



Full wwPDB X-ray Structure Validation Report ⓘ

Aug 22, 2023 – 02:09 PM JST

PDB ID : 8J0R
Title : Structure of human TFAP2A in complex with DNA
Authors : Liu, K.; Xiao, Y.Q.; Li, W.F.; Min, J.R.
Deposited on : 2023-04-11
Resolution : 2.10 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 : 4.02b-467
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtrriage (Phenix) : 1.13
EDS : 2.35
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) : 2.35

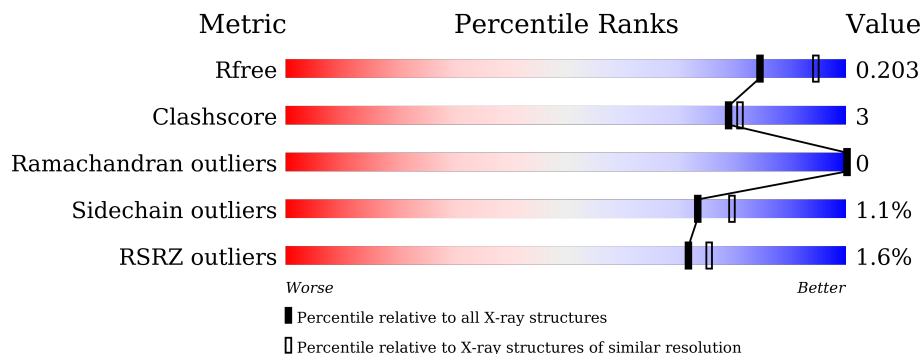
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.10 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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	219	 92% 6% 2% 0% 0%
1	B	219	 89% 6% 5% 0% 0%
2	C	13	 54% 38% 8% 0% 0%
3	D	13	 62% 31% 8% 0% 0%

2 Entry composition [i](#)

There are 8 unique types of molecules in this entry. The entry contains 4212 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 Transcription factor AP-2-alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	215	Total	C	N	O	S	0	0	0
			1642	1026	300	307	9			
1	B	209	Total	C	N	O	S	0	0	0
			1601	1002	292	298	9			

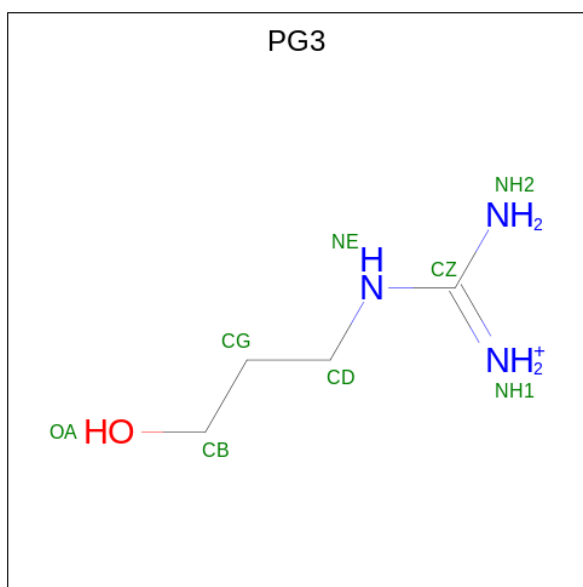
- Molecule 2 is a DNA chain called DNA (5'-D(*CP*TP*GP*CP*CP*TP*CP*AP*GP*GP*CP*AP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	13	Total	C	N	O	P	0	0	0
			259	124	47	76	12			

- Molecule 3 is a DNA chain called DNA (5'-D(*GP*TP*GP*CP*CP*TP*GP*AP*GP*GP*CP*AP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	D	13	Total	C	N	O	P	0	0	0
			268	127	53	76	12			

- Molecule 4 is GUANIDINE-3-PROPANOL (three-letter code: PG3) (formula: C₄H₁₂N₃O).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
4	A	1	8	4	3	1	0	0

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



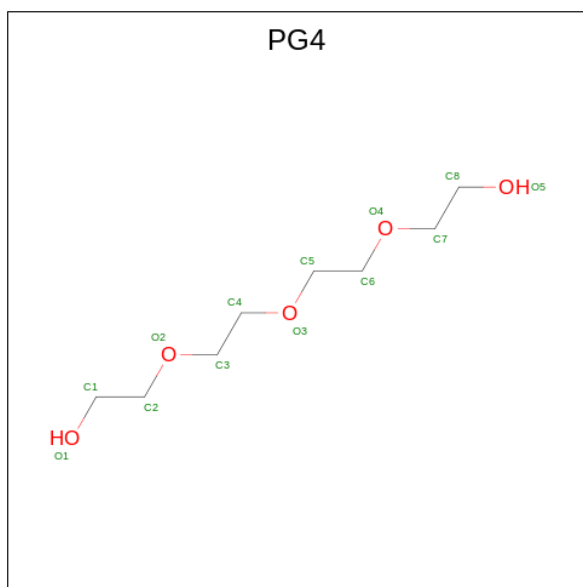
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
5	A	1	6	3	3	0	0
5	A	1	6	3	3	0	0
5	B	1	6	3	3	0	0

Continued on next page...

Continued from previous page...

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

- Molecule 6 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula: C₈H₁₈O₅).



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

- Molecule 7 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: C₄H₁₀O₃).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	B	1	Total C O 7 4 3	0	0
7	B	1	Total C O 7 4 3	0	0
7	B	1	Total C O 7 4 3	0	0

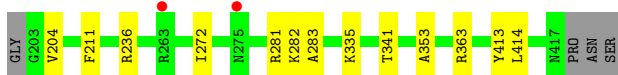
- Molecule 8 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	A	151	Total O 151 151	0	0
8	B	139	Total O 139 139	0	0
8	C	34	Total O 34 34	0	0
8	D	33	Total O 33 33	0	0

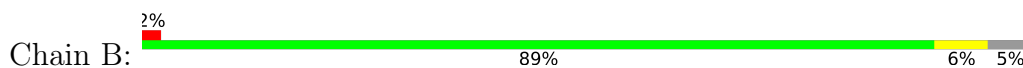
3 Residue-property plots [i](#)

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: Transcription factor AP-2-alpha



- Molecule 1: Transcription factor AP-2-alpha



- Molecule 2: DNA (5'-D(*CP*TP*GP*CP*CP*TP*CP*AP*GP*GP*CP*AP*C)-3')



- Molecule 3: DNA (5'-D(*GP*TP*GP*CP*CP*TP*GP*AP*GP*GP*CP*AP*G)-3')



4 Data and refinement statistics

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α , β , γ	124.79Å 124.79Å 96.92Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	19.99 – 2.10 19.98 – 2.10	Depositor EDS
% Data completeness (in resolution range)	99.8 (19.99-2.10) 99.9 (19.98-2.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.76 (at 2.09Å)	Xtrriage
Refinement program	REFMAC 5.8.0257	Depositor
R, R_{free}	0.162 , 0.194 0.172 , 0.203	Depositor DCC
R_{free} test set	2688 reflections (5.27%)	wwPDB-VP
Wilson B-factor (Å ²)	30.5	Xtrriage
Anisotropy	0.001	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.39 , 52.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.022 for -h,-k,l	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	4212	wwPDB-VP
Average B, all atoms (Å ²)	36.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.53% 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: PG4, GOL, PEG, PG3

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	0.86	0/1668	0.88	1/2256 (0.0%)
1	B	0.87	1/1627 (0.1%)	0.93	0/2200
2	C	1.14	2/289 (0.7%)	1.32	1/443 (0.2%)
3	D	1.10	2/301 (0.7%)	1.14	0/464
All	All	0.91	5/3885 (0.1%)	0.97	2/5363 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
2	C	0	1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	347	GLU	CD-OE1	14.34	1.41	1.25
3	D	2	DT	O3'-P	-9.44	1.49	1.61
2	C	8	DA	O3'-P	-8.78	1.50	1.61
3	D	8	DA	O3'-P	-6.05	1.53	1.61
2	C	10	DG	P-OP2	-5.63	1.39	1.49

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	11	DC	O5'-P-OP2	-11.86	95.02	105.70
1	A	236	ARG	NE-CZ-NH2	-5.03	117.78	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	4	DC	Sidechain

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	1642	0	1683	9	0
1	B	1601	0	1643	8	0
2	C	259	0	147	3	0
3	D	268	0	147	2	0
4	A	8	0	12	2	0
5	A	12	0	16	0	0
5	B	12	0	16	0	0
5	C	6	0	8	0	0
6	A	26	0	36	1	0
7	B	21	0	30	2	0
8	A	151	0	0	2	0
8	B	139	0	0	0	0
8	C	34	0	0	0	0
8	D	33	0	0	0	0
All	All	4212	0	3738	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 3.

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)
2:C:1:DC:H2'	2:C:2:DT:H71	1.78	0.64
1:B:353:ALA:HA	7:B:501:PEG:H41	1.83	0.60
1:A:363:ARG:NH1	8:A:601:HOH:O	2.34	0.59
1:A:413:TYR:HD2	1:A:414:LEU:HD23	1.76	0.51
1:A:281:ARG:HB2	4:A:501:PG3:HB1	1.92	0.50
1:B:315:LYS:O	1:B:319:GLU:HG2	2.13	0.48
2:C:1:DC:H2'	2:C:2:DT:C7	2.42	0.48

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:413:TYR:CD2	1:A:414:LEU:HD23	2.48	0.48
3:D:11:DC:H2''	3:D:12:DA:C8	2.49	0.47
1:B:335:LYS:HB2	1:B:409:MET:HE1	1.97	0.47
1:A:341:THR:HG23	1:B:341:THR:HG23	1.99	0.44
1:B:238:LEU:HD21	1:B:249:LEU:HD22	2.00	0.44
1:A:335:LYS:NZ	8:A:605:HOH:O	2.50	0.44
1:B:335:LYS:HB2	1:B:409:MET:CE	2.48	0.44
1:A:353:ALA:HA	6:A:504:PG4:H11	2.01	0.43
3:D:1:DG:H2'	3:D:2:DT:H72	2.02	0.41
1:A:211:PHE:CD1	1:A:272:ILE:HD13	2.55	0.41
1:A:283:ALA:HA	4:A:501:PG3:HE	1.84	0.40
1:B:226:LYS:HE3	2:C:11:DC:P	2.62	0.40
1:B:389:PRO:HD3	7:B:501:PEG:H32	2.03	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	213/219 (97%)	211 (99%)	2 (1%)	0	100	100
1	B	207/219 (94%)	205 (99%)	2 (1%)	0	100	100
All	All	420/438 (96%)	416 (99%)	4 (1%)	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	182/185 (98%)	180 (99%)	2 (1%)	73	79
1	B	177/185 (96%)	175 (99%)	2 (1%)	73	79
All	All	359/370 (97%)	355 (99%)	4 (1%)	73	79

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

Mol	Chain	Res	Type
1	A	204	VAL
1	A	282	LYS
1	B	380	ASN
1	B	401	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	329	ASN
1	B	343	GLN

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](#)

11 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
5	GOL	B	502	-	5,5,5	0.13	0	5,5,5	0.32	0
6	PG4	A	503	-	12,12,12	0.27	0	11,11,11	0.12	0
5	GOL	B	503	-	5,5,5	0.16	0	5,5,5	0.46	0
5	GOL	A	505	-	5,5,5	0.35	0	5,5,5	0.57	0
6	PG4	A	504	-	12,12,12	0.62	0	11,11,11	0.46	0
5	GOL	C	101	-	5,5,5	0.12	0	5,5,5	0.22	0
7	PEG	B	501	-	6,6,6	0.53	0	5,5,5	0.57	0
7	PEG	B	505	-	6,6,6	0.71	0	5,5,5	0.50	0
4	PG3	A	501	-	7,7,7	0.32	0	6,7,7	0.69	0
5	GOL	A	502	-	5,5,5	0.10	0	5,5,5	0.39	0
7	PEG	B	504	-	6,6,6	0.43	0	5,5,5	0.24	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	GOL	B	502	-	-	4/4/4/4	-
6	PG4	A	503	-	-	6/10/10/10	-
5	GOL	B	503	-	-	2/4/4/4	-
5	GOL	A	505	-	-	0/4/4/4	-
6	PG4	A	504	-	-	7/10/10/10	-
5	GOL	C	101	-	-	2/4/4/4	-
7	PEG	B	501	-	-	3/4/4/4	-
7	PEG	B	505	-	-	3/4/4/4	-
4	PG3	A	501	-	-	2/5/5/5	-
5	GOL	A	502	-	-	3/4/4/4	-
7	PEG	B	504	-	-	3/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (35) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	502	GOL	O1-C1-C2-O2
5	A	502	GOL	O1-C1-C2-C3
5	B	502	GOL	O1-C1-C2-C3
5	B	502	GOL	C1-C2-C3-O3
5	B	503	GOL	C1-C2-C3-O3
4	A	501	PG3	OA-CB-CG-CD
5	B	503	GOL	O2-C2-C3-O3
7	B	505	PEG	O1-C1-C2-O2
6	A	504	PG4	O1-C1-C2-O2
7	B	505	PEG	O2-C3-C4-O4
5	B	502	GOL	O2-C2-C3-O3
6	A	504	PG4	O4-C7-C8-O5
6	A	503	PG4	O1-C1-C2-O2
7	B	501	PEG	O1-C1-C2-O2
5	B	502	GOL	O1-C1-C2-O2
7	B	501	PEG	O2-C3-C4-O4
6	A	504	PG4	C1-C2-O2-C3
6	A	503	PG4	C5-C6-O4-C7
6	A	503	PG4	C1-C2-O2-C3
5	C	101	GOL	O1-C1-C2-O2
6	A	504	PG4	C4-C3-O2-C2
6	A	503	PG4	O2-C3-C4-O3
7	B	504	PEG	C1-C2-O2-C3
7	B	501	PEG	C4-C3-O2-C2
7	B	504	PEG	O1-C1-C2-O2
7	B	504	PEG	C4-C3-O2-C2
6	A	504	PG4	O3-C5-C6-O4
5	C	101	GOL	O1-C1-C2-C3
6	A	503	PG4	O3-C5-C6-O4
4	A	501	PG3	CG-CD-NE-CZ
6	A	504	PG4	C6-C5-O3-C4
7	B	505	PEG	C4-C3-O2-C2
6	A	503	PG4	C3-C4-O3-C5
5	A	502	GOL	C1-C2-C3-O3
6	A	504	PG4	O2-C3-C4-O3

There are no ring outliers.

3 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	A	504	PG4	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
7	B	501	PEG	2	0
4	A	501	PG3	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 [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	215/219 (98%)	-0.33	2 (0%) 84 86	24, 32, 55, 69	0
1	B	209/219 (95%)	-0.36	5 (2%) 59 64	23, 31, 56, 95	0
2	C	13/13 (100%)	-0.78	0 100 100	27, 35, 40, 58	0
3	D	13/13 (100%)	-0.63	0 100 100	26, 31, 52, 69	0
All	All	450/464 (96%)	-0.36	7 (1%) 72 75	23, 32, 55, 95	0

All (7) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	413	TYR	5.1
1	B	260	ASN	3.5
1	A	263	ARG	2.7
1	B	329	ASN	2.7
1	B	328	PRO	2.6
1	B	412	MET	2.3
1	A	275	ASN	2.3

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
5	GOL	A	505	6/6	0.55	0.27	53,67,75,78	0
7	PEG	B	501	7/7	0.72	0.17	43,52,53,59	0
6	PG4	A	503	13/13	0.76	0.28	58,80,90,93	0
7	PEG	B	505	7/7	0.78	0.28	59,61,66,68	0
7	PEG	B	504	7/7	0.79	0.15	53,68,75,77	0
4	PG3	A	501	8/8	0.81	0.15	56,62,73,75	0
6	PG4	A	504	13/13	0.83	0.18	46,64,74,76	0
5	GOL	B	503	6/6	0.83	0.28	59,66,69,72	0
5	GOL	A	502	6/6	0.92	0.23	48,62,67,69	0
5	GOL	C	101	6/6	0.93	0.19	54,59,75,77	0
5	GOL	B	502	6/6	0.94	0.15	56,64,67,79	0

6.5 Other polymers [i](#)

There are no such residues in this entry.