



# Full wwPDB X-ray Structure Validation Report ⓘ

Jul 5, 2017 – 06:31 PM EDT

PDB ID : 5LFE  
Title : Protruding domain of GI.4 human norovirus isolate 8-14 in complex with HBGA type B (triglycan)  
Authors : Singh, B.K.; Hansman, G.S.  
Deposited on : unknown  
Resolution : 2.30 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-20029824  
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-20029824

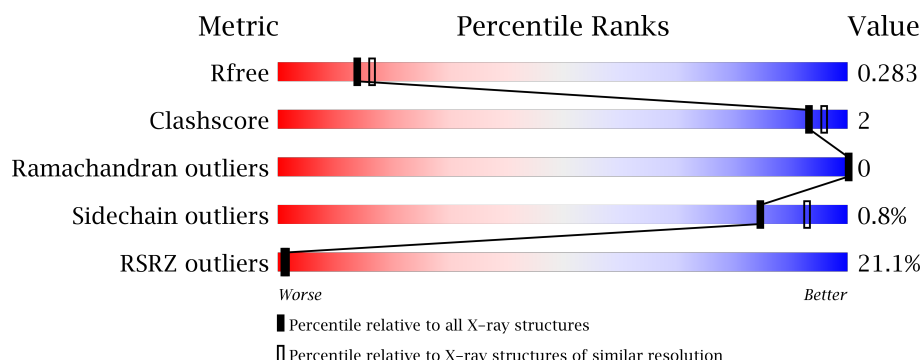
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.30 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	4130 (2.30-2.30)
Clashscore	112137	4751 (2.30-2.30)
Ramachandran outliers	110173	4705 (2.30-2.30)
Sidechain outliers	110143	4704 (2.30-2.30)
RSRZ outliers	101464	4156 (2.30-2.30)

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	309	<div> <div>25%</div> <div> <div></div> <div>95%</div> <div>.</div> </div> </div>
1	B	309	<div> <div>17%</div> <div> <div></div> <div>93%</div> <div>6% .</div> </div> </div>

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 9639 atoms, of which 4508 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.

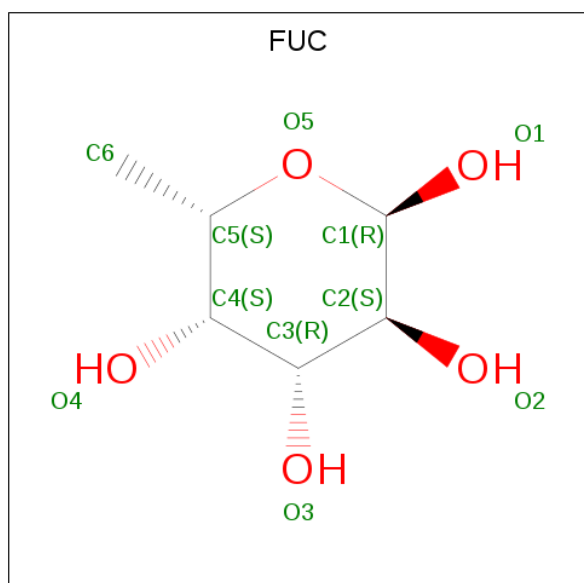
- Molecule 1 is a protein called VP1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	309	Total	C	H	N	O	S	0	0	0
			4667	1520	2273	409	454	11			
1	B	306	Total	C	H	N	O	S	0	0	0
			4606	1505	2235	404	451	11			

There are 6 discrepancies between the modelled and reference sequences:

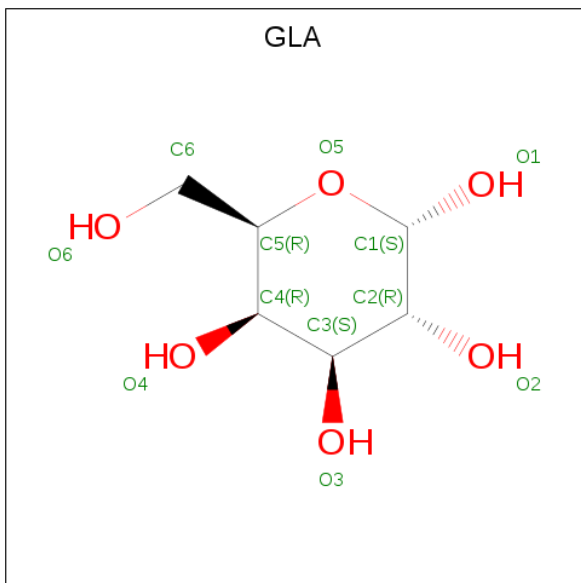
Chain	Residue	Modelled	Actual	Comment	Reference
A	222	PRO	-	expression tag	UNP M4QNL3
A	223	GLY	-	expression tag	UNP M4QNL3
A	224	SER	-	expression tag	UNP M4QNL3
B	222	PRO	-	expression tag	UNP M4QNL3
B	223	GLY	-	expression tag	UNP M4QNL3
B	224	SER	-	expression tag	UNP M4QNL3

- Molecule 2 is ALPHA-L-FUCOSE (three-letter code: FUC) (formula: C<sub>6</sub>H<sub>12</sub>O<sub>5</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			10	6	4		

- Molecule 3 is ALPHA D-GALACTOSE (three-letter code: GLA) (formula: C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			12	6	6		
3	A	1	Total	C	O	0	0
			11	6	5		

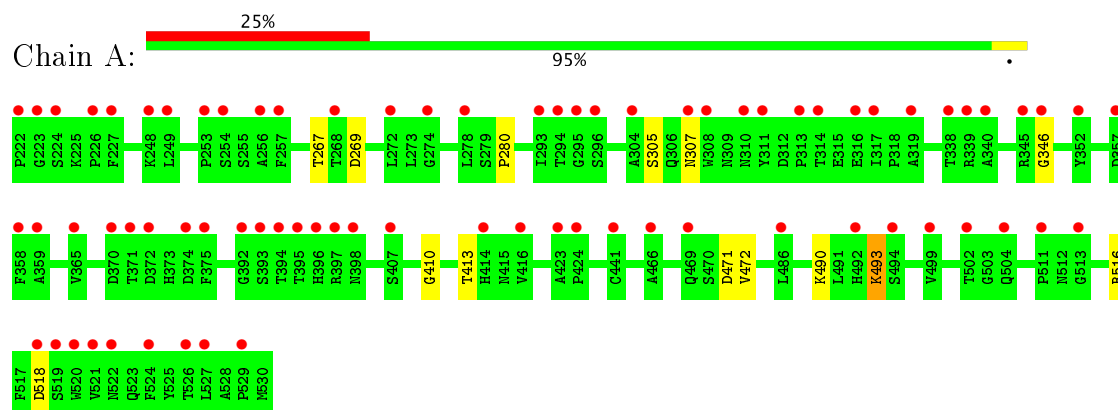
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	150	Total	O	0	0
			150	150		
4	B	183	Total	O	0	0
			183	183		

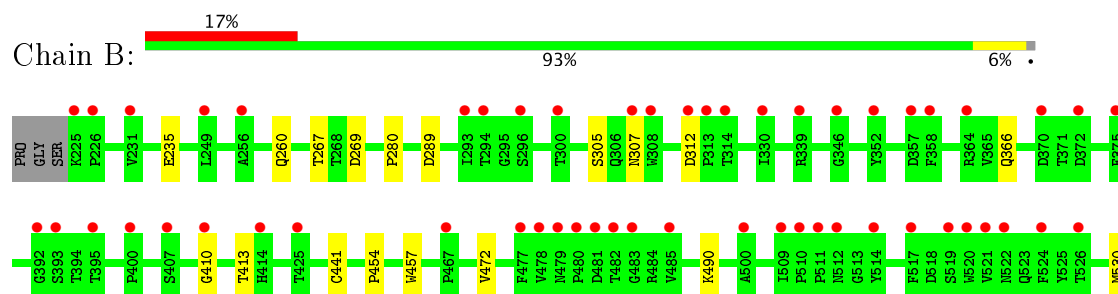
### 3 Residue-property plots

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 ( $\text{RSRZ} > 2$ ). 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: VP1



#### • Molecule 1: VP1



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	63.83Å 55.42Å 87.10Å 90.00° 99.38° 90.00°	Depositor
Resolution (Å)	46.58 – 2.30 47.26 – 2.27	Depositor EDS
% Data completeness (in resolution range)	94.4 (46.58-2.30) 94.0 (47.26-2.27)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.02 (at 2.27Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.233 , 0.277 0.234 , 0.283	Depositor DCC
$R_{free}$ test set	1270 reflections (4.99%)	DCC
Wilson B-factor (Å <sup>2</sup> )	24.8	Xtriage
Anisotropy	0.143	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 48.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.81	EDS
Total number of atoms	9639	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 78.73 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 6.4942e-07. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: GLA, 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	0.24	0/2466	0.45	0/3374
1	B	0.25	0/2442	0.45	0/3344
All	All	0.24	0/4908	0.45	0/6718

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

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	2394	2273	2275	8	0
1	B	2371	2235	2248	10	0
2	A	10	0	10	0	0
3	A	23	0	20	0	0
4	A	150	0	0	0	0
4	B	183	0	0	2	0
All	All	5131	4508	4553	16	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 2.

All (16) 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:305:SER:OG	1:B:307:ASN:OD1	2.04	0.76
1:A:267:THR:OG1	1:A:269:ASP:OD1	2.13	0.59
1:B:410:GLY:O	1:B:413:THR:OG1	2.22	0.56
1:B:260:GLN:NE2	4:B:608:HOH:O	2.39	0.51
1:A:305:SER:OG	1:A:307:ASN:OD1	2.27	0.50
1:A:280:PRO:HG2	1:B:280:PRO:HG2	1.93	0.49
1:A:472:VAL:HG11	1:A:490:LYS:HG2	1.94	0.49
1:A:516:ARG:NE	1:A:518:ASP:OD1	2.43	0.49
1:B:235:GLU:OE1	4:B:601:HOH:O	2.19	0.48
1:A:410:GLY:O	1:A:413:THR:OG1	2.29	0.45
1:B:267:THR:OG1	1:B:269:ASP:OD2	2.18	0.45
1:A:471:ASP:HA	1:A:493:LYS:HE2	2.01	0.42
1:A:346:GLY:N	1:B:441:CYS:O	2.53	0.42
1:B:472:VAL:HG11	1:B:490:LYS:HG2	2.03	0.41
1:B:312:ASP:N	1:B:312:ASP:OD1	2.54	0.40
1:B:454:PRO:HG2	1:B:457:TRP:CG	2.57	0.40

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	307/309 (99%)	303 (99%)	4 (1%)	0	100	100
1	B	304/309 (98%)	300 (99%)	4 (1%)	0	100	100
All	All	611/618 (99%)	603 (99%)	8 (1%)	0	100	100

There are no Ramachandran outliers to report.



### 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	262/267 (98%)	261 (100%)	1 (0%)	93	97
1	B	260/267 (97%)	257 (99%)	3 (1%)	75	87
All	All	522/534 (98%)	518 (99%)	4 (1%)	85	93

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	493	LYS
1	B	289	ASP
1	B	366	GLN
1	B	530	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	309	ASN
1	B	522	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](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry

3 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	FUC	A	601	3	9,10,11	0.66	0	13,14,16	0.90	0
3	GLA	A	602	3,2	12,12,12	0.57	0	17,17,17	0.98	1 (5%)
3	GLA	A	603	3	11,11,12	0.82	0	13,15,17	1.09	1 (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	FUC	A	601	3	-	0/0/17/20	0/1/1/1
3	GLA	A	602	3,2	-	0/2/22/22	0/1/1/1
3	GLA	A	603	3	-	0/2/19/22	0/1/1/1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	603	GLA	C1-O5-C5	-3.39	107.49	112.17
3	A	602	GLA	C1-O5-C5	-3.01	107.96	113.39

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	309/309 (100%)	1.48	76 (24%) ⓘ ⓘ	21, 33, 51, 81	0
1	B	306/309 (99%)	1.33	54 (17%) ⓘ ⓘ	22, 32, 54, 83	1 (0%)
All	All	615/618 (99%)	1.41	130 (21%) ⓘ ⓘ	21, 33, 53, 83	1 (0%)

All (130) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	295	GLY	7.4
1	B	226	PRO	5.0
1	A	394	THR	4.8
1	B	483	GLY	4.7
1	A	296	SER	4.6
1	B	481	ASP	4.5
1	A	340	ALA	4.5
1	A	293	ILE	4.3
1	A	359	ALA	4.2
1	A	374	ASP	4.2
1	B	293	ILE	4.2
1	B	480	PRO	4.0
1	B	294	THR	4.0
1	A	307	ASN	3.9
1	B	314	THR	3.9
1	B	370	ASP	3.9
1	B	312	ASP	3.8
1	A	392	GLY	3.8
1	A	314	THR	3.7
1	B	346	GLY	3.6
1	A	223	GLY	3.6
1	A	226	PRO	3.6
1	B	479	ASN	3.6
1	B	511	PRO	3.4

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	375	PHE	3.4
1	A	338	THR	3.3
1	B	313	PRO	3.3
1	B	482	THR	3.2
1	A	224	SER	3.2
1	A	524	PHE	3.2
1	A	529	PRO	3.2
1	B	485	VAL	3.1
1	B	517	PHE	3.0
1	A	424	PRO	3.0
1	B	514	TYR	3.0
1	B	392	GLY	3.0
1	A	372	ASP	2.9
1	B	339	ARG	2.9
1	A	227	PHE	2.9
1	A	395	THR	2.9
1	B	478	VAL	2.9
1	A	521	VAL	2.8
1	B	330	ILE	2.8
1	A	357	ASP	2.8
1	A	222	PRO	2.8
1	B	467	PRO	2.8
1	A	257	PHE	2.8
1	A	365	VAL	2.7
1	A	272	LEU	2.7
1	B	393	SER	2.7
1	A	469	GLN	2.7
1	B	300	THR	2.7
1	B	524	PHE	2.7
1	A	346	GLY	2.7
1	A	511	PRO	2.7
1	A	414	HIS	2.7
1	A	370	ASP	2.7
1	A	502	THR	2.7
1	B	357	ASP	2.6
1	A	294	THR	2.6
1	A	358	PHE	2.6
1	B	477	PHE	2.6
1	A	310	ASN	2.6
1	A	278	LEU	2.6
1	B	375	PHE	2.6
1	B	509	ILE	2.5

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>	<b>RSRZ</b>
1	A	352	TYR	2.5
1	A	423	ALA	2.5
1	B	425	THR	2.5
1	A	526	THR	2.5
1	A	397	ARG	2.5
1	A	393	SER	2.5
1	B	410	GLY	2.5
1	A	466	ALA	2.4
1	B	510	PRO	2.4
1	A	494	SER	2.4
1	B	521	VAL	2.4
1	B	395	THR	2.4
1	A	371	THR	2.4
1	A	308	TRP	2.4
1	B	520	TRP	2.4
1	B	407	SER	2.4
1	A	256	ALA	2.4
1	B	231	VAL	2.4
1	A	345	ARG	2.4
1	A	522	ASN	2.4
1	A	492	HIS	2.4
1	B	249	LEU	2.4
1	B	400	PRO	2.4
1	B	364	ARG	2.3
1	A	316	GLU	2.3
1	B	414	HIS	2.3
1	A	441	CYS	2.3
1	B	296	SER	2.3
1	B	372	ASP	2.3
1	B	352	TYR	2.2
1	A	513	GLY	2.2
1	A	248	LYS	2.2
1	A	407	SER	2.2
1	B	308	TRP	2.2
1	B	307	ASN	2.2
1	A	339	ARG	2.2
1	B	225	LYS	2.2
1	A	311	TYR	2.2
1	A	313	PRO	2.2
1	A	416	VAL	2.2
1	A	520	TRP	2.2
1	A	518	ASP	2.2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	274	GLY	2.2
1	A	249	LEU	2.1
1	A	504	GLN	2.1
1	A	319	ALA	2.1
1	B	256	ALA	2.1
1	B	526	THR	2.1
1	A	254	SER	2.1
1	A	527	LEU	2.1
1	A	268	THR	2.1
1	A	253	PRO	2.1
1	A	486	LEU	2.1
1	A	499	VAL	2.1
1	B	522	ASN	2.1
1	A	519	SER	2.1
1	B	358	PHE	2.1
1	A	396	HIS	2.1
1	A	398	ASN	2.1
1	A	317	ILE	2.0
1	A	304	ALA	2.0
1	B	519	SER	2.0
1	B	512	ASN	2.0
1	B	500	ALA	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( $\text{\AA}^2$ )	Q<0.9
3	GLA	A	602	12/12	0.73	0.32	1.58	30,34,37,39	0
3	GLA	A	603	11/12	0.62	0.39	1.19	37,42,44,46	0
2	FUC	A	601	10/11	0.81	0.27	0.20	26,31,33,33	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.