wwPDB X-ray Structure Validation Summary Report (i)

Oct 23, 2021 – 12:24 PM EDT

PDB ID	:	1D9D
Title	:	CRYSTALL STRUCTURE OF THE COMPLEX OF DNA POLYMERASE I KLENOW FRAGMENT WITH SHORT DNA FRAGMENT CARRYING 2'-0-AMINOPROPYL-RNA MODIFICATIONS 5'-D(TCG)-AP(AUC)-3'
Authors	:	Teplova, M.; Wallace, S.T.; Tereshko, V.; Minasov, G.; Simons, A.M.; Cook, P.D.; Manoharan, M.; Egli, M.
Deposited on	:	1999-10-27
Resolution	:	2.18 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 following versions of software and data (see references (i)) were used in the production of this report:

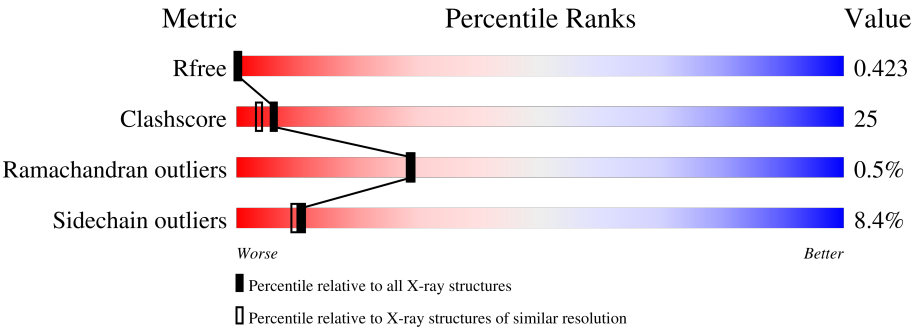
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.18 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	6864 (2.20-2.16)
Clashscore	141614	7689 (2.20-2.16)
Ramachandran outliers	138981	7564 (2.20-2.16)
Sidechain outliers	138945	7564 (2.20-2.16)

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	B	6	33% 17% 50%
2	A	605	61% 33% 5% •

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
1	U31	B	1005[A]	-	-	X	-
1	U31	B	1005[B]	-	-	X	-
1	C31	B	1006[B]	-	-	X	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	SO4	A	32	-	-	X	-
5	SO4	A	38	-	-	X	-

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 5149 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 DNA/RNA hybrid called 5'-D(*TP*CP*GP)-R(AP*(U31)P*(C31))-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	B	3	106	55	18	30	3	0	3	0

- Molecule 2 is a protein called DNA POLYMERASE I.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	A	601	4753	3008	830	899	16	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	324	MET	VAL	engineered mutation	UNP P00582

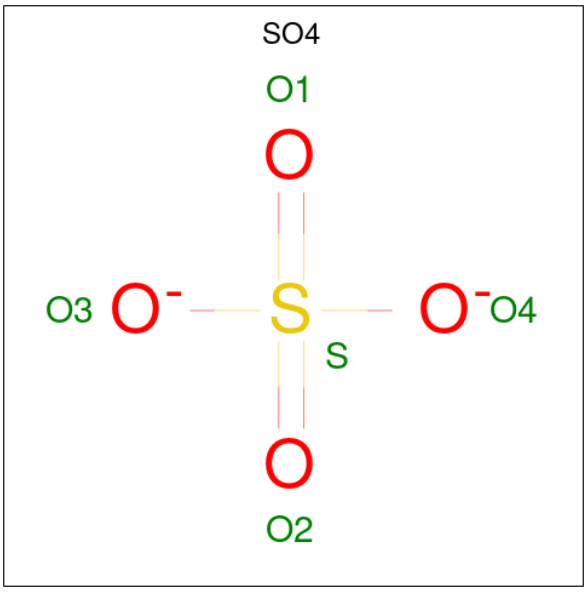
- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
3	A	4	4	4	0	0

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mg		
4	A	1	1	1	0	0

- Molecule 5 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total O S 5 4 1	0	0
5	A	1	Total O S 5 4 1	0	0
5	A	1	Total O S 5 4 1	0	0
5	A	1	Total O S 5 4 1	0	0

- Molecule 6 is water.

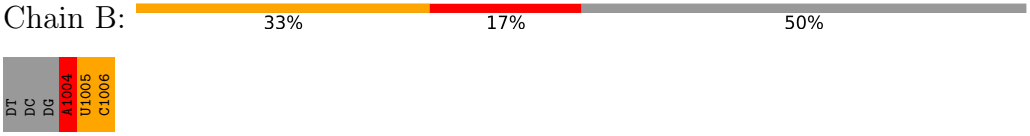
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	B	10	Total O 10 10	0	0
6	A	255	Total O 255 255	0	0

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. 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.

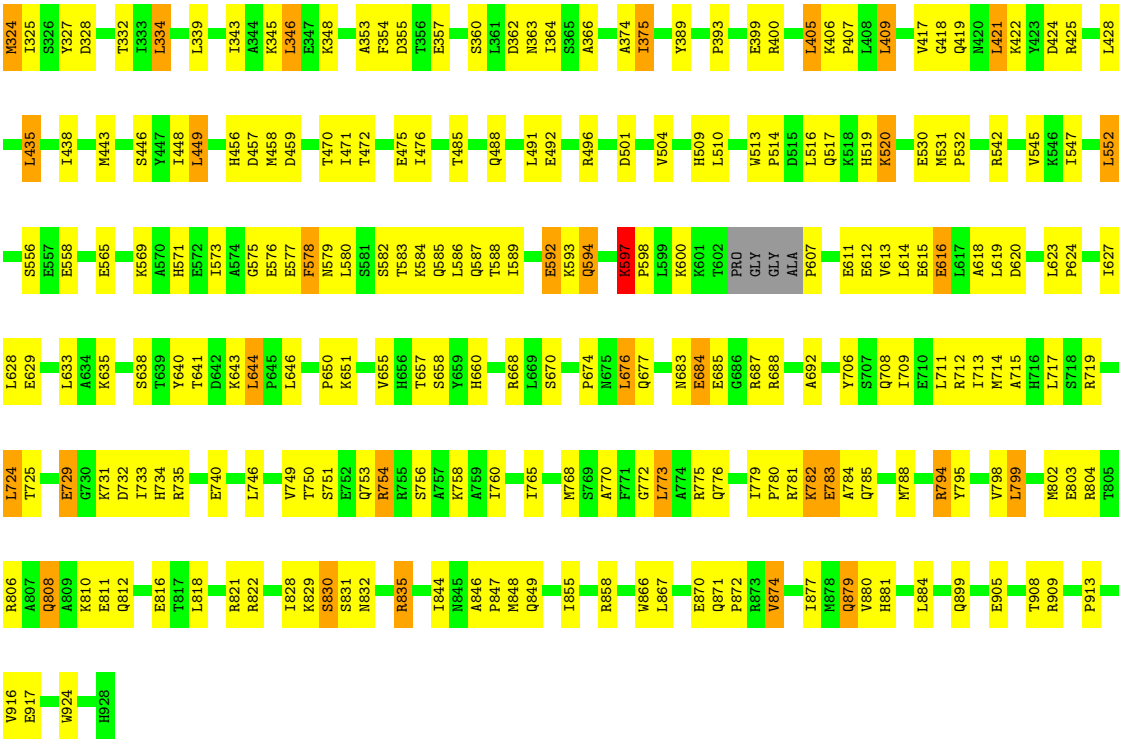
- Molecule 1: 5'-D(*TP*CP*GP)-R(AP*(U31)P*(C31))-3'

Chain B:



- Molecule 2: DNA POLYMERASE I

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43	Depositor
Cell constants a, b, c, α, β, γ	102.55Å 102.55Å 86.28Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.18 34.18 – 2.11	Depositor EDS
% Data completeness (in resolution range)	84.5 (20.00-2.18) 91.3 (34.18-2.11)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.56 (at 2.12Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.215 , 0.238 0.414 , 0.423	Depositor DCC
R_{free} test set	4799 reflections (10.21%)	wwPDB-VP
Wilson B-factor (Å^2)	26.1	Xtriage
Anisotropy	0.146	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.40 , 76.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.032 for h,-k,-l	Xtriage
F_o,F_c correlation	0.74	EDS
Total number of atoms	5149	wwPDB-VP
Average B, all atoms (Å^2)	33.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.13% 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, SO4, U31, ZN, C31

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	1.96	1/20 (5.0%)	2.55	0/29
2	A	0.33	0/4839	0.66	3/6547 (0.0%)
All	All	0.36	1/4859 (0.0%)	0.68	3/6576 (0.0%)

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	B	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	1004[A]	DA	C6-N1	-5.56	1.31	1.35

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	835	ARG	NE-CZ-NH1	7.13	123.86	120.30
2	A	607	PRO	N-CA-CB	5.57	109.99	103.30
2	A	597	LYS	N-CA-C	5.51	125.86	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	1004[A]	DA	Sidechain

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	106	0	73	65	0
2	A	4753	0	4752	202	13
3	A	4	0	0	0	0
4	A	1	0	0	0	0
5	A	20	0	0	6	0
6	A	255	0	0	12	11
6	B	10	0	0	10	0
All	All	5149	0	4825	239	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 25.

The worst 5 of 239 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:B:1005[A]:U31:H5	2:A:458:MET:CE	1.75	1.16
1:B:1005[A]:U31:HB'1	1:B:1005[A]:U31:H3'	1.23	1.14
1:B:1005[B]:U31:O5'	6:B:189:HOH:O	1.64	1.13
1:B:1005[A]:U31:HB'1	1:B:1005[A]:U31:C3'	1.86	1.05
1:B:1004[A]:DA:H2	2:A:422:LYS:NZ	1.55	1.05

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)
2:A:517:GLN:OE1	6:A:973:HOH:O[2_764]	1.67	0.53
2:A:565:GLU:OE2	6:A:1027:HOH:O[2_765]	1.87	0.33
2:A:751:SER:CA	6:A:1056:HOH:O[2_765]	1.95	0.25
6:A:1010:HOH:O	6:A:1026:HOH:O[3_654]	1.97	0.23
2:A:476:ILE:O	6:A:1128:HOH:O[3_654]	2.01	0.19

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
2	A	597/605 (99%)	563 (94%)	31 (5%)	3 (0%)	29 28

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	A	830	SER
2	A	594	GLN
2	A	597	LYS

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
2	A	500/510 (98%)	458 (92%)	42 (8%)	11 9

5 of 42 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	A	754	ARG
2	A	821	ARG
2	A	773	LEU
2	A	794	ARG
2	A	870	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
2	A	879	GLN
2	A	901	HIS
2	A	543	ASN
2	A	571	HIS
2	A	587	GLN

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	C31	B	1006[B]	1	18,25,26	1.23	1 (5%)	21,34,37	2.98	5 (23%)
1	U31	B	1005[A]	1	17,25,26	1.45	3 (17%)	18,34,37	2.93	8 (44%)
1	U31	B	1005[B]	1	14,17,26	1.22	2 (14%)	14,24,37	1.27	1 (7%)
1	C31	B	1006[A]	1,3	18,25,26	1.59	5 (27%)	21,34,37	3.09	5 (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	C31	B	1006[B]	1	-	7/10/30/31	0/2/2/2
1	U31	B	1005[A]	1	-	7/10/30/31	0/2/2/2
1	U31	B	1005[B]	1	-	2/3/18/31	0/2/2/2
1	C31	B	1006[A]	1,3	-	4/10/30/31	0/2/2/2

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	1006[B]	C31	C4-N4	-3.86	1.23	1.35
1	B	1006[A]	C31	C4-N4	-3.73	1.24	1.35
1	B	1006[A]	C31	O5'-C5'	3.32	1.52	1.44
1	B	1005[A]	U31	O2'-CA'	3.18	1.51	1.43
1	B	1005[B]	U31	C4-N3	2.78	1.37	1.33

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	1006[A]	C31	C2-N3-C4	11.04	127.52	116.34
1	B	1006[B]	C31	C2-N3-C4	10.55	127.03	116.34
1	B	1005[A]	U31	O2'-C2'-C1'	7.38	123.04	109.04
1	B	1006[B]	C31	C5-C4-N3	-6.66	114.03	121.72
1	B	1006[A]	C31	C5-C4-N3	-6.32	114.42	121.72

There are no chirality outliers.

5 of 20 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	B	1005[A]	U31	C3'-C2'-O2'-CA'
1	B	1005[A]	U31	CB'-CA'-O2'-C2'
1	B	1006[A]	C31	C4'-C5'-O5'-P
1	B	1006[B]	C31	C4'-C5'-O5'-P
1	B	1006[B]	C31	C3'-C4'-C5'-O5'

There are no ring outliers.

4 monomers are involved in 55 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	B	1006[B]	C31	17	0
1	B	1005[A]	U31	22	0
1	B	1005[B]	U31	18	0
1	B	1006[A]	C31	7	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 9 ligands modelled in this entry, 5 are monoatomic - leaving 4 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	$\#\|Z\| > 2$
5	SO4	A	38	-	4,4,4	0.43	0	6,6,6	0.14	0
5	SO4	A	32	-	4,4,4	0.52	0	6,6,6	0.17	0
5	SO4	A	40	-	4,4,4	0.49	0	6,6,6	0.14	0
5	SO4	A	94	-	4,4,4	0.45	0	6,6,6	0.15	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.

3 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	38	SO4	3	0
5	A	32	SO4	2	0
5	A	94	SO4	1	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.