W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report ⓘ

Apr 11, 2022 – 04:07 PM EDT

PDB ID	:	7RI3
Title	:	Crystal structure of Albireti Toxin, a diphtheria toxin homolog, from Streptomyces albireticuli
Authors	:	Sugiman-Marangos, S.N.; Gill, S.K.; Melnyk, R.A.
Deposited on	:	2021-07-19
Resolution	:	2.69 Å(reported)

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

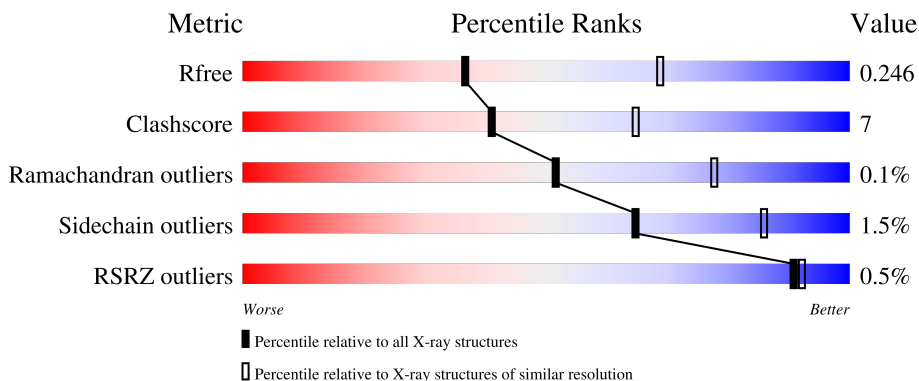|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.27
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

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 2.69 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		0.246
Clashscore		7
Ramachandran outliers		0.1%
Sidechain outliers		1.5%
RSRZ outliers		0.5%

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	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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	587	82% 15% • •
1	B	587	82% 16% • •
1	C	587	82% 17% •
1	D	587	83% 15% •

2   Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 18370 atoms, of which 84 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 Diphtheria_T domain-containing protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	572	Total	C	N	O	S	Se	0	1	0
			4408	2800	752	842	6	8			
1	B	575	Total	C	N	O	S	Se	0	1	0
			4413	2802	758	840	6	7			
1	C	583	Total	C	N	O	S	Se	0	0	0
			4452	2826	763	850	6	7			
1	D	578	Total	C	N	O	S	Se	0	1	0
			4430	2812	757	848	6	7			

- Molecule 2 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula: $C_8H_{18}O_5$).
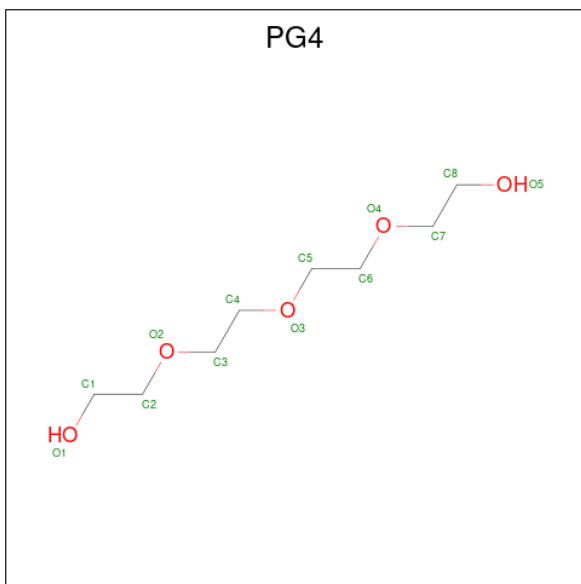
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	H	O	0	0
			31	8	18	5		
2	B	1	Total	C	H	O	0	0
			31	8	18	5		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	C	1	Total 31	C 8	H 18	O 5	0	0
2	D	1	Total 31	C 8	H 18	O 5	0	0

- Molecule 3 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).
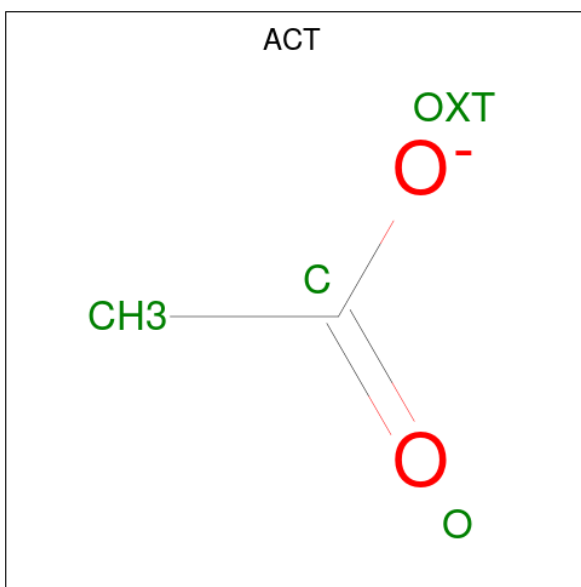
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total 7	C 2	H 3	O 2	0	0
3	B	1	Total 7	C 2	H 3	O 2	0	0
3	C	1	Total 7	C 2	H 3	O 2	0	0
3	D	1	Total 7	C 2	H 3	O 2	0	0

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

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

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	D	1	Total 1	Ca 1	0	0

- Molecule 5 is water.

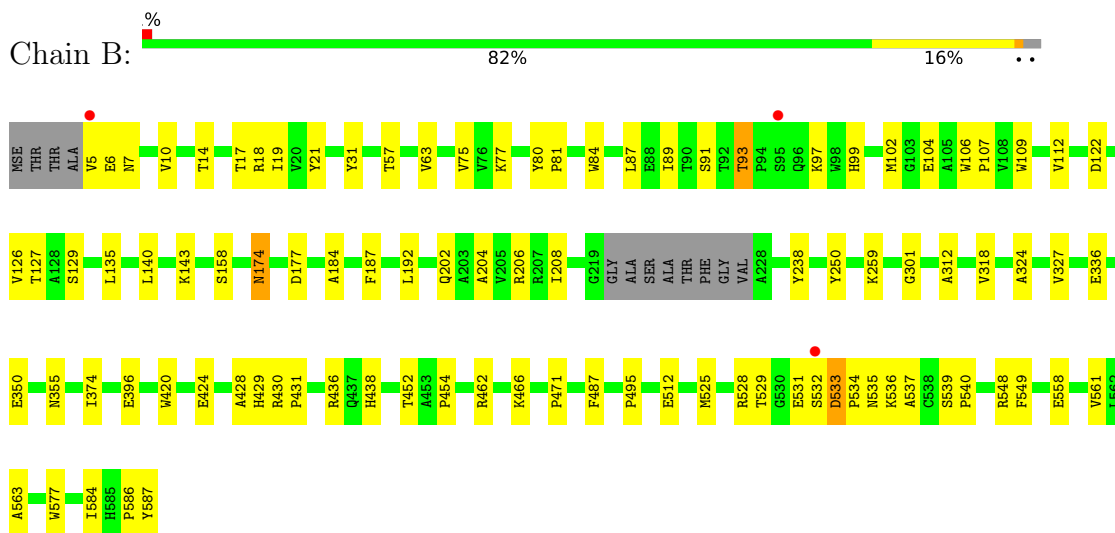| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	145	Total 145	O 145	0	0
5	B	120	Total 120	O 120	0	0
5	C	131	Total 131	O 131	0	0
5	D	115	Total 115	O 115	0	0

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.

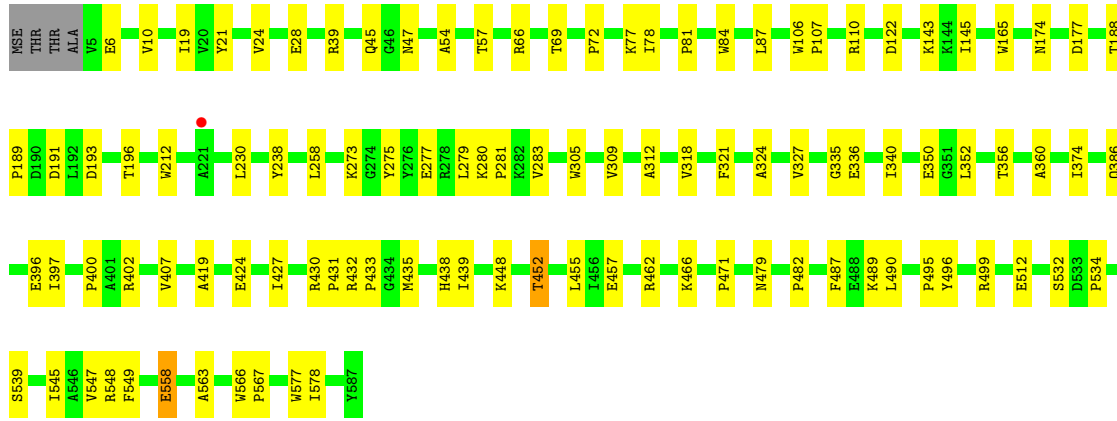- Molecule 1: Diphtheria_T domain-containing protein

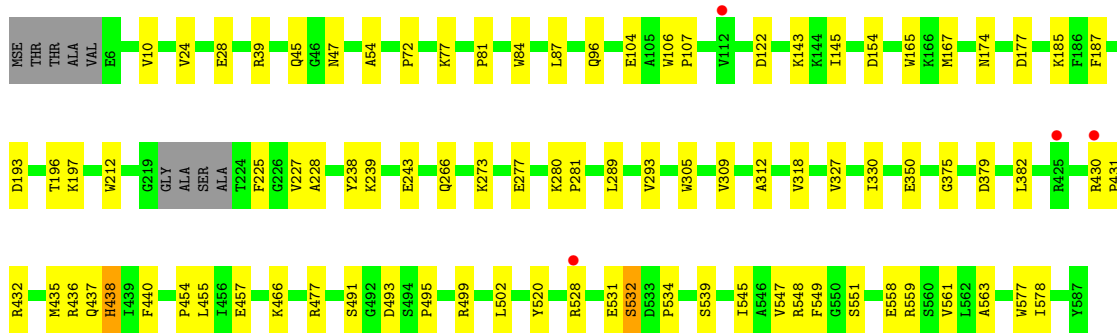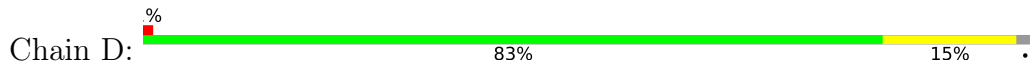- Molecule 1: Diphtheria_T domain-containing protein

- Molecule 1: Diphtheria_T domain-containing protein

MSE THR THR ALA V5 E6 V10 I19 V20 Y21 V24 E28 R39 Q45 G46 N47 A54 T57 R66 T69 P72 K77 I78 P81 W84 L87 W106 P107 R110 D122 K143 K144 I145 W165 N174 D177 T188

P189 D190 D191 L192 D193 T196 W212 A221 L230 Y238 L258 K273 G274 Y275 Y276 E277 R278 L279 K280 P281 K282 V283 W305 V309 A312 V318 F321 A324 V327 G335 E336 I340 E350 G351 L352 T356 A360 I374 Q386

E396 I397 P400 A401 R402 V407 A419 E424 I427 R430 P431 R432 P433 G434 M435 H438 I439 K448 T452 L455 I456 E457 R462 K466 P471 N479 P482 F487 E488 K489 L490 P495 Y496 R499 E512 S532 D533 P534

S539 I545 A546 V547 R548 F549 E558 A563 W566 P567 W577 I578 Y587

• Molecule 1: Diphtheria_T domain-containing protein

%

Chain D: 83%   15%

MSE THR THR ALA VAL E6 V10 V24 E28 R39 Q45 G46 N47 A54 P72 K77 P81 W84 L87 Q96 E104 A105 W106 P107 V112 D122 K143 K144 I145 D154 W165 K166 M167 N174 D177 K185 F186 F187

D193 T196 K197 W212 G219 GLY ALA SER ALA T224 F225 G226 V227 A228 Y238 K239 E243 Q266 K273 E277 K280 P281 L289 V293 W305 V309 A312 V318 V327 I330 E350 G375 D379 L382 R425 R430 P431

R432 M435 R436 Q437 H438 I439 F440 P454 L455 I456 E457 K466 R477 S491 G492 D493 S494 P495 R499 L502 Y520 R528 E531 S532 D533 P534 S539 I545 A546 V547 R548 F549 G550 S551 E558 R559 S560 V561 L562 A563 W577 I578 Y587

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	106.95Å 137.08Å 203.51Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	60.80 – 2.69 203.51 – 2.69	Depositor EDS
% Data completeness (in resolution range)	99.9 (60.80-2.69) 99.9 (203.51-2.69)	Depositor EDS
R_{merge}	0.19	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.60 (at 2.69Å)	Xtriage
Refinement program	PHENIX 1.19.1_4122+SVN	Depositor
R, R_{free}	0.203 , 0.248 0.199 , 0.246	Depositor DCC
R_{free} test set	4124 reflections (4.91%)	wwPDB-VP
Wilson B-factor (Å^2)	48.0	Xtriage
Anisotropy	0.292	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 44.2	EDS
L-test for twinning[2]	$< |L| > = 0.52$, $< L^2 > = 0.35$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	18370	wwPDB-VP
Average B, all atoms (Å^2)	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 43.56 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.7331e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[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: PG4, CA, ACT

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	0.37	0/4512	0.56	0/6138
1	B	0.35	0/4517	0.54	0/6147
1	C	0.36	0/4555	0.56	0/6203
1	D	0.34	0/4535	0.53	0/6175
All	All	0.36	0/18119	0.55	0/24663

There are no bond length outliers.

There are no bond angle outliers.

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	4408	0	4393	58	0
1	B	4413	0	4377	67	0
1	C	4452	0	4410	72	0
1	D	4430	0	4379	57	0
2	A	13	18	18	1	0
2	B	13	18	18	0	0
2	C	13	18	18	0	0
2	D	13	18	18	0	0
3	A	4	3	3	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	4	3	3	0	0
3	C	4	3	3	1	0
3	D	4	3	3	0	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
4	C	1	0	0	0	0
4	D	1	0	0	0	0
5	A	145	0	0	1	0
5	B	120	0	0	0	0
5	C	131	0	0	4	0
5	D	115	0	0	3	0
All	All	18286	84	17643	250	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 7.

The worst 5 of 250 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:291:ALA:HB1	1:B:5:VAL:HG22	1.36	1.02
1:B:127:THR:HG23	1:B:129:SER:H	1.22	1.01
1:B:318:VAL:HG12	1:B:374:ILE:HD11	1.46	0.96
1:D:77:LYS:HB3	1:D:177:ASP:HB3	1.58	0.84
1:A:233:ILE:HG21	1:A:353:ILE:CD1	2.09	0.82

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	569/587 (97%)	557 (98%)	11 (2%)	1 (0%)	47	73

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	572/587 (97%)	557 (97%)	13 (2%)	2 (0%)	41	66
1	C	581/587 (99%)	566 (97%)	15 (3%)	0	100	100
1	D	575/587 (98%)	558 (97%)	17 (3%)	0	100	100
All	All	2297/2348 (98%)	2238 (97%)	56 (2%)	3 (0%)	51	78

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	533	ASP
1	A	317	ASN
1	B	529	THR

5.3.2 Protein sidechains ⓘ

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	478/479 (100%)	468 (98%)	10 (2%)	53	80
1	B	473/479 (99%)	466 (98%)	7 (2%)	65	86
1	C	476/479 (99%)	470 (99%)	6 (1%)	69	87
1	D	475/479 (99%)	469 (99%)	6 (1%)	69	87
All	All	1902/1916 (99%)	1873 (98%)	29 (2%)	65	86

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	452	THR
1	D	551	SER
1	C	69	THR
1	D	491	SER
1	B	558	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	45	GLN
1	C	119	ASN
1	C	437	GLN
1	C	386	GLN
1	B	429	HIS

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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 12 ligands modelled in this entry, 4 are monoatomic - leaving 8 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
2	PG4	D	602	-	12,12,12	0.54	0	11,11,11	0.37	0
3	ACT	B	601	-	1,3,3	3.55	1 (100%)	0,3,3	-	-
3	ACT	A	602	-	1,3,3	3.60	1 (100%)	0,3,3	-	-
3	ACT	D	601	-	1,3,3	3.56	1 (100%)	0,3,3	-	-
3	ACT	C	602	-	1,3,3	3.75	1 (100%)	0,3,3	-	-
2	PG4	A	601	-	12,12,12	0.52	0	11,11,11	0.44	0
2	PG4	C	601	-	12,12,12	0.53	0	11,11,11	0.39	0
2	PG4	B	602	-	12,12,12	0.56	0	11,11,11	0.58	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
2	PG4	D	602	-	-	5/10/10/10	-
2	PG4	A	601	-	-	2/10/10/10	-
2	PG4	C	601	-	-	6/10/10/10	-
2	PG4	B	602	-	-	6/10/10/10	-

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	602	ACT	CH3-C	3.75	1.53	1.48
3	A	602	ACT	CH3-C	3.60	1.53	1.48
3	D	601	ACT	CH3-C	3.56	1.53	1.48
3	B	601	ACT	CH3-C	3.55	1.53	1.48

There are no bond angle outliers.

There are no chirality outliers.

5 of 19 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	C	601	PG4	O3-C5-C6-O4
2	B	602	PG4	O2-C3-C4-O3
2	B	602	PG4	O1-C1-C2-O2
2	C	601	PG4	O2-C3-C4-O3
2	B	602	PG4	O4-C7-C8-O5

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	602	ACT	1	0
2	A	601	PG4	1	0

5.7 Other polymers ⓘ

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	565/587 (96%)	-0.14	3 (0%)	91 92	29, 49, 74, 114	0
1	B	568/587 (96%)	-0.12	3 (0%)	91 92	32, 50, 80, 139	0
1	C	576/587 (98%)	-0.10	1 (0%)	95 96	29, 46, 81, 127	0
1	D	571/587 (97%)	-0.10	4 (0%)	87 89	33, 52, 79, 144	0
All	All	2280/2348 (97%)	-0.12	11 (0%)	91 92	29, 50, 78, 144	0

The worst 5 of 11 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	5	VAL	3.9
1	B	532	SER	3.3
1	A	531	GLU	2.9
1	A	528	ARG	2.6
1	B	95	SER	2.5

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, 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
2	PG4	B	602	13/13	0.89	0.20	48,62,72,78	0
2	PG4	D	602	13/13	0.91	0.16	55,68,79,80	0
2	PG4	C	601	13/13	0.92	0.17	49,61,69,70	0
3	ACT	C	602	4/4	0.92	0.20	47,50,58,58	0
3	ACT	B	601	4/4	0.93	0.14	64,66,81,81	0
2	PG4	A	601	13/13	0.94	0.16	46,58,67,71	0
3	ACT	A	602	4/4	0.94	0.15	55,55,67,67	0
4	CA	D	603	1/1	0.94	0.12	52,52,52,52	0
3	ACT	D	601	4/4	0.95	0.11	65,66,79,79	0
4	CA	B	603	1/1	0.97	0.10	48,48,48,48	0
4	CA	C	603	1/1	0.98	0.12	46,46,46,46	0
4	CA	A	603	1/1	0.99	0.10	47,47,47,47	0

6.5 Other polymers ⓘ

There are no such residues in this entry.