



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 15, 2017 – 12:37 am GMT

PDB ID : 5ABG  
Title : Structure of GH84 with ligand  
Authors : Bergeron-Brele, M.; Goodwin-Tindall, J.; Cekic, N.; Varghese, V.; Zandberg, W.F.; Shan, X.; Roth, C.; Chan, S.; Davies, G.J.; Vocadlo, D.J.; Britton, R.  
Deposited on : 2015-08-05  
Resolution : 2.00 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 : trunk28620  
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) : recalc28949

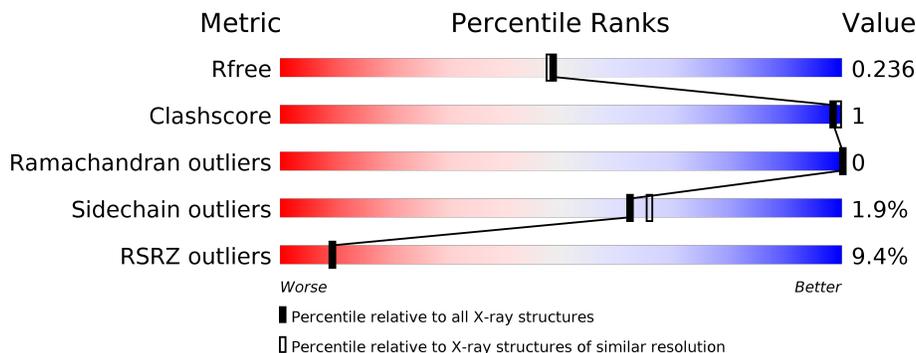
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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	6609 (2.00-2.00)
Clashscore	112137	7775 (2.00-2.00)
Ramachandran outliers	110173	7679 (2.00-2.00)
Sidechain outliers	110143	7678 (2.00-2.00)
RSRZ outliers	101464	6696 (2.00-2.00)

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	726	
1	B	726	

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	V0N	B	1718	-	-	-	X

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 21834 atoms, of which 10707 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.

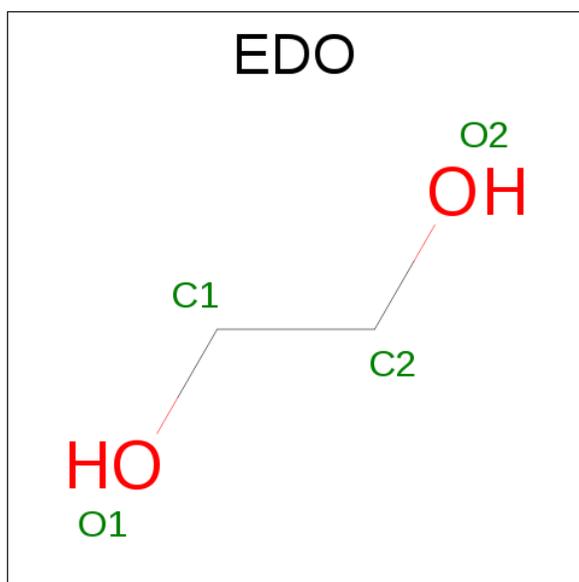
- Molecule 1 is a protein called O-GLCNACASE BT\_4395.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	663	10732	3460	5334	912	1007	19	0	3	0
1	B	662	10691	3445	5311	910	1006	19	0	1	0

There are 20 discrepancies between the modelled and reference sequences:

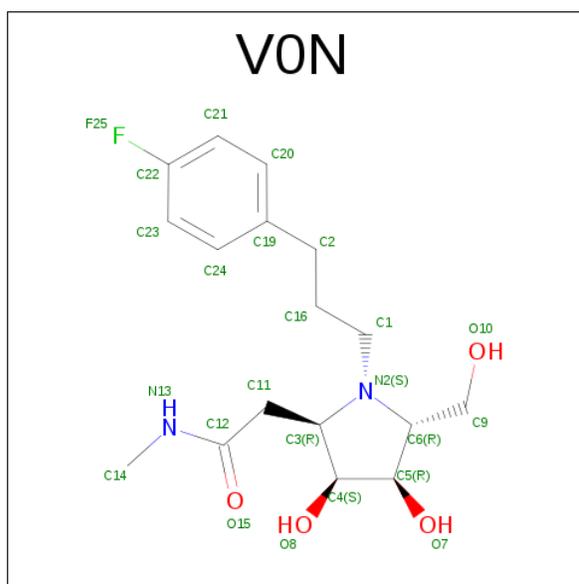
Chain	Residue	Modelled	Actual	Comment	Reference
A	-9	MET	-	EXPRESSION TAG	UNP Q89ZI2
A	-8	GLY	-	EXPRESSION TAG	UNP Q89ZI2
A	-7	SER	-	EXPRESSION TAG	UNP Q89ZI2
A	-6	SER	-	EXPRESSION TAG	UNP Q89ZI2
A	-5	HIS	-	EXPRESSION TAG	UNP Q89ZI2
A	-4	HIS	-	EXPRESSION TAG	UNP Q89ZI2
A	-3	HIS	-	EXPRESSION TAG	UNP Q89ZI2
A	-2	HIS	-	EXPRESSION TAG	UNP Q89ZI2
A	-1	HIS	-	EXPRESSION TAG	UNP Q89ZI2
A	0	HIS	-	EXPRESSION TAG	UNP Q89ZI2
B	-9	MET	-	EXPRESSION TAG	UNP Q89ZI2
B	-8	GLY	-	EXPRESSION TAG	UNP Q89ZI2
B	-7	SER	-	EXPRESSION TAG	UNP Q89ZI2
B	-6	SER	-	EXPRESSION TAG	UNP Q89ZI2
B	-5	HIS	-	EXPRESSION TAG	UNP Q89ZI2
B	-4	HIS	-	EXPRESSION TAG	UNP Q89ZI2
B	-3	HIS	-	EXPRESSION TAG	UNP Q89ZI2
B	-2	HIS	-	EXPRESSION TAG	UNP Q89ZI2
B	-1	HIS	-	EXPRESSION TAG	UNP Q89ZI2
B	0	HIS	-	EXPRESSION TAG	UNP Q89ZI2

- Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



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

- Molecule 3 is 2-[(2R,3S,4R,5R)-1-[3-(4-FLUOROPHENYL)PROPYL]-5-(HYDROXY METHYL)-3,4-BIS(OXIDANYL)PYRROLIDIN-2-YL]-N-METHYL-ETHANAMIDE (three-letter code: VON) (formula: C<sub>17</sub>H<sub>25</sub>FN<sub>2</sub>O<sub>4</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	
			Total	C	F	H	N			O
3	A	1	49	17	1	25	2	4	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	
3	B	1	Total	C	F	H	N	O	0	0
			49	17	1	25	2	4		

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Ca	0	0
			1	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	148	Total	O	0	0
			148	148		
5	B	144	Total	O	0	0
			144	144		



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 2 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	51.53Å 161.18Å 223.98Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.44 – 2.00 48.44 – 2.00	Depositor EDS
% Data completeness (in resolution range)	98.7 (48.44-2.00) 84.6 (48.44-2.00)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.85 (at 2.00Å)	Xtrriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, $R_{free}$	0.204 , 0.235 0.202 , 0.236	Depositor DCC
$R_{free}$ test set	12015 reflections (5.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	32.7	Xtrriage
Anisotropy	0.760	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.39 , 51.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	21834	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	69.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 21.90 % 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 6.3617e-03. 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 [i](#)

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

Bond lengths and bond angles in the following residue types are not validated in this section: CA, V0N, EDO

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	0.22	0/5544	0.39	0/7506
1	B	0.22	0/5512	0.39	0/7460
All	All	0.22	0/11056	0.39	0/14966

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	5398	5334	5325	6	0
1	B	5380	5311	5310	9	0
2	A	4	6	6	0	0
2	B	4	6	6	0	0
3	A	24	25	25	0	0
3	B	24	25	25	0	0
4	B	1	0	0	0	0
5	A	148	0	0	0	0
5	B	144	0	0	0	0
All	All	11127	10707	10697	15	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 1.

All (15) 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:A:662:SER:OG	1:A:664:ASP:O	2.20	0.60
1:B:676:LYS:CB	1:B:677:GLU:HA	2.34	0.58
1:A:360:THR:OG1	1:A:394:PRO:O	2.21	0.58
1:B:408:ARG:NH1	1:B:419:GLU:OE2	2.38	0.57
1:A:32:GLU:OE2	1:A:65:LYS:NZ	2.29	0.55
1:B:676:LYS:HB3	1:B:677:GLU:HA	1.94	0.50
1:A:664:ASP:OD1	1:A:665:GLY:N	2.47	0.47
1:B:562:LYS:HB3	1:B:563:PRO:HD3	1.98	0.44
1:A:562[B]:LYS:HB3	1:A:563:PRO:HD3	2.01	0.43
1:A:442:ASP:N	1:A:442:ASP:OD1	2.53	0.42
1:B:444:GLN:N	1:B:445:PRO:HD2	2.35	0.42
1:B:660:GLU:HB2	1:B:693:ARG:HB3	2.02	0.41
1:B:660:GLU:CG	1:B:661:ILE:H	2.33	0.41
1:B:242:ASP:N	1:B:242:ASP:OD1	2.54	0.41
1:B:676:LYS:CB	1:B:677:GLU:CA	2.99	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	654/726 (90%)	625 (96%)	29 (4%)	0	100	100
1	B	649/726 (89%)	626 (96%)	23 (4%)	0	100	100
All	All	1303/1452 (90%)	1251 (96%)	52 (4%)	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	587/639 (92%)	577 (98%)	10 (2%)	66	70
1	B	585/639 (92%)	571 (98%)	14 (2%)	54	56
All	All	1172/1278 (92%)	1148 (98%)	24 (2%)	62	64

All (24) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	14	VAL
1	A	17	LYS
1	A	49	GLN
1	A	120	GLU
1	A	371	THR
1	A	423[A]	MET
1	A	423[B]	MET
1	A	519	ARG
1	A	669	LYS
1	A	694	PHE
1	B	39	LYS
1	B	48	LYS
1	B	119	VAL
1	B	371	THR
1	B	423[A]	MET
1	B	423[B]	MET
1	B	454	PHE
1	B	456	GLU
1	B	595	MET
1	B	617	SER
1	B	673	LEU
1	B	674	LYS
1	B	685	GLN
1	B	686	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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 [i](#)

Of 5 ligands modelled in this entry, 1 is 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
2	EDO	A	1716	-	3,3,3	0.44	0	2,2,2	0.30	0
3	V0N	A	1717	-	25,25,25	0.59	0	30,34,34	1.08	1 (3%)
2	EDO	B	1717	-	3,3,3	0.45	0	2,2,2	0.25	0
3	V0N	B	1718	-	25,25,25	0.63	0	30,34,34	0.91	1 (3%)

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	EDO	A	1716	-	-	0/1/1/1	0/0/0/0
3	V0N	A	1717	-	-	0/14/34/34	0/2/2/2
2	EDO	B	1717	-	-	0/1/1/1	0/0/0/0
3	V0N	B	1718	-	-	0/14/34/34	0/2/2/2

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	1717	V0N	C23-C22-C21	-2.12	119.91	122.86
3	B	1718	V0N	C23-C22-C21	-2.01	120.07	122.86

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 [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 [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	663/726 (91%)	0.64	59 (8%) 10   10	36, 57, 111, 144	0
1	B	662/726 (91%)	0.63	66 (9%) 8   8	34, 54, 113, 164	0
All	All	1325/1452 (91%)	0.63	125 (9%) 9   9	34, 56, 112, 164	0

All (125) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	646	ILE	9.1
1	B	685	GLN	8.1
1	B	692	VAL	8.1
1	B	684	LEU	7.5
1	A	53	LYS	7.5
1	A	616	ILE	7.3
1	B	673	LEU	6.8
1	B	596	ILE	6.4
1	A	604	ASN	6.3
1	B	631	VAL	6.3
1	B	659	LEU	6.2
1	B	683	GLY	6.2
1	B	709	PHE	5.9
1	A	682	ALA	5.8
1	B	682	ALA	5.7
1	B	644	ILE	5.7
1	A	603	LYS	5.7
1	B	633	ILE	5.5
1	A	680	LEU	5.4
1	B	674	LYS	5.3
1	B	688	PRO	5.3
1	B	661	ILE	5.2
1	B	691	PHE	5.2
1	A	709	PHE	5.1

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>	<b>RSRZ</b>
1	B	618	PRO	5.0
1	B	694	PHE	5.0
1	A	14	VAL	4.8
1	B	603	LYS	4.8
1	B	658	ARG	4.8
1	B	583	HIS	4.8
1	A	706	LEU	4.7
1	A	631	VAL	4.7
1	B	675	GLN	4.6
1	B	687	ALA	4.6
1	A	661	ILE	4.5
1	B	641	GLY	4.5
1	B	290	ASN	4.4
1	B	595	MET	4.4
1	A	596	ILE	4.3
1	B	693	ARG	4.3
1	B	689	VAL	4.3
1	A	49	GLN	4.2
1	B	715	LYS	4.1
1	B	672	ASP	4.0
1	B	616	ILE	4.0
1	A	607	LEU	3.9
1	B	604	ASN	3.9
1	B	581	ASN	3.8
1	A	51	SER	3.7
1	A	606	PRO	3.6
1	B	632	GLU	3.5
1	A	21	LEU	3.5
1	A	24	VAL	3.5
1	A	644	ILE	3.4
1	B	605	LEU	3.4
1	B	676	LYS	3.3
1	A	683	GLY	3.3
1	A	595	MET	3.3
1	A	15	GLN	3.2
1	B	681	SER	3.2
1	B	594	LYS	3.1
1	A	605	LEU	3.1
1	A	85	VAL	3.1
1	B	48	LYS	3.1
1	A	646	ILE	3.1
1	B	660	GLU	3.1

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>	<b>RSRZ</b>
1	B	642	GLU	3.1
1	A	518	GLY	3.0
1	A	668	TRP	3.0
1	B	640	PRO	3.0
1	A	27	LEU	3.0
1	A	681	SER	2.9
1	A	54	GLY	2.9
1	A	119	VAL	2.8
1	A	602	ILE	2.8
1	B	606	PRO	2.8
1	B	519	ARG	2.8
1	B	46	SER	2.8
1	A	17	LYS	2.7
1	B	711	LEU	2.7
1	A	694	PHE	2.7
1	A	594	LYS	2.6
1	A	707	ARG	2.6
1	B	582	ALA	2.6
1	A	684	LEU	2.6
1	A	618	PRO	2.6
1	A	632	GLU	2.6
1	A	65	LYS	2.5
1	A	635	LEU	2.5
1	B	638	ILE	2.5
1	A	13	ILE	2.4
1	A	19	ILE	2.4
1	B	639	TYR	2.4
1	B	611	ALA	2.4
1	A	633	ILE	2.4
1	B	712	THR	2.3
1	A	456	GLU	2.3
1	B	454	PHE	2.3
1	A	88	LYS	2.3
1	B	645	GLN	2.3
1	B	53	LYS	2.3
1	A	454	PHE	2.3
1	A	48	LYS	2.3
1	B	678	SER	2.3
1	B	677	GLU	2.3
1	B	3	VAL	2.2
1	A	641	GLY	2.2
1	B	455	LYS	2.2

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	647	ASN	2.2
1	A	574	LYS	2.2
1	B	713	ILE	2.1
1	A	711	LEU	2.1
1	A	576	PHE	2.1
1	B	540	TYR	2.1
1	A	667	GLU	2.1
1	A	660	GLU	2.1
1	A	423[A]	MET	2.1
1	B	456	GLU	2.1
1	A	608	GLN	2.1
1	A	74	ILE	2.0
1	A	692	VAL	2.0
1	A	118	GLU	2.0
1	B	52	LYS	2.0
1	B	686	LYS	2.0
1	B	708	GLN	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
3	VON	B	1718	24/24	0.91	0.19	2.38	40,53,111,116	0
2	EDO	B	1717	4/4	0.92	0.20	1.33	50,60,67,67	0
4	CA	B	1716	1/1	0.97	0.14	1.18	50,50,50,50	0
3	VON	A	1717	24/24	0.94	0.17	-0.12	37,52,98,106	0
2	EDO	A	1716	4/4	0.85	0.16	-0.52	48,58,70,70	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.