



# Full wwPDB X-ray Structure Validation Report ⓘ

Aug 21, 2020 – 08:46 PM BST

PDB ID : 2HS5  
Title : Structural Genomics, the crystal structure of a putative transcriptional regulator GntR from *Rhodococcus* sp. RHA1  
Authors : Tan, K.; Skarina, T.; Onopriyenko, A.; Savchenko, A.; Edwards, A.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)  
Deposited on : 2006-07-21  
Resolution : 2.20 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 : 2.13.1  
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) : 2.13.1

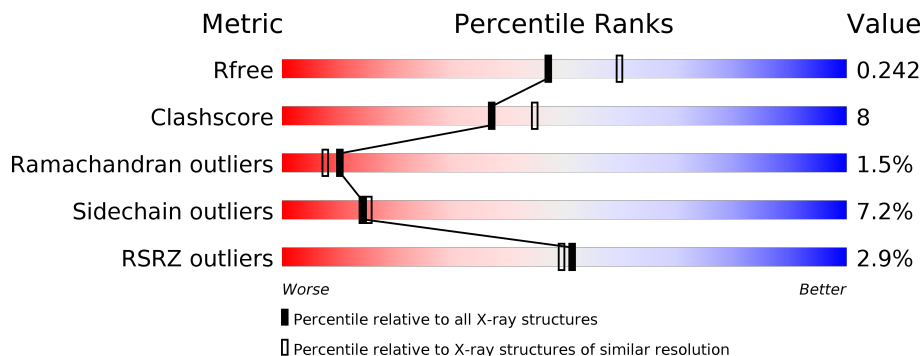
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	239	

## 2 Entry composition [i](#)

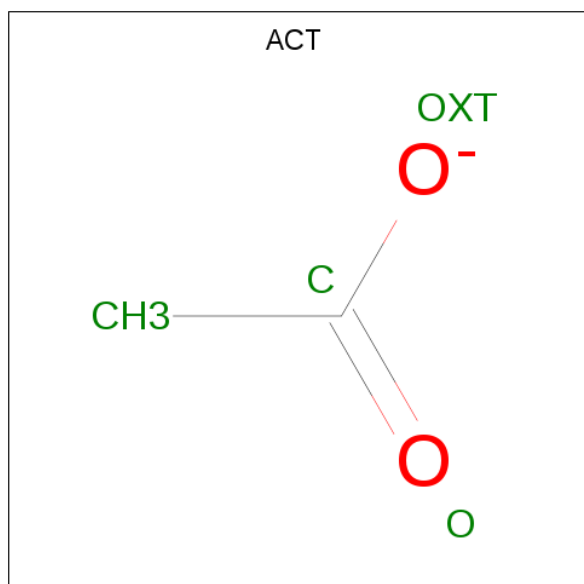
There are 3 unique types of molecules in this entry. The entry contains 1710 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 putative transcriptional regulator GntR.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	207	1618	1005	296	312	3	2	0	0	0

- Molecule 2 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
			Total	C	O		
2	A	1	4	2	2	0	0

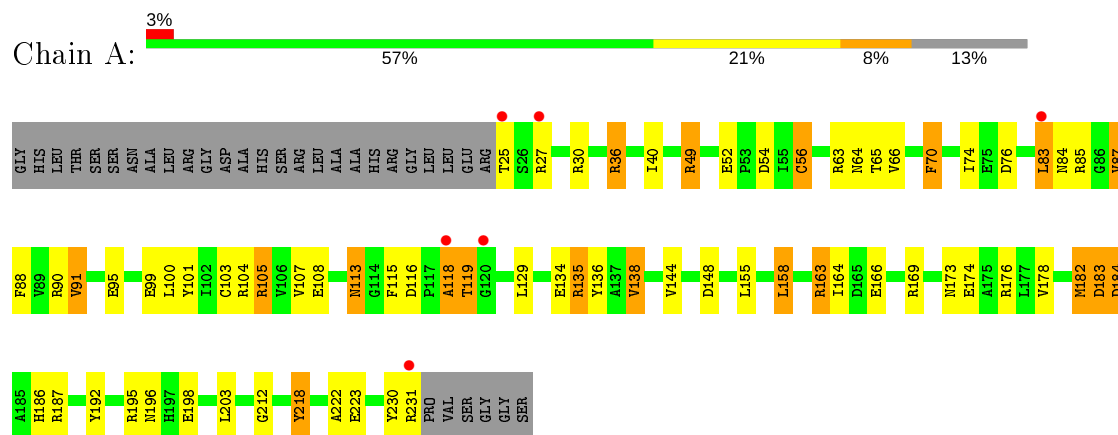
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	88	88	88	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: putative transcriptional regulator GntR



## 4 Data and refinement statistics

Property	Value	Source
Space group	I 4 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	108.05Å 108.05Å 107.26Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.21 – 2.20 38.20 – 2.20	Depositor EDS
% Data completeness (in resolution range)	97.9 (38.21-2.20) 97.9 (38.20-2.20)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.74 (at 2.20Å)	Xtrriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.194 , 0.249 0.195 , 0.242	Depositor DCC
$R_{free}$ test set	815 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	40.0	Xtrriage
Anisotropy	0.124	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 37.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.009 for l,-k,h 0.018 for -h,-l,-k	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	1710	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	43.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.98% 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:  
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	1.80	38/1644 (2.3%)	1.54	26/2231 (1.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

All (38) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	166	GLU	CG-CD	9.17	1.65	1.51
1	A	230	TYR	CD2-CE2	8.03	1.51	1.39
1	A	198	GLU	CD-OE1	7.97	1.34	1.25
1	A	101	TYR	CD2-CE2	7.97	1.51	1.39
1	A	136	TYR	CD1-CE1	7.66	1.50	1.39
1	A	144	VAL	CB-CG2	7.22	1.68	1.52
1	A	222	ALA	CA-CB	-7.21	1.37	1.52
1	A	108	GLU	CG-CD	7.09	1.62	1.51
1	A	105	ARG	CZ-NH1	6.94	1.42	1.33
1	A	52	GLU	CG-CD	6.93	1.62	1.51
1	A	230	TYR	CE1-CZ	6.92	1.47	1.38
1	A	195	ARG	CD-NE	-6.54	1.35	1.46
1	A	49	ARG	CG-CD	6.43	1.68	1.51
1	A	134	GLU	CB-CG	-6.35	1.40	1.52
1	A	99	GLU	CG-CD	6.21	1.61	1.51
1	A	101	TYR	CB-CG	5.90	1.60	1.51
1	A	88	PHE	CE1-CZ	5.89	1.48	1.37
1	A	70	PHE	C-O	5.81	1.34	1.23

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	87	VAL	CB-CG2	-5.79	1.40	1.52
1	A	169	ARG	CG-CD	5.70	1.66	1.51
1	A	113	ASN	CB-CG	5.67	1.64	1.51
1	A	134	GLU	CD-OE1	5.62	1.31	1.25
1	A	134	GLU	CG-CD	5.61	1.60	1.51
1	A	103	CYS	CB-SG	-5.57	1.72	1.81
1	A	105	ARG	CG-CD	5.52	1.65	1.51
1	A	218	TYR	CD2-CE2	5.51	1.47	1.39
1	A	54	ASP	N-CA	5.44	1.57	1.46
1	A	198	GLU	CG-CD	5.41	1.60	1.51
1	A	166	GLU	CD-OE2	5.36	1.31	1.25
1	A	223	GLU	CG-CD	5.34	1.59	1.51
1	A	192	TYR	CD2-CE2	5.30	1.47	1.39
1	A	192	TYR	CE1-CZ	5.29	1.45	1.38
1	A	230	TYR	CZ-OH	5.23	1.46	1.37
1	A	101	TYR	CE1-CZ	5.22	1.45	1.38
1	A	115	PHE	CE1-CZ	5.18	1.47	1.37
1	A	95	GLU	CG-CD	5.16	1.59	1.51
1	A	56	CYS	CB-SG	-5.12	1.73	1.81
1	A	107	VAL	CB-CG1	-5.05	1.42	1.52

All (26) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	49	ARG	NE-CZ-NH2	-23.45	108.57	120.30
1	A	195	ARG	NE-CZ-NH2	-16.39	112.10	120.30
1	A	49	ARG	NE-CZ-NH1	13.74	127.17	120.30
1	A	195	ARG	NE-CZ-NH1	10.38	125.49	120.30
1	A	105	ARG	NE-CZ-NH2	-9.82	115.39	120.30
1	A	104	ARG	NE-CZ-NH2	-7.92	116.34	120.30
1	A	163	ARG	NE-CZ-NH2	-7.65	116.47	120.30
1	A	49	ARG	CD-NE-CZ	7.59	134.23	123.60
1	A	138	VAL	CA-CB-CG1	7.44	122.06	110.90
1	A	49	ARG	CG-CD-NE	-7.00	97.09	111.80
1	A	195	ARG	CB-CG-CD	-6.53	94.63	111.60
1	A	135	ARG	NE-CZ-NH2	-6.52	117.04	120.30
1	A	63	ARG	NE-CZ-NH2	-6.39	117.10	120.30
1	A	163	ARG	NE-CZ-NH1	6.17	123.39	120.30
1	A	63	ARG	CG-CD-NE	-6.17	98.84	111.80
1	A	138	VAL	CG1-CB-CG2	6.13	120.71	110.90
1	A	183	ASP	CB-CG-OD1	-5.95	112.95	118.30
1	A	182	MSE	CG-SE-CE	5.75	111.54	98.90

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	90	ARG	NE-CZ-NH2	-5.64	117.48	120.30
1	A	138	VAL	CB-CA-C	5.55	121.94	111.40
1	A	129	LEU	CB-CG-CD1	-5.44	101.76	111.00
1	A	144	VAL	CG1-CB-CG2	-5.21	102.57	110.90
1	A	184	ASP	CB-CG-OD1	-5.10	113.71	118.30
1	A	91	VAL	CG1-CB-CG2	5.05	118.98	110.90
1	A	36	ARG	NE-CZ-NH1	5.02	122.81	120.30
1	A	203	LEU	CB-CG-CD1	-5.02	102.47	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	184	ASP	Peptide

## 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	1618	0	1569	25	0
2	A	4	0	3	1	0
3	A	88	0	0	4	0
All	All	1710	0	1572	25	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 (25) 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:105:ARG:HB3	3:A:309:HOH:O	1.71	0.87
1:A:40:ILE:HD12	1:A:176:ARG:HD3	1.71	0.71
1:A:116:ASP:HB3	1:A:119:THR:HG23	1.88	0.56
1:A:105:ARG:CB	3:A:309:HOH:O	2.42	0.55
1:A:118:ALA:O	1:A:119:THR:HB	2.07	0.55

Continued on next page...



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:155:LEU:HA	1:A:158:LEU:HD22	1.93	0.51
1:A:49:ARG:HD2	1:A:85:ARG:O	2.13	0.48
1:A:76:ASP:HB3	1:A:173:ASN:ND2	2.31	0.46
1:A:27:ARG:NH2	1:A:65:THR:OG1	2.39	0.46
1:A:113:ASN:HD21	1:A:212:GLY:HA3	1.81	0.46
1:A:27:ARG:HH21	1:A:65:THR:HG1	1.63	0.45
1:A:164:ILE:HD13	1:A:164:ILE:HG21	1.79	0.44
1:A:148:ASP:OD2	2:A:238:ACT:O	2.35	0.44
1:A:36:ARG:HD2	3:A:243:HOH:O	2.17	0.44
1:A:70:PHE:O	1:A:74:ILE:HG13	2.17	0.44
1:A:187:ARG:HH11	1:A:187:ARG:HB3	1.84	0.43
1:A:100:LEU:HD21	1:A:174:GLU:HB3	2.01	0.43
1:A:84:ASN:ND2	3:A:310:HOH:O	2.43	0.42
1:A:83:LEU:HD23	1:A:83:LEU:HA	1.86	0.42
1:A:135:ARG:HD2	1:A:135:ARG:N	2.35	0.42
1:A:118:ALA:O	1:A:119:THR:CB	2.69	0.41
1:A:56:CYS:SG	1:A:66:VAL:HG21	2.60	0.41
1:A:116:ASP:OD1	1:A:119:THR:HG22	2.21	0.41
1:A:174:GLU:O	1:A:178:VAL:HG23	2.21	0.41
1:A:196:ASN:OD1	1:A:218:TYR:CE2	2.75	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	205/239 (86%)	199 (97%)	3 (2%)	3 (2%)	<b>10</b>   <b>8</b>

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	83	LEU

Continued on next page...

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	118	ALA
1	A	119	THR

### 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	167/188 (89%)	155 (93%)	12 (7%)	14 15

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	25	THR
1	A	30	ARG
1	A	64	ASN
1	A	87	VAL
1	A	91	VAL
1	A	138	VAL
1	A	158	LEU
1	A	163	ARG
1	A	182	MSE
1	A	183	ASP
1	A	186	HIS
1	A	231	ARG

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

Mol	Chain	Res	Type
1	A	113	ASN
1	A	225	GLN

### 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](#)

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	ACT	A	238	-	1,3,3	5.86	1 (100%)	0,3,3	0.00	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	238	ACT	CH3-C	5.86	1.56	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
2	A	238	ACT	1	0

## 5.7 Other polymers [i](#)

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 [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	205/239 (85%)	-0.13	6 (2%) 51 49	29, 40, 63, 81	0

All (6) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	118	ALA	5.7
1	A	83	LEU	5.1
1	A	120	GLY	3.3
1	A	25	THR	2.9
1	A	27	ARG	2.8
1	A	231	ARG	2.5

### 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	ACT	A	238	4/4	0.98	0.10	33,35,36,37	0

## 6.5 Other polymers

There are no such residues in this entry.