wwPDB X-ray Structure Validation Summary Report (i)

Jun 13, 2024 – 12:44 AM EDT

PDB ID	:	1GKF
Title	:	Crystal structures of penicillin acylase enzyme-substrate complexes: Structural insights into the catalytic mechanism
Authors	:	McVey, C.E.; Walsh, M.A.; Dodson, G.G.; Wilson, K.S.; Brannigan, J.A.
Deposited on	:	2001-08-13
Resolution	:	1.41 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:

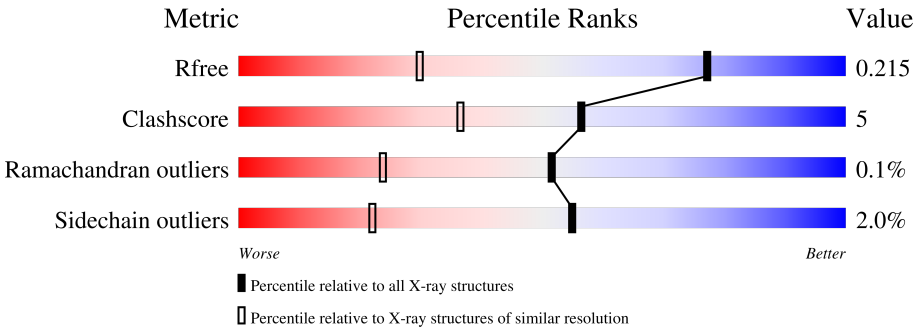
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 1.41 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)

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%

Mol	Chain	Length	Quality of chain
1	A	260	66% 11% 20%
2	B	557	84% 15%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 7283 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 PENICILLIN G ACYLASE ALPHA SUBUNIT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	208	1764	1121	300	333	10	5	13	1

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	16	SME	MET	modified residue	UNP P06875

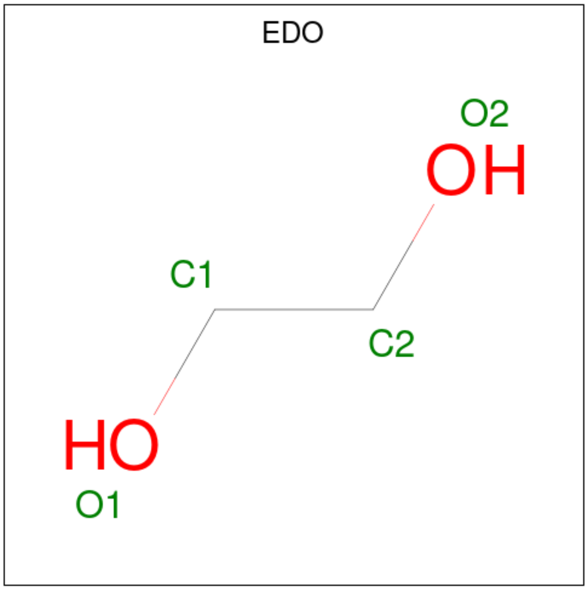
- Molecule 2 is a protein called PENICILLIN G ACYLASE BETA SUBUNIT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	557	4562	2896	788	867	11	15	18	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	241	ALA	ASN	engineered mutation	UNP P06875

- Molecule 3 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 7 4 3	0	1
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0

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

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

- Molecule 5 is water.

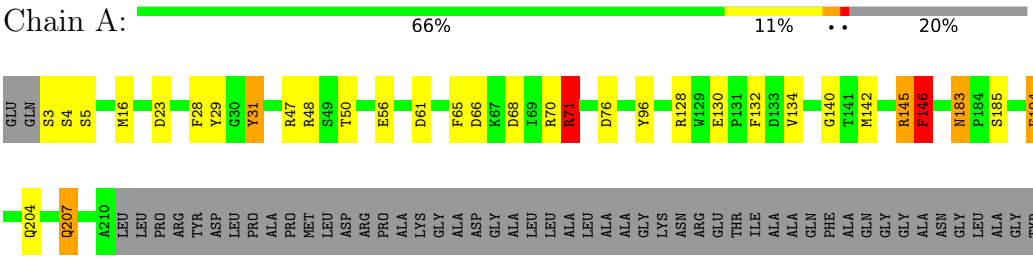
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	241	Total O 241 241	2	5
5	B	660	Total O 660 660	0	10

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. 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. 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.

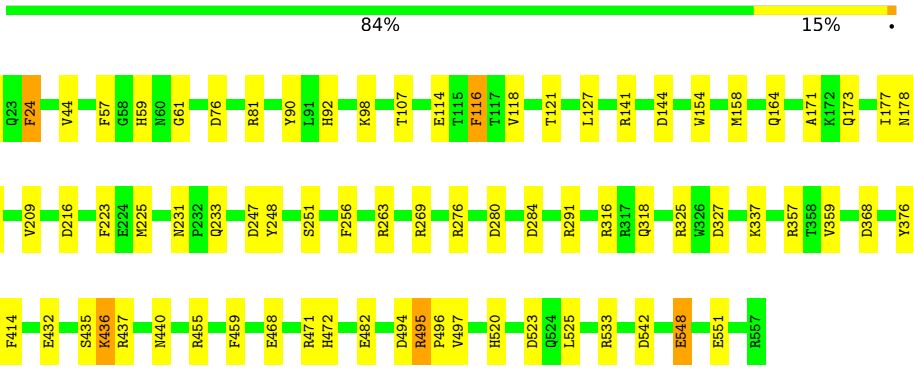
- Molecule 1: PENICILLIN G ACYLASE ALPHA SUBUNIT

Chain A:



- Molecule 2: PENICILLIN G ACYLASE BETA SUBUNIT

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	51.20Å 131.70Å 63.90Å 90.00° 105.70° 90.00°	Depositor
Resolution (Å)	25.00 – 1.41 24.91 – 1.41	Depositor EDS
% Data completeness (in resolution range)	97.4 (25.00-1.41) 97.2 (24.91-1.41)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.83 (at 1.41Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.148 , 0.168 0.204 , 0.215	Depositor DCC
R_{free} test set	4584 reflections (3.01%)	wwPDB-VP
Wilson B-factor (Å^2)	10.9	Xtriage
Anisotropy	0.178	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 41.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	7283	wwPDB-VP
Average B, all atoms (Å^2)	14.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.39% 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: CA, EDO, SME

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	0.96	1/1797 (0.1%)	1.64	33/2435 (1.4%)
2	B	0.98	4/4691 (0.1%)	1.52	55/6396 (0.9%)
All	All	0.97	5/6488 (0.1%)	1.55	88/8831 (1.0%)

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
1	A	0	3
2	B	0	3
All	All	0	6

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	98	LYS	CD-CE	18.66	1.98	1.51
2	B	81	ARG	CD-NE	-11.34	1.27	1.46
1	A	200	LYS	CG-CD	-8.39	1.24	1.52
2	B	432	GLU	CG-CD	-6.50	1.42	1.51
2	B	482	GLU	CD-OE2	5.77	1.31	1.25

The worst 5 of 88 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	455	ARG	NE-CZ-NH2	-18.45	111.08	120.30
2	B	533	ARG	NE-CZ-NH1	17.12	128.86	120.30
2	B	471	ARG	NE-CZ-NH1	13.79	127.19	120.30
1	A	48	ARG	NE-CZ-NH1	13.75	127.18	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	199	ARG	NE-CZ-NH2	-12.82	113.89	120.30

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	146[B]	PHE	Mainchain
1	A	5[B]	SER	Mainchain
1	A	71[B]	ARG	Mainchain
2	B	318[B]	GLN	Mainchain
2	B	376	TYR	Mainchain

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	1764	0	1702	29	0
2	B	4562	0	4364	45	0
3	B	55	0	80	5	0
4	B	1	0	0	0	0
5	A	241	0	0	6	0
5	B	660	0	0	10	2
All	All	7283	0	6146	63	2

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 5.

The worst 5 of 63 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:B:495[A]:ARG:NH2	5:B:2577[A]:HOH:O	1.70	1.24
2:B:114[A]:GLU:OE1	5:B:2154[A]:HOH:O	1.81	0.98
1:A:207:GLN:H	1:A:207:GLN:HE21	1.18	0.91
2:B:59:HIS:HD2	2:B:61:GLY:H	1.22	0.83
1:A:146[B]:PHE:HB2	5:A:2151:HOH:O	1.79	0.82

All (2) 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)
5:B:2126:HOH:O	5:B:2508:HOH:O[1_656]	2.04	0.16
5:B:2170:HOH:O	5:B:2378:HOH:O[1_656]	2.17	0.03

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	218/260 (84%)	213 (98%)	5 (2%)	0	100	100
2	B	573/557 (103%)	558 (97%)	14 (2%)	1 (0%)	47	22
All	All	791/817 (97%)	771 (98%)	19 (2%)	1 (0%)	51	24

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	251	SER

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	188/210 (90%)	182 (97%)	6 (3%)	39	8
2	B	477/459 (104%)	469 (98%)	8 (2%)	60	29
All	All	665/669 (99%)	651 (98%)	14 (2%)	55	20

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

Mol	Chain	Res	Type
2	B	154	TRP
2	B	173	GLN
2	B	436	LYS
2	B	378[B]	THR
2	B	404	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	401	GLN
2	B	472	HIS
2	B	520	HIS
2	B	473	GLN
2	B	59	HIS

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

1 non-standard protein/DNA/RNA residue is 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	SME	A	16	1	7,8,9	1.09	1 (14%)	4,9,11	0.60	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
1	SME	A	16	1	-	1/6/7/9	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	16	SME	OE-S	-2.16	1.44	1.50

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	16	SME	C-CA-CB-CG

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 15 ligands modelled in this entry, 1 is monoatomic - leaving 14 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	$\#\|Z\| > 2$
3	EDO	B	1563	-	3,3,3	0.58	0	2,2,2	1.33	0
3	EDO	B	1567	-	3,3,3	0.54	0	2,2,2	0.82	0
3	EDO	B	1559[A]	-	3,3,3	0.53	0	2,2,2	0.50	0
3	EDO	B	1564	-	3,3,3	0.27	0	2,2,2	0.40	0
3	EDO	B	1565	-	3,3,3	0.54	0	2,2,2	0.45	0
3	EDO	B	1561	-	3,3,3	0.61	0	2,2,2	0.18	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
3	EDO	B	1569	-	3,3,3	1.42	0	2,2,2	0.71	0
3	EDO	B	1558	-	3,3,3	0.58	0	2,2,2	0.21	0
3	EDO	B	1568	-	3,3,3	0.87	0	2,2,2	2.18	1 (50%)
3	EDO	B	1560	-	3,3,3	0.65	0	2,2,2	0.20	0
3	EDO	B	1570	-	3,3,3	0.65	0	2,2,2	0.54	0
3	EDO	B	1562	-	3,3,3	0.43	0	2,2,2	1.33	0
3	EDO	B	1559[B]	-	3,3,3	0.56	0	2,2,2	0.34	0
3	EDO	B	1566	-	3,3,3	0.61	0	2,2,2	0.45	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
3	EDO	B	1563	-	-	0/1/1/1	-
3	EDO	B	1567	-	-	0/1/1/1	-
3	EDO	B	1559[A]	-	-	0/1/1/1	-
3	EDO	B	1564	-	-	0/1/1/1	-
3	EDO	B	1565	-	-	0/1/1/1	-
3	EDO	B	1561	-	-	0/1/1/1	-
3	EDO	B	1569	-	-	1/1/1/1	-
3	EDO	B	1558	-	-	0/1/1/1	-
3	EDO	B	1568	-	-	1/1/1/1	-
3	EDO	B	1560	-	-	0/1/1/1	-
3	EDO	B	1570	-	-	0/1/1/1	-
3	EDO	B	1562	-	-	1/1/1/1	-
3	EDO	B	1559[B]	-	-	1/1/1/1	-
3	EDO	B	1566	-	-	0/1/1/1	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	B	1568	EDO	O1-C1-C2	2.58	130.47	111.91

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	B	1562	EDO	O1-C1-C2-O2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
3	B	1559[B]	EDO	O1-C1-C2-O2
3	B	1569	EDO	O1-C1-C2-O2
3	B	1568	EDO	O1-C1-C2-O2

There are no ring outliers.

4 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	1569	EDO	1	0
3	B	1558	EDO	1	0
3	B	1568	EDO	1	0
3	B	1562	EDO	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

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.