



# Full wwPDB X-ray Structure Validation Report ⓘ

May 3, 2017 – 10:49 AM EDT

PDB ID : 5G4E  
Title : S. enterica HisA mutant D10G, Dup13-15, Q24L, G102A, V106L  
Authors : Soderholm, A.; Guo, X.; Newton, M.; Nasvall, J.; Duarte, F.; Andersson, D.;  
Patrick, W.; Selmer, M.  
Deposited on : 2016-05-12  
Resolution : 2.65 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 : rb-20029077  
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) : rb-20029077

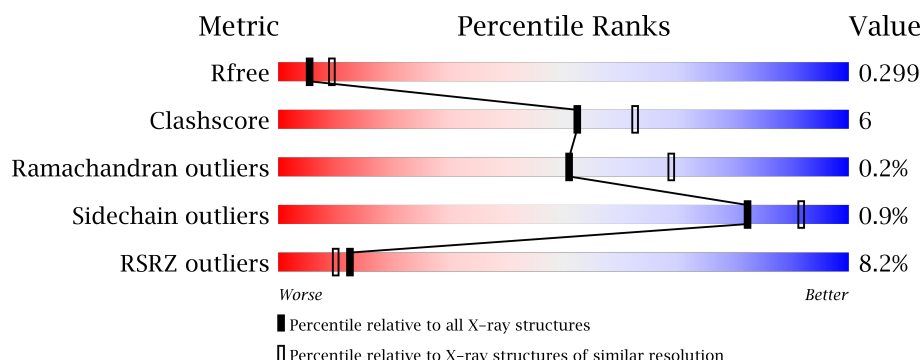
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.65 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	3491 (2.70-2.62)
Clashscore	112137	1026 (2.68-2.64)
Ramachandran outliers	110173	1010 (2.68-2.64)
Sidechain outliers	110143	1010 (2.68-2.64)
RSRZ outliers	101464	3511 (2.70-2.62)

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	256	<div> <div>8%</div> <div>76%</div> <div>10%</div> <div>14%</div> </div>
1	B	256	<div> <div>6%</div> <div>68%</div> <div>12%</div> <div>20%</div> </div>

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 3230 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	220	Total	C	N	O	S	0	0	0
			1639	1044	290	301	4			
1	B	206	Total	C	N	O	S	0	0	0
			1537	980	268	285	4			

There are 30 discrepancies between the modelled and reference sequences:

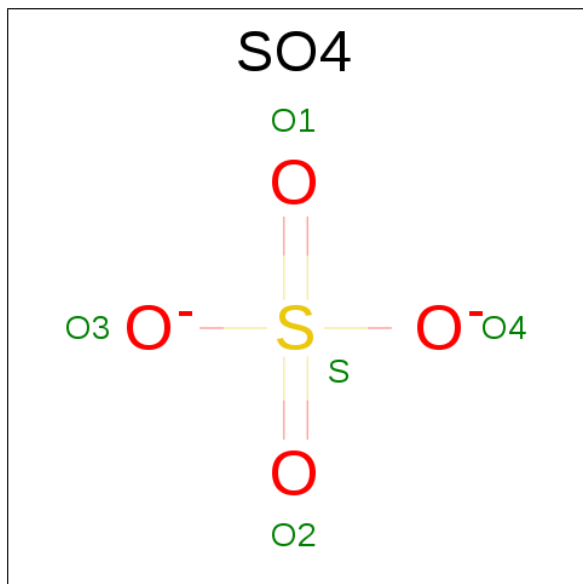
Chain	Residue	Modelled	Actual	Comment	Reference
A	246	LYS	-	expression tag	UNP V7IJE3
A	247	GLY	-	expression tag	UNP V7IJE3
A	248	HIS	-	expression tag	UNP V7IJE3
A	249	HIS	-	expression tag	UNP V7IJE3
A	250	HIS	-	expression tag	UNP V7IJE3
A	251	HIS	-	expression tag	UNP V7IJE3
A	252	HIS	-	expression tag	UNP V7IJE3
A	253	HIS	-	expression tag	UNP V7IJE3
A	10	GLY	ASP	engineered mutation	UNP V7IJE3
A	15A	VAL	-	insertion	UNP V7IJE3
A	15B	VAL	-	insertion	UNP V7IJE3
A	15C	ARG	-	insertion	UNP V7IJE3
A	24	LEU	GLN	engineered mutation	UNP V7IJE3
A	102	ALA	GLY	engineered mutation	UNP V7IJE3
A	106	LEU	VAL	engineered mutation	UNP V7IJE3
B	246	LYS	-	expression tag	UNP V7IJE3
B	247	GLY	-	expression tag	UNP V7IJE3
B	248	HIS	-	expression tag	UNP V7IJE3
B	249	HIS	-	expression tag	UNP V7IJE3
B	250	HIS	-	expression tag	UNP V7IJE3
B	251	HIS	-	expression tag	UNP V7IJE3
B	252	HIS	-	expression tag	UNP V7IJE3
B	253	HIS	-	expression tag	UNP V7IJE3
B	10	GLY	ASP	engineered mutation	UNP V7IJE3
B	15A	VAL	-	insertion	UNP V7IJE3

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	15B	VAL	-	insertion	UNP V7IJE3
B	15C	ARG	-	insertion	UNP V7IJE3
B	24	LEU	GLN	engineered mutation	UNP V7IJE3
B	102	ALA	GLY	engineered mutation	UNP V7IJE3
B	106	LEU	VAL	engineered mutation	UNP V7IJE3

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

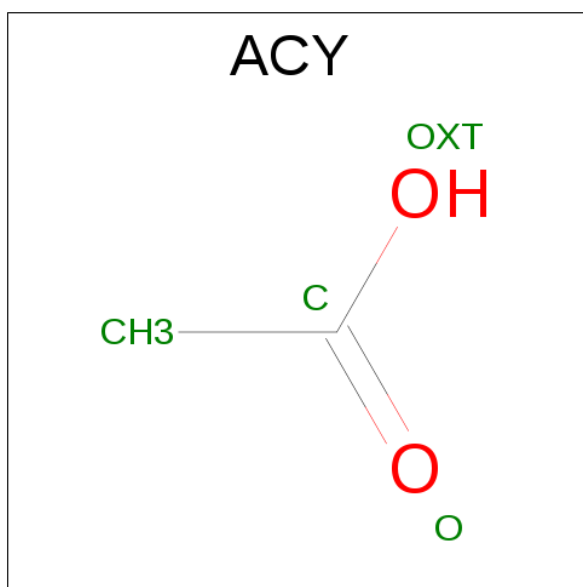


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total O S 5 4 1	0	0
2	B	1	Total O S 5 4 1	0	0

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

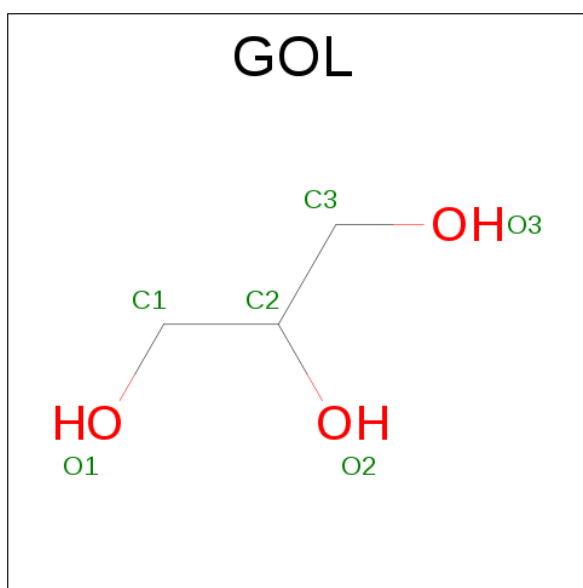
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	2	Total Na 2 2	0	0
3	A	2	Total Na 2 2	0	0

- Molecule 4 is ACETIC ACID (three-letter code: ACY) (formula: C<sub>2</sub>H<sub>4</sub>O<sub>2</sub>).



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

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	B	1	Total	C	O	0	0
			6	3	3		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	18	Total 18	O 18	0	0
6	B	8	Total 8	O 8	0	0



- Molecule 1:



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	48.82Å 88.83Å 117.78Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.10 – 2.65 45.10 – 2.65	Depositor EDS
% Data completeness (in resolution range)	98.5 (45.10-2.65) 98.9 (45.10-2.65)	Depositor EDS
$R_{merge}$	0.17	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.85 (at 2.65Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.240 , 0.299 0.244 , 0.299	Depositor DCC
$R_{free}$ test set	789 reflections (5.16%)	DCC
Wilson B-factor (Å <sup>2</sup> )	34.9	Xtriage
Anisotropy	1.139	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 46.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	3230	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	45.0	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 30.35 % 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 1.3272e-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: GOL, ACY, SO4, 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	0.22	0/1660	0.39	0/2252
1	B	0.27	0/1557	0.51	2/2114 (0.1%)
All	All	0.25	0/3217	0.45	2/4366 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	15	ARG	NE-CZ-NH1	7.16	123.88	120.30
1	B	15	ARG	NE-CZ-NH2	-5.65	117.47	120.30

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	1639	0	1689	16	0
1	B	1537	0	1593	26	0
2	A	5	0	0	0	0
2	B	5	0	0	0	0
3	A	2	0	0	0	0
3	B	2	0	0	0	0
4	A	4	0	3	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	B	4	0	3	0	0
5	B	6	0	8	0	0
6	A	18	0	0	0	0
6	B	8	0	0	0	0
All	All	3230	0	3296	41	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 6.

All (41) 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:B:171:THR:HG22	1:B:202:SER:HB3	1.72	0.72
1:B:15:ARG:HG2	1:B:15:ARG:HH11	1.57	0.70
1:B:60:ARG:NH2	1:B:80:GLY:O	2.22	0.70
1:A:60:ARG:HG2	1:A:62:ILE:HG12	1.76	0.68
1:B:77:GLN:HG2	1:B:98:ARG:HB2	1.75	0.67
1:A:77:GLN:HG2	1:A:98:ARG:HB2	1.78	0.65
1:B:51:LEU:O	1:B:55:LYS:NZ	2.30	0.64
1:B:188:GLU:HG3	1:B:189:GLU:HG2	1.78	0.63
1:A:171:THR:HG22	1:A:202:SER:HB3	1.84	0.60
1:A:60:ARG:NH2	1:A:80:GLY:O	2.35	0.60
1:B:154:GLU:OE1	1:B:194:TYR:OH	2.15	0.56
1:B:60:ARG:HD3	1:B:91:LEU:HD21	1.88	0.56
1:A:83:ARG:HH21	1:A:104:THR:HG21	1.73	0.53
1:A:101:ILE:HD11	1:A:124:LEU:HD22	1.89	0.53
1:B:188:GLU:HG3	1:B:189:GLU:N	2.24	0.52
1:B:15:ARG:HG2	1:B:25:ARG:N	2.24	0.52
1:B:53:GLY:HA2	1:B:59:LYS:O	2.08	0.52
1:A:81:GLY:HA2	1:A:83:ARG:NH1	2.24	0.52
1:B:15:ARG:HD3	1:B:25:ARG:HG2	1.92	0.52
1:A:128:LEU:HD21	1:A:156:LEU:HD13	1.92	0.51
1:B:104:THR:O	1:B:107:LYS:HG3	2.12	0.50
1:B:15:ARG:HB3	1:B:25:ARG:HG2	1.95	0.49
1:A:62:ILE:O	1:A:66:LYS:HG2	2.12	0.48
1:A:85:GLU:OE2	1:A:118:ARG:NH1	2.41	0.48
1:B:185:SER:HA	1:B:188:GLU:HG2	1.95	0.47
1:B:60:ARG:O	1:B:62:ILE:HG12	2.14	0.47
1:A:25:ARG:HH12	1:B:26:ASP:CG	2.18	0.46
1:B:15:ARG:NH1	1:B:25:ARG:N	2.62	0.46
1:A:9:ILE:HG12	1:A:51:LEU:HB2	1.97	0.46

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:60:ARG:O	1:B:62:ILE:N	2.42	0.45
1:B:61:GLN:HA	1:B:61:GLN:OE1	2.16	0.45
1:B:161:LEU:N	1:B:162:PRO:HD2	2.32	0.45
1:B:25:ARG:NE	1:B:27:TYR:HE2	2.15	0.44
1:A:81:GLY:HA2	1:A:83:ARG:HH11	1.83	0.43
1:B:81:GLY:HA2	1:B:102:ALA:HB2	2.01	0.43
1:B:30:ASP:N	1:B:30:ASP:OD1	2.52	0.42
1:A:226:ARG:NE	1:A:230:GLU:OE2	2.53	0.42
1:A:226:ARG:HE	1:A:230:GLU:CD	2.24	0.42
1:A:32:LEU:HB3	1:A:33:PRO:HD3	2.02	0.41
1:B:32:LEU:HB3	1:B:33:PRO:HD3	2.02	0.41
1:B:15:ARG:CG	1:B:15:ARG:HH11	2.31	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	210/256 (82%)	202 (96%)	8 (4%)	0	100	100
1	B	198/256 (77%)	188 (95%)	9 (4%)	1 (0%)	32	49
All	All	408/512 (80%)	390 (96%)	17 (4%)	1 (0%)	51	69

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	53	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	170/202 (84%)	169 (99%)	1 (1%)	89	95
1	B	162/202 (80%)	160 (99%)	2 (1%)	75	89
All	All	332/404 (82%)	329 (99%)	3 (1%)	82	91

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	84	THR
1	B	15	ARG
1	B	56	ASP

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 9 ligands modelled in this entry, 4 are monoatomic - leaving 5 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	SO4	A	1245	-	4,4,4	0.16	0	6,6,6	0.07	0
4	ACY	A	1248	-	1,3,3	1.40	0	0,3,3	0.00	-
4	ACY	B	1245	-	1,3,3	1.35	0	0,3,3	0.00	-
2	SO4	B	1246	-	4,4,4	0.15	0	6,6,6	0.06	0
5	GOL	B	1247	-	5,5,5	0.34	0	5,5,5	0.20	0

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	SO4	A	1245	-	-	0/0/0/0	0/0/0/0
4	ACY	A	1248	-	-	0/0/0/0	0/0/0/0
4	ACY	B	1245	-	-	0/0/0/0	0/0/0/0
2	SO4	B	1246	-	-	0/0/0/0	0/0/0/0
5	GOL	B	1247	-	-	0/4/4/4	0/0/0/0

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	220/256 (85%)	0.65	20 (9%) 10 8	26, 40, 73, 95	0
1	B	206/256 (80%)	0.72	15 (7%) 16 13	25, 40, 73, 89	0
All	All	426/512 (83%)	0.68	35 (8%) 12 10	25, 40, 73, 95	0

All (35) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	25	ARG	6.2
1	B	182	SER	4.6
1	A	105	ALA	3.8
1	A	21	TYR	3.6
1	A	15(A)	VAL	3.6
1	B	204	GLY	3.4
1	B	15	ARG	3.3
1	B	183	ASN	3.0
1	B	14	VAL	3.0
1	A	213	ALA	3.0
1	B	58	ALA	3.0
1	B	213	ALA	2.9
1	A	58	ALA	2.9
1	B	57	PRO	2.9
1	B	185	SER	2.8
1	A	16	LEU	2.7
1	A	83	ARG	2.6
1	A	15(C)	ARG	2.6
1	A	141	ALA	2.5
1	A	15(B)	VAL	2.5
1	A	172	ASP	2.5
1	A	22	ALA	2.5
1	B	187	TYR	2.5
1	A	140	VAL	2.5

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	239	ILE	2.4
1	A	17	HIS	2.3
1	A	233	PHE	2.3
1	A	57	PRO	2.2
1	A	14	VAL	2.2
1	A	235	VAL	2.1
1	B	233	PHE	2.1
1	B	28	GLY	2.1
1	B	189	GLU	2.0
1	A	56	ASP	2.0
1	B	51	LEU	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	NA	B	1248	1/1	0.87	0.23	0.94	43,43,43,43	0
2	SO4	B	1246	5/5	0.80	0.26	0.08	61,64,83,84	0
2	SO4	A	1245	5/5	0.84	0.24	-0.04	45,49,69,71	0
4	ACY	A	1248	4/4	0.90	0.21	-0.21	25,40,43,46	0
5	GOL	B	1247	6/6	0.88	0.20	-0.45	37,40,43,44	0
3	NA	A	1246	1/1	0.85	0.13	-3.81	37,37,37,37	0
3	NA	B	1249	1/1	0.96	0.12	-	34,34,34,34	0
4	ACY	B	1245	4/4	0.81	0.26	-	36,41,41,41	0
3	NA	A	1247	1/1	0.92	0.15	-	28,28,28,28	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.