



# Full wwPDB X-ray Structure Validation Report ⓘ

May 13, 2020 – 05:06 am BST

PDB ID : 2HQ0  
Title : Structure of Rafe from Streptococcus pneumoniae  
Authors : Paterson, N.G.; Riboldi-Tunncliffe, A.; Mitchell, T.J.; Isaacs, N.W.  
Deposited on : 2006-07-18  
Resolution : 1.40 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 : 2.11  
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) : 2.11

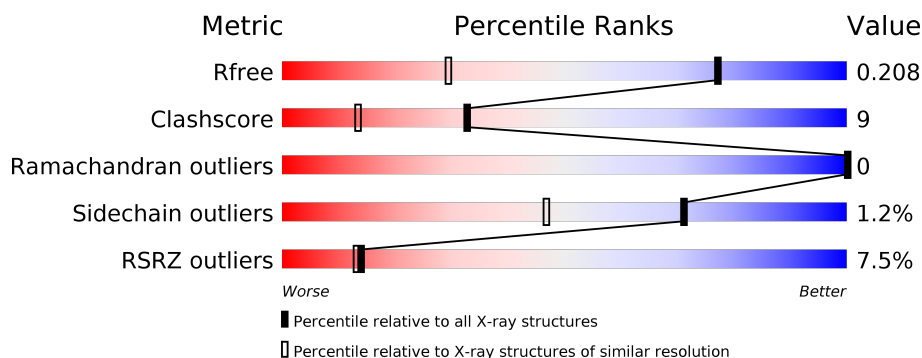
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.40 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	1714 (1.40-1.40)
Clashscore	141614	1812 (1.40-1.40)
Ramachandran outliers	138981	1763 (1.40-1.40)
Sidechain outliers	138945	1762 (1.40-1.40)
RSRZ outliers	127900	1674 (1.40-1.40)

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 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	388	<div> <div>7%</div> <div>84%</div> <div>15%</div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3753 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 Sugar ABC transporter, sugar-binding protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	388	Total	C	N	O	S	3	31	0
			3296	2125	538	621	12			

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	CLONING ARTIFACT	UNP Q97NW2
A	2	PRO	-	CLONING ARTIFACT	UNP Q97NW2
A	3	SER	-	CLONING ARTIFACT	UNP Q97NW2
A	4	SER	-	CLONING ARTIFACT	UNP Q97NW2
A	388	MET	-	CLONING ARTIFACT	UNP Q97NW2

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

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

- Molecule 3 is ACETATE ION (three-letter code: ACT) (formula: C<sub>2</sub>H<sub>3</sub>O<sub>2</sub>).



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

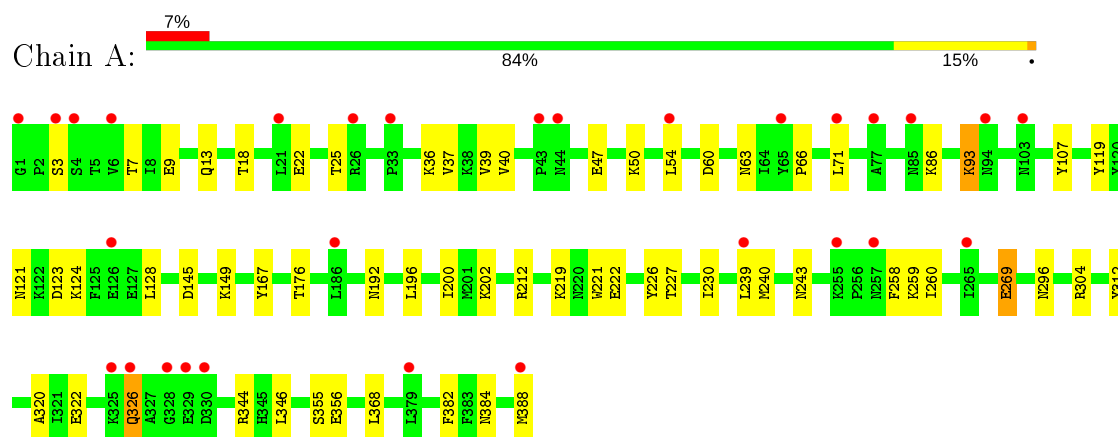
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	442	Total	O	0	2
			444	444		

### 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: Sugar ABC transporter, sugar-binding protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	73.75Å 77.70Å 79.23Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.13 – 1.40 18.87 – 1.40	Depositor EDS
% Data completeness (in resolution range)	99.0 (19.13-1.40) 99.0 (18.87-1.40)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.09 (at 1.40Å)	Xtriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.161 , 0.201 0.171 , 0.208	Depositor DCC
$R_{free}$ test set	4460 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	22.0	Xtriage
Anisotropy	0.355	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.42 , 54.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.35$	Xtriage
Estimated twinning fraction	0.003 for -h,l,k	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	3753	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.02% 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: NA, ACT

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	1.19	7/3425 (0.2%)	1.06	14/4623 (0.3%)

All (7) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	269	GLU	CG-CD	-7.55	1.40	1.51
1	A	47	GLU	CG-CD	6.33	1.61	1.51
1	A	326	GLN	CD-OE1	6.30	1.37	1.24
1	A	7	THR	C-O	5.67	1.34	1.23
1	A	355	SER	CB-OG	-5.63	1.34	1.42
1	A	107	TYR	CG-CD1	-5.39	1.32	1.39
1	A	356	GLU	CB-CG	-5.24	1.42	1.52

All (14) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	212	ARG	NE-CZ-NH2	-7.93	116.33	120.30
1	A	212	ARG	NE-CZ-NH1	7.87	124.23	120.30
1	A	346	LEU	CB-CG-CD1	7.00	122.90	111.00
1	A	123	ASP	CB-CG-OD1	6.44	124.09	118.30
1	A	304	ARG	NE-CZ-NH2	-6.18	117.21	120.30
1	A	382	PHE	CB-CG-CD2	5.54	124.68	120.80
1	A	344	ARG	NE-CZ-NH2	5.50	123.05	120.30
1	A	167	TYR	CG-CD2-CE2	5.40	125.62	121.30
1	A	60	ASP	CB-CG-OD2	-5.35	113.48	118.30
1	A	3	SER	N-CA-C	-5.26	96.78	111.00
1	A	167	TYR	CZ-CE2-CD2	-5.19	115.12	119.80
1	A	212	ARG	CG-CD-NE	-5.16	100.96	111.80
1	A	346	LEU	CA-CB-CG	5.07	126.96	115.30
1	A	196	LEU	CB-CG-CD1	5.05	119.59	111.00

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	3296	0	3343	58	1
2	A	1	0	0	0	0
3	A	12	0	9	1	0
4	A	444	0	0	14	1
All	All	3753	0	3352	58	1

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 9.

All (58) 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:368[B]:LEU:HD11	4:A:957:HOH:O	1.04	1.18
1:A:124[B]:LYS:HD3	1:A:239[B]:LEU:CD2	1.76	1.15
1:A:222[B]:GLU:HG3	1:A:368[B]:LEU:CD2	1.84	1.06
1:A:66:PRO:HA	1:A:71[B]:LEU:HD13	1.15	1.06
1:A:222[B]:GLU:HG3	1:A:368[B]:LEU:HD21	1.42	1.02
1:A:66:PRO:CA	1:A:71[B]:LEU:HD13	1.98	0.94
1:A:22[B]:GLU:OE2	1:A:39:VAL:HG21	1.69	0.91
1:A:124[B]:LYS:HD3	1:A:239[B]:LEU:HD21	1.50	0.90
1:A:66:PRO:HA	1:A:71[B]:LEU:CD1	2.02	0.90
1:A:50[A]:LYS:NZ	4:A:1009:HOH:O	2.00	0.89
1:A:124[B]:LYS:HD3	1:A:239[B]:LEU:HD22	1.52	0.89
1:A:259[B]:LYS:CA	1:A:260:ILE:N	2.36	0.88
1:A:13:GLN:H	1:A:63:ASN:HD22	1.26	0.84
1:A:200[A]:ILE:HD13	4:A:921:HOH:O	1.78	0.82
1:A:50[B]:LYS:NZ	4:A:655:HOH:O	2.11	0.81
1:A:22[B]:GLU:OE2	1:A:39:VAL:CG2	2.32	0.78
1:A:50[B]:LYS:HB3	1:A:50[B]:LYS:NZ	1.97	0.78
1:A:50[B]:LYS:HB3	1:A:50[B]:LYS:HZ3	1.53	0.74
1:A:222[B]:GLU:HG3	1:A:368[B]:LEU:HD22	1.72	0.72

*Continued on next page...*



*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:9[B]:GLU:CD	1:A:40:VAL:HG21	2.11	0.71
1:A:192:ASN:ND2	1:A:384:ASN:HD21	1.93	0.66
1:A:22[B]:GLU:OE2	1:A:22[B]:GLU:HA	1.95	0.65
1:A:258:PHE:C	1:A:259[B]:LYS:CA	2.64	0.65
1:A:9[B]:GLU:OE2	1:A:40:VAL:HG21	1.96	0.65
1:A:50[B]:LYS:CB	1:A:50[B]:LYS:NZ	2.62	0.62
1:A:145:ASP:O	1:A:149[B]:LYS:HG3	2.03	0.58
1:A:326:GLN:NE2	4:A:643:HOH:O	2.39	0.56
1:A:128:LEU:HD22	1:A:149[B]:LYS:HD3	1.87	0.56
1:A:93[B]:LYS:HG3	1:A:320:ALA:O	2.07	0.54
1:A:124[B]:LYS:CD	1:A:239[B]:LEU:CD2	2.67	0.54
1:A:227[B]:THR:HG22	4:A:1034:HOH:O	2.08	0.54
1:A:221:TRP:CH2	1:A:368[B]:LEU:HD23	2.44	0.52
1:A:9[B]:GLU:HG3	1:A:40:VAL:HG23	1.92	0.52
1:A:86:LYS:HD3	3:A:601:ACT:H3	1.93	0.50
1:A:18:THR:O	1:A:22[A]:GLU:HG3	2.12	0.49
1:A:25[A]:THR:HG23	1:A:37:VAL:HB	1.94	0.49
1:A:50[B]:LYS:NZ	4:A:614:HOH:O	2.46	0.49
1:A:66:PRO:HA	1:A:71[A]:LEU:HD22	1.93	0.48
1:A:226:TYR:CZ	1:A:230[B]:ILE:HD11	2.49	0.47
1:A:54:LEU:HD12	4:A:658:HOH:O	2.13	0.47
1:A:9[B]:GLU:HG3	1:A:40:VAL:CG2	2.45	0.47
1:A:176:THR:HG22	1:A:200[B]:ILE:HD13	1.99	0.45
1:A:192:ASN:HD21	1:A:384:ASN:HD21	1.64	0.45
1:A:312:TYR:CE2	1:A:326:GLN:HA	2.52	0.45
1:A:312:TYR:CZ	1:A:326:GLN:HG2	2.53	0.43
1:A:124[B]:LYS:CD	1:A:239[B]:LEU:HD22	2.35	0.43
1:A:50[B]:LYS:CB	1:A:50[B]:LYS:HZ2	2.30	0.43
1:A:202:LYS:NZ	4:A:622:HOH:O	2.14	0.43
1:A:119:TYR:O	1:A:240:MET:HA	2.17	0.43
1:A:219[B]:LYS:CD	4:A:611:HOH:O	2.67	0.43
1:A:322:GLU:HG3	4:A:648:HOH:O	2.17	0.43
1:A:121:ASN:ND2	1:A:124[B]:LYS:HD2	2.35	0.42
1:A:86:LYS:HE3	1:A:296:ASN:OD1	2.19	0.42
1:A:149[A]:LYS:HG3	4:A:1020:HOH:O	2.19	0.41
1:A:50[B]:LYS:HB3	1:A:50[B]:LYS:HZ2	1.83	0.41
1:A:25[B]:THR:HG21	4:A:910:HOH:O	2.21	0.41
1:A:219[B]:LYS:HD2	4:A:611:HOH:O	2.21	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:36[B]:LYS:NZ	4:A:608:HOH:O[3_755]	2.18	0.02

### 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	420/388 (108%)	411 (98%)	9 (2%)	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	358/324 (110%)	353 (99%)	5 (1%)	67	40

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	93[A]	LYS
1	A	93[B]	LYS
1	A	243	ASN
1	A	269	GLU
1	A	388	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (6) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	63	ASN
1	A	151	GLN
1	A	192	ASN
1	A	220	ASN
1	A	326	GLN
1	A	380	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 [i](#)

Of 4 ligands modelled in this entry, 1 is monoatomic - leaving 3 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	$\# Z  > 2$
3	ACT	A	601	-	1,3,3	2.21	1 (100%)	0,3,3	0.00	-
3	ACT	A	602	-	1,3,3	1.75	0	0,3,3	0.00	-
3	ACT	A	603	-	1,3,3	2.34	1 (100%)	0,3,3	0.00	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	603	ACT	CH3-C	2.34	1.51	1.48

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	601	ACT	CH3-C	2.21	1.51	1.48

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	601	ACT	1	0

## 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	388/388 (100%)	0.35	29 (7%)	14 13	18, 28, 43, 52	3 (0%)

All (29) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	65	TYR	7.9
1	A	54	LEU	5.0
1	A	328	GLY	4.1
1	A	388	MET	3.5
1	A	103	ASN	3.3
1	A	1	GLY	3.3
1	A	330	ASP	3.3
1	A	126[A]	GLU	3.3
1	A	94	ASN	2.8
1	A	257[A]	ASN	2.8
1	A	325[A]	LYS	2.7
1	A	33	PRO	2.6
1	A	379	LEU	2.6
1	A	4	SER	2.5
1	A	26[A]	ARG	2.5
1	A	329	GLU	2.5
1	A	44	ASN	2.5
1	A	71[A]	LEU	2.4
1	A	186	LEU	2.4
1	A	326	GLN	2.3
1	A	3	SER	2.3
1	A	43	PRO	2.2
1	A	239[A]	LEU	2.2
1	A	85	ASN	2.1
1	A	21	LEU	2.1
1	A	77	ALA	2.1
1	A	265	ILE	2.0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	255	LYS	2.0
1	A	6	VAL	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. 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( $\text{\AA}^2$ )	Q<0.9
3	ACT	A	601	4/4	0.81	0.20	45,46,47,47	0
3	ACT	A	602	4/4	0.86	0.14	31,33,34,36	4
3	ACT	A	603	4/4	0.90	0.15	31,31,31,34	4
2	NA	A	501	1/1	0.95	0.16	34,34,34,34	1

## 6.5 Other polymers [i](#)

There are no such residues in this entry.