



# Full wwPDB X-ray Structure Validation Report ⓘ

Oct 17, 2017 – 04:56 AM EDT

PDB ID : 3C6F  
Title : Crystal structure of protein Bsu07140 from *Bacillus subtilis*  
Authors : Patskovsky, Y.; Min, T.; Zhang, A.; Adams, J.; Groshong, C.; Wasserman, S.R.; Sauder, J.M.; Burley, S.K.; Almo, S.C.; New York SGX Research Center for Structural Genomics (NYSGXRC)  
Deposited on : unknown  
Resolution : 2.50 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](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	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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-20030345

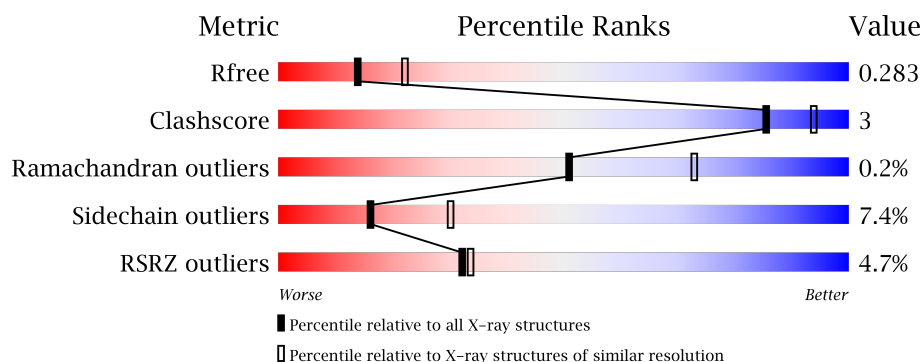
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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))
$R_{free}$	100719	3846 (2.50-2.50)
Clashscore	112137	4554 (2.50-2.50)
Ramachandran outliers	110173	4463 (2.50-2.50)
Sidechain outliers	110143	4465 (2.50-2.50)
RSRZ outliers	101464	3876 (2.50-2.50)

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	153	<div> <div>3%</div> <div> <div></div> <div>78%</div> <div>10%</div> <div>10%</div> </div> </div>
1	B	153	<div> <div>8%</div> <div> <div></div> <div>75%</div> <div>14%</div> <div>10%</div> </div> </div>
1	C	153	<div> <div>3%</div> <div> <div></div> <div>78%</div> <div>11%</div> <div>10%</div> </div> </div>
1	D	153	<div> <div>3%</div> <div> <div></div> <div>78%</div> <div>11%</div> <div>10%</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4566 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 YetF protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	137	Total	C	N	O	S	0	1	0
			1115	706	188	215	6			
1	B	137	Total	C	N	O	S	0	3	0
			1131	717	193	214	7			
1	C	137	Total	C	N	O	S	0	1	0
			1118	708	190	214	6			
1	D	138	Total	C	N	O	S	0	2	0
			1131	717	192	216	6			

There are 24 discrepancies between the modelled and reference sequences:

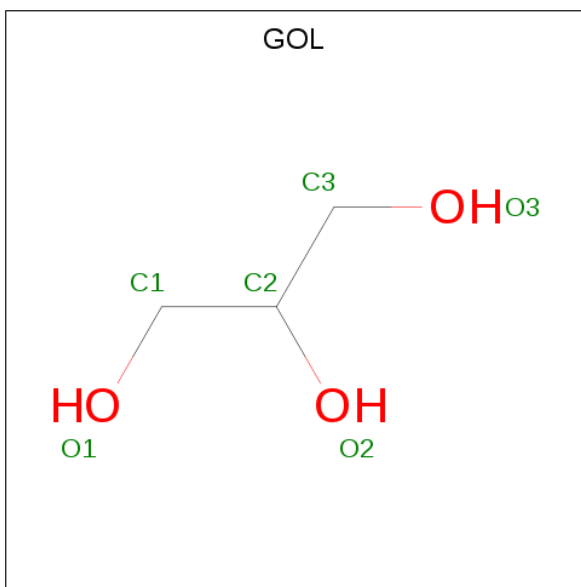
Chain	Residue	Modelled	Actual	Comment	Reference
A	79	SER	-	EXPRESSION TAG	UNP O31533
A	80	LEU	-	EXPRESSION TAG	UNP O31533
A	121	MET	LEU	ENGINEERED	UNP O31533
A	138	MET	LEU	ENGINEERED	UNP O31533
A	160	MET	LEU	ENGINEERED	UNP O31533
A	198	MET	LEU	ENGINEERED	UNP O31533
B	79	SER	-	EXPRESSION TAG	UNP O31533
B	80	LEU	-	EXPRESSION TAG	UNP O31533
B	121	MET	LEU	ENGINEERED	UNP O31533
B	138	MET	LEU	ENGINEERED	UNP O31533
B	160	MET	LEU	ENGINEERED	UNP O31533
B	198	MET	LEU	ENGINEERED	UNP O31533
C	79	SER	-	EXPRESSION TAG	UNP O31533
C	80	LEU	-	EXPRESSION TAG	UNP O31533
C	121	MET	LEU	ENGINEERED	UNP O31533
C	138	MET	LEU	ENGINEERED	UNP O31533
C	160	MET	LEU	ENGINEERED	UNP O31533
C	198	MET	LEU	ENGINEERED	UNP O31533
D	79	SER	-	EXPRESSION TAG	UNP O31533
D	80	LEU	-	EXPRESSION TAG	UNP O31533
D	121	MET	LEU	ENGINEERED	UNP O31533

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	138	MET	LEU	ENGINEERED	UNP O31533
D	160	MET	LEU	ENGINEERED	UNP O31533
D	198	MET	LEU	ENGINEERED	UNP O31533

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	D	1	Total C O 6 3 3	0	0

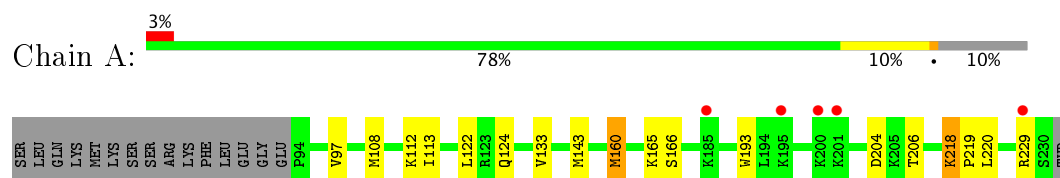
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	16	Total O 16 16	0	0
3	B	17	Total O 17 17	0	0
3	C	14	Total O 14 14	0	0
3	D	18	Total O 18 18	0	0

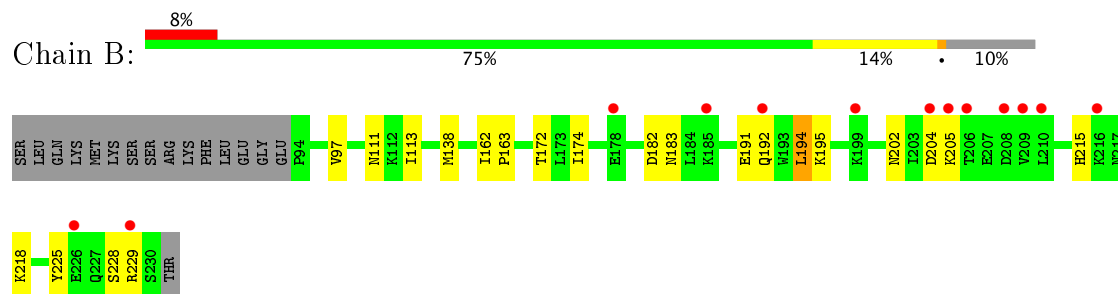
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes 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.

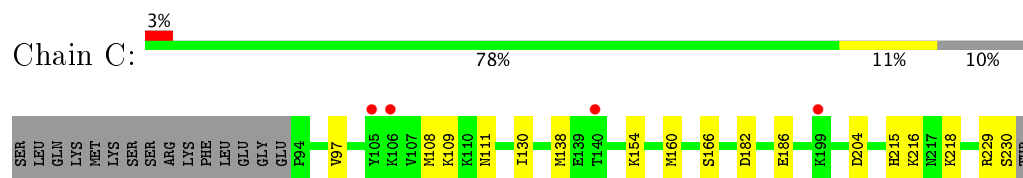
- Molecule 1: YetF protein



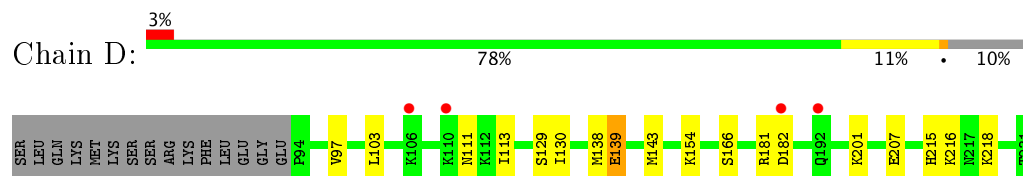
- Molecule 1: YetF protein



- Molecule 1: YetF protein



- Molecule 1: YetF protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	68.27Å 71.33Å 70.09Å 90.00° 111.93° 90.00°	Depositor
Resolution (Å)	20.00 – 2.50 30.60 – 2.21	Depositor EDS
% Data completeness (in resolution range)	98.9 (20.00-2.50) 92.5 (30.60-2.21)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.27 (at 2.22Å)	Xtriage
Refinement program	REFMAC 5.3.0034	Depositor
R, $R_{free}$	0.233 , 0.288 0.229 , 0.283	Depositor DCC
$R_{free}$ test set	681 reflections (3.28%)	DCC
Wilson B-factor (Å <sup>2</sup> )	57.0	Xtriage
Anisotropy	0.072	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 53.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.021 for l,-k,h	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	4566	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	72.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.23% 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, 0.333 respectively for untwinned datasets, and 0.375, 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: GOL

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	0.37	0/1135	0.55	0/1525
1	B	0.37	0/1157	0.58	0/1552
1	C	0.36	0/1138	0.53	0/1528
1	D	0.39	0/1154	0.56	0/1549
All	All	0.37	0/4584	0.55	0/6154

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	1115	0	1146	10	0
1	B	1131	0	1175	7	0
1	C	1118	0	1153	4	0
1	D	1131	0	1173	5	0
2	D	6	0	8	0	0
3	A	16	0	0	0	0
3	B	17	0	0	0	0
3	C	14	0	0	0	0
3	D	18	0	0	0	0
All	All	4566	0	4655	24	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 (24) 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:C:215:HIS:HB3	1:C:218:LYS:HB2	1.82	0.62
1:A:218:LYS:HG3	1:A:219:PRO:HD2	1.83	0.61
1:A:218:LYS:CG	1:A:219:PRO:HD2	2.31	0.61
1:C:229:ARG:O	1:C:230:SER:C	2.42	0.57
1:D:215:HIS:HB3	1:D:218:LYS:HB2	1.87	0.56
1:B:225:TYR:HA	1:B:228:SER:HB3	1.90	0.54
1:B:174:ILE:HD11	1:B:194:LEU:HD11	1.88	0.54
1:B:172:THR:O	1:B:183:ASN:ND2	2.36	0.53
1:A:160:MET:CE	1:A:160:MET:HA	2.39	0.52
1:B:215:HIS:HB3	1:B:218:LYS:HB2	1.92	0.51
1:C:97:VAL:HG23	1:C:108:MET:HG3	1.91	0.51
1:A:124:GLN:HE21	1:D:139:GLU:HG2	1.77	0.49
1:A:97:VAL:HG23	1:A:108:MET:HG2	1.95	0.48
1:B:97:VAL:HG11	1:B:113:ILE:HD11	1.96	0.47
1:C:154:LYS:HB2	1:D:154:LYS:HE3	1.98	0.46
1:A:193:TRP:CZ3	1:A:220:LEU:HD13	2.52	0.45
1:D:103:LEU:HD21	1:D:130:ILE:HD13	1.97	0.45
1:A:97:VAL:CG2	1:A:108:MET:HG2	2.47	0.44
1:B:191:GLU:HG2	1:B:195:LYS:HE2	2.01	0.42
1:A:218:LYS:CB	1:A:219:PRO:HD2	2.51	0.41
1:D:97:VAL:HG11	1:D:113:ILE:HD11	2.02	0.41
1:A:122:LEU:HD22	1:A:133:VAL:HG21	2.03	0.41
1:B:162:ILE:HA	1:B:163:PRO:HD3	1.93	0.40
1:A:97:VAL:HG21	1:A:113:ILE:HD11	2.03	0.40

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	136/153 (89%)	133 (98%)	3 (2%)	0	100	100
1	B	138/153 (90%)	133 (96%)	5 (4%)	0	100	100
1	C	136/153 (89%)	133 (98%)	3 (2%)	0	100	100
1	D	138/153 (90%)	133 (96%)	4 (3%)	1 (1%)	25	43
All	All	548/612 (90%)	532 (97%)	15 (3%)	1 (0%)	51	73

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	181	ARG

### 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	129/143 (90%)	120 (93%)	9 (7%)	18	33
1	B	131/143 (92%)	122 (93%)	9 (7%)	18	34
1	C	129/143 (90%)	119 (92%)	10 (8%)	15	28
1	D	131/143 (92%)	121 (92%)	10 (8%)	15	29
All	All	520/572 (91%)	482 (93%)	38 (7%)	16	31

All (38) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	112	LYS
1	A	143	MET
1	A	160	MET
1	A	165	LYS
1	A	166	SER
1	A	204	ASP
1	A	206	THR
1	A	218	LYS
1	A	229	ARG
1	B	111	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	138	MET
1	B	182	ASP
1	B	192	GLN
1	B	194	LEU
1	B	202	ASN
1	B	204	ASP
1	B	205	LYS
1	B	229	ARG
1	C	109	LYS
1	C	111	ASN
1	C	130	ILE
1	C	138	MET
1	C	160	MET
1	C	166	SER
1	C	182	ASP
1	C	186	GLU
1	C	204	ASP
1	C	216	LYS
1	D	111	ASN
1	D	129	SER
1	D	138	MET
1	D	139	GLU
1	D	143	MET
1	D	166	SER
1	D	182	ASP
1	D	201	LYS
1	D	207	GLU
1	D	216	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (8) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	124	GLN
1	A	192	GLN
1	A	196	GLN
1	A	217	ASN
1	A	227	GLN
1	B	117	GLN
1	B	124	GLN
1	D	196	GLN

### 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 ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

1 ligand 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	$\# Z  > 2$
2	GOL	D	301	-	5,5,5	0.35	0	5,5,5	0.27	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
2	GOL	D	301	-	-	0/4/4/4	0/0/0/0

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.

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<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	137/153 (89%)	0.38	5 (3%) 43 45	34, 67, 109, 115	0
1	B	137/153 (89%)	0.48	13 (9%) 9 8	36, 70, 115, 122	0
1	C	137/153 (89%)	0.34	4 (2%) 52 55	44, 70, 106, 110	0
1	D	138/153 (90%)	0.14	4 (2%) 52 55	42, 68, 101, 114	0
All	All	549/612 (89%)	0.34	26 (4%) 32 34	34, 69, 108, 122	0

All (26) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	209	VAL	4.7
1	B	204	ASP	4.2
1	B	206	THR	4.2
1	C	140	THR	3.9
1	C	105	TYR	3.9
1	B	226	GLU	3.9
1	B	205	LYS	3.8
1	A	201	LYS	3.3
1	B	210	LEU	3.2
1	D	182	ASP	3.2
1	B	192	GLN	2.9
1	C	199	LYS	2.9
1	A	229	ARG	2.8
1	A	195	LYS	2.8
1	B	208	ASP	2.6
1	A	185	LYS	2.6
1	D	106	LYS	2.5
1	D	192	GLN	2.4
1	B	185	LYS	2.3
1	B	199	LYS	2.3
1	A	200	LYS	2.3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	C	106	LYS	2.3
1	B	216	LYS	2.2
1	B	229	ARG	2.0
1	B	178	GLU	2.0
1	D	110	LYS	2.0

## 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](#)

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<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(Å <sup>2</sup> )	Q<0.9
2	GOL	D	301	6/6	0.80	0.17	-0.28	63,80,81,83	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.