



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 23, 2020 – 10:38 pm BST

PDB ID : 4BGF  
Title : The 3D-structure of arylamine-N-acetyltransferase from *M. tuberculosis*  
Authors : Abuhammad, A.; Lowe, E.D.; McDonough, M.A.; Shaw Stewart, P.D.; Kolek, S.A.; Sim, E.; Garman, E.F.  
Deposited on : 2013-03-26  
Resolution : 2.10 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  
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

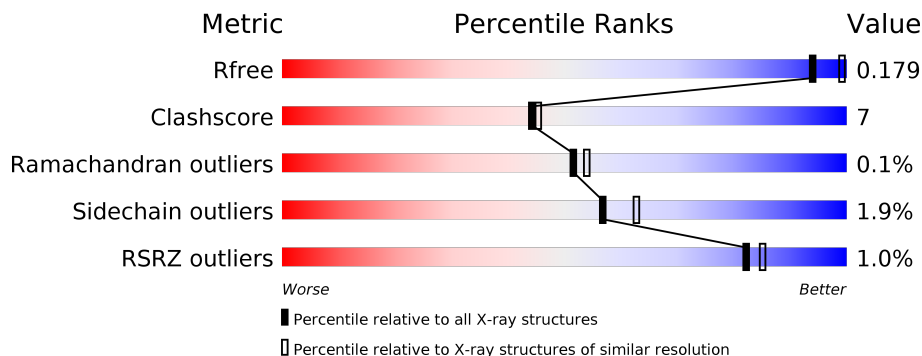
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.10 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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	283	 % 82% 13% 5%
1	B	283	 % 78% 16% 5%
1	C	283	 % 81% 14% 5%
1	D	283	 % 83% 12% 5%
1	E	283	 % 79% 16% 5%
1	F	283	 % 82% 14% 5%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	G	283	 82% 13% 5%
1	H	283	 80% 15% 5%

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 33783 atoms, of which 16242 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 ARYLAMINE N-ACETYLTRANSFERASE NAT.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	270	Total 4095	C 1314	H 2029	N 359	O 389	S 4	0	0	0
1	B	270	Total 4109	C 1316	H 2038	N 362	O 389	S 4	0	0	0
1	C	269	Total 4052	C 1304	H 2000	N 354	O 390	S 4	0	0	0
1	D	270	Total 4117	C 1318	H 2042	N 362	O 391	S 4	0	0	0
1	E	270	Total 4111	C 1316	H 2039	N 363	O 389	S 4	0	0	0
1	F	271	Total 4111	C 1321	H 2039	N 357	O 390	S 4	0	0	0
1	G	270	Total 4088	C 1313	H 2022	N 356	O 393	S 4	0	0	0
1	H	270	Total 4101	C 1316	H 2033	N 359	O 389	S 4	0	0	0

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	122	Total 122	O 122	0	0
2	B	100	Total 100	O 100	0	0
2	C	122	Total 122	O 122	0	0
2	D	142	Total 142	O 142	0	0
2	E	115	Total 115	O 115	0	0
2	F	123	Total 123	O 123	0	0

*Continued on next page...*

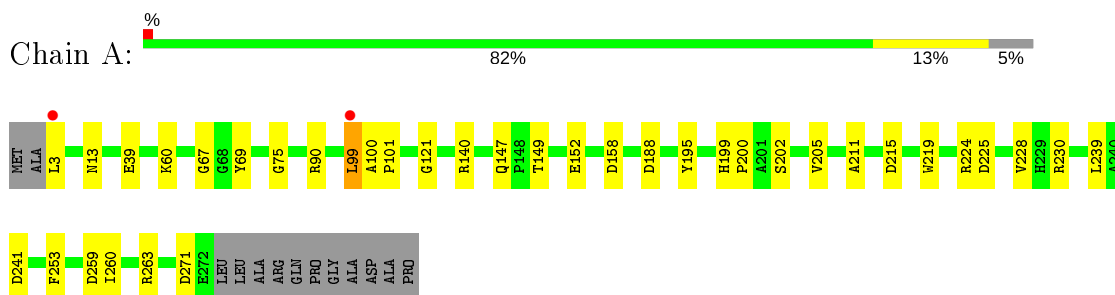
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
2	G	153	Total 153	O 153	0	0
2	H	122	Total 122	O 122	0	0

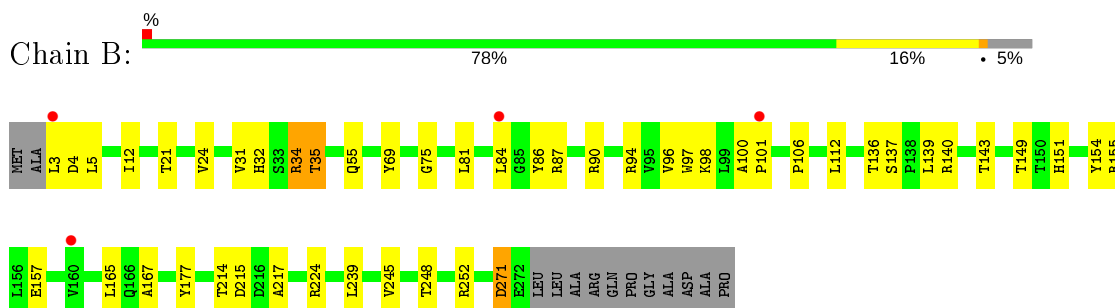
### 3 Residue-property plots [i](#)

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.

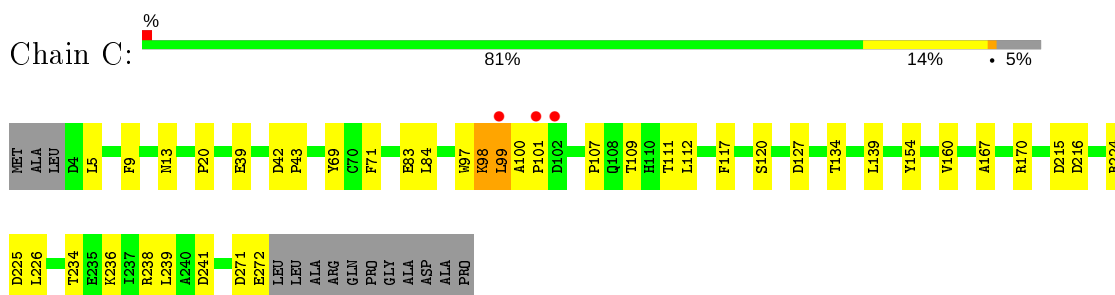
- Molecule 1: ARYLAMINE N-ACETYLTRANSFERASE NAT



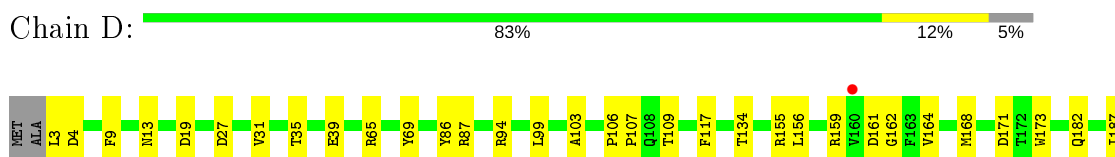
- Molecule 1: ARYLAMINE N-ACETYLTRANSFERASE NAT

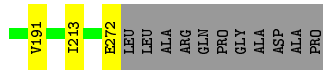


- Molecule 1: ARYLAMINE N-ACETYLTRANSFERASE NAT

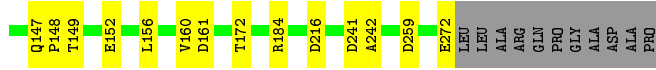
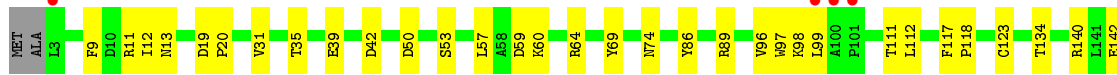
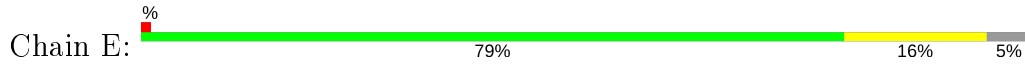


- Molecule 1: ARYLAMINE N-ACETYLTRANSFERASE NAT

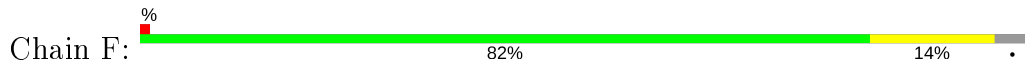




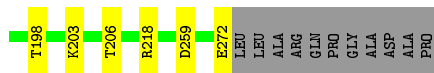
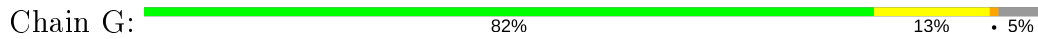
• Molecule 1: ARYLAMINE N-ACETYLTRANSFERASE NAT



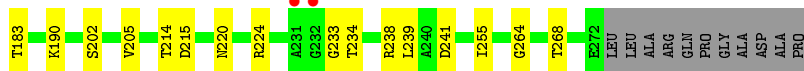
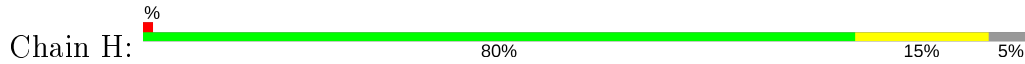
• Molecule 1: ARYLAMINE N-ACETYLTRANSFERASE NAT



• Molecule 1: ARYLAMINE N-ACETYLTRANSFERASE NAT



• Molecule 1: ARYLAMINE N-ACETYLTRANSFERASE NAT



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	96.49Å 139.15Å 96.51Å 90.00° 91.18° 90.00°	Depositor
Resolution (Å)	29.63 – 2.10 29.63 – 2.10	Depositor EDS
% Data completeness (in resolution range)	91.7 (29.63-2.10) 87.3 (29.63-2.10)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.13 (at 2.10Å)	Xtrriage
Refinement program	PHENIX (PHENIX.REFINE: 1.8.2_1309)	Depositor
R, $R_{free}$	0.161 , 0.187 0.158 , 0.179	Depositor DCC
$R_{free}$ test set	6821 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	25.1	Xtrriage
Anisotropy	0.395	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.39 , 36.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.40$ , $\langle L^2 \rangle = 0.23$	Xtrriage
Estimated twinning fraction	0.076 for l,k,-h 0.078 for h,-k,-l 0.429 for l,-k,h	Xtrriage
Reported twinning fraction	0.570 for L,-K,H	Depositor
Outliers	0 of 136337 reflections	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	33783	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	36.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 47.37 % 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 1.0023e-04. 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](#)

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	0.64	0/2114	0.77	1/2892 (0.0%)
1	B	0.65	0/2119	0.78	3/2898 (0.1%)
1	C	0.66	0/2100	0.77	1/2875 (0.0%)
1	D	0.68	0/2123	0.79	0/2903
1	E	0.64	0/2120	0.73	0/2898
1	F	0.67	0/2119	0.75	0/2898
1	G	0.70	1/2114 (0.0%)	0.75	1/2893 (0.0%)
1	H	0.65	0/2116	0.75	1/2894 (0.0%)
All	All	0.66	1/16925 (0.0%)	0.76	7/23151 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	G	79	TYR	CD2-CE2	-5.01	1.31	1.39

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	B	140	ARG	NE-CZ-NH1	-5.96	117.32	120.30
1	B	140	ARG	NE-CZ-NH2	5.64	123.12	120.30
1	G	3	LEU	CA-CB-CG	5.51	127.97	115.30
1	A	188	ASP	CB-CG-OD1	5.40	123.16	118.30
1	H	94	ARG	NE-CZ-NH1	5.37	122.98	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	2066	2029	2023	27	0
1	B	2071	2038	2032	34	1
1	C	2052	2000	1994	26	0
1	D	2075	2042	2036	30	0
1	E	2072	2039	2031	25	1
1	F	2072	2039	2033	30	0
1	G	2066	2022	2016	24	0
1	H	2068	2033	2027	24	0
2	A	122	0	0	7	0
2	B	100	0	0	2	4
2	C	122	0	0	4	4
2	D	142	0	0	7	0
2	E	115	0	0	6	0
2	F	123	0	0	8	0
2	G	153	0	0	8	0
2	H	122	0	0	4	0
All	All	17541	16242	16192	214	5

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 7.

The worst 5 of 214 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:H:121:GLY:N	2:H:2069:HOH:O	2.09	0.86
1:F:39:GLU:OE2	1:F:67:GLY:N	2.09	0.85
1:G:66:ARG:NH1	2:G:2068:HOH:O	2.14	0.79
1:A:149:THR:OG1	1:A:152:GLU:O	2.00	0.78
1:F:99:LEU:O	2:F:2055:HOH:O	2.02	0.77

All (5) 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)
2:B:2005:HOH:O	2:C:2001:HOH:O[1_456]	2.04	0.16
1:B:87:ARG:HH22	1:E:259:ASP:O[1_556]	1.45	0.15
2:B:2003:HOH:O	2:C:2003:HOH:O[1_456]	2.12	0.08
2:B:2006:HOH:O	2:C:2002:HOH:O[1_456]	2.13	0.07

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:2004:HOH:O	2:C:2002:HOH:O[1_456]	2.18	0.02

## 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	268/283 (95%)	250 (93%)	18 (7%)	0	100	100
1	B	268/283 (95%)	255 (95%)	13 (5%)	0	100	100
1	C	267/283 (94%)	252 (94%)	13 (5%)	2 (1%)	22	18
1	D	268/283 (95%)	259 (97%)	9 (3%)	0	100	100
1	E	268/283 (95%)	257 (96%)	11 (4%)	0	100	100
1	F	267/283 (94%)	258 (97%)	9 (3%)	0	100	100
1	G	268/283 (95%)	260 (97%)	7 (3%)	1 (0%)	34	32
1	H	268/283 (95%)	259 (97%)	9 (3%)	0	100	100
All	All	2142/2264 (95%)	2050 (96%)	89 (4%)	3 (0%)	51	54

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	98	LYS
1	C	99	LEU
1	G	160	VAL

### 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	215/230 (94%)	210 (98%)	5 (2%)	50	55
1	B	216/230 (94%)	212 (98%)	4 (2%)	57	63
1	C	213/230 (93%)	208 (98%)	5 (2%)	50	55
1	D	217/230 (94%)	215 (99%)	2 (1%)	78	84
1	E	215/230 (94%)	207 (96%)	8 (4%)	34	35
1	F	216/230 (94%)	213 (99%)	3 (1%)	67	73
1	G	216/230 (94%)	213 (99%)	3 (1%)	67	73
1	H	215/230 (94%)	212 (99%)	3 (1%)	67	73
All	All	1723/1840 (94%)	1690 (98%)	33 (2%)	57	63

5 of 33 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	39	GLU
1	E	69	TYR
1	H	39	GLU
1	D	69	TYR
1	E	19	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](#)

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 [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	270/283 (95%)	-0.12	2 (0%) 87 89	16, 32, 55, 69	3 (1%)
1	B	270/283 (95%)	0.00	4 (1%) 73 77	18, 38, 60, 74	3 (1%)
1	C	269/283 (95%)	-0.09	3 (1%) 80 84	15, 30, 52, 68	3 (1%)
1	D	270/283 (95%)	-0.24	1 (0%) 92 93	16, 27, 48, 61	3 (1%)
1	E	270/283 (95%)	-0.16	4 (1%) 73 77	18, 32, 52, 64	3 (1%)
1	F	271/283 (95%)	-0.10	3 (1%) 80 84	15, 31, 55, 69	3 (1%)
1	G	270/283 (95%)	-0.17	1 (0%) 92 93	13, 26, 50, 62	3 (1%)
1	H	270/283 (95%)	-0.13	4 (1%) 73 77	18, 32, 54, 73	3 (1%)
All	All	2160/2264 (95%)	-0.13	22 (1%) 82 85	13, 31, 54, 74	24 (1%)

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	84	LEU	4.3
1	C	101	PRO	3.9
1	A	99	LEU	3.7
1	F	264	GLY	3.3
1	E	101	PRO	3.3

### 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

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.