wwPDB X-ray Structure Validation Summary Report (i)

Nov 5, 2023 – 01:50 AM EST

PDB ID	:	5DPH
Title	:	sfGFP mutant - 149 p-cyano-L-phenylalanine
Authors	:	Dippel, A.B.; Olenginski, G.M.; Maurici, N.; Liskov, M.T.; Brewer, S.H.; Phillips-Piro, C.M.
Deposited on	:	2015-09-12
Resolution	:	1.42 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:

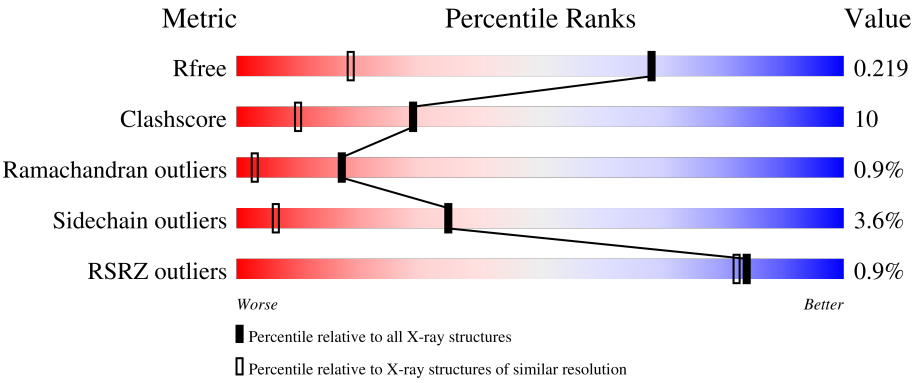
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 1.42 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	2579 (1.44-1.40)
Clashscore	141614	2696 (1.44-1.40)
Ramachandran outliers	138981	2632 (1.44-1.40)
Sidechain outliers	138945	2631 (1.44-1.40)
RSRZ outliers	127900	2528 (1.44-1.40)

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	237	81% 12% • 5%
1	B	237	2% 78% 19% ••

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	CO2	A	301	-	-	X	-

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 7710 atoms, of which 3465 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.

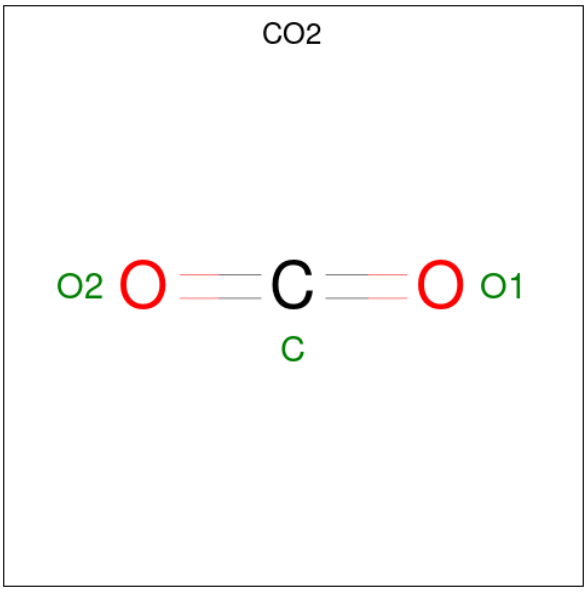
- Molecule 1 is a protein called Green fluorescent protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	225	Total	C	H	N	O	S	0	1	0
			3500	1141	1703	307	344	5			
1	B	233	Total	C	H	N	O	S	0	1	0
			3630	1186	1762	319	357	6			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	initiating methionine	UNP A0A059PIQ0
A	1	VAL	MET	engineered mutation	UNP A0A059PIQ0
A	2	SER	ARG	engineered mutation	UNP A0A059PIQ0
A	30	ARG	SER	engineered mutation	UNP A0A059PIQ0
A	?	CRO	THR	chromophore	UNP A0A059PIQ0
A	?	CRO	TYR	chromophore	UNP A0A059PIQ0
A	66	CRO	GLY	chromophore	UNP A0A059PIQ0
A	72	SER	ALA	engineered mutation	UNP A0A059PIQ0
A	80	ARG	GLN	engineered mutation	UNP A0A059PIQ0
A	149	4CF	ASN	engineered mutation	UNP A0A059PIQ0
A	206	VAL	ALA	engineered mutation	UNP A0A059PIQ0
B	0	MET	-	initiating methionine	UNP A0A059PIQ0
B	1	VAL	MET	engineered mutation	UNP A0A059PIQ0
B	2	SER	ARG	engineered mutation	UNP A0A059PIQ0
B	30	ARG	SER	engineered mutation	UNP A0A059PIQ0
B	?	CRO	THR	chromophore	UNP A0A059PIQ0
B	?	CRO	TYR	chromophore	UNP A0A059PIQ0
B	66	CRO	GLY	chromophore	UNP A0A059PIQ0
B	72	SER	ALA	engineered mutation	UNP A0A059PIQ0
B	80	ARG	GLN	engineered mutation	UNP A0A059PIQ0
B	149	4CF	ASN	engineered mutation	UNP A0A059PIQ0
B	206	VAL	ALA	engineered mutation	UNP A0A059PIQ0

- Molecule 2 is CARBON DIOXIDE (three-letter code: CO2) (formula: CO_2).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 3 1 2	0	0
2	B	1	Total C O 3 1 2	0	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	5	Total Mg 5 5	0	0
3	B	2	Total Mg 2 2	0	0

- Molecule 4 is water.

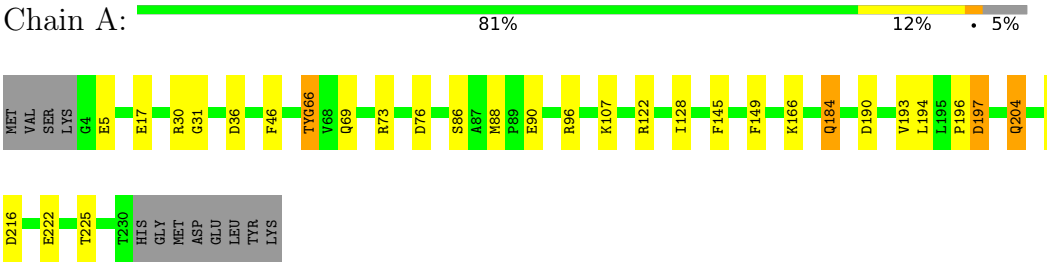
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	278	Total O 278 278	0	0
4	B	289	Total O 289 289	0	0

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.

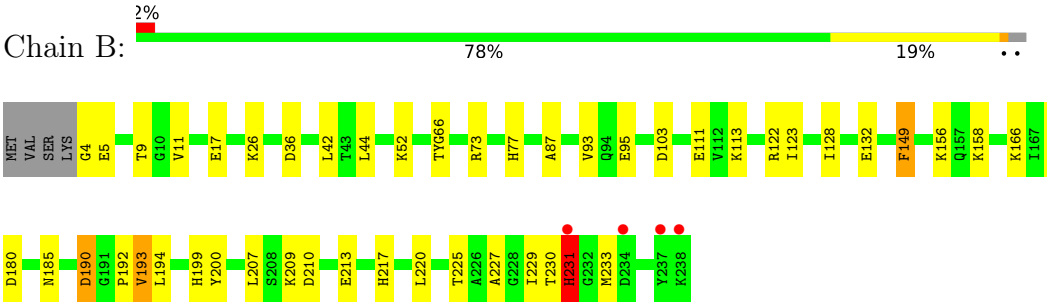
• Molecule 1: Green fluorescent protein

Chain A:



• Molecule 1: Green fluorescent protein

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 61	Depositor
Cell constants a, b, c, α, β, γ	47.27Å 47.27Å 344.55Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	57.00 – 1.42 57.42 – 1.42	Depositor EDS
% Data completeness (in resolution range)	99.4 (57.00-1.42) 99.4 (57.42-1.42)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.97 (at 1.42Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.180 , 0.208 0.188 , 0.219	Depositor DCC
R_{free} test set	4159 reflections (5.13%)	wwPDB-VP
Wilson B-factor (Å^2)	21.8	Xtriage
Anisotropy	0.267	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 53.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	0.489 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.98	EDS
Total number of atoms	7710	wwPDB-VP
Average B, all atoms (Å^2)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.03% 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: MG, CO2, 4CF, CRO

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	0.43	0/1799	0.61	0/2432
1	B	0.43	0/1872	0.62	0/2529
All	All	0.43	0/3671	0.61	0/4961

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	1797	1703	1734	30	0
1	B	1868	1762	1805	39	0
2	A	3	0	0	2	0
2	B	3	0	0	1	0
3	A	5	0	0	0	0
3	B	2	0	0	0	0
4	A	278	0	0	21	14
4	B	289	0	0	22	14
All	All	4245	3465	3539	69	16

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 69 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:30:ARG:NH1	4:A:405:HOH:O	2.07	0.88
1:A:36:ASP:OD1	4:A:401:HOH:O	1.92	0.86
1:A:197:ASP:OD1	4:A:402:HOH:O	1.95	0.83
1:A:212:ASN:ND2	4:A:408:HOH:O	2.13	0.80
1:A:88:MET:O	4:A:403:HOH:O	2.01	0.78

The worst 5 of 16 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)
4:A:634:HOH:O	4:B:662:HOH:O[6_755]	1.99	0.21
4:B:603:HOH:O	4:B:653:HOH:O[1_455]	2.02	0.18
4:A:457:HOH:O	4:B:612:HOH:O[6_765]	2.06	0.14
4:A:539:HOH:O	4:B:582:HOH:O[6_755]	2.06	0.14
4:A:564:HOH:O	4:B:637:HOH:O[6_755]	2.07	0.13

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	220/237 (93%)	206 (94%)	13 (6%)	1 (0%)	29 8
1	B	228/237 (96%)	212 (93%)	13 (6%)	3 (1%)	12 1
All	All	448/474 (94%)	418 (93%)	26 (6%)	4 (1%)	17 3

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	193	VAL
1	B	231	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	192	PRO
1	B	193	VAL

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	193/206 (94%)	187 (97%)	6 (3%)	40	9
1	B	201/206 (98%)	193 (96%)	8 (4%)	31	4
All	All	394/412 (96%)	380 (96%)	14 (4%)	35	6

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	123	ILE
1	B	128	ILE
1	B	231	HIS
1	B	190	ASP
1	B	194	LEU

Sometimes 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)

4 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	2.07	6 (26%)	30,32,34	2.70	11 (36%)
1	4CF	B	149	1	12,13,14	3.74	6 (50%)	13,16,18	0.90	0
1	4CF	A	149	1	12,13,14	3.70	7 (58%)	13,16,18	0.77	0
1	CRO	A	66	1	23,23,24	2.13	6 (26%)	30,32,34	2.96	10 (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
1	CRO	B	66	1	-	2/12/31/32	0/2/2/2
1	4CF	B	149	1	-	0/7/8/10	0/1/1/1
1	4CF	A	149	1	-	0/7/8/10	0/1/1/1
1	CRO	A	66	1	-	0/12/31/32	0/2/2/2

The worst 5 of 25 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	149	4CF	CE1-CD1	7.09	1.51	1.38
1	A	149	4CF	CE1-CD1	7.00	1.51	1.38
1	A	66	CRO	C1-N2	6.16	1.41	1.32
1	B	149	4CF	CD2-CG	6.05	1.51	1.38
1	A	149	4CF	CD2-CG	5.95	1.51	1.38

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	66	CRO	O2-C2-CA2	-9.75	125.49	130.96
1	B	66	CRO	O2-C2-CA2	-9.39	125.69	130.96
1	A	66	CRO	CA2-C2-N3	7.57	106.95	103.37
1	B	66	CRO	CA2-C2-N3	6.04	106.22	103.37
1	A	66	CRO	CA1-C1-N3	-4.38	119.50	124.75

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	B	66	CRO	C3-CA3-N3-C2
1	B	66	CRO	C3-CA3-N3-C1

There are no ring outliers.

2 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	B	149	4CF	2	0
1	A	66	CRO	5	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 9 ligands modelled in this entry, 7 are monoatomic - leaving 2 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	CO2	B	301	-	2,2,2	1.11	0	1,1,1	0.34	0
2	CO2	A	301	-	2,2,2	0.98	0	1,1,1	0.21	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	301	CO2	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	301	CO2	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	223/237 (94%)	-0.27	0 100 100	14, 24, 45, 58	0
1	B	231/237 (97%)	-0.17	4 (1%) 70 69	14, 25, 52, 76	0
All	All	454/474 (95%)	-0.22	4 (0%) 84 82	14, 25, 50, 76	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	238	LYS	2.5
1	B	231	HIS	2.2
1	B	237	TYR	2.2
1	B	234	ASP	2.1

6.2 Non-standard residues in protein, DNA, RNA chains

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
1	4CF	A	149	13/14	0.95	0.08	20,23,28,29	0
1	CRO	B	66	22/23	0.96	0.09	12,16,18,19	0
1	CRO	A	66	22/23	0.96	0.08	13,16,19,21	0
1	4CF	B	149	13/14	0.96	0.08	17,27,31,34	0

6.3 Carbohydrates

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, 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
2	CO2	A	301	3/3	0.92	0.10	24,24,28,30	0
3	MG	A	304	1/1	0.92	0.08	32,32,32,32	0
3	MG	A	302	1/1	0.93	0.09	31,31,31,31	0
3	MG	A	303	1/1	0.94	0.10	29,29,29,29	0
2	CO2	B	301	3/3	0.95	0.09	26,26,29,29	0
3	MG	B	303	1/1	0.96	0.05	34,34,34,34	0
3	MG	A	306	1/1	0.97	0.04	32,32,32,32	0
3	MG	B	302	1/1	0.99	0.08	19,19,19,19	0
3	MG	A	305	1/1	0.99	0.09	23,23,23,23	0

6.5 Other polymers (i)

There are no such residues in this entry.