



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 07:00 PM GMT

PDB ID : 4NDZ  
Title : Structure of Maltose Binding Protein fusion to 2-O-Sulfotransferase with bound heptasaccharide and PAP  
Authors : Liu, C.; Sheng, J.; Krahn, J.M.; Perera, L.; Xu, Y.; Hsieh, P.; Liu, J.; Pedersen, L.C.  
Deposited on : 2013-10-28  
Resolution : 3.45 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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
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) : trunk26865

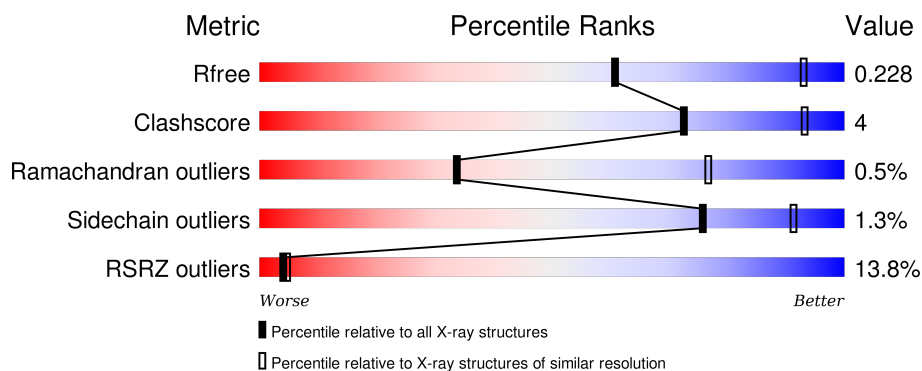
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.45 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	1000 (3.56-3.36)
Clashscore	102246	1090 (3.56-3.36)
Ramachandran outliers	100387	1057 (3.56-3.36)
Sidechain outliers	100360	1058 (3.56-3.36)
RSRZ outliers	91569	1005 (3.56-3.36)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	658	<div> <div>38%</div> <div>6%</div> <div>56%</div> </div>
1	B	658	<div> <div>28%</div> <div>86%</div> <div>14%</div> <div>•</div> </div>
1	C	658	<div> <div>10%</div> <div>87%</div> <div>12%</div> </div>
1	D	658	<div> <div>7%</div> <div>86%</div> <div>12%</div> <div>••</div> </div>
1	E	658	<div> <div>5%</div> <div>87%</div> <div>13%</div> <div>•</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	F	658	<p>24% 86% 13%</p>

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
3	GNS	F	2009	-	-	-	X
4	NPO	A	2011	-	-	-	X
4	NPO	E	2011	-	-	-	X

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 28764 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 Maltose-binding periplasmic protein, Heparan sulfate 2-O-sulfotransferase 1 fusion.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	287	Total	C	N	O	S	0	0	1
			2352	1522	399	422	9			
1	B	654	Total	C	N	O	S	0	0	1
			5136	3314	852	955	15			
1	C	655	Total	C	N	O	S	0	0	1
			5128	3311	848	955	14			
1	D	654	Total	C	N	O	S	0	0	1
			5136	3313	853	955	15			
1	E	654	Total	C	N	O	S	0	0	1
			5110	3298	842	955	15			
1	F	654	Total	C	N	O	S	0	1	0
			5124	3307	846	956	15			

There are 42 discrepancies between the modelled and reference sequences:

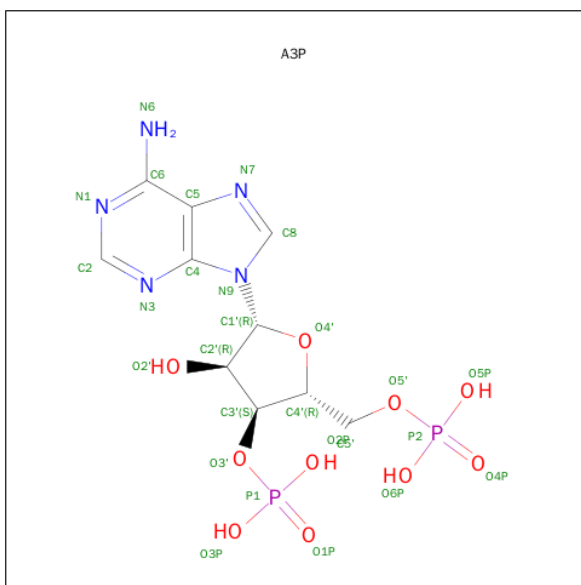
Chain	Residue	Modelled	Actual	Comment	Reference
A	359	ALA	GLU	ENGINEERED MUTATION	UNP P0AEX9
A	362	ALA	LYS	ENGINEERED MUTATION	UNP P0AEX9
A	363	ALA	ASP	ENGINEERED MUTATION	UNP P0AEX9
A	367	ASN	ARG	ENGINEERED MUTATION	UNP P0AEX9
A	368	ALA	-	LINKER	UNP P0AEX9
A	369	ALA	-	LINKER	UNP P0AEX9
A	370	ALA	-	LINKER	UNP P0AEX9
B	359	ALA	GLU	ENGINEERED MUTATION	UNP P0AEX9
B	362	ALA	LYS	ENGINEERED MUTATION	UNP P0AEX9
B	363	ALA	ASP	ENGINEERED MUTATION	UNP P0AEX9
B	367	ASN	ARG	ENGINEERED MUTATION	UNP P0AEX9
B	368	ALA	-	LINKER	UNP P0AEX9
B	369	ALA	-	LINKER	UNP P0AEX9
B	370	ALA	-	LINKER	UNP P0AEX9
C	359	ALA	GLU	ENGINEERED MUTATION	UNP P0AEX9
C	362	ALA	LYS	ENGINEERED MUTATION	UNP P0AEX9

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	363	ALA	ASP	ENGINEERED MUTATION	UNP P0AEX9
C	367	ASN	ARG	ENGINEERED MUTATION	UNP P0AEX9
C	368	ALA	-	LINKER	UNP P0AEX9
C	369	ALA	-	LINKER	UNP P0AEX9
C	370	ALA	-	LINKER	UNP P0AEX9
D	359	ALA	GLU	ENGINEERED MUTATION	UNP P0AEX9
D	362	ALA	LYS	ENGINEERED MUTATION	UNP P0AEX9
D	363	ALA	ASP	ENGINEERED MUTATION	UNP P0AEX9
D	367	ASN	ARG	ENGINEERED MUTATION	UNP P0AEX9
D	368	ALA	-	LINKER	UNP P0AEX9
D	369	ALA	-	LINKER	UNP P0AEX9
D	370	ALA	-	LINKER	UNP P0AEX9
E	359	ALA	GLU	ENGINEERED MUTATION	UNP P0AEX9
E	362	ALA	LYS	ENGINEERED MUTATION	UNP P0AEX9
E	363	ALA	ASP	ENGINEERED MUTATION	UNP P0AEX9
E	367	ASN	ARG	ENGINEERED MUTATION	UNP P0AEX9
E	368	ALA	-	LINKER	UNP P0AEX9
E	369	ALA	-	LINKER	UNP P0AEX9
E	370	ALA	-	LINKER	UNP P0AEX9
F	359	ALA	GLU	ENGINEERED MUTATION	UNP P0AEX9
F	362	ALA	LYS	ENGINEERED MUTATION	UNP P0AEX9
F	363	ALA	ASP	ENGINEERED MUTATION	UNP P0AEX9
F	367	ASN	ARG	ENGINEERED MUTATION	UNP P0AEX9
F	368	ALA	-	LINKER	UNP P0AEX9
F	369	ALA	-	LINKER	UNP P0AEX9
F	370	ALA	-	LINKER	UNP P0AEX9

- Molecule 2 is ADENOSINE-3'-5'-DIPHOSPHATE (three-letter code: A3P) (formula:  $C_{10}H_{15}N_5O_{10}P_2$ ).

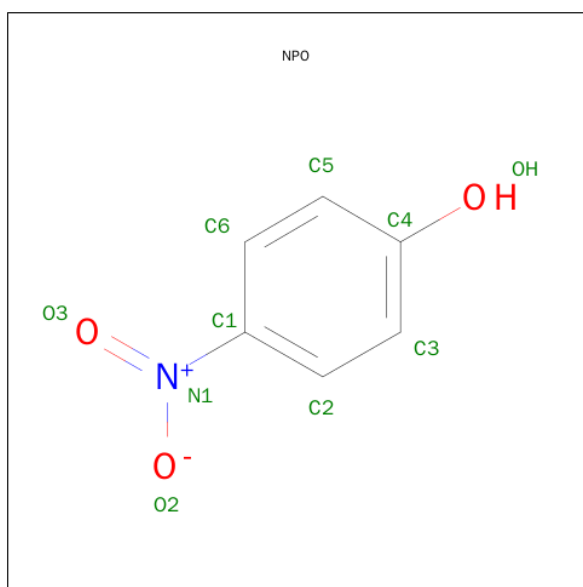


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	P	0	0
			27	10	5	10	2		
2	B	1	Total	C	N	O	P	0	0
			27	10	5	10	2		
2	C	1	Total	C	N	O	P	0	0
			27	10	5	10	2		
2	D	1	Total	C	N	O	P	0	0
			27	10	5	10	2		
2	E	1	Total	C	N	O	P	0	0
			27	10	5	10	2		
2	F	1	Total	C	N	O	P	0	0
			27	10	5	10	2		

- Molecule 3 is a polymer of unknown type called SUGAR (7-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	7	Total	C	N	O	S	0	0
			92	44	3	43	2		
3	B	7	Total	C	N	O	S	0	0
			92	44	3	43	2		
3	D	7	Total	C	N	O	S	0	0
			92	44	3	43	2		
3	E	7	Total	C	N	O	S	0	0
			92	44	3	43	2		
3	F	7	Total	C	N	O	S	0	0
			93	44	3	44	2		

- Molecule 4 is P-NITROPHENOL (three-letter code: NPO) (formula:  $C_6H_5NO_3$ ).



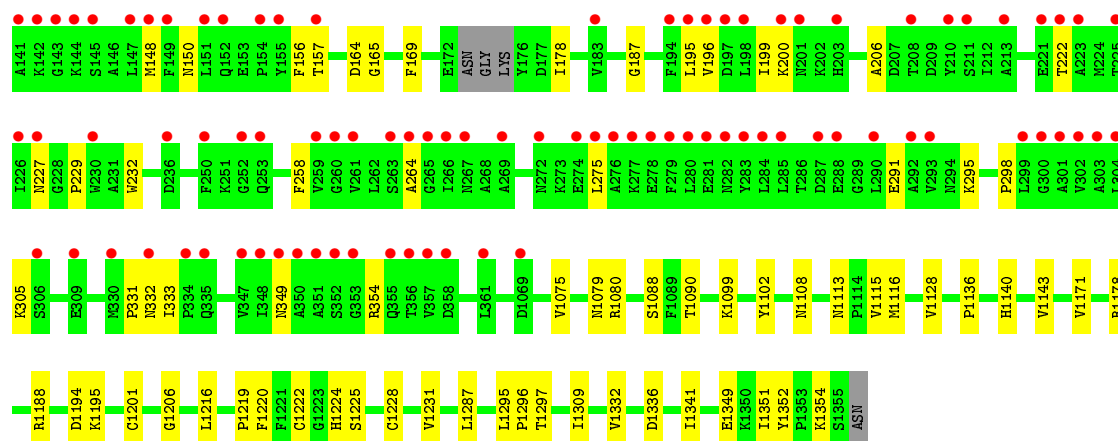
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	N	O	0	0
			10	6	1	3		
4	B	1	Total	C	N	O	0	0
			10	6	1	3		
4	D	1	Total	C	N	O	0	0
			10	6	1	3		
4	E	1	Total	C	N	O	0	0
			10	6	1	3		

- Molecule 5 is a polymer of unknown type called SUGAR (2-MER).

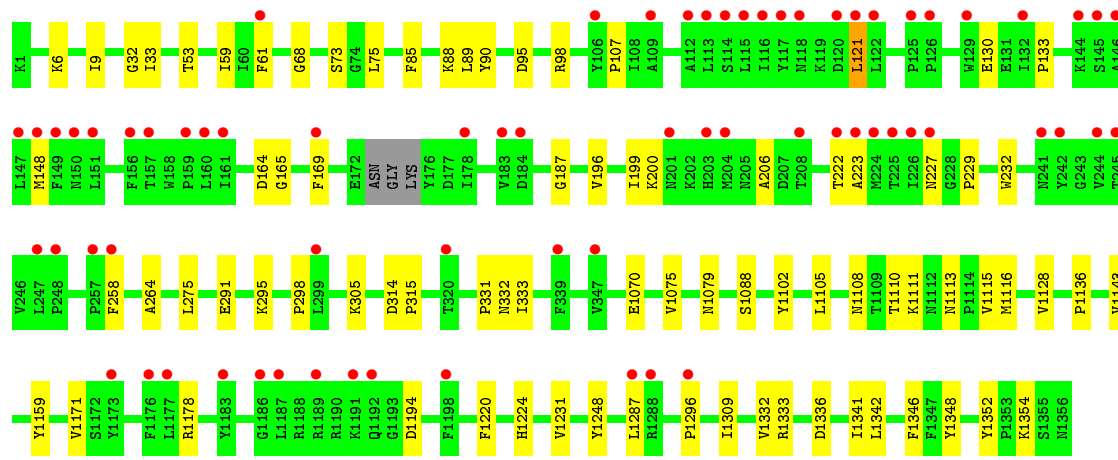
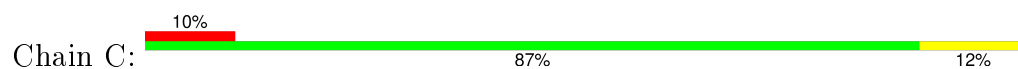
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	B	2	Total	C	O	0	0
			23	12	11		
5	C	2	Total	C	O	0	0
			23	12	11		
5	D	2	Total	C	O	0	0
			23	12	11		
5	E	2	Total	C	O	0	0
			23	12	11		
5	F	2	Total	C	O	0	0
			23	12	11		



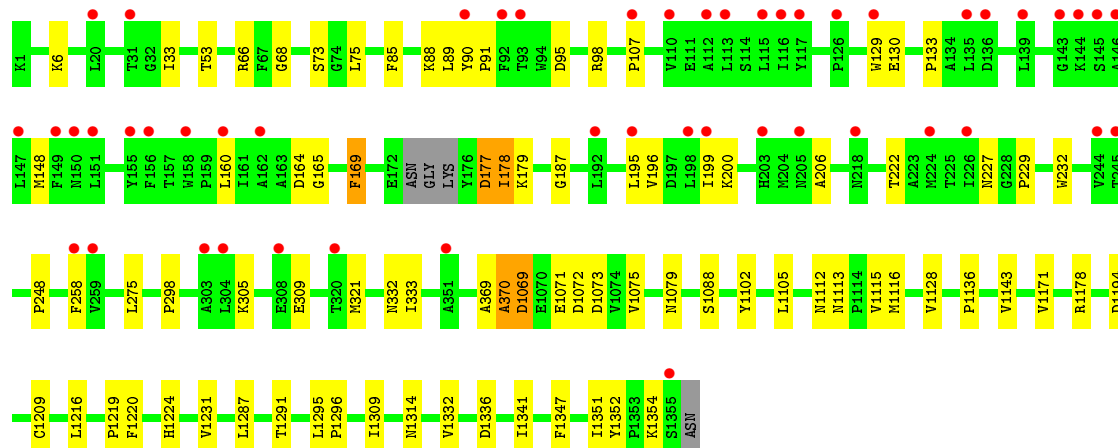
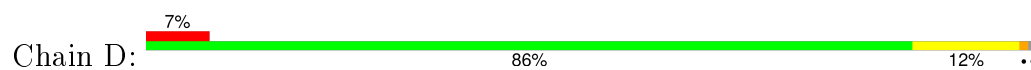




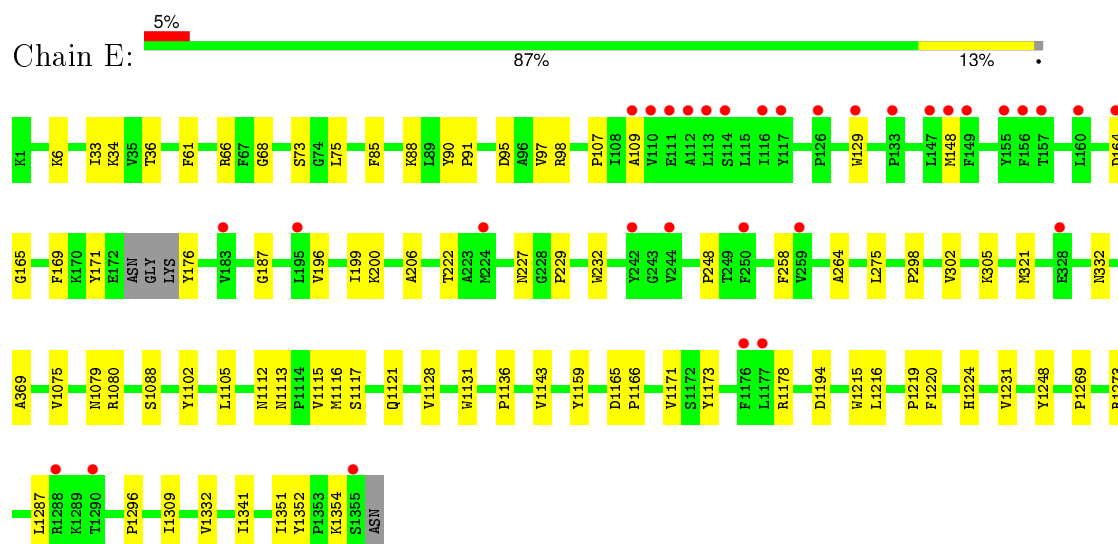
- Molecule 1: Maltose-binding periplasmic protein, Heparan sulfate 2-O-sulfotransferase 1 fusion



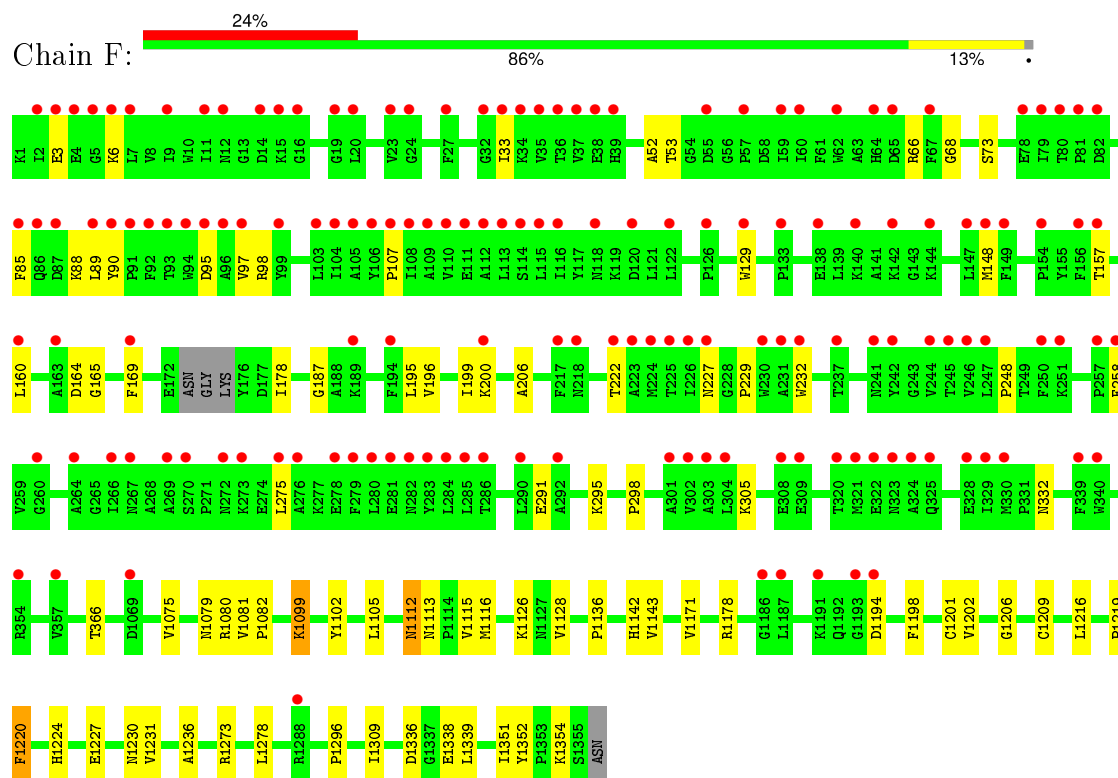
- Molecule 1: Maltose-binding periplasmic protein, Heparan sulfate 2-O-sulfotransferase 1 fusion



- Molecule 1: Maltose-binding periplasmic protein, Heparan sulfate 2-O-sulfotransferase 1 fusion



- Molecule 1: Maltose-binding periplasmic protein, Heparan sulfate 2-O-sulfotransferase 1 fusion



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	155.72Å 170.69Å 183.96Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.66 – 3.45 49.80 – 3.45	Depositor EDS
% Data completeness (in resolution range)	98.5 (49.66-3.45) 91.1 (49.80-3.45)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.11	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.65 (at 3.48Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.4_1496)	Depositor
R, $R_{free}$	0.176 , 0.228 0.186 , 0.228	Depositor DCC
$R_{free}$ test set	1962 reflections (3.41%)	DCC
Wilson B-factor (Å <sup>2</sup> )	90.6	Xtriage
Anisotropy	0.167	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 85.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	1 of 64251 reflections (0.002%)	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	28764	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	138.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.93% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: GNS, NPO, A3P, GLC, IDR, NDG, BDP

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	0.38	0/2416	0.47	0/3270
1	B	0.30	0/5267	0.42	0/7153
1	C	0.29	0/5259	0.43	0/7144
1	D	0.30	0/5267	0.43	0/7152
1	E	0.30	0/5241	0.44	0/7123
1	F	0.29	0/5256	0.42	0/7142
All	All	0.30	0/28706	0.43	0/38984

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 [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	2352	0	2265	21	0
1	B	5136	0	4988	49	0
1	C	5128	0	4973	41	0
1	D	5136	0	4988	45	0
1	E	5110	0	4933	47	0
1	F	5124	0	4950	48	0
2	A	27	0	11	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	27	0	11	0	0
2	C	27	0	11	0	0
2	D	27	0	11	0	0
2	E	27	0	11	0	0
2	F	27	0	11	0	0
3	A	92	0	59	0	0
3	B	92	0	59	2	0
3	D	92	0	59	0	0
3	E	92	0	59	2	0
3	F	93	0	61	2	0
4	A	10	0	4	0	0
4	B	10	0	4	0	0
4	D	10	0	4	0	0
4	E	10	0	4	0	0
5	B	23	0	21	0	0
5	C	23	0	21	0	0
5	D	23	0	21	1	0
5	E	23	0	21	1	0
5	F	23	0	21	1	0
All	All	28764	0	27581	239	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 4.

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:D:369:ALA:O	1:D:1069:ASP:N	2.24	0.71
1:E:1131:TRP:NE1	1:F:366:THR:HG21	2.08	0.68
1:E:1131:TRP:CD1	1:F:366:THR:HG21	2.30	0.66
1:D:1071:GLU:O	1:D:1073:ASP:N	2.30	0.65
1:A:1120:ASP:OD1	1:A:1123:ARG:NH1	2.31	0.63

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	285/658 (43%)	272 (95%)	13 (5%)	0	100	100
1	B	650/658 (99%)	616 (95%)	31 (5%)	3 (0%)	34	77
1	C	651/658 (99%)	619 (95%)	28 (4%)	4 (1%)	30	74
1	D	650/658 (99%)	614 (94%)	30 (5%)	6 (1%)	21	67
1	E	650/658 (99%)	619 (95%)	28 (4%)	3 (0%)	34	77
1	F	651/658 (99%)	619 (95%)	29 (4%)	3 (0%)	34	77
All	All	3537/3948 (90%)	3359 (95%)	159 (4%)	19 (0%)	34	77

5 of 19 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	370	ALA
1	D	1072	ASP
1	E	369	ALA
1	B	1336	ASP
1	B	169	PHE

### 5.3.2 Protein sidechains

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	245/548 (45%)	240 (98%)	5 (2%)	63	86
1	B	524/548 (96%)	518 (99%)	6 (1%)	80	92
1	C	522/548 (95%)	518 (99%)	4 (1%)	86	95

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	D	524/548 (96%)	515 (98%)	9 (2%)	68	89
1	E	519/548 (95%)	514 (99%)	5 (1%)	82	93
1	F	521/548 (95%)	512 (98%)	9 (2%)	68	89
All	All	2855/3288 (87%)	2817 (99%)	38 (1%)	76	91

5 of 38 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	227	ASN
1	D	1209	CYS
1	F	1126	LYS
1	D	321	MET
1	D	1220	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	1127	ASN
1	F	1127	ASN
1	E	1127	ASN
1	C	1106	HIS
1	D	1127	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

## 5.5 Carbohydrates ⓘ

45 carbohydrates 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	$\# Z  > 2$
3	BDP	A	2004	3	9,12,13	2.81	5 (55%)	13,17,19	1.25	2 (15%)
3	NDG	A	2005	3	14,14,15	0.59	1 (7%)	15,19,21	0.39	0
3	BDP	A	2006	3	9,12,13	2.67	3 (33%)	13,17,19	1.57	1 (7%)
3	GNS	A	2007	3	13,15,16	2.45	8 (61%)	14,22,24	1.86	3 (21%)
3	IDR	A	2008	3	9,12,13	2.65	4 (44%)	13,17,19	0.95	0
3	GNS	A	2009	3	13,15,16	2.57	7 (53%)	14,22,24	1.40	1 (7%)
3	BDP	A	2010	3,4	9,12,13	3.00	4 (44%)	13,17,19	1.65	2 (15%)
5	GLC	B	2001	5	11,11,12	0.60	0	14,15,17	0.70	0
5	GLC	B	2002	5	12,12,12	0.53	0	17,17,17	0.66	0
3	BDP	B	2004	3	9,12,13	2.88	4 (44%)	13,17,19	1.18	1 (7%)
3	NDG	B	2005	3	14,14,15	0.57	0	15,19,21	0.61	0
3	BDP	B	2006	3	9,12,13	2.67	4 (44%)	13,17,19	1.33	1 (7%)
3	GNS	B	2007	3	13,15,16	2.61	7 (53%)	14,22,24	1.85	2 (14%)
3	IDR	B	2008	3	9,12,13	2.62	5 (55%)	13,17,19	1.01	1 (7%)
3	GNS	B	2009	3	13,15,16	2.57	6 (46%)	14,22,24	1.36	1 (7%)
3	BDP	B	2010	3,4	9,12,13	3.00	5 (55%)	13,17,19	1.33	2 (15%)
5	GLC	C	2001	5	11,11,12	0.58	0	14,15,17	0.74	0
5	GLC	C	2002	5	12,12,12	0.53	0	17,17,17	0.64	0
5	GLC	D	2001	5	11,11,12	0.63	0	14,15,17	0.76	0
5	GLC	D	2002	5	12,12,12	0.53	0	17,17,17	0.65	0
3	BDP	D	2004	3	9,12,13	2.73	5 (55%)	13,17,19	1.26	2 (15%)
3	NDG	D	2005	3	14,14,15	0.48	0	15,19,21	0.61	0
3	BDP	D	2006	3	9,12,13	2.57	3 (33%)	13,17,19	1.39	1 (7%)
3	GNS	D	2007	3	13,15,16	2.56	6 (46%)	14,22,24	1.94	3 (21%)
3	IDR	D	2008	3	9,12,13	2.60	4 (44%)	13,17,19	0.64	0
3	GNS	D	2009	3	13,15,16	2.68	7 (53%)	14,22,24	1.42	3 (21%)
3	BDP	D	2010	3,4	9,12,13	2.67	5 (55%)	13,17,19	1.57	1 (7%)
5	GLC	E	2001	5	11,11,12	0.63	0	14,15,17	0.84	0
5	GLC	E	2002	5	12,12,12	0.53	0	17,17,17	0.73	1 (5%)
3	BDP	E	2004	3	9,12,13	2.83	5 (55%)	13,17,19	1.25	1 (7%)
3	NDG	E	2005	3	14,14,15	0.58	1 (7%)	15,19,21	0.39	0
3	BDP	E	2006	3	9,12,13	2.86	3 (33%)	13,17,19	0.95	1 (7%)
3	GNS	E	2007	3	13,15,16	2.40	6 (46%)	14,22,24	1.72	2 (14%)
3	IDR	E	2008	3	9,12,13	2.44	4 (44%)	13,17,19	0.91	0



Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	GNS	E	2009	3	13,15,16	2.60	7 (53%)	14,22,24	1.63	1 (7%)
3	BDP	E	2010	3,4	9,12,13	2.92	4 (44%)	13,17,19	1.35	2 (15%)
5	GLC	F	2001	5	11,11,12	0.62	0	14,15,17	0.63	0
5	GLC	F	2002	5	12,12,12	0.51	0	17,17,17	0.71	0
3	BDP	F	2004	3	9,12,13	2.70	5 (55%)	13,17,19	1.29	2 (15%)
3	NDG	F	2005	3	14,14,15	0.56	0	15,19,21	0.40	0
3	BDP	F	2006	3	9,12,13	2.87	3 (33%)	13,17,19	1.22	1 (7%)
3	GNS	F	2007	3	13,15,16	2.42	7 (53%)	14,22,24	1.64	1 (7%)
3	IDR	F	2008	3	9,12,13	2.47	4 (44%)	13,17,19	0.85	0
3	GNS	F	2009	3	13,15,16	2.61	8 (61%)	14,22,24	1.32	1 (7%)
3	BDP	F	2010	3	10,13,13	1.98	3 (30%)	15,19,19	1.38	1 (6%)

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	BDP	A	2004	3	-	0/0/21/24	0/1/1/1
3	NDG	A	2005	3	-	0/6/23/26	0/1/1/1
3	BDP	A	2006	3	-	0/0/21/24	0/1/1/1
3	GNS	A	2007	3	-	0/7/24/27	0/1/1/1
3	IDR	A	2008	3	-	0/0/21/24	0/1/1/1
3	GNS	A	2009	3	-	0/7/24/27	0/1/1/1
3	BDP	A	2010	3,4	-	0/0/21/24	0/1/1/1
5	GLC	B	2001	5	-	0/2/19/22	0/1/1/1
5	GLC	B	2002	5	-	0/2/22/22	0/1/1/1
3	BDP	B	2004	3	-	0/0/21/24	0/1/1/1
3	NDG	B	2005	3	-	0/6/23/26	0/1/1/1
3	BDP	B	2006	3	-	0/0/21/24	0/1/1/1
3	GNS	B	2007	3	-	0/7/24/27	0/1/1/1
3	IDR	B	2008	3	-	0/0/21/24	0/1/1/1
3	GNS	B	2009	3	-	0/7/24/27	0/1/1/1
3	BDP	B	2010	3,4	-	0/0/21/24	0/1/1/1
5	GLC	C	2001	5	-	0/2/19/22	0/1/1/1
5	GLC	C	2002	5	-	0/2/22/22	0/1/1/1
5	GLC	D	2001	5	-	0/2/19/22	0/1/1/1
5	GLC	D	2002	5	-	0/2/22/22	0/1/1/1
3	BDP	D	2004	3	-	0/0/21/24	0/1/1/1
3	NDG	D	2005	3	-	0/6/23/26	0/1/1/1
3	BDP	D	2006	3	-	0/0/21/24	0/1/1/1

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	GNS	D	2007	3	-	0/7/24/27	0/1/1/1
3	IDR	D	2008	3	-	0/0/21/24	0/1/1/1
3	GNS	D	2009	3	-	0/7/24/27	0/1/1/1
3	BDP	D	2010	3,4	-	0/0/21/24	0/1/1/1
5	GLC	E	2001	5	-	0/2/19/22	0/1/1/1
5	GLC	E	2002	5	-	0/2/22/22	0/1/1/1
3	BDP	E	2004	3	-	0/0/21/24	0/1/1/1
3	NDG	E	2005	3	-	0/6/23/26	0/1/1/1
3	BDP	E	2006	3	-	0/0/21/24	0/1/1/1
3	GNS	E	2007	3	-	0/7/24/27	0/1/1/1
3	IDR	E	2008	3	-	0/0/21/24	0/1/1/1
3	GNS	E	2009	3	-	0/7/24/27	0/1/1/1
3	BDP	E	2010	3,4	-	0/0/21/24	0/1/1/1
5	GLC	F	2001	5	-	0/2/19/22	0/1/1/1
5	GLC	F	2002	5	-	0/2/22/22	0/1/1/1
3	BDP	F	2004	3	-	0/0/21/24	0/1/1/1
3	NDG	F	2005	3	-	0/6/23/26	0/1/1/1
3	BDP	F	2006	3	-	0/0/21/24	0/1/1/1
3	GNS	F	2007	3	-	0/7/24/27	0/1/1/1
3	IDR	F	2008	3	-	0/0/21/24	0/1/1/1
3	GNS	F	2009	3	-	0/7/24/27	0/1/1/1
3	BDP	F	2010	3	-	0/0/24/24	0/1/1/1

The worst 5 of 153 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	2010	BDP	O5-C5	-6.77	1.36	1.43
3	B	2010	BDP	O5-C5	-6.45	1.37	1.43
3	E	2010	BDP	O5-C5	-6.44	1.37	1.43
3	F	2006	BDP	O5-C5	-6.40	1.37	1.43
3	B	2004	BDP	O5-C5	-6.35	1.37	1.43

The worst 5 of 41 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	2010	BDP	O5-C1-C2	-2.74	106.41	110.86
3	D	2004	BDP	O5-C1-C2	-2.53	106.76	110.86
3	E	2010	BDP	O5-C1-C2	-2.38	107.00	110.86
3	B	2010	BDP	O5-C1-C2	-2.21	107.28	110.86
3	A	2007	GNS	C3-C4-C5	-2.20	106.36	110.20

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

9 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	2005	NDG	1	0
3	B	2007	GNS	1	0
5	D	2001	GLC	1	0
5	E	2001	GLC	1	0
3	E	2007	GNS	1	0
3	E	2010	BDP	1	0
5	F	2001	GLC	1	0
3	F	2007	GNS	1	0
3	F	2010	BDP	1	0

## 5.6 Ligand geometry

10 ligands 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	A3P	A	2003	-	24,29,29	2.69	11 (45%)	28,45,45	1.96	2 (7%)
4	NPO	A	2011	3	8,10,10	0.44	0	11,13,13	0.86	1 (9%)
2	A3P	B	2003	-	24,29,29	2.65	11 (45%)	28,45,45	2.04	1 (3%)
4	NPO	B	2011	3	8,10,10	0.34	0	11,13,13	0.71	0
2	A3P	C	2003	-	24,29,29	2.71	11 (45%)	28,45,45	2.01	2 (7%)
2	A3P	D	2003	-	24,29,29	2.68	11 (45%)	28,45,45	2.01	2 (7%)
4	NPO	D	2011	3	8,10,10	0.30	0	11,13,13	0.53	0
2	A3P	E	2003	-	24,29,29	2.72	11 (45%)	28,45,45	2.10	2 (7%)
4	NPO	E	2011	3	8,10,10	0.33	0	11,13,13	0.49	0
2	A3P	F	2003	-	24,29,29	2.75	11 (45%)	28,45,45	2.00	2 (7%)

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	A3P	A	2003	-	-	0/11/31/31	0/3/3/3
4	NPO	A	2011	3	-	0/4/4/4	0/1/1/1
2	A3P	B	2003	-	-	0/11/31/31	0/3/3/3
4	NPO	B	2011	3	-	0/4/4/4	0/1/1/1
2	A3P	C	2003	-	-	0/11/31/31	0/3/3/3
2	A3P	D	2003	-	-	0/11/31/31	0/3/3/3
4	NPO	D	2011	3	-	0/4/4/4	0/1/1/1
2	A3P	E	2003	-	-	0/11/31/31	0/3/3/3
4	NPO	E	2011	3	-	0/4/4/4	0/1/1/1
2	A3P	F	2003	-	-	0/11/31/31	0/3/3/3

The worst 5 of 66 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	2003	A3P	O4'-C4'	-3.23	1.37	1.45
2	A	2003	A3P	O4'-C4'	-3.19	1.37	1.45
2	C	2003	A3P	O4'-C4'	-3.03	1.38	1.45
2	F	2003	A3P	O4'-C4'	-3.03	1.38	1.45
2	D	2003	A3P	O4'-C4'	-3.01	1.38	1.45

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	2003	A3P	N3-C2-N1	-10.11	121.16	128.89
2	B	2003	A3P	N3-C2-N1	-9.86	121.34	128.89
2	D	2003	A3P	N3-C2-N1	-9.76	121.42	128.89
2	C	2003	A3P	N3-C2-N1	-9.64	121.52	128.89
2	F	2003	A3P	N3-C2-N1	-9.45	121.66	128.89

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 [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	287/658 (43%)	-0.07	2 (0%) 89 84	48, 83, 133, 185	0
1	B	654/658 (99%)	1.47	182 (27%) 1 1	54, 196, 294, 399	0
1	C	655/658 (99%)	0.50	69 (10%) 8 9	63, 129, 207, 293	0
1	D	654/658 (99%)	0.38	49 (7%) 17 16	54, 128, 195, 252	0
1	E	654/658 (99%)	0.28	32 (4%) 33 28	58, 122, 191, 274	0
1	F	654/658 (99%)	1.11	156 (23%) 1 1	57, 170, 263, 333	0
All	All	3558/3948 (90%)	0.68	490 (13%) 4 5	48, 127, 250, 399	0

The worst 5 of 490 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	266	ILE	25.3
1	B	7	LEU	17.8
1	B	51	ALA	16.3
1	B	35	VAL	11.8
1	F	111	GLU	10.0

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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

### 6.3 Carbohydrates [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<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
3	GNS	F	2009	15/16	0.91	0.52	1.73	177,183,197,203	0
3	BDP	D	2004	12/13	0.81	0.25	0.89	172,199,213,219	0
3	GNS	F	2007	15/16	0.90	0.28	0.24	112,150,186,198	0
5	GLC	E	2002	12/12	0.96	0.33	-0.22	88,98,112,113	0
3	IDR	A	2008	12/13	0.97	0.24	-0.24	68,88,103,113	0
3	IDR	E	2008	12/13	0.95	0.31	-0.31	128,148,209,213	0
5	GLC	C	2002	12/12	0.96	0.30	-0.33	95,124,137,149	0
3	GNS	E	2007	15/16	0.96	0.24	-0.50	123,144,168,187	0
5	GLC	D	2002	12/12	0.90	0.28	-0.52	135,142,153,154	0
5	GLC	B	2001	11/12	0.72	0.35	-0.53	218,235,256,257	0
3	GNS	E	2009	15/16	0.94	0.27	-0.56	150,164,191,196	0
5	GLC	D	2001	11/12	0.93	0.24	-0.65	92,109,127,127	0
3	IDR	F	2008	12/13	0.87	0.23	-0.80	142,162,187,200	0
3	IDR	B	2008	12/13	0.94	0.20	-0.82	100,112,121,124	0
3	GNS	A	2007	15/16	0.98	0.20	-1.03	72,78,98,101	0
3	BDP	E	2010	12/13	0.85	0.18	-1.11	146,191,216,218	0
5	GLC	F	2001	11/12	0.89	0.31	-1.12	149,166,181,185	0
3	GNS	A	2009	15/16	0.97	0.17	-1.17	63,75,99,100	0
3	GNS	D	2009	15/16	0.94	0.18	-1.22	100,120,139,152	0
3	GNS	B	2007	15/16	0.98	0.18	-1.24	79,93,110,133	0
5	GLC	F	2002	12/12	0.91	0.24	-1.32	165,176,181,186	0
5	GLC	B	2002	12/12	0.78	0.27	-1.34	197,218,230,239	0
3	IDR	D	2008	12/13	0.97	0.19	-1.42	103,126,151,165	0
5	GLC	E	2001	11/12	0.98	0.20	-1.57	79,95,124,133	0
3	GNS	D	2007	15/16	0.96	0.19	-1.77	88,117,142,145	0
3	GNS	B	2009	15/16	0.96	0.18	-1.83	84,89,108,118	0
5	GLC	C	2001	11/12	0.98	0.13	-2.03	93,97,110,112	0
3	NDG	E	2005	14/15	0.87	0.29	-	139,188,202,212	0
3	BDP	A	2006	12/13	0.88	0.15	-	95,130,153,167	0
3	BDP	F	2010	13/13	0.69	0.39	-	120,143,173,179	0
3	BDP	D	2006	12/13	0.83	0.20	-	126,144,163,173	0
3	BDP	A	2010	12/13	0.96	0.15	-	68,93,105,108	0
3	BDP	F	2004	12/13	0.72	0.40	-	195,218,230,230	0
3	BDP	F	2006	12/13	0.68	0.24	-	136,206,212,214	0
3	BDP	B	2010	12/13	0.97	0.18	-	77,94,107,109	0
3	BDP	D	2010	12/13	0.93	0.14	-	103,127,150,155	0
3	BDP	B	2004	12/13	0.88	0.16	-	148,161,172,176	0
3	BDP	E	2004	12/13	0.77	0.45	-	186,205,213,218	0
3	NDG	B	2005	14/15	0.93	0.13	-	109,128,158,160	0
3	NDG	A	2005	14/15	0.92	0.17	-	107,132,150,170	0
3	NDG	F	2005	14/15	0.83	0.22	-	145,173,202,210	0
3	BDP	A	2004	12/13	0.78	0.38	-	179,197,225,229	0
3	BDP	B	2006	12/13	0.91	0.14	-	108,137,152,153	0

Continued on next page...

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	BDP	E	2006	12/13	0.75	0.44	-	145,194,213,217	0
3	NDG	D	2005	14/15	0.90	0.18	-	104,145,162,178	0

## 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<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
4	NPO	E	2011	10/10	0.73	0.37	7.26	157,172,185,189	0
4	NPO	A	2011	10/10	0.92	0.32	2.12	103,138,146,154	0
4	NPO	B	2011	10/10	0.96	0.30	1.71	105,120,169,173	0
4	NPO	D	2011	10/10	0.89	0.28	0.53	159,170,199,200	0
2	A3P	E	2003	27/27	0.94	0.27	-0.51	80,104,114,119	0
2	A3P	D	2003	27/27	0.97	0.20	-0.89	57,83,108,112	0
2	A3P	B	2003	27/27	0.97	0.18	-0.94	66,80,103,148	0
2	A3P	C	2003	27/27	0.93	0.21	-0.95	74,106,135,145	0
2	A3P	F	2003	27/27	0.96	0.20	-0.99	67,89,104,115	0
2	A3P	A	2003	27/27	0.97	0.19	-1.02	53,68,87,93	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.