



# wwPDB X-ray Structure Validation Summary Report

Jun 16, 2024 – 11:32 AM EDT

PDB ID : 2B5F  
Title : Crystal structure of the spinach aquaporin SoPIP2;1 in an open conformation to 3.9 resolution  
Authors : Tornroth-Horsefield, S.; Wang, Y.; Hedfalk, K.; Johanson, U.; Karlsson, M.; Tajkhorshid, E.; Neutze, R.; Kjellbom, P.  
Deposited on : 2005-09-28  
Resolution : 3.90 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

The following versions of software and data (see [references](#) ) were used in the production of this report:

MolProbity : 4.02b-467  
Xtriage (Phenix) : 1.20.1  
EDS : 2.37.1  
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) : 2.37.1

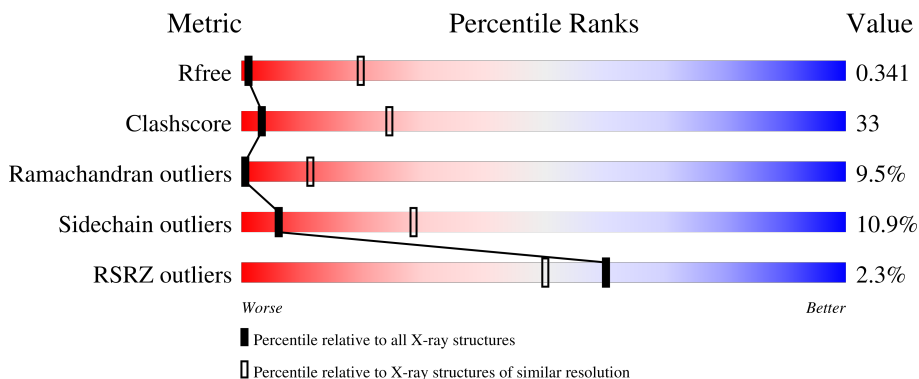
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.90 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	1002 (4.14-3.66)
Clashscore	141614	1004 (4.12-3.68)
Ramachandran outliers	138981	1021 (4.14-3.66)
Sidechain outliers	138945	1014 (4.14-3.66)
RSRZ outliers	127900	1275 (4.20-3.60)

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 of 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	303	
1	B	303	
1	C	303	
1	D	303	

## 2 Entry composition i

There is only 1 type of molecule in this entry. The entry contains 6854 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 aquaporin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	236	Total 1767	C 1175	N 288	O 296	S 8	0	0	0
1	B	224	Total 1671	C 1116	N 267	O 280	S 8	0	0	0
1	C	227	Total 1696	C 1134	N 271	O 283	S 8	0	0	0
1	D	230	Total 1720	C 1150	N 276	O 286	S 8	0	0	0

There are 88 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	282	LEU	-	EXPRESSION TAG	UNP Q41372
A	283	GLU	-	EXPRESSION TAG	UNP Q41372
A	284	GLN	-	EXPRESSION TAG	UNP Q41372
A	285	LYS	-	EXPRESSION TAG	UNP Q41372
A	286	LEU	-	EXPRESSION TAG	UNP Q41372
A	287	ILE	-	EXPRESSION TAG	UNP Q41372
A	288	SER	-	EXPRESSION TAG	UNP Q41372
A	289	GLU	-	EXPRESSION TAG	UNP Q41372
A	290	GLU	-	EXPRESSION TAG	UNP Q41372
A	291	ASP	-	EXPRESSION TAG	UNP Q41372
A	292	LEU	-	EXPRESSION TAG	UNP Q41372
A	293	ASN	-	EXPRESSION TAG	UNP Q41372
A	294	SER	-	EXPRESSION TAG	UNP Q41372
A	295	ALA	-	EXPRESSION TAG	UNP Q41372
A	296	VAL	-	EXPRESSION TAG	UNP Q41372
A	297	ASP	-	EXPRESSION TAG	UNP Q41372
A	298	HIS	-	EXPRESSION TAG	UNP Q41372
A	299	HIS	-	EXPRESSION TAG	UNP Q41372
A	300	HIS	-	EXPRESSION TAG	UNP Q41372
A	301	HIS	-	EXPRESSION TAG	UNP Q41372
A	302	HIS	-	EXPRESSION TAG	UNP Q41372

*Continued on next page...*

*Continued from previous page...*

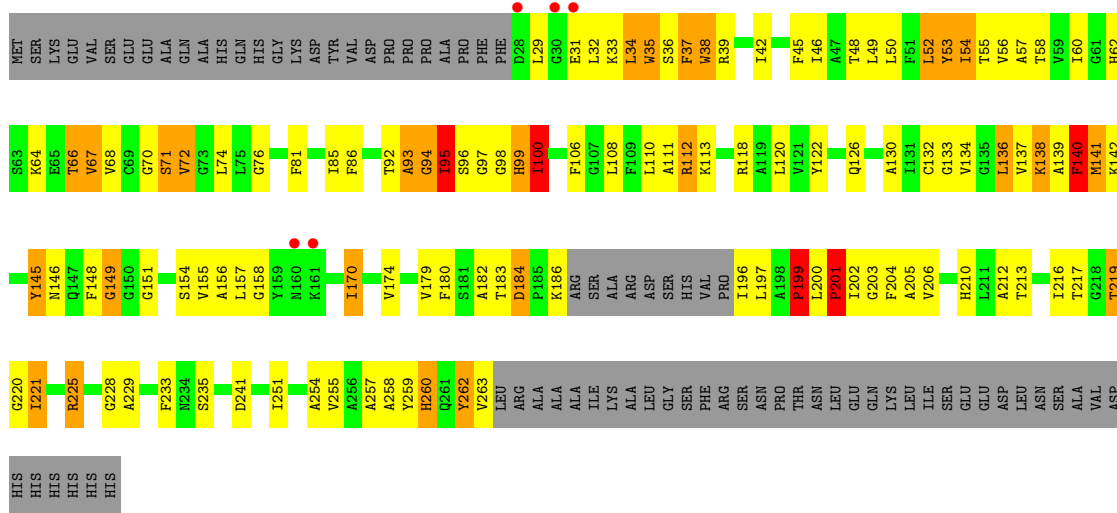
Chain	Residue	Modelled	Actual	Comment	Reference
A	303	HIS	-	EXPRESSION TAG	UNP Q41372
B	282	LEU	-	EXPRESSION TAG	UNP Q41372
B	283	GLU	-	EXPRESSION TAG	UNP Q41372
B	284	GLN	-	EXPRESSION TAG	UNP Q41372
B	285	LYS	-	EXPRESSION TAG	UNP Q41372
B	286	LEU	-	EXPRESSION TAG	UNP Q41372
B	287	ILE	-	EXPRESSION TAG	UNP Q41372
B	288	SER	-	EXPRESSION TAG	UNP Q41372
B	289	GLU	-	EXPRESSION TAG	UNP Q41372
B	290	GLU	-	EXPRESSION TAG	UNP Q41372
B	291	ASP	-	EXPRESSION TAG	UNP Q41372
B	292	LEU	-	EXPRESSION TAG	UNP Q41372
B	293	ASN	-	EXPRESSION TAG	UNP Q41372
B	294	SER	-	EXPRESSION TAG	UNP Q41372
B	295	ALA	-	EXPRESSION TAG	UNP Q41372
B	296	VAL	-	EXPRESSION TAG	UNP Q41372
B	297	ASP	-	EXPRESSION TAG	UNP Q41372
B	298	HIS	-	EXPRESSION TAG	UNP Q41372
B	299	HIS	-	EXPRESSION TAG	UNP Q41372
B	300	HIS	-	EXPRESSION TAG	UNP Q41372
B	301	HIS	-	EXPRESSION TAG	UNP Q41372
B	302	HIS	-	EXPRESSION TAG	UNP Q41372
B	303	HIS	-	EXPRESSION TAG	UNP Q41372
C	282	LEU	-	EXPRESSION TAG	UNP Q41372
C	283	GLU	-	EXPRESSION TAG	UNP Q41372
C	284	GLN	-	EXPRESSION TAG	UNP Q41372
C	285	LYS	-	EXPRESSION TAG	UNP Q41372
C	286	LEU	-	EXPRESSION TAG	UNP Q41372
C	287	ILE	-	EXPRESSION TAG	UNP Q41372
C	288	SER	-	EXPRESSION TAG	UNP Q41372
C	289	GLU	-	EXPRESSION TAG	UNP Q41372
C	290	GLU	-	EXPRESSION TAG	UNP Q41372
C	291	ASP	-	EXPRESSION TAG	UNP Q41372
C	292	LEU	-	EXPRESSION TAG	UNP Q41372
C	293	ASN	-	EXPRESSION TAG	UNP Q41372
C	294	SER	-	EXPRESSION TAG	UNP Q41372
C	295	ALA	-	EXPRESSION TAG	UNP Q41372
C	296	VAL	-	EXPRESSION TAG	UNP Q41372
C	297	ASP	-	EXPRESSION TAG	UNP Q41372
C	298	HIS	-	EXPRESSION TAG	UNP Q41372
C	299	HIS	-	EXPRESSION TAG	UNP Q41372
C	300	HIS	-	EXPRESSION TAG	UNP Q41372

*Continued on next page...*

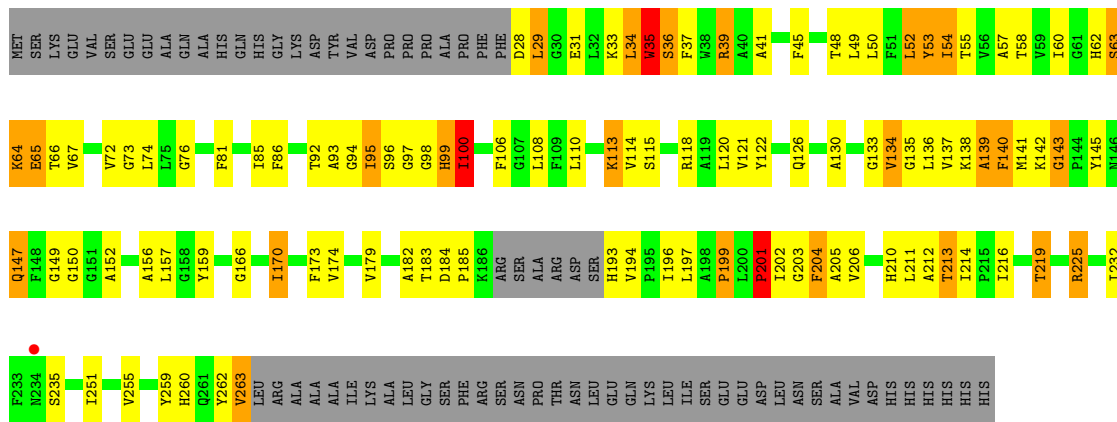
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	301	HIS	-	EXPRESSION TAG	UNP Q41372
C	302	HIS	-	EXPRESSION TAG	UNP Q41372
C	303	HIS	-	EXPRESSION TAG	UNP Q41372
D	282	LEU	-	EXPRESSION TAG	UNP Q41372
D	283	GLU	-	EXPRESSION TAG	UNP Q41372
D	284	GLN	-	EXPRESSION TAG	UNP Q41372
D	285	LYS	-	EXPRESSION TAG	UNP Q41372
D	286	LEU	-	EXPRESSION TAG	UNP Q41372
D	287	ILE	-	EXPRESSION TAG	UNP Q41372
D	288	SER	-	EXPRESSION TAG	UNP Q41372
D	289	GLU	-	EXPRESSION TAG	UNP Q41372
D	290	GLU	-	EXPRESSION TAG	UNP Q41372
D	291	ASP	-	EXPRESSION TAG	UNP Q41372
D	292	LEU	-	EXPRESSION TAG	UNP Q41372
D	293	ASN	-	EXPRESSION TAG	UNP Q41372
D	294	SER	-	EXPRESSION TAG	UNP Q41372
D	295	ALA	-	EXPRESSION TAG	UNP Q41372
D	296	VAL	-	EXPRESSION TAG	UNP Q41372
D	297	ASP	-	EXPRESSION TAG	UNP Q41372
D	298	HIS	-	EXPRESSION TAG	UNP Q41372
D	299	HIS	-	EXPRESSION TAG	UNP Q41372
D	300	HIS	-	EXPRESSION TAG	UNP Q41372
D	301	HIS	-	EXPRESSION TAG	UNP Q41372
D	302	HIS	-	EXPRESSION TAG	UNP Q41372
D	303	HIS	-	EXPRESSION TAG	UNP Q41372





• Molecule 1: aquaporin



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	181.30Å 103.97Å 67.07Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 3.90 32.37 – 3.80	Depositor EDS
% Data completeness (in resolution range)	(Not available) (50.00-3.90) 89.9 (32.37-3.80)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.50 (at 3.75Å)	Xtrriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.290 , 0.332 0.283 , 0.341	Depositor DCC
$R_{free}$ test set	558 reflections (4.75%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	103.8	Xtrriage
Anisotropy	0.488	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.17 , 60.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.27$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.87	EDS
Total number of atoms	6854	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	68.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.73% 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](#)

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	1.11	4/1817 (0.2%)	0.97	6/2479 (0.2%)
1	B	1.12	4/1718 (0.2%)	1.03	7/2344 (0.3%)
1	C	1.12	4/1743 (0.2%)	1.01	6/2377 (0.3%)
1	D	1.12	5/1769 (0.3%)	1.00	7/2414 (0.3%)
All	All	1.12	17/7047 (0.2%)	1.00	26/9614 (0.3%)

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
1	B	0	1
1	C	0	1
1	D	0	1
All	All	0	4

The worst 5 of 17 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	122	TYR	CE2-CZ	-12.34	1.22	1.38
1	C	122	TYR	CE2-CZ	-12.30	1.22	1.38
1	A	122	TYR	CE2-CZ	-12.30	1.22	1.38
1	B	122	TYR	CE2-CZ	-12.27	1.22	1.38
1	D	122	TYR	CG-CD1	-10.79	1.25	1.39

The worst 5 of 26 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	201	PRO	CA-N-CD	-15.52	89.77	111.50
1	D	201	PRO	CA-N-CD	-15.52	89.77	111.50
1	C	201	PRO	CA-N-CD	-15.48	89.82	111.50

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	201	PRO	CA-N-CD	-15.48	89.83	111.50
1	A	122	TYR	CB-CG-CD1	6.54	124.93	121.00

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	99	HIS	Sidechain
1	B	99	HIS	Sidechain
1	C	99	HIS	Sidechain
1	D	99	HIS	Sidechain

## 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	1767	0	1796	126	0
1	B	1671	0	1692	106	0
1	C	1696	0	1727	144	0
1	D	1720	0	1750	121	0
All	All	6854	0	6965	452	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 33.

The worst 5 of 452 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:C:39:ARG:HD3	1:D:263:VAL:HG13	1.42	1.01
1:C:66:THR:HG23	1:C:67:VAL:H	1.28	0.98
1:C:134:VAL:HG11	1:C:229:ALA:HA	1.48	0.95
1:C:202:ILE:O	1:C:203:GLY:C	2.02	0.95
1:A:202:ILE:O	1:A:203:GLY:C	2.02	0.95

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	234/303 (77%)	170 (73%)	42 (18%)	22 (9%)	0	12
1	B	220/303 (73%)	173 (79%)	29 (13%)	18 (8%)	1	14
1	C	223/303 (74%)	175 (78%)	21 (9%)	27 (12%)	0	6
1	D	226/303 (75%)	180 (80%)	27 (12%)	19 (8%)	1	13
All	All	903/1212 (74%)	698 (77%)	119 (13%)	86 (10%)	0	11

5 of 86 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	36	SER
1	A	96	SER
1	A	145	TYR
1	A	187	ARG
1	A	195	PRO

### 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	176/234 (75%)	157 (89%)	19 (11%)	6	29
1	B	165/234 (70%)	147 (89%)	18 (11%)	6	28
1	C	168/234 (72%)	151 (90%)	17 (10%)	7	30
1	D	171/234 (73%)	151 (88%)	20 (12%)	5	26
All	All	680/936 (73%)	606 (89%)	74 (11%)	6	28

5 of 74 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	52	LEU
1	D	260	HIS
1	D	66	THR
1	D	157	LEU
1	B	72	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	153	ASN
1	B	210	HIS
1	D	210	HIS
1	C	210	HIS
1	D	99	HIS

### 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](#)

There are no ligands in this entry.

### 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	236/303 (77%)	-0.02	11 (4%) 31 25	34, 60, 87, 97	236 (100%)
1	B	224/303 (73%)	-0.14	4 (1%) 68 59	41, 66, 89, 103	224 (100%)
1	C	227/303 (74%)	-0.17	5 (2%) 62 51	44, 68, 94, 103	227 (100%)
1	D	230/303 (75%)	-0.16	1 (0%) 92 87	52, 77, 93, 107	230 (100%)
All	All	917/1212 (75%)	-0.12	21 (2%) 60 50	34, 68, 92, 107	917 (100%)

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	28	ASP	4.5
1	A	191	ASP	4.2
1	A	190	ARG	3.9
1	C	31	GLU	3.8
1	A	192	SER	3.4

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.