Full wwPDB X-ray Structure Validation Report (i)

Feb 19, 2016 – 11:04 PM GMT

PDB ID	:	5FIS
Title	:	Exonuclease domain-containing 1 (Exd1) in the Gd bound conformation
Authors	:	Yang, Z.; Chen, K.M.; Pandey, R.R.; Homolka, D.; Reuter, M.; Rodino Janeiro, B.K.; Sachidanandam, R.; Fauvarque, M.O.; McCarthy, A.A.; Pillai, R.S.
Deposited on	:	2015-10-02
Resolution	:	1.60 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

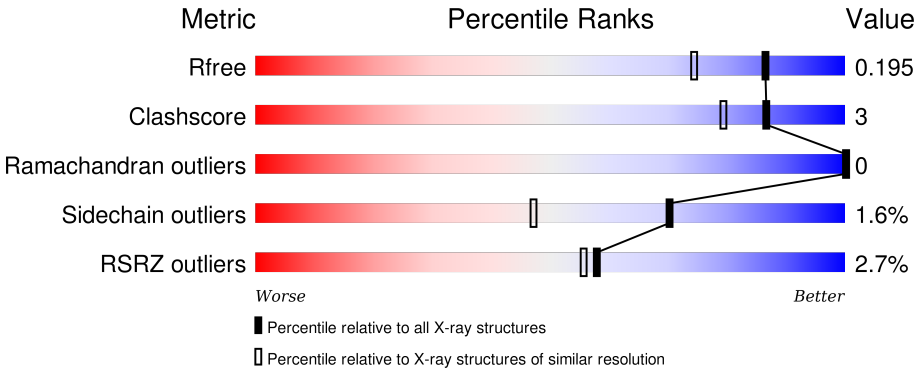
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026982
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20026982

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 1.60 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}	91344	2475 (1.60-1.60)
Clashscore	102246	2732 (1.60-1.60)
Ramachandran outliers	100387	2654 (1.60-1.60)
Sidechain outliers	100360	2653 (1.60-1.60)
RSRZ outliers	91569	2479 (1.60-1.60)

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. 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	243	2% 85% 13%
2	B	243	3% 85% 14%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 4342 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 EXD1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	241	Total	C	N	O	S	0	3	0
			1976	1263	334	369	10			

- Molecule 2 is a protein called EXD1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	241	Total	C	N	O	S	0	0	0
			1952	1242	334	366	10			

- Molecule 3 is GADOLINIUM ATOM (three-letter code: GD) (formula: Gd).

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

- Molecule 4 is water.

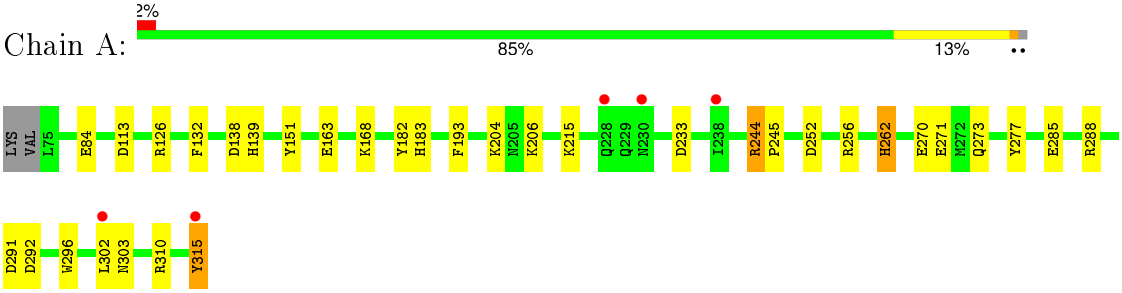
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	233	Total	O	0	0
			233	233		
4	B	178	Total	O	0	0
			178	178		

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 errors 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.

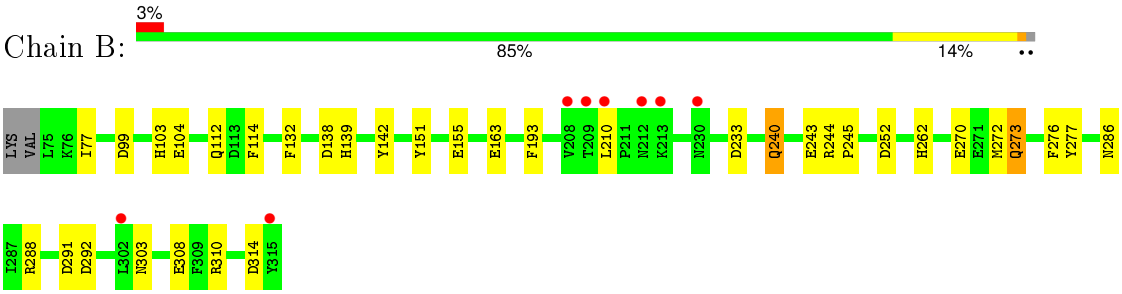
- Molecule 1: EXD1

Chain A:



- Molecule 2: EXD1

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	82.86Å 135.70Å 51.46Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.01 – 1.60 48.12 – 1.60	Depositor EDS
% Data completeness (in resolution range)	99.2 (50.01-1.60) 99.2 (48.12-1.60)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.95 (at 1.60Å)	Xtriage
Refinement program	REFMAC 5.8.0131	Depositor
R, R_{free}	0.157 , 0.182 0.172 , 0.195	Depositor DCC
R_{free} test set	3895 reflections (5.34%)	DCC
Wilson B-factor (Å^2)	19.3	Xtriage
Anisotropy	0.108	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 38.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Outliers	0 of 76779 reflections	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	4342	wwPDB-VP
Average B, all atoms (Å^2)	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.91% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: CSO, GD

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	1.40	9/2013 (0.4%)	1.33	19/2705 (0.7%)
2	B	1.38	11/1978 (0.6%)	1.27	15/2658 (0.6%)
All	All	1.39	20/3991 (0.5%)	1.30	34/5363 (0.6%)

All (20) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	271	GLU	CD-OE2	13.43	1.40	1.25
2	B	308	GLU	CD-OE1	9.46	1.36	1.25
2	B	163	GLU	CD-OE1	6.56	1.32	1.25
1	A	270	GLU	CG-CD	6.33	1.61	1.51
1	A	163	GLU	CD-OE1	6.12	1.32	1.25
2	B	104	GLU	CG-CD	5.77	1.60	1.51
1	A	285	GLU	CD-OE1	5.73	1.31	1.25
2	B	99	ASP	CB-CG	5.63	1.63	1.51
1	A	277	TYR	CZ-OH	5.62	1.47	1.37
1	A	151	TYR	CD1-CE1	5.60	1.47	1.39
1	A	84	GLU	CD-OE2	5.54	1.31	1.25
2	B	243	GLU	CD-OE1	5.35	1.31	1.25
2	B	273	GLN	CG-CD	5.29	1.63	1.51
1	A	182	TYR	CZ-OH	5.26	1.46	1.37
2	B	155	GLU	CD-OE2	5.23	1.31	1.25
1	A	296	TRP	CE3-CZ3	5.15	1.47	1.38
2	B	310	ARG	CZ-NH1	-5.14	1.26	1.33
2	B	142	TYR	CE1-CZ	-5.05	1.31	1.38
2	B	151	TYR	CZ-OH	5.04	1.46	1.37
2	B	314	ASP	CG-OD1	5.03	1.36	1.25

All (34) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	138	ASP	CB-CG-OD1	11.52	128.67	118.30
1	A	291	ASP	CB-CG-OD1	11.22	128.40	118.30
2	B	310	ARG	NE-CZ-NH2	-10.59	115.00	120.30
1	A	292	ASP	CB-CG-OD1	8.85	126.26	118.30
2	B	138	ASP	CB-CG-OD1	8.58	126.02	118.30
2	B	291	ASP	CB-CG-OD1	7.61	125.15	118.30
1	A	252	ASP	CB-CG-OD1	7.38	124.94	118.30
1	A	113	ASP	CB-CG-OD2	-7.33	111.70	118.30
1	A	244	ARG	NE-CZ-NH2	-7.32	116.64	120.30
2	B	277	TYR	CB-CG-CD1	-7.12	116.73	121.00
2	B	252	ASP	CB-CG-OD1	7.10	124.69	118.30
1	A	193	PHE	CB-CG-CD1	-6.95	115.94	120.80
1	A	132	PHE	CB-CG-CD2	-6.95	115.94	120.80
2	B	314	ASP	CB-CG-OD1	6.48	124.13	118.30
1	A	277	TYR	CE1-CZ-OH	-6.35	102.95	120.10
1	A	288	ARG	NE-CZ-NH2	-6.30	117.15	120.30
2	B	233	ASP	CB-CG-OD1	6.20	123.88	118.30
1	A	277	TYR	CB-CG-CD1	-5.98	117.42	121.00
1	A	256	ARG	NE-CZ-NH1	5.97	123.29	120.30
2	B	151	TYR	CG-CD2-CE2	-5.93	116.56	121.30
2	B	104	GLU	OE1-CD-OE2	-5.79	116.36	123.30
1	A	271	GLU	CG-CD-OE1	-5.67	106.96	118.30
1	A	138	ASP	CB-CG-OD2	-5.48	113.37	118.30
2	B	314	ASP	CB-CG-OD2	-5.40	113.44	118.30
2	B	292	ASP	CB-CG-OD1	5.34	123.11	118.30
1	A	132	PHE	CB-CG-CD1	5.26	124.48	120.80
2	B	276	PHE	CB-CG-CD1	5.26	124.48	120.80
1	A	126	ARG	NE-CZ-NH2	-5.23	117.69	120.30
1	A	302	LEU	CB-CG-CD1	-5.22	102.13	111.00
2	B	114	PHE	CB-CG-CD1	-5.21	117.15	120.80
1	A	233	ASP	CB-CG-OD1	5.13	122.91	118.30
2	B	193	PHE	CB-CG-CD1	-5.11	117.23	120.80
2	B	272	MET	CG-SD-CE	-5.07	92.09	100.20
1	A	277	TYR	CE1-CZ-CE2	5.03	127.84	119.80

There are no chirality outliers.

There are no planarity outliers.

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	1976	0	2004	13	0
2	B	1952	0	1974	13	0
3	A	2	0	0	0	0
3	B	1	0	0	0	0
4	A	233	0	0	7	0
4	B	178	0	0	3	0
All	All	4342	0	3978	21	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 (21) 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:139:HIS:HD2	4:A:2060:HOH:O	1.82	0.62
1:A:310:ARG:HG2	2:B:210:LEU:HD12	1.83	0.60
2:B:139:HIS:HD2	4:B:2046:HOH:O	1.86	0.58
4:A:2057:HOH:O	2:B:139:HIS:HE1	1.86	0.57
2:B:240:GLN:HE21	2:B:240:GLN:HA	1.68	0.57
1:A:204:LYS:NZ	4:A:2151:HOH:O	2.26	0.57
1:A:215:LYS:HE3	4:A:2148:HOH:O	2.05	0.56
2:B:103:HIS:HD2	4:B:2087:HOH:O	1.93	0.51
1:A:139:HIS:HE1	4:A:2010:HOH:O	1.93	0.49
1:A:315:TYR:CD1	1:A:315:TYR:C	2.86	0.48
1:A:244:ARG:HA	1:A:245:PRO:C	2.35	0.48
1:A:168:LYS:NZ	2:B:270:GLU:OE2	2.46	0.47
4:A:2136:HOH:O	2:B:286:ASN:ND2	2.45	0.47
2:B:112:GLN:NE2	4:B:2043:HOH:O	2.46	0.46
1:A:262:HIS:HD2	4:A:2093:HOH:O	1.98	0.45
2:B:244:ARG:HA	2:B:245:PRO:C	2.38	0.43
1:A:273[A]:GLN:NE2	2:B:273:GLN:NE2	2.67	0.43
1:A:183:HIS:HE1	2:B:288:ARG:O	2.01	0.43
1:A:310:ARG:HG2	2:B:210:LEU:CD1	2.49	0.42
1:A:206:LYS:HB3	1:A:206:LYS:HE3	1.77	0.42
2:B:77:ILE:HG21	2:B:77:ILE:HD13	1.80	0.41

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	241/243 (99%)	230 (95%)	11 (5%)	0	100	100
2	B	238/243 (98%)	228 (96%)	10 (4%)	0	100	100
All	All	479/486 (99%)	458 (96%)	21 (4%)	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	224/223 (100%)	221 (99%)	3 (1%)	76	56
2	B	220/223 (99%)	216 (98%)	4 (2%)	66	41
All	All	444/446 (100%)	437 (98%)	7 (2%)	70	47

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	262	HIS
1	A	303	ASN
1	A	315	TYR
2	B	132	PHE
2	B	240	GLN
2	B	262	HIS
2	B	303	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (13) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	112	GLN
1	A	139	HIS
1	A	147	GLN
1	A	183	HIS
1	A	262	HIS
1	A	303	ASN
2	B	103	HIS
2	B	112	GLN
2	B	139	HIS
2	B	147	GLN
2	B	196	GLN
2	B	240	GLN
2	B	311	ASN

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	CSO	A	128	1	3,6,7	1.70	1 (33%)	2,6,8	1.87	1 (50%)
2	CSO	B	128	2	3,6,7	2.13	1 (33%)	2,6,8	2.82	2 (100%)

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	CSO	A	128	1	-	0/1/5/7	0/0/0/0
2	CSO	B	128	2	-	0/1/5/7	0/0/0/0

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	128	CSO	CB-SG	-3.27	1.77	1.82
1	A	128	CSO	CB-SG	-2.54	1.78	1.82

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	128	CSO	O-C-CA	-3.34	116.77	125.72
1	A	128	CSO	O-C-CA	-2.63	118.67	125.72
2	B	128	CSO	CA-CB-SG	-2.18	106.51	114.09

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

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.

No monomer is involved in short contacts.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

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	240/243 (98%)	-0.33	5 (2%) 67 65	14, 21, 44, 90	0
2	B	240/243 (98%)	-0.29	8 (3%) 50 47	14, 22, 47, 78	0
All	All	480/486 (98%)	-0.31	13 (2%) 58 56	14, 21, 45, 90	0

All (13) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	315	TYR	7.8
2	B	208	VAL	4.3
2	B	212	ASN	3.8
2	B	209	THR	3.5
1	A	302	LEU	3.3
1	A	238	ILE	3.2
2	B	213	LYS	3.0
2	B	230	ASN	2.8
2	B	210	LEU	2.6
1	A	230	ASN	2.4
1	A	228	GLN	2.3
2	B	315	TYR	2.1
2	B	302	LEU	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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å^2)	Q<0.9
2	CSO	B	128	7/8	0.91	0.09	-	20,21,39,40	0
1	CSO	A	128	7/8	0.93	0.07	-	19,19,37,38	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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å^2)	Q<0.9
3	GD	A	1316	1/1	1.00	0.07	0.22	17,17,17,17	0
3	GD	B	1316	1/1	1.00	0.07	0.08	17,17,17,17	0
3	GD	A	1317	1/1	0.79	0.13	-	20,20,20,20	1

6.5 Other polymers (i)

There are no such residues in this entry.