



wwPDB X-ray Structure Validation Summary Report ⓘ

May 14, 2020 – 10:23 am BST

PDB ID : 5MAK
Title : GFP-binding DARPin fusion gc_R7
Authors : Hansen, S.; Stueber, J.; Ernst, P.; Koch, A.; Bojar, D.; Batyuk, A.; Plueckthun, A.
Deposited on : 2016-11-03
Resolution : 2.50 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 ⓘ symbol.

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	:	2.11
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

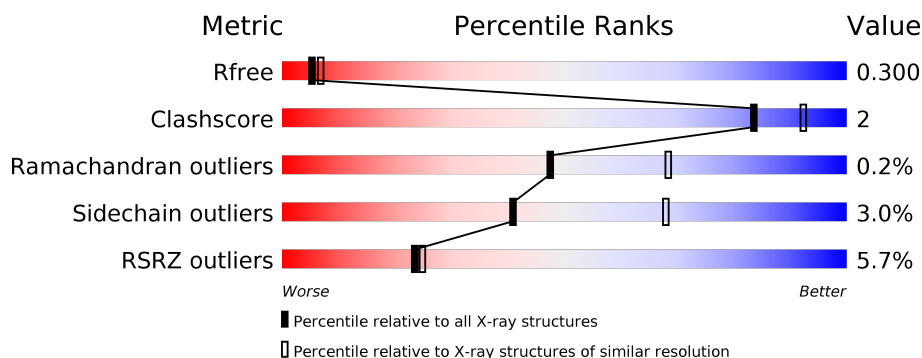
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 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 $\leq 5\%$. 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	B	243	<div> <div>2%</div> <div> <div></div> <div>88%</div> <div>5% • 6%</div> </div> </div>
1	D	243	<div> <div>6%</div> <div> <div></div> <div>86%</div> <div>7% • 7%</div> </div> </div>
2	A	298	<div> <div>6%</div> <div> <div></div> <div>87%</div> <div>8% • •</div> </div> </div>
2	C	298	<div> <div>7%</div> <div> <div></div> <div>90%</div> <div>5% • •</div> </div> </div>

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 15890 atoms, of which 7716 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 Green fluorescent protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	B	228	Total	C	H	N	O	S	0	0	0
			3568	1161	1744	307	350	6			
1	D	227	Total	C	H	N	O	S	0	1	0
			3564	1160	1740	307	351	6			

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-4	GLY	-	expression tag	UNP P42212
B	-3	PRO	-	expression tag	UNP P42212
B	-2	GLY	-	expression tag	UNP P42212
B	-1	SER	-	expression tag	UNP P42212
B	0	MET	-	expression tag	UNP P42212
B	1	VAL	-	expression tag	UNP P42212
B	64	LEU	PHE	conflict	UNP P42212
B	?	-	SER	deletion	UNP P42212
B	?	-	TYR	deletion	UNP P42212
B	66	CRO	GLY	chromophore	UNP P42212
B	231	LEU	HIS	conflict	UNP P42212
B	239	GLN	-	expression tag	UNP P42212
B	240	ALA	-	expression tag	UNP P42212
D	-4	GLY	-	expression tag	UNP P42212
D	-3	PRO	-	expression tag	UNP P42212
D	-2	GLY	-	expression tag	UNP P42212
D	-1	SER	-	expression tag	UNP P42212
D	0	MET	-	expression tag	UNP P42212
D	1	VAL	-	expression tag	UNP P42212
D	64	LEU	PHE	conflict	UNP P42212
D	?	-	SER	deletion	UNP P42212
D	?	-	TYR	deletion	UNP P42212
D	66	CRO	GLY	chromophore	UNP P42212
D	231	LEU	HIS	conflict	UNP P42212
D	239	GLN	-	expression tag	UNP P42212

Continued on next page...

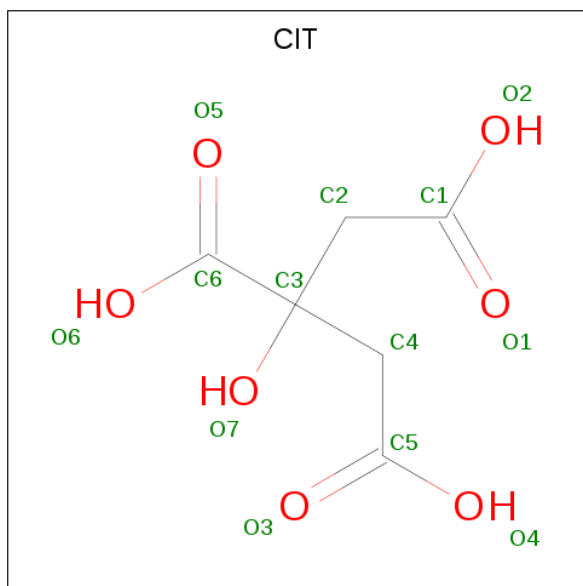
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	240	ALA	-	expression tag	UNP P42212

- Molecule 2 is a protein called R7.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	A	286	Total	C	H	N	O	S	0	0	0
			4263	1368	2104	371	419	1			
2	C	286	Total	C	H	N	O	S	0	0	0
			4263	1368	2104	371	419	1			

- Molecule 3 is CITRIC ACID (three-letter code: CIT) (formula: $C_6H_8O_7$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	H	O	0	0
			18	6	5	7		
3	A	1	Total	C	H	O	0	0
			18	6	5	7		
3	C	1	Total	C	H	O	0	0
			18	6	5	7		
3	C	1	Total	C	H	O	0	0
			18	6	5	7		
3	C	1	Total	C	H	O	0	0
			17	6	4	7		

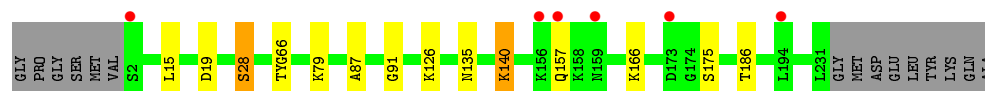
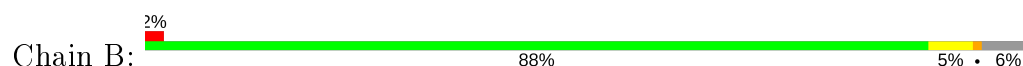
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	48	Total 48	O 48	0	0
4	A	38	Total 38	O 38	0	0
4	D	32	Total 32	O 32	0	0
4	C	25	Total 25	O 25	0	0

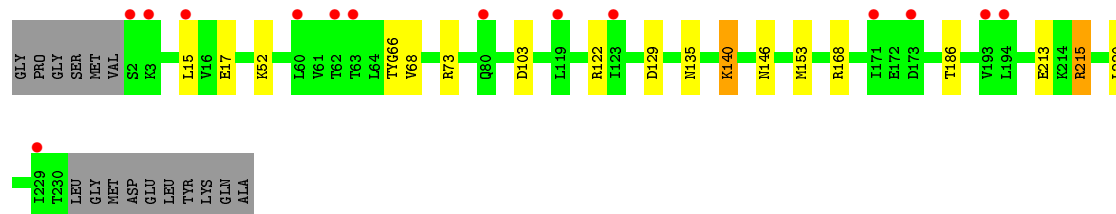
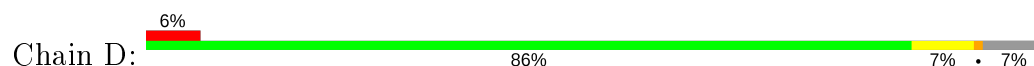
3 Residue-property plots [i](#)

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.

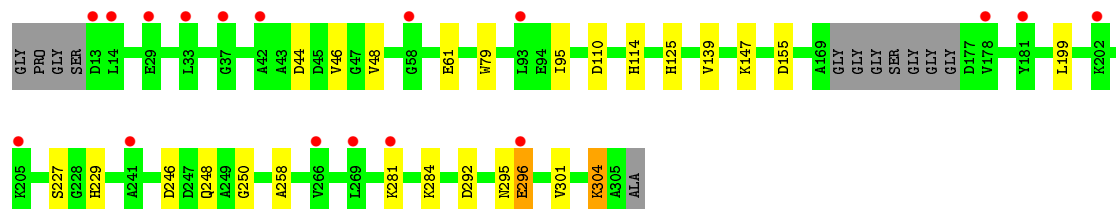
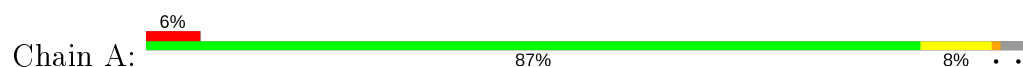
- Molecule 1: Green fluorescent protein



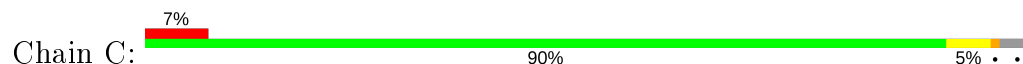
- Molecule 1: Green fluorescent protein

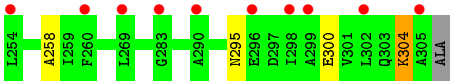


- Molecule 2: R7



- Molecule 2: R7





4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	57.96Å 61.38Å 89.38Å 93.12° 102.74° 94.76°	Depositor
Resolution (Å)	43.52 – 2.50 43.53 – 2.50	Depositor EDS
% Data completeness (in resolution range)	95.1 (43.52-2.50) 95.2 (43.53-2.50)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.30 (at 2.51Å)	Xtriage
Refinement program	PHENIX (1.11_2567: ???)	Depositor
R, R_{free}	0.255 , 0.304 0.251 , 0.300	Depositor DCC
R_{free} test set	1896 reflections (4.82%)	wwPDB-VP
Wilson B-factor (Å ²)	36.0	Xtriage
Anisotropy	0.556	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 40.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.35$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	15890	wwPDB-VP
Average B, all atoms (Å ²)	63.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.58% 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: CRO, CIT

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	B	0.25	0/1842	0.46	0/2489
1	D	0.25	0/1842	0.46	0/2489
2	A	0.24	0/2198	0.40	0/2994
2	C	0.24	0/2198	0.39	0/2994
All	All	0.24	0/8080	0.43	0/10966

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
1	B	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	157	GLN	Peptide

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	B	1824	1744	1775	5	0
1	D	1824	1740	1769	12	1
2	A	2159	2104	2114	13	1
2	C	2159	2104	2114	8	0
3	A	26	10	10	2	0
3	C	39	14	15	0	0
4	A	38	0	0	1	0
4	B	48	0	0	1	0
4	C	25	0	0	0	0
4	D	32	0	0	1	0
All	All	8174	7716	7797	37	1

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 2.

The worst 5 of 37 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:D:66:CRO:C3	1:D:68:VAL:CA	2.41	0.97
1:D:66:CRO:CA3	1:D:68:VAL:N	2.31	0.93
2:A:296:GLU:OE2	4:A:501:HOH:O	2.04	0.76
2:C:300:GLU:OE1	2:C:304:LYS:NZ	2.21	0.73
1:D:66:CRO:C3	1:D:68:VAL:HA	2.20	0.70

All (1) 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)
2:A:61:GLU:H	1:D:129:ASP:OD2[1_455]	1.59	0.01

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	B	223/243 (92%)	215 (96%)	7 (3%)	1 (0%)	34	54
1	D	223/243 (92%)	215 (96%)	8 (4%)	0	100	100
2	A	282/298 (95%)	269 (95%)	12 (4%)	1 (0%)	34	54
2	C	282/298 (95%)	270 (96%)	12 (4%)	0	100	100
All	All	1010/1082 (93%)	969 (96%)	39 (4%)	2 (0%)	47	68

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	175	SER
2	A	246	ASP

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	B	199/210 (95%)	193 (97%)	6 (3%)	41	68
1	D	199/210 (95%)	192 (96%)	7 (4%)	36	62
2	A	217/220 (99%)	212 (98%)	5 (2%)	50	76
2	C	217/220 (99%)	210 (97%)	7 (3%)	39	65
All	All	832/860 (97%)	807 (97%)	25 (3%)	41	68

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	15	LEU
1	D	103	ASP
2	C	246	ASP
1	D	52	LYS
1	D	140	LYS

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

Mol	Chain	Res	Type
2	A	196	HIS
1	D	157	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

2 non-standard protein/DNA/RNA residues 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	$\# Z > 2$
1	CRO	B	66	1	23,23,24	2.67	8 (34%)	30,32,34	2.98	9 (30%)
1	CRO	D	66	1	23,23,24	2.69	8 (34%)	30,32,34	2.81	7 (23%)

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
1	CRO	B	66	1	-	1/12/31/32	0/2/2/2
1	CRO	D	66	1	-	1/12/31/32	0/2/2/2

The worst 5 of 16 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	66	CRO	C1-N2	7.32	1.43	1.32
1	B	66	CRO	C1-N2	6.98	1.42	1.32
1	B	66	CRO	CA2-C2	6.19	1.54	1.48
1	D	66	CRO	CA2-C2	6.08	1.54	1.48
1	D	66	CRO	C1-N3	5.01	1.45	1.37

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	66	CRO	O2-C2-CA2	-8.72	126.07	130.96
1	D	66	CRO	CA2-C2-N3	8.47	107.37	103.37
1	B	66	CRO	CA2-C2-N3	8.35	107.32	103.37
1	D	66	CRO	O2-C2-CA2	-7.99	126.47	130.96
1	B	66	CRO	C2-N3-C1	-6.06	104.90	107.97

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	D	66	CRO	C3-CA3-N3-C2
1	B	66	CRO	C3-CA3-N3-C2

There are no ring outliers.

2 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	B	66	CRO	1	0
1	D	66	CRO	6	0

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

5 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
3	CIT	C	402	-	3,12,12	1.26	0	3,17,17	2.43	2 (66%)
3	CIT	C	403	-	3,12,12	1.21	0	3,17,17	1.45	1 (33%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
3	CIT	C	401	-	3,12,12	1.28	0	3,17,17	2.48	2 (66%)
3	CIT	A	402	-	3,12,12	1.45	0	3,17,17	1.85	1 (33%)
3	CIT	A	401	-	3,12,12	1.28	0	3,17,17	1.93	1 (33%)

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	CIT	C	402	-	-	3/6/16/16	-
3	CIT	C	403	-	-	1/6/16/16	-
3	CIT	C	401	-	-	3/6/16/16	-
3	CIT	A	402	-	-	3/6/16/16	-
3	CIT	A	401	-	-	1/6/16/16	-

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	401	CIT	C3-C4-C5	-3.51	109.37	114.98
3	C	402	CIT	C3-C2-C1	-3.10	110.02	114.98
3	A	402	CIT	C3-C4-C5	-2.68	110.69	114.98
3	C	402	CIT	C3-C4-C5	-2.50	110.98	114.98
3	C	401	CIT	C3-C2-C1	-2.47	111.03	114.98

There are no chirality outliers.

5 of 11 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	C	402	CIT	C2-C3-C4-C5
3	C	402	CIT	O7-C3-C4-C5
3	C	402	CIT	C6-C3-C4-C5
3	C	401	CIT	O7-C3-C4-C5
3	C	401	CIT	C6-C3-C4-C5

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	402	CIT	1	0
3	A	401	CIT	1	0

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	D	2
1	B	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	D	66:CRO	C3	68:VAL	N	1.70
1	B	66:CRO	C3	68:VAL	N	1.11
1	D	64:LEU	C	66:CRO	N1	1.04

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	B	227/243 (93%)	0.43	6 (2%) 56 59	24, 44, 73, 140	0
1	D	226/243 (93%)	0.71	14 (6%) 20 21	30, 55, 88, 137	0
2	A	286/298 (95%)	0.56	17 (5%) 22 23	30, 51, 76, 107	0
2	C	286/298 (95%)	0.79	21 (7%) 15 15	30, 62, 105, 241	0
All	All	1025/1082 (94%)	0.63	58 (5%) 23 25	24, 53, 90, 241	0

The worst 5 of 58 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	62	THR	5.4
1	D	63	THR	5.4
2	C	305	ALA	4.5
2	C	241	ALA	4.1
2	C	299	ALA	3.9

6.2 Non-standard residues in protein, DNA, RNA chains [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
1	CRO	B	66	22/23	0.92	0.19	26,31,37,39	0
1	CRO	D	66	22/23	0.94	0.24	33,55,106,122	0

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(\AA^2)	Q<0.9
3	CIT	C	403	13/13	0.49	0.33	78,81,97,97	0
3	CIT	C	402	13/13	0.70	0.26	90,91,111,111	0
3	CIT	C	401	13/13	0.74	0.21	74,77,93,94	0
3	CIT	A	401	13/13	0.82	0.19	60,62,74,74	0
3	CIT	A	402	13/13	0.88	0.19	58,67,81,81	0

6.5 Other polymers [i](#)

There are no such residues in this entry.