



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 13, 2020 – 06:30 PM BST

PDB ID : 4X2S  
Title : Crystal structure of 276S/M395R-GltPh in inward-facing conformation  
Authors : Akyuz, N.; Boudker, O.  
Deposited on : 2014-11-26  
Resolution : 4.21 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 : 2.14.4.dev1  
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) : 2.14.4.dev1

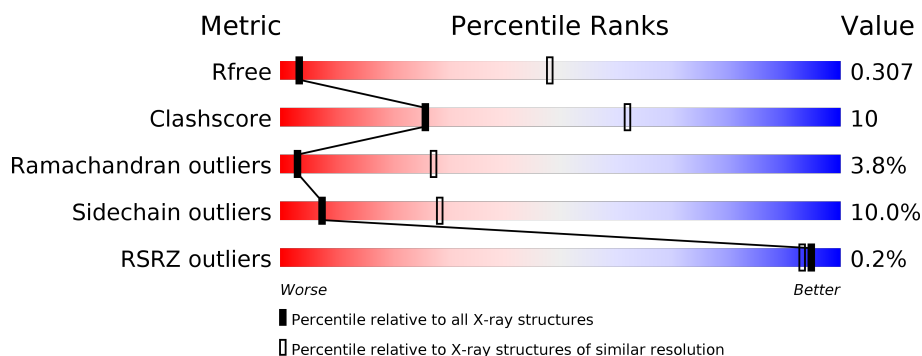
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 4.21 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	1008 (4.66-3.78)
Clashscore	141614	1047 (4.62-3.80)
Ramachandran outliers	138981	1003 (4.62-3.80)
Sidechain outliers	138945	1010 (4.66-3.78)
RSRZ outliers	127900	1064 (4.72-3.70)

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	422	
1	B	422	
1	C	422	

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9097 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 425aa long hypothetical proton glutamate symport protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	405	Total	C	N	O	S	0	0	0
			2972	1956	479	522	15			
1	B	411	Total	C	N	O	S	0	0	0
			3033	1999	487	532	15			
1	C	415	Total	C	N	O	S	0	0	0
			3059	2016	491	537	15			

There are 42 discrepancies between the modelled and reference sequences:

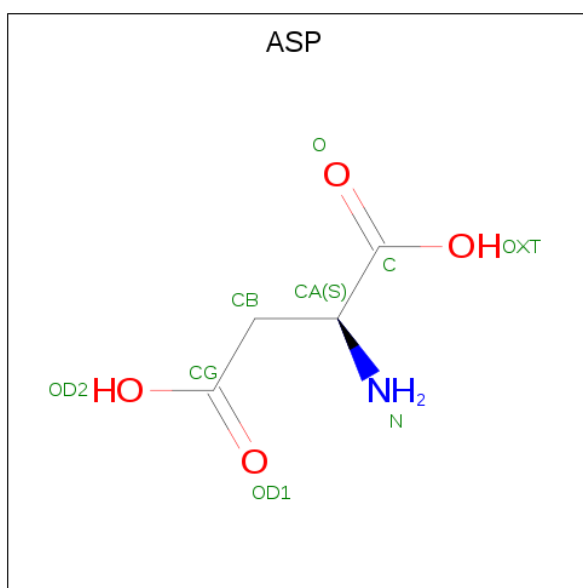
Chain	Residue	Modelled	Actual	Comment	Reference
A	37	HIS	ASP	engineered mutation	UNP O59010
A	40	HIS	LYS	engineered mutation	UNP O59010
A	125	HIS	LYS	engineered mutation	UNP O59010
A	132	HIS	LYS	engineered mutation	UNP O59010
A	223	HIS	LYS	engineered mutation	UNP O59010
A	264	HIS	LYS	engineered mutation	UNP O59010
A	276	SER	ARG	engineered mutation	UNP O59010
A	321	ALA	CYS	engineered mutation	UNP O59010
A	368	HIS	GLU	engineered mutation	UNP O59010
A	395	ARG	MET	engineered mutation	UNP O59010
A	418	THR	GLU	engineered mutation	UNP O59010
A	420	VAL	-	expression tag	UNP O59010
A	421	PRO	-	expression tag	UNP O59010
A	422	ARG	-	expression tag	UNP O59010
B	37	HIS	ASP	engineered mutation	UNP O59010
B	40	HIS	LYS	engineered mutation	UNP O59010
B	125	HIS	LYS	engineered mutation	UNP O59010
B	132	HIS	LYS	engineered mutation	UNP O59010
B	223	HIS	LYS	engineered mutation	UNP O59010
B	264	HIS	LYS	engineered mutation	UNP O59010
B	276	SER	ARG	engineered mutation	UNP O59010
B	321	ALA	CYS	engineered mutation	UNP O59010
B	368	HIS	GLU	engineered mutation	UNP O59010

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	395	ARG	MET	engineered mutation	UNP O59010
B	418	THR	GLU	engineered mutation	UNP O59010
B	420	VAL	-	expression tag	UNP O59010
B	421	PRO	-	expression tag	UNP O59010
B	422	ARG	-	expression tag	UNP O59010
C	37	HIS	ASP	engineered mutation	UNP O59010
C	40	HIS	LYS	engineered mutation	UNP O59010
C	125	HIS	LYS	engineered mutation	UNP O59010
C	132	HIS	LYS	engineered mutation	UNP O59010
C	223	HIS	LYS	engineered mutation	UNP O59010
C	264	HIS	LYS	engineered mutation	UNP O59010
C	276	SER	ARG	engineered mutation	UNP O59010
C	321	ALA	CYS	engineered mutation	UNP O59010
C	368	HIS	GLU	engineered mutation	UNP O59010
C	395	ARG	MET	engineered mutation	UNP O59010
C	418	THR	GLU	engineered mutation	UNP O59010
C	420	VAL	-	expression tag	UNP O59010
C	421	PRO	-	expression tag	UNP O59010
C	422	ARG	-	expression tag	UNP O59010

- Molecule 2 is ASPARTIC ACID (three-letter code: ASP) (formula:  $C_4H_7NO_4$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			9	4	1	4		
2	B	1	Total	C	N	O	0	0
			9	4	1	4		

Continued on next page...

*Continued from previous page...*

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	C	1	Total	C	N	O	0	0
			9	4	1	4		

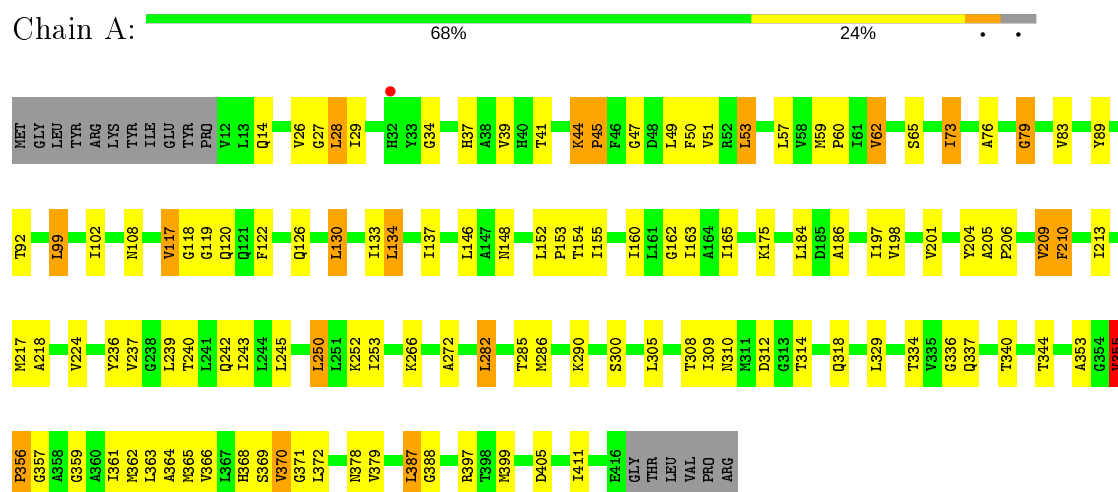
- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

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

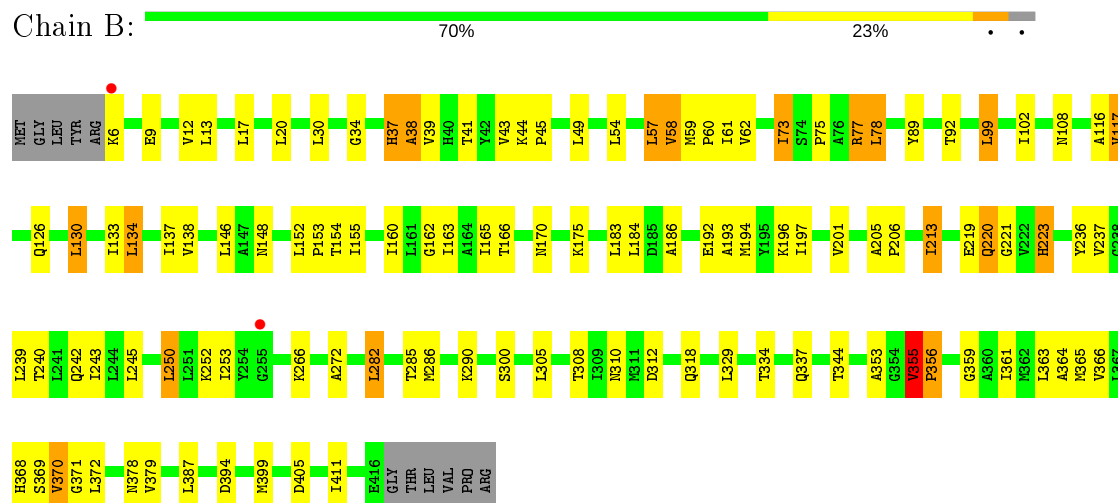
### 3 Residue-property plots [i](#)

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: 425aa long hypothetical proton glutamate symport protein

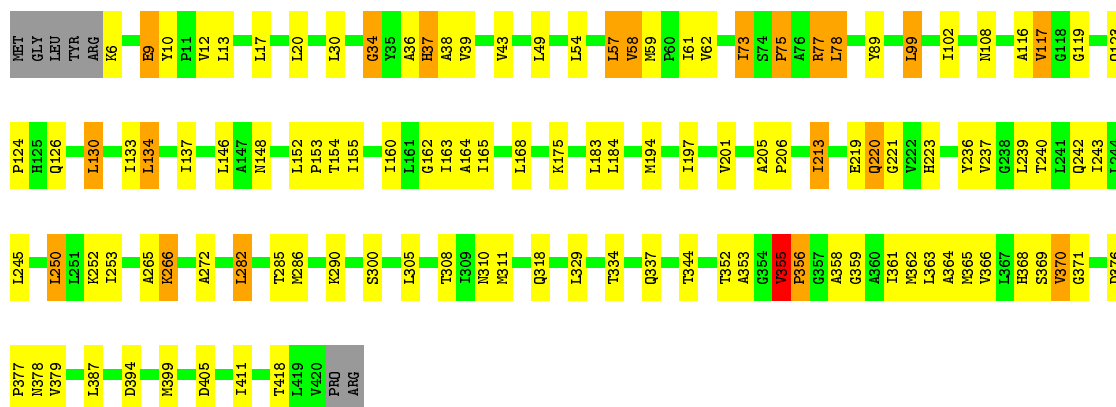


- Molecule 1: 425aa long hypothetical proton glutamate symport protein



- Molecule 1: 425aa long hypothetical proton glutamate symport protein





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	121.65Å 121.65Å 578.70Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	20.00 – 4.21 19.97 – 4.21	Depositor EDS
% Data completeness (in resolution range)	83.2 (20.00-4.21) 84.3 (19.97-4.21)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.60 (at 4.21Å)	Xtriage
Refinement program	REFMAC 5.8.0069	Depositor
R, $R_{free}$	0.278 , 0.314 0.271 , 0.307	Depositor DCC
$R_{free}$ test set	824 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	132.2	Xtriage
Anisotropy	0.501	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.23 , 114.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	9097	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	199.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.22% 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	0.35	0/3028	0.54	0/4134
1	B	0.37	0/3092	0.54	0/4220
1	C	0.41	0/3118	0.56	0/4256
All	All	0.38	0/9238	0.55	0/12610

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	3
1	B	0	4
1	C	0	4
All	All	0	11

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 11 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	210	PHE	Peptide
1	A	49	LEU	Peptide
1	A	79	GLY	Peptide
1	B	213	ILE	Peptide
1	B	75	PRO	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	2972	0	3136	75	0
1	B	3033	0	3203	64	0
1	C	3059	0	3232	66	0
2	A	9	0	3	1	0
2	B	9	0	3	2	0
2	C	9	0	3	2	0
3	A	2	0	0	0	0
3	B	2	0	0	0	0
3	C	2	0	0	0	0
All	All	9097	0	9580	196	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 10.

The worst 5 of 196 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:205:ALA:O	1:A:209:VAL:HG12	1.76	0.86
1:A:210:PHE:HA	1:A:213:ILE:HB	1.63	0.79
1:B:250:LEU:HD13	1:B:411:ILE:HD11	1.69	0.75
1:C:250:LEU:HD13	1:C:411:ILE:HD11	1.69	0.75
1:A:250:LEU:HD13	1:A:411:ILE:HD11	1.68	0.75

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	403/422 (96%)	341 (85%)	50 (12%)	12 (3%)	4	32
1	B	409/422 (97%)	352 (86%)	41 (10%)	16 (4%)	3	27
1	C	413/422 (98%)	351 (85%)	44 (11%)	18 (4%)	2	25
All	All	1225/1266 (97%)	1044 (85%)	135 (11%)	46 (4%)	3	27

5 of 46 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	117	VAL
1	A	266	LYS
1	A	355	VAL
1	A	371	GLY
1	B	37	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	310/329 (94%)	281 (91%)	29 (9%)	8	30
1	B	317/329 (96%)	284 (90%)	33 (10%)	7	27
1	C	320/329 (97%)	287 (90%)	33 (10%)	7	28
All	All	947/987 (96%)	852 (90%)	95 (10%)	7	28

5 of 95 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	126	GLN
1	B	300	SER
1	C	305	LEU
1	B	134	LEU
1	B	242	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	108	ASN
1	B	170	ASN
1	C	170	ASN
1	A	368	HIS
1	C	318	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](#)

Of 9 ligands modelled in this entry, 6 are 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	$\# Z  > 2$
2	ASP	A	501	-	2,8,8	0.34	0	1,10,10	0.06	0
2	ASP	C	501	-	2,8,8	0.16	0	1,10,10	0.45	0
2	ASP	B	501	-	2,8,8	0.35	0	1,10,10	0.29	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	ASP	A	501	-	-	0/2/8/8	-
2	ASP	C	501	-	-	0/2/8/8	-
2	ASP	B	501	-	-	0/2/8/8	-

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.

3 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	501	ASP	1	0
2	C	501	ASP	2	0
2	B	501	ASP	2	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	405/422 (95%)	-0.54	1 (0%) 95 93	118, 222, 300, 356	0
1	B	411/422 (97%)	-0.53	2 (0%) 91 86	89, 200, 312, 378	0
1	C	415/422 (98%)	-0.62	0 100 100	71, 157, 297, 410	0
All	All	1231/1266 (97%)	-0.56	3 (0%) 95 93	71, 194, 306, 410	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	32	HIS	2.7
1	B	255	GLY	2.6
1	B	6	LYS	2.4

### 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
3	NA	B	502	1/1	0.89	0.27	162,162,162,162	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
3	NA	A	502	1/1	0.90	0.29	153,153,153,153	0
2	ASP	A	501	9/9	0.94	0.16	185,200,218,220	0
3	NA	A	503	1/1	0.94	0.24	172,172,172,172	0
3	NA	C	502	1/1	0.95	0.23	124,124,124,124	0
2	ASP	B	501	9/9	0.95	0.23	153,164,180,186	0
3	NA	C	503	1/1	0.96	0.35	85,85,85,85	0
2	ASP	C	501	9/9	0.97	0.22	108,120,130,133	0
3	NA	B	503	1/1	0.98	0.24	140,140,140,140	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.