



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 14, 2020 – 08:02 pm BST

PDB ID : 3AUE  
Title : A simplified BPTI variant with poly His amino acid tag (C5H) at the C-terminus  
Authors : Islam, M.M.; Kato, A.; Khan, M.M.A.; Noguchi, K.; Yohda, M.; Kidokoro, S.I.; Kuroda, Y.  
Deposited on : 2011-02-03  
Resolution : 2.28 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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

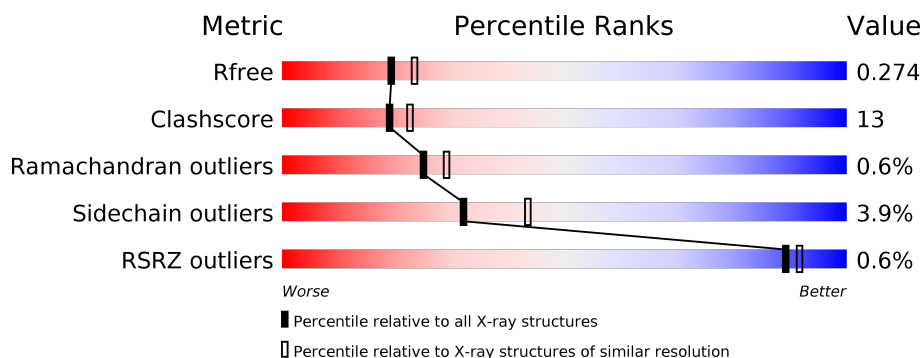
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.28 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	6980 (2.30-2.26)
Clashscore	141614	7711 (2.30-2.26)
Ramachandran outliers	138981	7597 (2.30-2.26)
Sidechain outliers	138945	7598 (2.30-2.26)
RSRZ outliers	127900	6849 (2.30-2.26)

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	65	<div> <div>80%</div> <div>5% • • 12%</div> </div>
1	B	65	<div> <div>72%</div> <div>18% • 8%</div> </div>
1	C	65	<div> <div>3%</div> <div>82%</div> <div>17% •</div> </div>
1	D	65	<div> <div>71%</div> <div>20% 5% • •</div> </div>
1	E	65	<div> <div>2%</div> <div>58%</div> <div>32% 6% •</div> </div>
1	F	65	<div> <div>65%</div> <div>23% 12%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	65	<div><div></div><div>63%23%14%</div></div>
1	H	65	<div><div></div><div>72%23%••</div></div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 3927 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 Bovine pancreatic trypsin inhibitor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	57	Total	C	N	O	S	0	0	0
			417	267	78	70	2			
1	B	60	Total	C	N	O	S	0	0	0
			430	274	81	73	2			
1	C	64	Total	C	N	O	S	0	0	0
			470	298	93	77	2			
1	D	63	Total	C	N	O	S	0	0	0
			460	292	90	76	2			
1	E	63	Total	C	N	O	S	0	0	0
			460	292	90	76	2			
1	F	57	Total	C	N	O	S	0	0	0
			417	267	78	70	2			
1	G	56	Total	C	N	O	S	0	0	0
			412	264	77	69	2			
1	H	63	Total	C	N	O	S	0	1	0
			468	297	93	76	2			

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		
2	C	1	Total	O	S	0	0
			5	4	1		
2	C	1	Total	O	S	0	0
			5	4	1		
2	D	1	Total	O	S	0	0
			5	4	1		
2	D	1	Total	O	S	0	0
			5	4	1		
2	E	1	Total	O	S	0	0
			5	4	1		
2	E	1	Total	O	S	0	0
			5	4	1		
2	F	1	Total	O	S	0	0
			5	4	1		
2	G	1	Total	O	S	0	0
			5	4	1		
2	H	1	Total	O	S	0	0
			5	4	1		


- Molecule 3 is water.

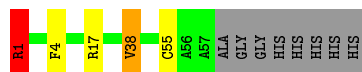
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	33	Total 33	O 33	0	0
3	B	44	Total 44	O 44	0	0
3	C	44	Total 44	O 44	0	0
3	D	40	Total 40	O 40	0	0
3	E	51	Total 51	O 51	0	0
3	F	35	Total 35	O 35	0	0
3	G	45	Total 45	O 45	0	0
3	H	36	Total 36	O 36	0	0

### 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: Bovine pancreatic trypsin inhibitor

Chain A: 




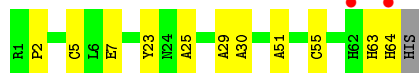
- Molecule 1: Bovine pancreatic trypsin inhibitor

Chain B: 



- Molecule 1: Bovine pancreatic trypsin inhibitor

Chain C: 



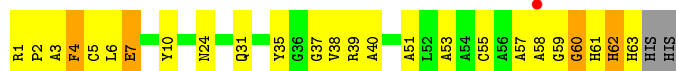
- Molecule 1: Bovine pancreatic trypsin inhibitor

Chain D: 



- Molecule 1: Bovine pancreatic trypsin inhibitor

Chain E: 



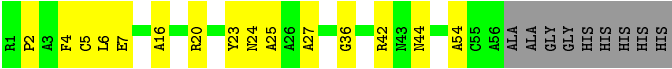
- Molecule 1: Bovine pancreatic trypsin inhibitor

Chain F:  65% 23% 12%



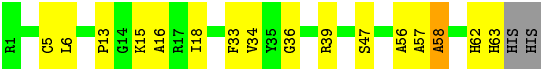
● Molecule 1: Bovine pancreatic trypsin inhibitor

Chain G:  63% 23% 14%



● Molecule 1: Bovine pancreatic trypsin inhibitor

Chain H:  72% 23% . .





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	39.01Å 120.33Å 53.32Å 90.00° 106.55° 90.00°	Depositor
Resolution (Å)	38.95 – 2.28 38.95 – 2.28	Depositor EDS
% Data completeness (in resolution range)	76.3 (38.95-2.28) 76.3 (38.95-2.28)	Depositor EDS
$R_{merge}$	0.23	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.95 (at 2.29Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.163 , 0.274 0.163 , 0.274	Depositor DCC
$R_{free}$ test set	824 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	20.0	Xtriage
Anisotropy	0.885	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 40.6	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	0.97	EDS
Total number of atoms	3927	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	19.0	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 86.47 % 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 9.2954e-08. 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: SO4

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	1.21	1/428 (0.2%)	1.14	4/580 (0.7%)
1	B	1.17	1/441 (0.2%)	1.04	2/597 (0.3%)
1	C	1.24	1/485 (0.2%)	0.97	0/657
1	D	1.37	2/474 (0.4%)	1.19	4/642 (0.6%)
1	E	1.24	2/474 (0.4%)	1.07	1/642 (0.2%)
1	F	1.28	2/428 (0.5%)	1.09	1/580 (0.2%)
1	G	1.14	0/423	0.95	0/573
1	H	1.20	1/485 (0.2%)	1.14	2/656 (0.3%)
All	All	1.24	10/3638 (0.3%)	1.08	14/4927 (0.3%)

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	55	CYS	CB-SG	-7.14	1.70	1.82
1	D	7	GLU	CG-CD	6.81	1.62	1.51
1	E	10	TYR	CD1-CE1	5.78	1.48	1.39
1	H	5	CYS	CB-SG	-5.71	1.72	1.81
1	E	4	PHE	CE2-CZ	5.33	1.47	1.37

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	17	ARG	NE-CZ-NH2	-7.26	116.67	120.30
1	D	62	HIS	C-N-CA	6.57	138.13	121.70
1	A	1	ARG	NE-CZ-NH2	6.36	123.48	120.30
1	H	39	ARG	NE-CZ-NH2	-5.94	117.33	120.30
1	D	42	ARG	NE-CZ-NH2	-5.92	117.34	120.30

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	417	0	409	1	0
1	B	430	0	420	7	0
1	C	470	0	448	10	0
1	D	460	0	441	19	0
1	E	460	0	441	25	0
1	F	417	0	409	8	0
1	G	412	0	404	9	0
1	H	468	0	454	24	0
2	A	5	0	0	0	0
2	B	15	0	0	0	0
2	C	10	0	0	0	0
2	D	10	0	0	0	0
2	E	10	0	0	1	0
2	F	5	0	0	0	0
2	G	5	0	0	0	0
2	H	5	0	0	0	0
3	A	33	0	0	0	3
3	B	44	0	0	0	6
3	C	44	0	0	0	0
3	D	40	0	0	5	2
3	E	51	0	0	5	5
3	F	35	0	0	0	1
3	G	45	0	0	3	3
3	H	36	0	0	2	4
All	All	3927	0	3426	91	12

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 13.

The worst 5 of 91 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:E:4:PHE:HA	3:E:68:HOH:O	1.53	1.08
1:D:52:LEU:HB2	3:D:330:HOH:O	1.57	1.04
1:H:56:ALA:C	1:H:58:ALA:HB2	1.79	1.03

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:H:56:ALA:O	1:H:58:ALA:HB2	1.66	0.96
1:H:57:ALA:N	1:H:58:ALA:HB2	1.81	0.94

The worst 5 of 12 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)
3:B:262:HOH:O	3:E:536:HOH:O[2_545]	1.35	0.85
3:B:538:HOH:O	3:E:610:HOH:O[2_545]	1.46	0.74
3:G:404:HOH:O	3:H:537:HOH:O[2_546]	1.47	0.73
3:B:76:HOH:O	3:G:274:HOH:O[1_554]	1.53	0.67
3:A:436:HOH:O	3:H:271:HOH:O[1_655]	1.56	0.64

## 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	55/65 (85%)	54 (98%)	1 (2%)	0	100	100
1	B	58/65 (89%)	56 (97%)	2 (3%)	0	100	100
1	C	62/65 (95%)	59 (95%)	3 (5%)	0	100	100
1	D	61/65 (94%)	57 (93%)	4 (7%)	0	100	100
1	E	61/65 (94%)	49 (80%)	10 (16%)	2 (3%)	4	2
1	F	55/65 (85%)	53 (96%)	2 (4%)	0	100	100
1	G	54/65 (83%)	53 (98%)	1 (2%)	0	100	100
1	H	62/65 (95%)	58 (94%)	3 (5%)	1 (2%)	9	8
All	All	468/520 (90%)	439 (94%)	26 (6%)	3 (1%)	25	29

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	H	58	ALA
1	E	60	GLY
1	E	62	HIS

### 5.3.2 Protein sidechains ⓘ

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	34/39 (87%)	31 (91%)	3 (9%)	10	11
1	B	34/39 (87%)	30 (88%)	4 (12%)	5	5
1	C	38/39 (97%)	38 (100%)	0	100	100
1	D	37/39 (95%)	35 (95%)	2 (5%)	22	28
1	E	37/39 (95%)	36 (97%)	1 (3%)	44	59
1	F	34/39 (87%)	34 (100%)	0	100	100
1	G	34/39 (87%)	33 (97%)	1 (3%)	42	56
1	H	38/39 (97%)	38 (100%)	0	100	100
All	All	286/312 (92%)	275 (96%)	11 (4%)	32	44

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	6	LEU
1	B	38	VAL
1	D	63	HIS
1	B	1	ARG
1	D	1	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	61	HIS
1	H	61	HIS
1	E	31	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	C	64	HIS
1	D	63	HIS

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

13 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	SO4	D	67	-	4,4,4	0.19	0	6,6,6	0.40	0
2	SO4	H	66	-	4,4,4	0.22	0	6,6,6	0.44	0
2	SO4	G	66	-	4,4,4	0.24	0	6,6,6	0.83	0
2	SO4	E	67	-	4,4,4	0.26	0	6,6,6	0.72	0
2	SO4	C	67	-	4,4,4	0.17	0	6,6,6	1.06	1 (16%)
2	SO4	D	66	-	4,4,4	0.30	0	6,6,6	0.46	0
2	SO4	C	66	-	4,4,4	0.13	0	6,6,6	0.53	0
2	SO4	F	66	-	4,4,4	0.41	0	6,6,6	0.73	0
2	SO4	E	66	-	4,4,4	0.12	0	6,6,6	0.25	0
2	SO4	B	66	-	4,4,4	0.29	0	6,6,6	0.67	0
2	SO4	A	66	-	4,4,4	0.15	0	6,6,6	0.69	0
2	SO4	B	68	-	4,4,4	0.19	0	6,6,6	0.22	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
2	SO4	B	67	-	4,4,4	0.47	0	6,6,6	0.66	0

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
2	C	67	SO4	O4-S-O2	-2.05	98.62	109.31

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	E	67	SO4	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 [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	57/65 (87%)	-0.63	0 100 100	9, 16, 28, 33	0
1	B	60/65 (92%)	-0.67	0 100 100	9, 15, 25, 32	0
1	C	64/65 (98%)	-0.47	2 (3%) 49 54	10, 16, 31, 48	0
1	D	63/65 (96%)	-0.45	0 100 100	10, 18, 28, 43	0
1	E	63/65 (96%)	-0.42	1 (1%) 72 77	11, 19, 42, 47	0
1	F	57/65 (87%)	-0.48	0 100 100	10, 18, 29, 39	0
1	G	56/65 (86%)	-0.65	0 100 100	11, 18, 27, 37	0
1	H	63/65 (96%)	-0.46	0 100 100	11, 18, 32, 48	0
All	All	483/520 (92%)	-0.53	3 (0%) 89 91	9, 18, 31, 48	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	64	HIS	3.0
1	C	62	HIS	2.4
1	E	58	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. 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	SO4	H	66	5/5	0.88	0.25	78,80,80,80	0
2	SO4	D	66	5/5	0.92	0.13	66,66,68,69	0
2	SO4	C	67	5/5	0.95	0.13	41,42,45,49	0
2	SO4	D	67	5/5	0.95	0.12	67,68,70,70	0
2	SO4	F	66	5/5	0.95	0.15	54,56,57,57	0
2	SO4	B	68	5/5	0.95	0.11	68,69,70,70	0
2	SO4	E	66	5/5	0.96	0.14	71,71,72,72	0
2	SO4	C	66	5/5	0.98	0.10	22,23,25,28	0
2	SO4	G	66	5/5	0.99	0.08	31,33,35,36	0
2	SO4	B	66	5/5	0.99	0.09	19,22,25,28	0
2	SO4	A	66	5/5	0.99	0.08	28,31,33,34	0
2	SO4	E	67	5/5	0.99	0.16	34,34,37,39	0
2	SO4	B	67	5/5	0.99	0.09	23,25,29,30	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.