 3 3	0	0
3	B	1	Total C O 6 3 3	0	0
3	B	1	Total C O 6 3 3	0	0

- Molecule 4 is water.

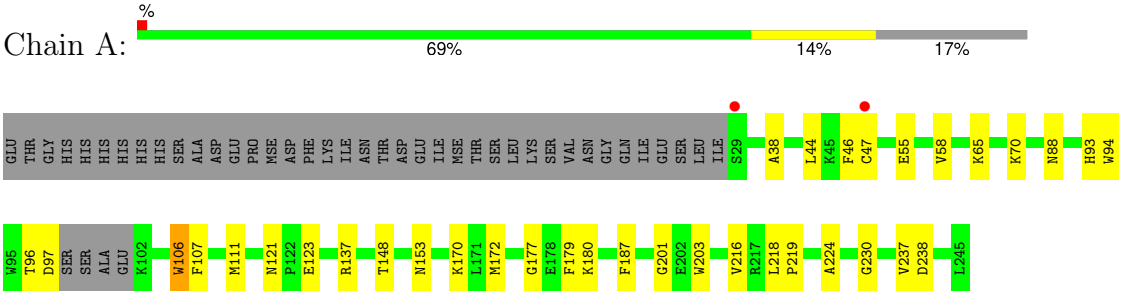
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	74	Total O 74 74	0	0
4	B	76	Total O 76 76	0	0
4	C	29	Total O 29 29	0	0

3 Residue-property plots (i)

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.

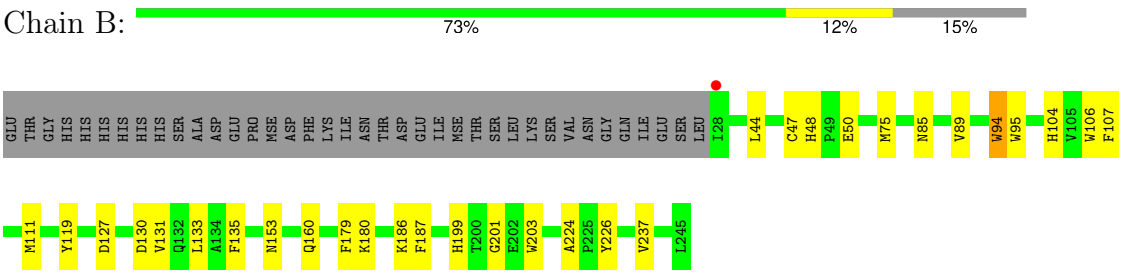
• Molecule 1: COLLAGEN ALPHA-1(III) CHAIN

Chain A:



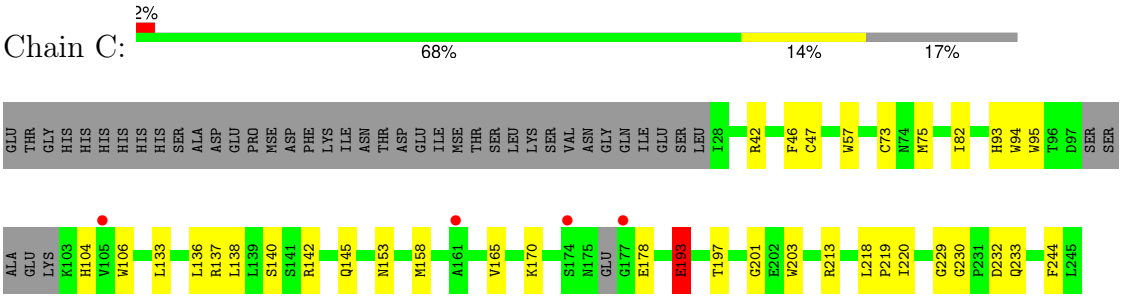
• Molecule 1: COLLAGEN ALPHA-1(III) CHAIN

Chain B:



• Molecule 1: COLLAGEN ALPHA-1(III) CHAIN

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	83.93Å 89.27Å 101.46Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.41 – 2.21 67.02 – 2.21	Depositor EDS
% Data completeness (in resolution range)	100.0 (39.41-2.21) 100.0 (67.02-2.21)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.97 (at 2.22Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.201 , 0.237 0.209 , 0.239	Depositor DCC
R_{free} test set	1932 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	31.0	Xtriage
Anisotropy	0.283	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 33.9	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	0.93	EDS
Total number of atoms	5277	wwPDB-VP
Average B, all atoms (Å^2)	33.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.64% 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: GOL, CA

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	0.62	2/1716 (0.1%)	0.66	0/2316
1	B	0.65	3/1770 (0.2%)	0.68	0/2385
1	C	0.67	4/1684 (0.2%)	0.66	1/2272 (0.0%)
All	All	0.65	9/5170 (0.2%)	0.67	1/6973 (0.0%)

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	95	TRP	CD2-CE2	5.42	1.47	1.41
1	C	57	TRP	CD2-CE2	5.40	1.47	1.41
1	A	106	TRP	CD2-CE2	5.31	1.47	1.41
1	A	203	TRP	CD2-CE2	5.27	1.47	1.41
1	B	203	TRP	CD2-CE2	5.18	1.47	1.41

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	193	GLU	N-CA-CB	5.11	119.80	110.60

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	1677	0	1574	30	0
1	B	1730	0	1647	35	2
1	C	1646	0	1549	40	2
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
3	A	30	0	40	0	0
3	B	12	0	16	4	0
4	A	74	0	0	1	0
4	B	76	0	0	3	0
4	C	29	0	0	3	0
All	All	5277	0	4826	96	2

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 10.

The worst 5 of 96 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:135:PHE:CE1	1:C:138:LEU:CD1	2.33	1.11
1:B:94:TRP:HB3	1:B:111:MSE:CE	1.82	1.08
1:A:94:TRP:HB3	1:A:111:MSE:CE	1.89	1.02
1:B:135:PHE:CD1	1:C:138:LEU:HD11	1.98	0.97
1:B:94:TRP:HB3	1:B:111:MSE:HE3	1.42	0.96

All (2) 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)
1:B:48:HIS:ND1	1:C:193:GLU:OE2[2_455]	1.74	0.46
1:B:50:GLU:OE2	1:C:193:GLU:OE2[2_455]	2.09	0.11

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	210/256 (82%)	203 (97%)	7 (3%)	0	100	100
1	B	218/256 (85%)	214 (98%)	4 (2%)	0	100	100
1	C	206/256 (80%)	197 (96%)	9 (4%)	0	100	100
All	All	634/768 (83%)	614 (97%)	20 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	179/217 (82%)	177 (99%)	2 (1%)	73	84
1	B	188/217 (87%)	185 (98%)	3 (2%)	62	75
1	C	174/217 (80%)	171 (98%)	3 (2%)	60	73
All	All	541/651 (83%)	533 (98%)	8 (2%)	65	76

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	193	GLU
1	C	47	CYS
1	B	160	GLN
1	B	85	ASN
1	C	46	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	153	ASN
1	C	104	HIS
1	C	153	ASN
1	C	145	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	104	HIS

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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 10 ligands modelled in this entry, 3 are monoatomic - leaving 7 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	GOL	A	1249	-	5,5,5	0.38	0	5,5,5	0.39	0
3	GOL	B	1246	-	5,5,5	0.15	0	5,5,5	0.57	0
3	GOL	A	1247	-	5,5,5	0.39	0	5,5,5	0.32	0
3	GOL	A	1250	-	5,5,5	0.24	0	5,5,5	0.57	0
3	GOL	A	1246	-	5,5,5	0.29	0	5,5,5	0.44	0
3	GOL	A	1248	-	5,5,5	0.20	0	5,5,5	0.89	0
3	GOL	B	1247	-	5,5,5	0.45	0	5,5,5	0.53	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
3	GOL	A	1249	-	-	0/4/4/4	-
3	GOL	B	1246	-	-	2/4/4/4	-
3	GOL	A	1247	-	-	3/4/4/4	-
3	GOL	A	1250	-	-	3/4/4/4	-
3	GOL	A	1246	-	-	0/4/4/4	-
3	GOL	A	1248	-	-	1/4/4/4	-
3	GOL	B	1247	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 11 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	1247	GOL	C1-C2-C3-O3
3	A	1250	GOL	C1-C2-C3-O3
3	B	1246	GOL	O1-C1-C2-O2
3	B	1246	GOL	O1-C1-C2-C3
3	A	1247	GOL	O2-C2-C3-O3

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	1246	GOL	2	0
3	B	1247	GOL	2	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	209/256 (81%)	-0.29	2 (0%)	82	81	17, 29, 52, 66	0
1	B	214/256 (83%)	-0.26	1 (0%)	91	90	17, 25, 43, 52	0
1	C	208/256 (81%)	0.02	4 (1%)	66	65	22, 41, 65, 79	0
All	All	631/768 (82%)	-0.18	7 (1%)	80	79	17, 32, 58, 79	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	161	ALA	3.5
1	A	47	CYS	2.5
1	C	177	GLY	2.5
1	C	174	SER	2.4
1	B	28	ILE	2.2

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 monosaccharides 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
3	GOL	A	1249	6/6	0.78	0.18	41,46,49,50	0
3	GOL	B	1247	6/6	0.86	0.32	41,43,45,47	0
3	GOL	A	1247	6/6	0.87	0.22	37,39,42,42	0
3	GOL	B	1246	6/6	0.90	0.18	43,44,46,49	0
3	GOL	A	1246	6/6	0.94	0.14	27,30,35,41	0
3	GOL	A	1250	6/6	0.94	0.13	30,36,37,39	0
3	GOL	A	1248	6/6	0.95	0.12	33,34,35,36	0
2	CA	C	246	1/1	0.99	0.09	23,23,23,23	0
2	CA	B	246	1/1	0.99	0.17	19,19,19,19	0
2	CA	A	246	1/1	1.00	0.11	21,21,21,21	0

6.5 Other polymers (i)

There are no such residues in this entry.