



# Full wwPDB X-ray Structure Validation Report ⓘ

Jan 23, 2021 – 05:44 PM EST

PDB ID : 2034  
Title : Crystal structure of protein DVU1097 from *Desulfovibrio vulgaris* Hildenborough, Pfam DUF375  
Authors : Malashkevich, V.N.; Toro, R.; Sauder, J.M.; Schwinn, K.D.; Thompson, D.A.; Rutter, M.E.; Dickey, M.; Groshong, C.; Bain, K.T.; Adams, J.M.; Reyes, C.; Rooney, I.; Powell, A.; Boice, A.; Gheyi, T.; Ozyurt, S.; Atwell, S.; Wasserman, S.R.; Emtage, S.; Burley, S.K.; Almo, S.C.; New York SGX Research Center for Structural Genomics (NYSGXRC)  
Deposited on : 2006-11-30  
Resolution : 1.95 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

<https://www.wwpdb.org/validation/2017/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 : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.16  
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) : 2.16

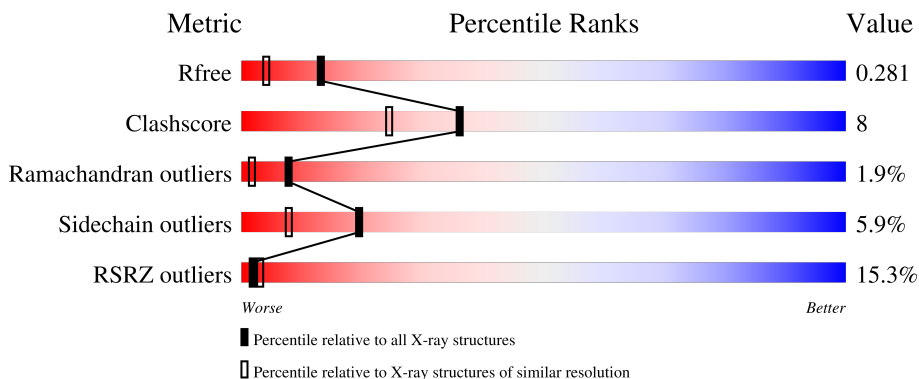
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.95 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))
$R_{free}$	130704	2580 (1.96-1.96)
Clashscore	141614	2705 (1.96-1.96)
Ramachandran outliers	138981	2678 (1.96-1.96)
Sidechain outliers	138945	2678 (1.96-1.96)
RSRZ outliers	127900	2539 (1.96-1.96)

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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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  $\leq 5\%$ . 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	250	
1	B	250	

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 3943 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 Hypothetical protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	249	1858	1161	338	347	3	9	0	1	0
1	B	231	1708	1069	308	322	2	7	0	1	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	SER	-	cloning artifact	UNP Q72D34
A	1	LEU	-	cloning artifact	UNP Q72D34
A	3	MSE	MET	modified residue	UNP Q72D34
A	48	MSE	MET	modified residue	UNP Q72D34
A	63	MSE	MET	modified residue	UNP Q72D34
A	89	MSE	MET	modified residue	UNP Q72D34
A	111	MSE	MET	modified residue	UNP Q72D34
A	152	MSE	MET	modified residue	UNP Q72D34
A	205	MSE	MET	modified residue	UNP Q72D34
A	244	MSE	MET	modified residue	UNP Q72D34
B	0	SER	-	cloning artifact	UNP Q72D34
B	1	LEU	-	cloning artifact	UNP Q72D34
B	3	MSE	MET	modified residue	UNP Q72D34
B	48	MSE	MET	modified residue	UNP Q72D34
B	63	MSE	MET	modified residue	UNP Q72D34
B	89	MSE	MET	modified residue	UNP Q72D34
B	111	MSE	MET	modified residue	UNP Q72D34
B	152	MSE	MET	modified residue	UNP Q72D34
B	205	MSE	MET	modified residue	UNP Q72D34
B	244	MSE	MET	modified residue	UNP Q72D34

- Molecule 2 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total 1	Na 1	0	0
2	A	2	Total 2	Na 2	0	0

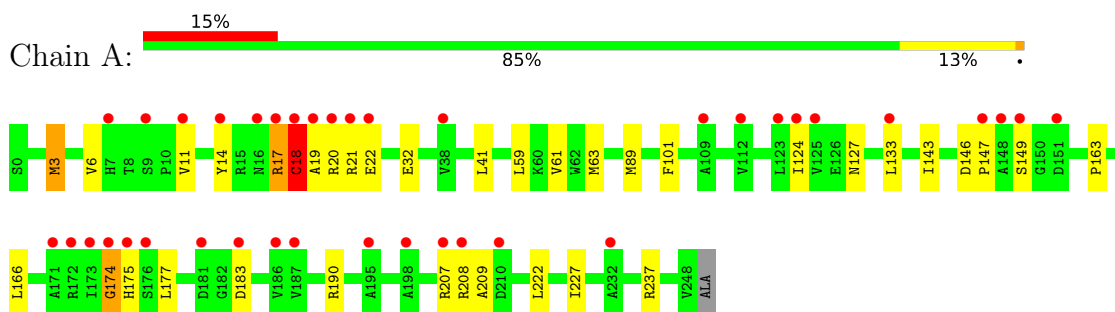
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	226	Total 226	O 226	0	0
3	B	148	Total 148	O 148	0	0

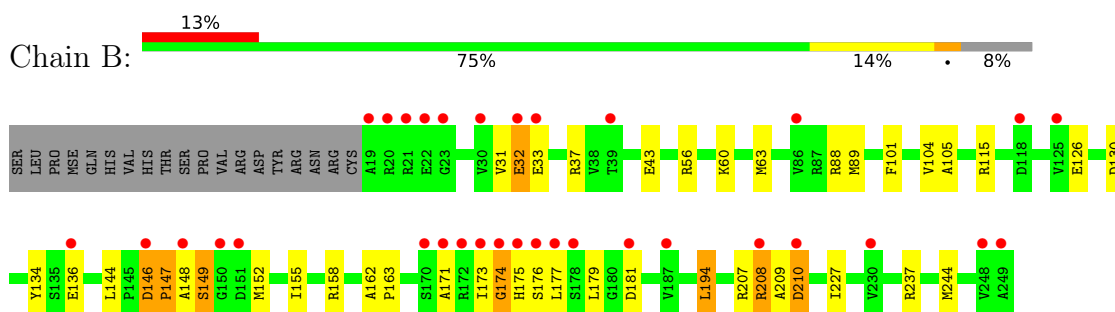
### 3 Residue-property plots [i](#)

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 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: Hypothetical protein



- Molecule 1: Hypothetical protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	60.48Å 81.48Å 151.17Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.98 – 1.95 19.98 – 1.95	Depositor EDS
% Data completeness (in resolution range)	(Not available) (19.98-1.95) 60.4 (19.98-1.95)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.28 (at 1.96Å)	Xtrriage
Refinement program	REFMAC 5.2.0019, SHELX	Depositor
R, $R_{free}$	0.210 , 0.273 0.212 , 0.281	Depositor DCC
$R_{free}$ test set	1717 reflections (5.18%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	39.6	Xtrriage
Anisotropy	0.044	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 46.8	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	0.94	EDS
Total number of atoms	3943	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	49.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.66% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: NA

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	0.68	0/1884	0.75	0/2547
1	B	0.63	0/1730	0.73	0/2339
All	All	0.66	0/3614	0.74	0/4886

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	1858	0	1861	28	0
1	B	1708	0	1716	30	0
2	A	2	0	0	0	0
2	B	1	0	0	0	0
3	A	226	0	0	4	0
3	B	148	0	0	4	0
All	All	3943	0	3577	57	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 8.

All (57) 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:174:GLY:H	1:A:175:HIS:HA	0.94	1.08
1:B:60:LYS:HA	1:B:63:MSE:HE2	1.40	1.01
1:A:174:GLY:N	1:A:175:HIS:HA	1.70	0.95
1:B:146:ASP:N	1:B:147:PRO:HD2	1.85	0.92
1:B:209:ALA:HB1	1:B:237[B]:ARG:HG3	1.59	0.83
1:B:207:ARG:HD2	3:B:639:HOH:O	1.79	0.81
1:B:146:ASP:H	1:B:147:PRO:CD	1.94	0.80
1:B:144:LEU:O	1:B:152:MSE:HB2	1.83	0.77
1:A:174:GLY:H	1:A:175:HIS:CA	1.86	0.76
1:B:146:ASP:N	1:B:147:PRO:CD	2.48	0.75
1:B:147:PRO:HA	1:B:149:SER:H	1.48	0.75
1:A:208:ARG:HB2	3:A:707:HOH:O	1.88	0.74
1:B:146:ASP:H	1:B:147:PRO:HD2	1.49	0.73
1:B:147:PRO:HA	1:B:149:SER:N	2.08	0.69
1:A:59:LEU:C	1:A:63:MSE:HE3	2.14	0.68
1:B:173:ILE:H	1:B:174:GLY:HA3	1.59	0.67
1:A:174:GLY:N	1:A:175:HIS:CA	2.53	0.66
1:B:60:LYS:HA	1:B:63:MSE:CE	2.24	0.65
1:A:59:LEU:HB3	1:A:63:MSE:CE	2.28	0.64
1:A:209:ALA:HB1	1:A:237:ARG:HG3	1.83	0.60
1:B:37:ARG:NH1	1:B:126:GLU:OE1	2.34	0.59
1:A:59:LEU:HB3	1:A:63:MSE:HE2	1.86	0.58
1:A:19:ALA:O	1:A:21:ARG:N	2.37	0.57
1:A:124:ILE:HG12	1:A:133:LEU:CD1	2.36	0.56
1:A:177:LEU:O	1:B:208:ARG:NH1	2.38	0.56
1:A:21:ARG:HD2	1:A:41:LEU:HB3	1.88	0.55
1:B:147:PRO:HG3	1:B:149:SER:O	2.08	0.53
1:B:173:ILE:H	1:B:174:GLY:CA	2.23	0.49
1:B:158:ARG:NH1	3:B:577:HOH:O	2.46	0.49
1:A:61:VAL:HG23	3:A:564:HOH:O	2.14	0.48
1:A:207:ARG:O	1:A:208:ARG:HD2	2.15	0.47
1:A:147:PRO:HA	3:A:720:HOH:O	2.14	0.46
1:B:194:LEU:HD22	1:B:227:ILE:HD11	1.98	0.46
1:B:32:GLU:HG2	1:B:33:GLU:OE2	2.16	0.45
1:B:179:LEU:C	1:B:181:ASP:H	2.20	0.44
1:A:59:LEU:O	1:A:63:MSE:HE3	2.17	0.44
1:B:162:ALA:HB1	1:B:163:PRO:HA	1.99	0.44
1:A:222:LEU:HB2	1:A:227:ILE:HD12	2.00	0.44
1:B:210:ASP:N	1:B:210:ASP:OD1	2.48	0.43
1:B:60:LYS:HE2	3:B:644:HOH:O	2.17	0.43
1:B:155:ILE:HG12	1:B:244:MSE:HE3	1.99	0.43

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:143:ILE:HG21	1:A:143:ILE:HD13	1.79	0.43
1:A:3:MSE:HE1	1:A:143:ILE:HG12	2.01	0.43
1:A:222:LEU:CB	1:A:227:ILE:HD12	2.49	0.43
1:B:134:TYR:CE2	1:B:163:PRO:HB3	2.54	0.42
1:A:124:ILE:HG12	1:A:133:LEU:HD12	2.00	0.42
1:B:175:HIS:HA	1:B:176:SER:HA	1.77	0.42
1:B:171:ALA:HB3	3:B:583:HOH:O	2.19	0.42
1:A:133:LEU:O	1:A:163:PRO:HA	2.20	0.42
1:A:190:ARG:NH1	3:A:533:HOH:O	2.54	0.41
1:B:105:ALA:HB3	1:B:130:ASP:OD2	2.21	0.41
1:A:146:ASP:HA	1:A:147:PRO:HD2	1.95	0.41
1:A:17:ARG:HG2	1:A:18:CYS:N	2.36	0.40
1:A:146:ASP:HB3	1:A:149:SER:HB2	2.02	0.40
1:A:59:LEU:HD21	1:A:89:MSE:SE	2.71	0.40
1:B:31:VAL:HG22	1:B:56:ARG:HD3	2.03	0.40
1:B:89:MSE:HE3	1:B:104:VAL:HB	2.04	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	248/250 (99%)	236 (95%)	8 (3%)	4 (2%)	9	2
1	B	230/250 (92%)	211 (92%)	14 (6%)	5 (2%)	6	1
All	All	478/500 (96%)	447 (94%)	22 (5%)	9 (2%)	8	2

All (9) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	20	ARG
1	B	147	PRO

Continued on next page...

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	177	LEU
1	A	14	TYR
1	B	148	ALA
1	A	18	CYS
1	B	146	ASP
1	A	174	GLY
1	B	174	GLY

### 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	189/181 (104%)	178 (94%)	11 (6%)	20	8
1	B	171/181 (94%)	161 (94%)	10 (6%)	20	8
All	All	360/362 (99%)	339 (94%)	21 (6%)	19	8

All (21) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	3	MSE
1	A	6	VAL
1	A	11	VAL
1	A	17	ARG
1	A	18	CYS
1	A	22	GLU
1	A	32	GLU
1	A	101	PHE
1	A	127	ASN
1	A	166	LEU
1	A	183	ASP
1	B	32	GLU
1	B	43	GLU
1	B	88	ARG
1	B	101	PHE
1	B	115	ARG

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	136	GLU
1	B	149	SER
1	B	194	LEU
1	B	208	ARG
1	B	210	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	4	GLN
1	A	127	ASN
1	A	204	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 monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

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 [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

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	241/250 (96%)	0.94	38 (15%) <b>2</b>   <b>2</b>	28, 41, 75, 90	0
1	B	224/250 (89%)	1.04	33 (14%) <b>2</b>   <b>3</b>	32, 47, 87, 99	0
All	All	465/500 (93%)	0.99	71 (15%) <b>2</b>   <b>3</b>	28, 43, 79, 99	0

All (71) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	176	SER	10.1
1	B	249	ALA	9.3
1	B	173	ILE	9.1
1	A	20	ARG	8.9
1	A	174	GLY	8.9
1	A	175	HIS	8.6
1	B	175	HIS	7.8
1	B	174	GLY	7.8
1	A	176	SER	7.6
1	B	171	ALA	7.0
1	A	148	ALA	6.9
1	B	22	GLU	6.5
1	B	151	ASP	5.8
1	A	21	ARG	5.7
1	B	23	GLY	5.7
1	A	19	ALA	5.3
1	A	171	ALA	5.1
1	A	147	PRO	5.1
1	B	172	ARG	5.1
1	A	208	ARG	5.0
1	A	17	ARG	4.8
1	B	20	ARG	4.7
1	A	172	ARG	4.7
1	A	173	ILE	4.5

*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	18	CYS	4.5
1	B	136	GLU	4.2
1	A	16	ASN	4.0
1	A	22	GLU	4.0
1	B	177	LEU	3.8
1	A	181	ASP	3.7
1	B	30	VAL	3.4
1	B	146	ASP	3.4
1	A	7	HIS	3.3
1	B	248	VAL	3.3
1	B	148	ALA	3.2
1	B	150	GLY	3.1
1	A	186	VAL	3.1
1	B	170	SER	3.1
1	B	118	ASP	3.1
1	B	19	ALA	3.0
1	B	21	ARG	3.0
1	A	112	VAL	2.9
1	A	123	LEU	2.8
1	A	11	VAL	2.7
1	B	32	GLU	2.7
1	A	133	LEU	2.7
1	A	210	ASP	2.6
1	A	9	SER	2.6
1	A	151	ASP	2.6
1	A	195	ALA	2.5
1	B	181	ASP	2.5
1	A	207	ARG	2.5
1	A	125	VAL	2.4
1	B	187	VAL	2.3
1	A	14	TYR	2.2
1	A	183	ASP	2.3
1	A	109	ALA	2.2
1	A	149	SER	2.2
1	B	208	ARG	2.2
1	A	124	ILE	2.2
1	A	187	VAL	2.2
1	B	39	THR	2.2
1	B	210	ASP	2.1
1	A	198	ALA	2.1
1	A	38	VAL	2.1
1	B	178	SER	2.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	86	VAL	2.0
1	B	230	VAL	2.0
1	B	33	GLU	2.0
1	B	125	VAL	2.0
1	A	232	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 monosaccharides 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. 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	B-factors(Å <sup>2</sup> )	Q<0.9
2	NA	B	502	1/1	0.95	0.09	37,37,37,37	0
2	NA	A	501	1/1	0.98	0.07	34,34,34,34	0
2	NA	A	503	1/1	1.00	0.06	43,43,43,43	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.