wwPDB X-ray Structure Validation Summary Report (i)

Aug 9, 2020 – 06:00 AM BST

PDB ID	:	1FUJ
Title	:	PR3 (MYELOBLASTIN)
Authors	:	Fujinaga, M.; Chernaia, M.M.; Halenbeck, R.; Koths, K.; James, M.N.G.
Deposited on	:	1996-01-25
Resolution	:	2.20 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 (1)) were used in the production of this report:

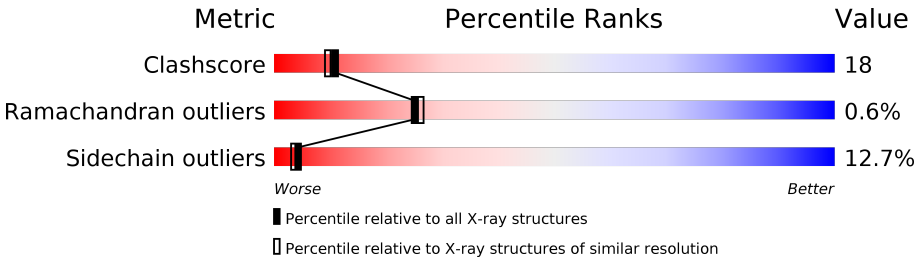
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13.1

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 2.20 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))
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)

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 <=5%

Mol	Chain	Length	Quality of chain
1	A	221	65% 30%
1	B	221	56% 38% 5%
1	C	221	61% 33% 5%
1	D	221	60% 34% 5%
2	E	2	50% 50%
2	F	2	50% 50%
2	G	2	50% 50%
2	H	2	50% 50%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 7174 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 PR3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	221	Total	C	N	O	S	0	0	0
			1708	1083	311	303	11			
1	B	221	Total	C	N	O	S	0	0	0
			1708	1083	311	303	11			
1	C	221	Total	C	N	O	S	0	0	0
			1708	1083	311	303	11			
1	D	221	Total	C	N	O	S	0	0	0
			1708	1083	311	303	11			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	103	ILE	VAL	conflict	UNP P24158
B	103	ILE	VAL	conflict	UNP P24158
C	103	ILE	VAL	conflict	UNP P24158
D	103	ILE	VAL	conflict	UNP P24158

- Molecule 2 is an oligosaccharide called alpha-L-fucopyranose-(1-6)-2-acetamido-2-deoxy-beta-D-glucopyranose.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	E	2	Total	C	N	O	0	0	0
			24	14	1	9			
2	F	2	Total	C	N	O	0	0	0
			24	14	1	9			
2	G	2	Total	C	N	O	0	0	0
			24	14	1	9			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	H	2	Total	C	N	O	0	0	0
			24	14	1	9			

- Molecule 3 is water.

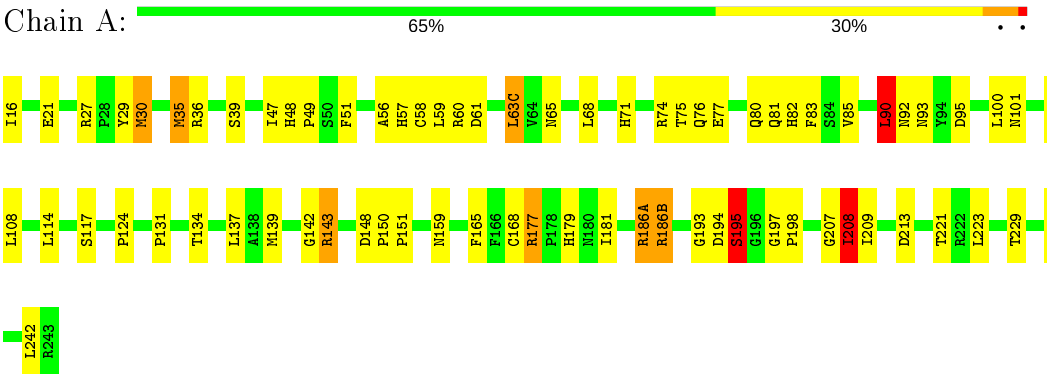
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	74	Total	O	0	0
			74	74		
3	B	48	Total	O	0	0
			48	48		
3	C	67	Total	O	0	0
			67	67		
3	D	57	Total	O	0	0
			57	57		

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.

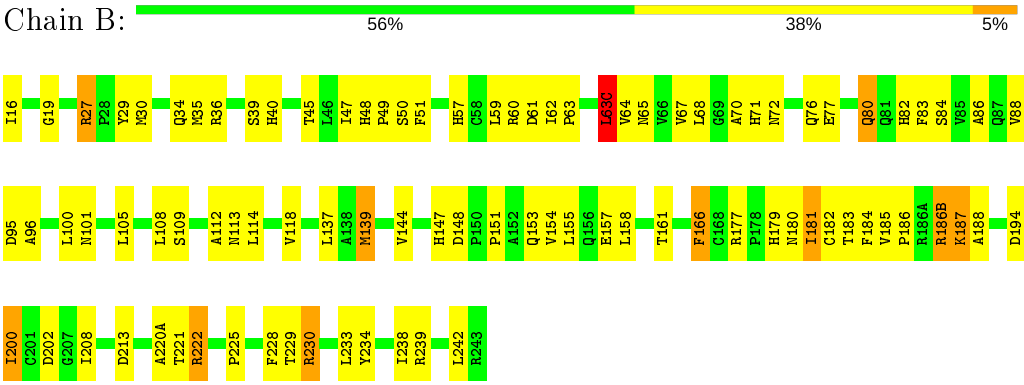
- Molecule 1: PR3

Chain A:



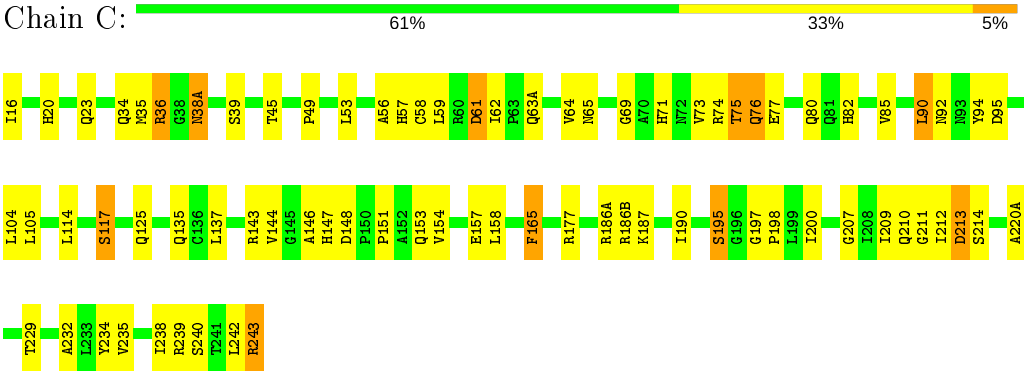
- Molecule 1: PR3

Chain B:



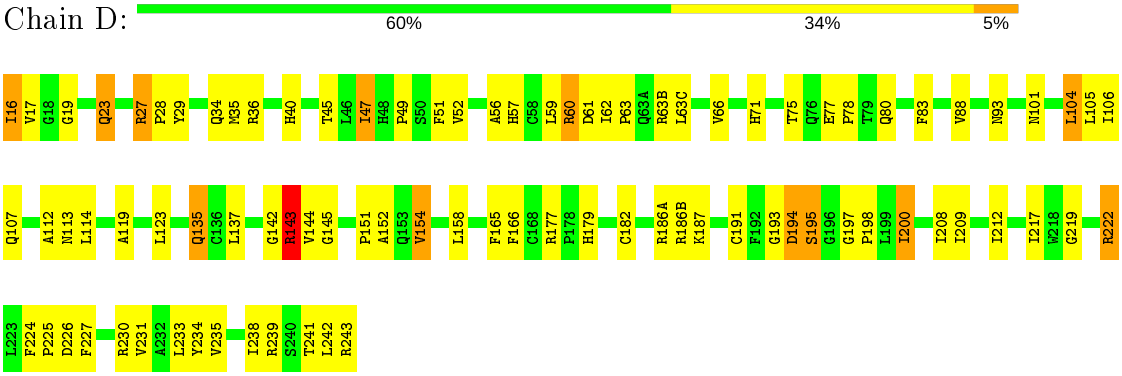
- Molecule 1: PR3

Chain C:



• Molecule 1: PR3

Chain D:



• Molecule 2: alpha-L-fucopyranose-(1-6)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain E:

• Molecule 2: alpha-L-fucopyranose-(1-6)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain F:

• Molecule 2: alpha-L-fucopyranose-(1-6)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain G:

• Molecule 2: alpha-L-fucopyranose-(1-6)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain H:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	85.60Å 54.07Å 113.51Å 90.00° 90.69° 90.00°	Depositor
Resolution (Å)	10.00 – 2.20 68.74 – 1.98	Depositor EDS
% Data completeness (in resolution range)	93.4 (10.00-2.20) 89.6 (68.74-1.98)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.22 (at 1.98Å)	Xtriage
Refinement program	TNT	Depositor
R, R_{free}	(Not available) , (Not available) 0.179 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	26.4	Xtriage
Anisotropy	0.716	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 124.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.043 for h,-k,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	7174	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 4.50% 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: NAG, FUC

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	0.97	2/1754 (0.1%)	1.09	6/2393 (0.3%)
1	B	0.92	0/1754	1.08	3/2393 (0.1%)
1	C	0.90	0/1754	1.01	1/2393 (0.0%)
1	D	0.91	0/1754	1.03	6/2393 (0.3%)
All	All	0.92	2/7016 (0.0%)	1.05	16/9572 (0.2%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	195	SER	CA-CB	5.68	1.61	1.52
1	A	195	SER	CB-OG	5.40	1.49	1.42

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	90	LEU	CA-CB-CG	7.78	133.19	115.30
1	A	63(C)	LEU	CA-CB-CG	-6.91	99.41	115.30
1	A	208	ILE	CG1-CB-CG2	-6.29	97.56	111.40
1	A	108	LEU	CB-CG-CD2	-5.91	100.96	111.00
1	D	200	ILE	CG1-CB-CG2	-5.83	98.56	111.40

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	1708	0	1672	54	0
1	B	1708	0	1672	69	0
1	C	1708	0	1672	59	0
1	D	1708	0	1672	70	0
2	E	24	0	22	1	0
2	F	24	0	22	2	0
2	G	24	0	22	1	0
2	H	24	0	22	1	0
3	A	74	0	0	3	0
3	B	48	0	0	3	0
3	C	67	0	0	2	0
3	D	57	0	0	2	0
All	All	7174	0	6776	249	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 18.

The worst 5 of 249 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:49:PRO:HG3	1:D:114:LEU:HD11	1.46	0.98
1:C:59:LEU:HD22	1:C:62:ILE:HD11	1.51	0.92
1:D:63:PRO:HD2	1:D:63(C):LEU:HD12	1.53	0.91
1:B:230:ARG:HG2	1:B:233:LEU:HD13	1.56	0.86
1:D:123:LEU:HD23	1:D:209:ILE:HD11	1.60	0.83

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	219/221 (99%)	204 (93%)	14 (6%)	1 (0%)	29	31
1	B	219/221 (99%)	197 (90%)	20 (9%)	2 (1%)	17	16
1	C	219/221 (99%)	201 (92%)	17 (8%)	1 (0%)	29	31
1	D	219/221 (99%)	205 (94%)	13 (6%)	1 (0%)	29	31
All	All	876/884 (99%)	807 (92%)	64 (7%)	5 (1%)	25	26

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	151	PRO
1	C	146	ALA
1	D	151	PRO
1	A	150	PRO
1	B	39	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	187/187 (100%)	166 (89%)	21 (11%)	6	5
1	B	187/187 (100%)	161 (86%)	26 (14%)	3	3
1	C	187/187 (100%)	162 (87%)	25 (13%)	4	3
1	D	187/187 (100%)	164 (88%)	23 (12%)	4	4
All	All	748/748 (100%)	653 (87%)	95 (13%)	4	3

5 of 95 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	213	ASP
1	C	75	THR
1	D	166	PHE
1	B	222	ARG
1	C	34	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 29 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	113	ASN
1	C	63(A)	GLN
1	D	135	GLN
1	B	210	GLN
1	C	65	ASN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

8 monosaccharides 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	$\#\|Z\| > 2$
2	NAG	E	1	1,2	14,14,15	1.47	2 (14%)	17,19,21	1.95	6 (35%)
2	FUC	E	2	2	10,10,11	0.63	0	14,14,16	2.61	9 (64%)
2	NAG	F	1	1,2	14,14,15	1.33	2 (14%)	17,19,21	2.75	8 (47%)
2	FUC	F	2	2	10,10,11	0.78	0	14,14,16	2.96	6 (42%)
2	NAG	G	1	1,2	14,14,15	1.60	2 (14%)	17,19,21	1.65	4 (23%)
2	FUC	G	2	2	10,10,11	0.85	0	14,14,16	1.31	3 (21%)
2	NAG	H	1	1,2	14,14,15	1.65	3 (21%)	17,19,21	2.02	6 (35%)
2	FUC	H	2	2	10,10,11	0.71	0	14,14,16	2.49	6 (42%)

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
2	NAG	E	1	1,2	-	4/6/23/26	0/1/1/1
2	FUC	E	2	2	-	-	0/1/1/1
2	NAG	F	1	1,2	-	2/6/23/26	0/1/1/1
2	FUC	F	2	2	-	-	0/1/1/1
2	NAG	G	1	1,2	-	4/6/23/26	0/1/1/1
2	FUC	G	2	2	-	-	0/1/1/1
2	NAG	H	1	1,2	-	4/6/23/26	0/1/1/1
2	FUC	H	2	2	-	-	0/1/1/1

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	H	1	NAG	C8-C7	4.53	1.60	1.50
2	G	1	NAG	C8-C7	4.01	1.58	1.50
2	E	1	NAG	C8-C7	3.87	1.58	1.50
2	F	1	NAG	C8-C7	2.63	1.56	1.50
2	F	1	NAG	O5-C1	-2.62	1.39	1.43

The worst 5 of 48 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	F	2	FUC	O5-C1-C2	7.59	122.48	110.77
2	F	1	NAG	O5-C5-C6	-6.12	97.61	107.20
2	H	2	FUC	O5-C1-C2	5.98	120.00	110.77
2	E	2	FUC	O5-C1-C2	5.87	119.84	110.77
2	F	2	FUC	C1-C2-C3	-4.57	104.05	109.67

There are no chirality outliers.

5 of 14 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	E	1	NAG	O7-C7-N2-C2
2	G	1	NAG	O5-C5-C6-O6
2	E	1	NAG	C8-C7-N2-C2
2	H	1	NAG	C8-C7-N2-C2
2	H	1	NAG	O7-C7-N2-C2

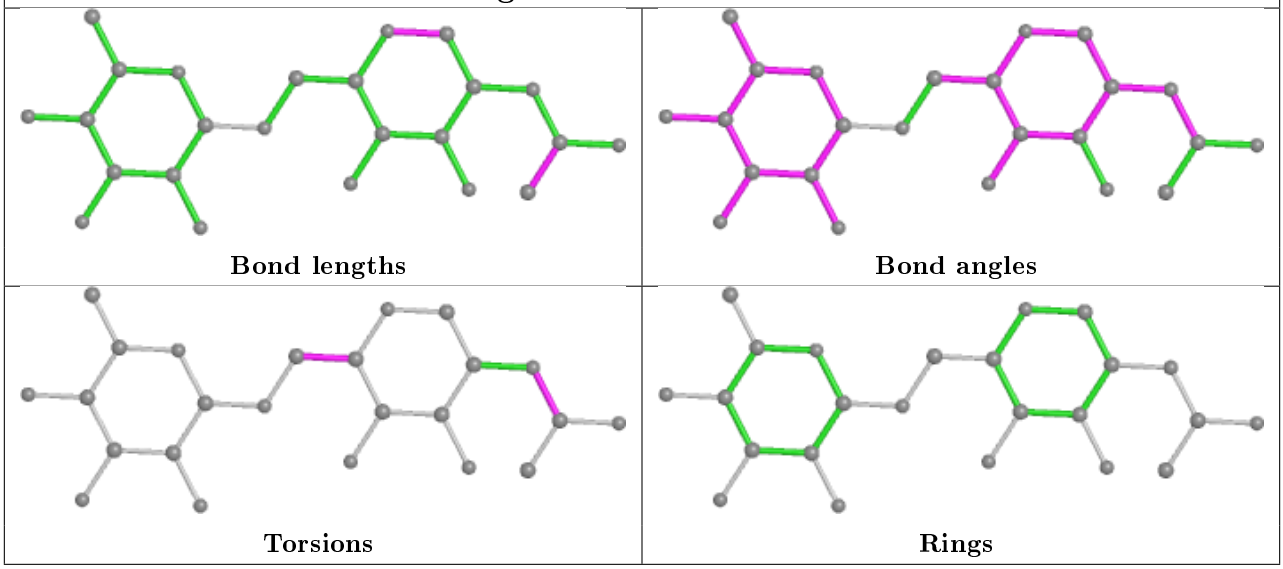
There are no ring outliers.

4 monomers are involved in 5 short contacts:

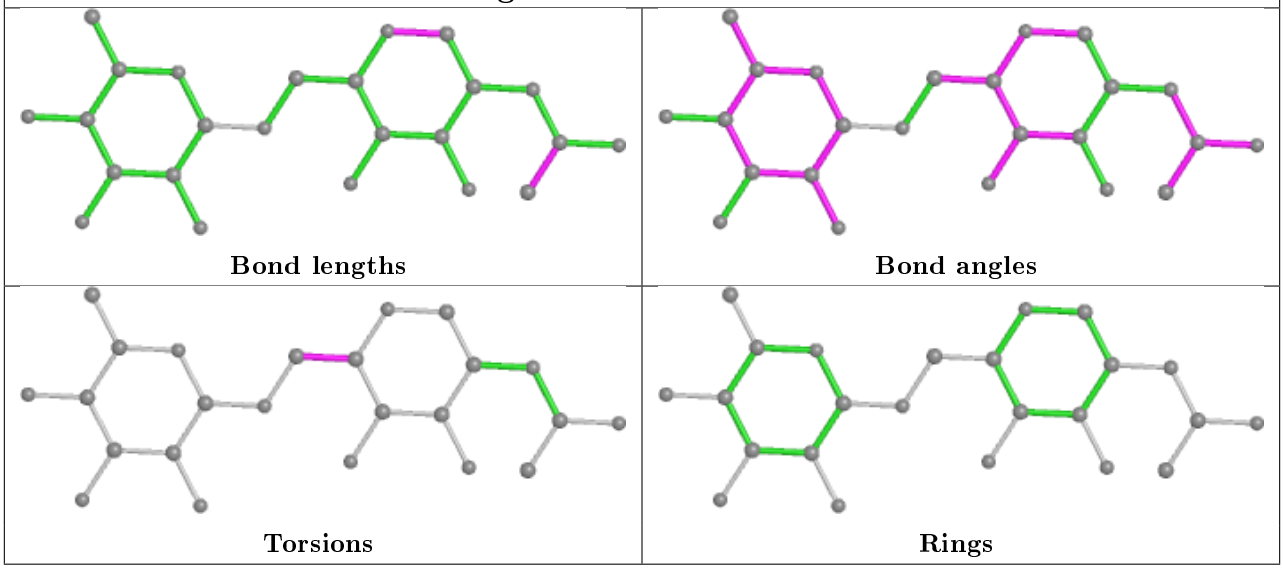
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	F	2	FUC	2	0
2	G	2	FUC	1	0
2	E	2	FUC	1	0
2	H	2	FUC	1	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.

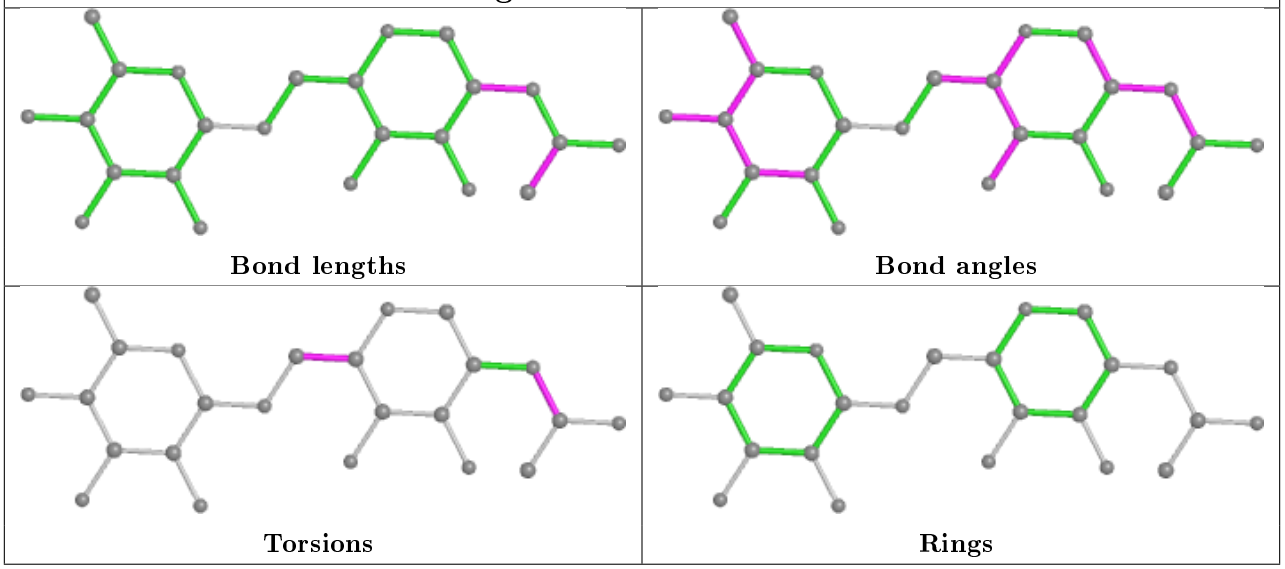
Oligosaccharide Chain E



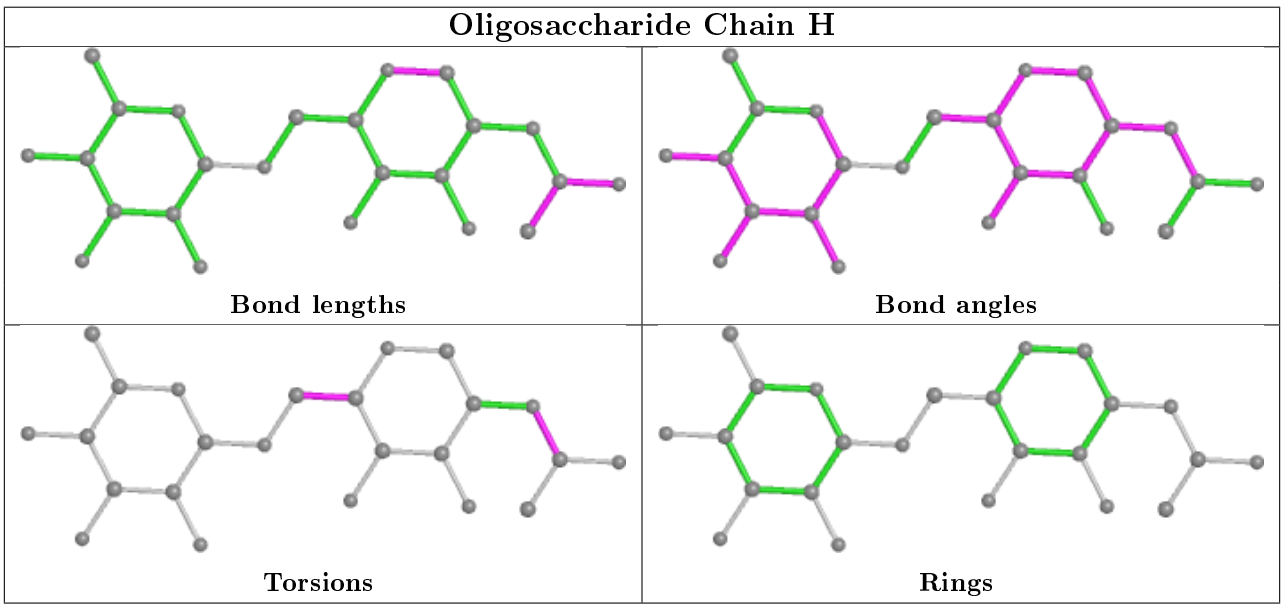
Oligosaccharide Chain F



Oligosaccharide Chain G



Oligosaccharide Chain H



5.6 Ligand geometry (i)

There are no ligands in this entry.

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.

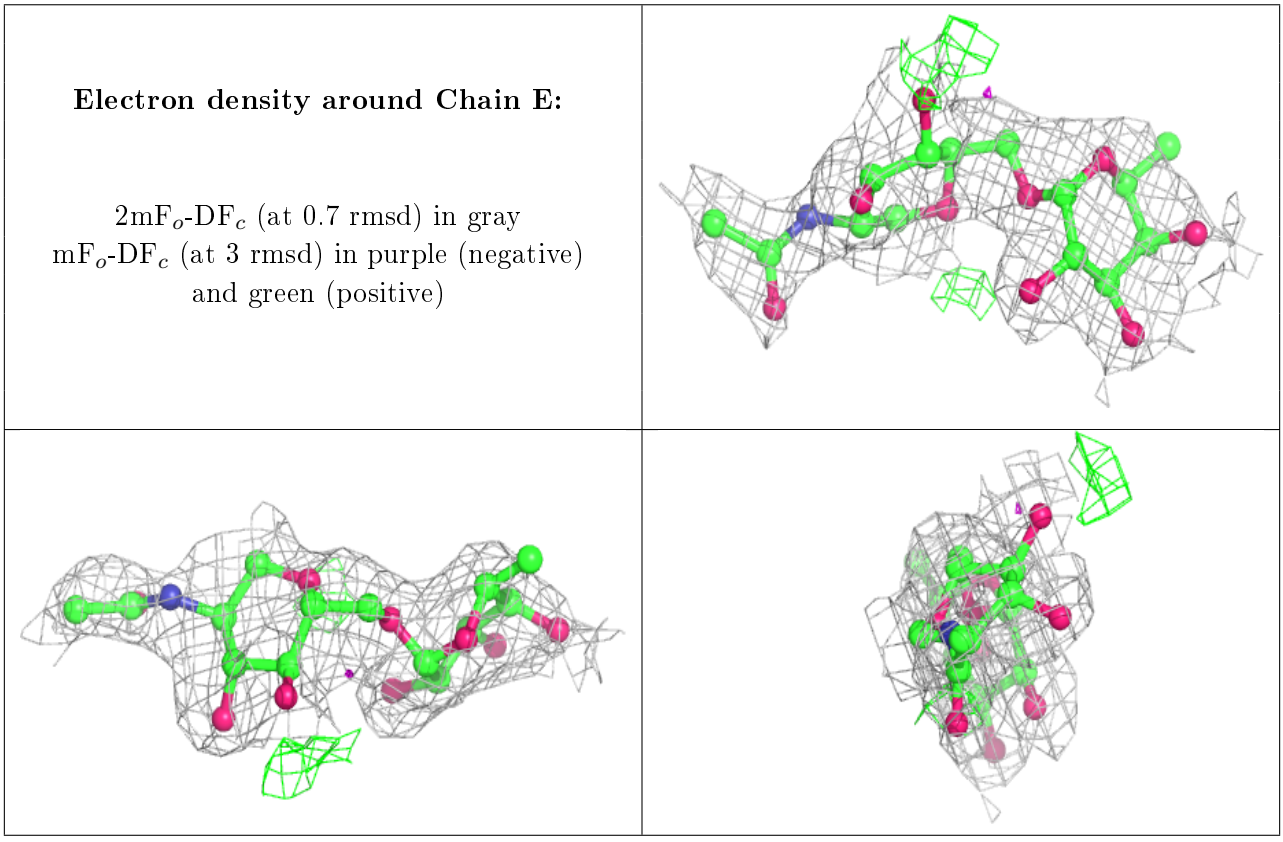
6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

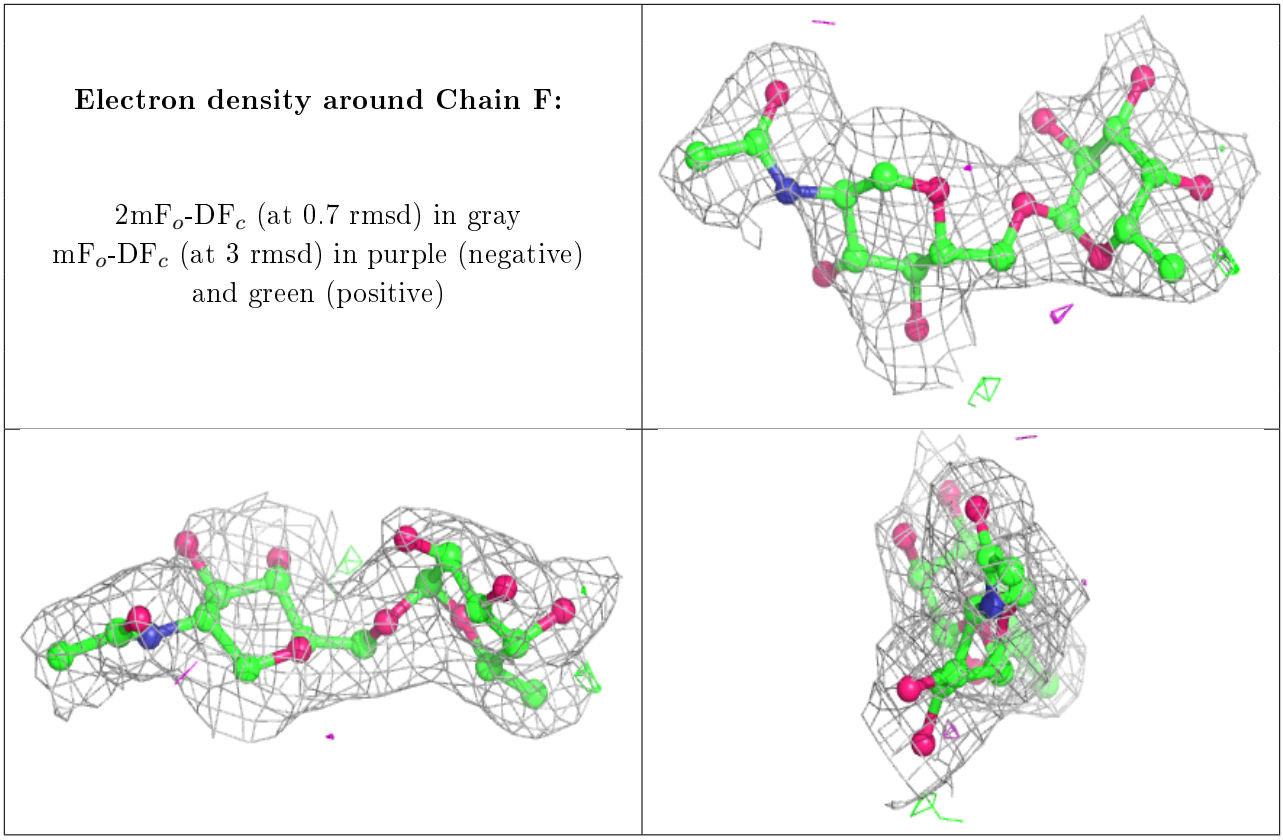
Unable to reproduce the depositors R factor - this section is therefore empty.

The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.



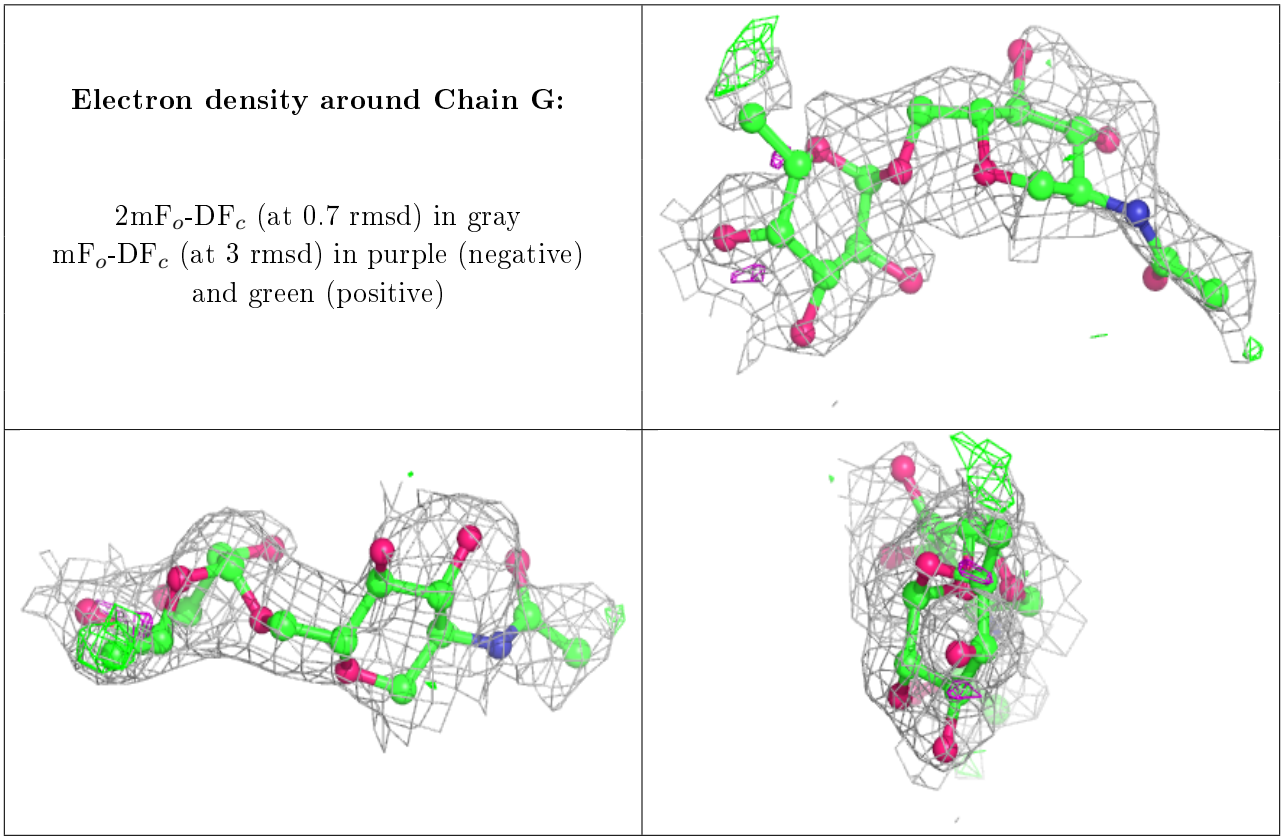
Electron density around Chain F:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)



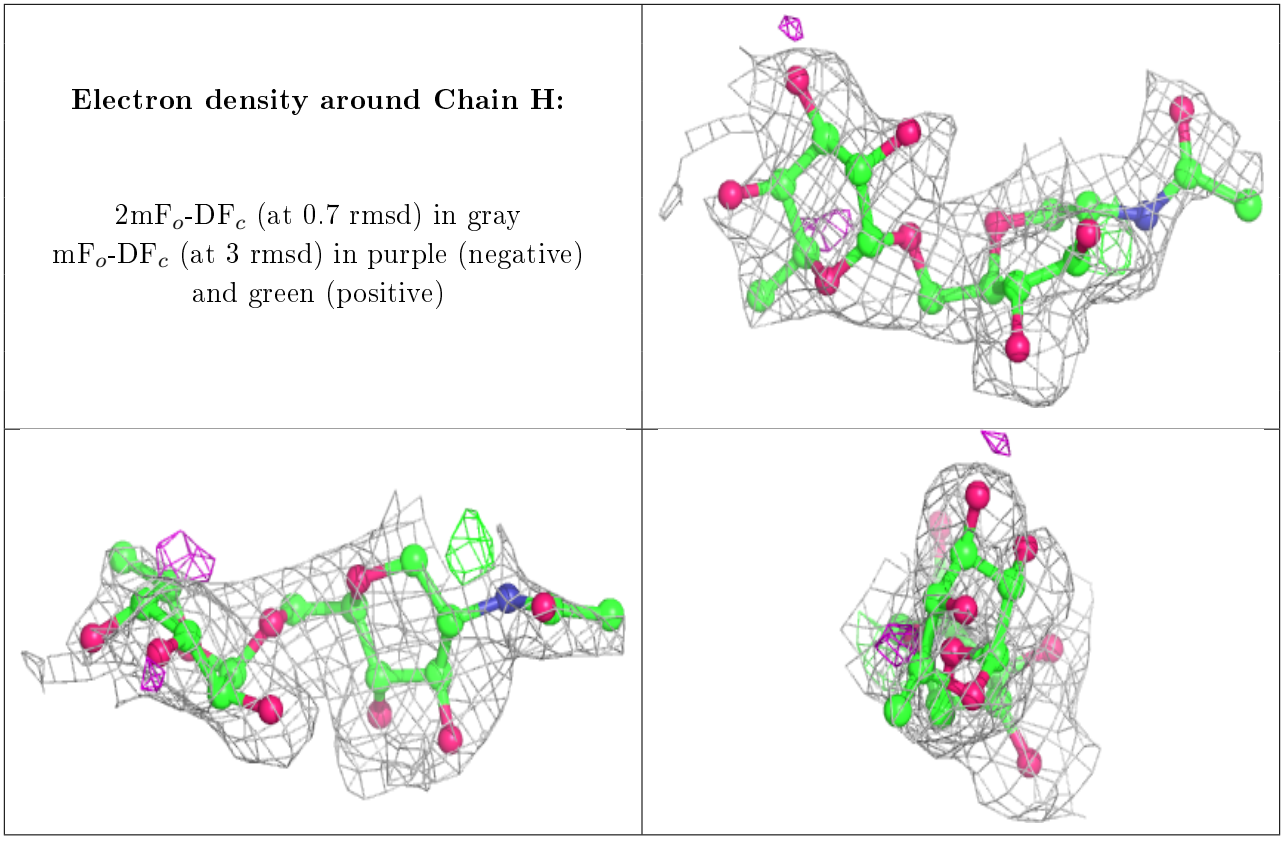
Electron density around Chain G:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)



Electron density around Chain H:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)



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.