wwPDB X-ray Structure Validation Summary Report (i)

May 15, 2020 – 12:20 pm BST

PDB ID	:	3MAY
Title	:	Crystal structure of a secreted Mycobacterium tuberculosis heme-binding protein
Authors	:	Goulding, C.W.; Chim, N.
Deposited on	:	2010-03-24
Resolution	:	2.50 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

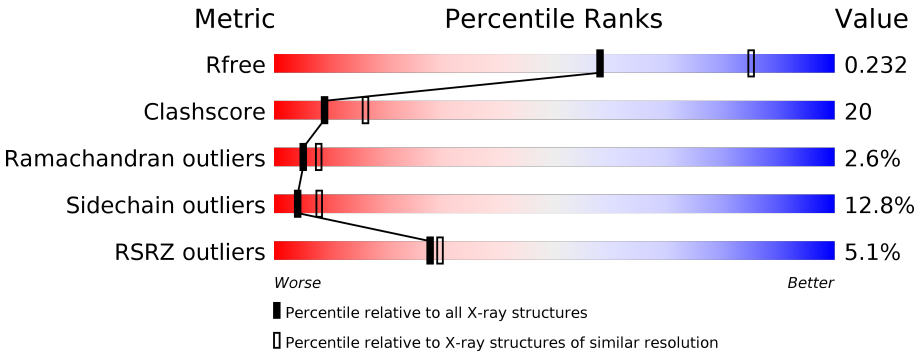
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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.



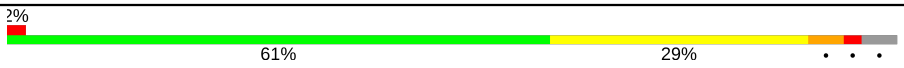
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 >=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 <=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	101	5% 57% 19% 7% • 15%
1	B	101	2% 61% 25% 5% • 8%
1	C	101	2% 65% 24% 6% • •
1	D	101	9% 72% 16% • • 7%
1	E	101	4% 68% 21% • 7%
1	F	101	8% 60% 15% 8% • 15%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	101	
1	H	101	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 5379 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 POSSIBLE EXPORTED PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	86	Total	C	N	O	S	0	0	0
			618	380	109	125	4			
1	B	93	Total	C	N	O	S	0	0	0
			668	411	118	134	5			
1	C	97	Total	C	N	O	S	0	0	0
			693	426	122	140	5			
1	D	94	Total	C	N	O	S	0	0	0
			673	414	119	135	5			
1	E	94	Total	C	N	O	S	0	0	0
			675	416	118	136	5			
1	G	97	Total	C	N	O	S	0	0	0
			690	425	122	138	5			
1	F	86	Total	C	N	O	S	0	0	0
			618	380	109	125	4			
1	H	96	Total	C	N	O	S	0	0	0
			688	423	121	139	5			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	9	Total	O	0	0
			9	9		
2	B	5	Total	O	0	0
			5	5		
2	C	3	Total	O	0	0
			3	3		
2	D	9	Total	O	0	0
			9	9		
2	E	5	Total	O	0	0
			5	5		
2	G	6	Total	O	0	0
			6	6		

Continued on next page...

Continued from previous page...

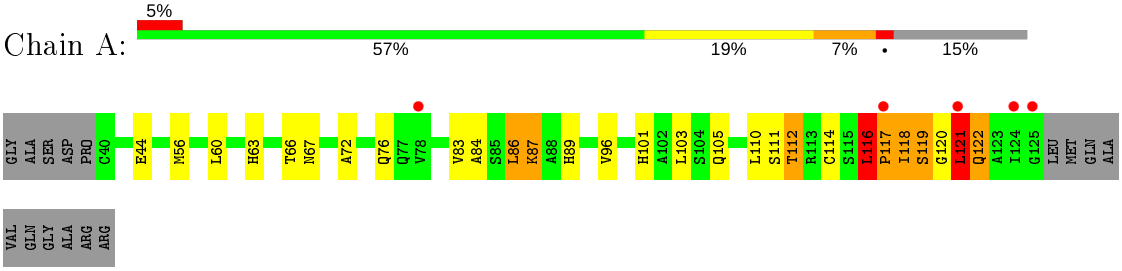
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	F	9	Total O 9 9	0	0
2	H	10	Total O 10 10	0	0

3 Residue-property plots

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.

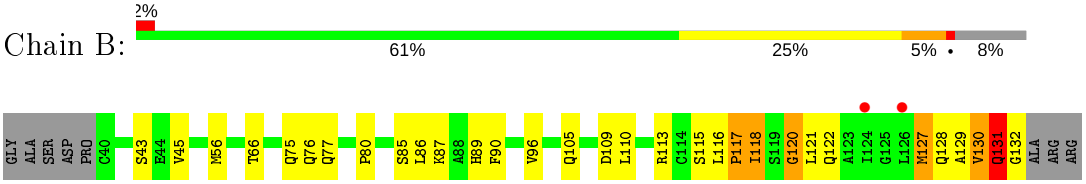
- Molecule 1: POSSIBLE EXPORTED PROTEIN

Chain A:



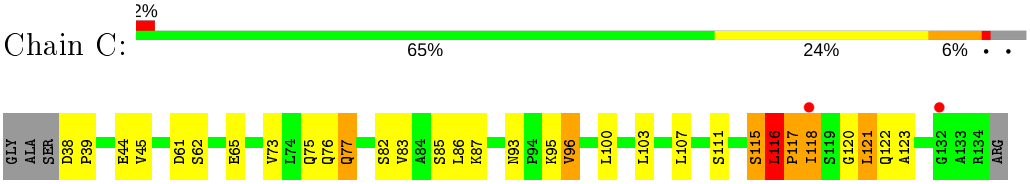
- Molecule 1: POSSIBLE EXPORTED PROTEIN

Chain B:



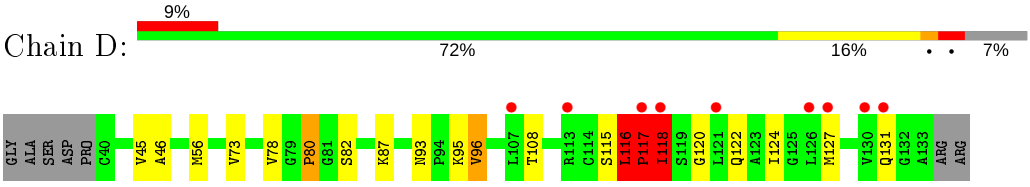
- Molecule 1: POSSIBLE EXPORTED PROTEIN

Chain C:



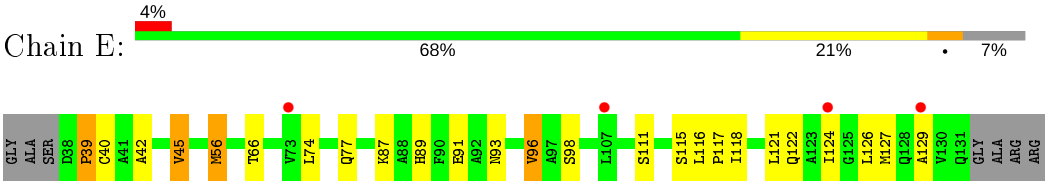
- Molecule 1: POSSIBLE EXPORTED PROTEIN

Chain D:



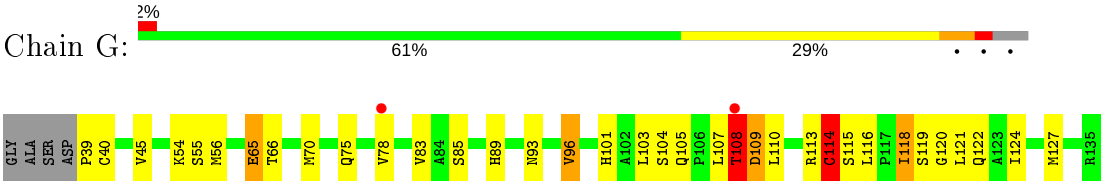
- Molecule 1: POSSIBLE EXPORTED PROTEIN

Chain E:



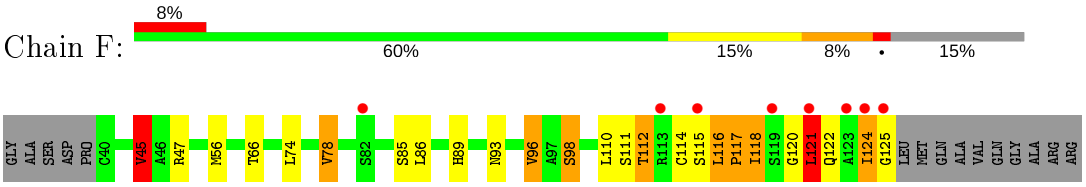
• Molecule 1: POSSIBLE EXPORTED PROTEIN

Chain G:



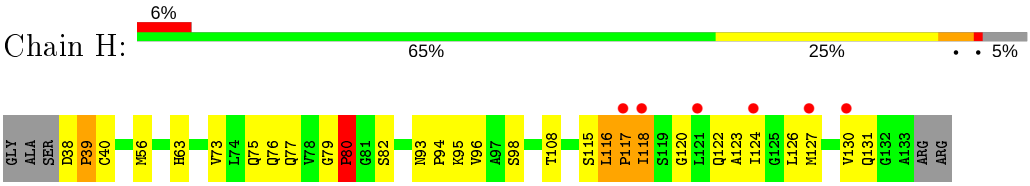
• Molecule 1: POSSIBLE EXPORTED PROTEIN

Chain F:



• Molecule 1: POSSIBLE EXPORTED PROTEIN

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41	Depositor
Cell constants a, b, c, α, β, γ	72.67Å 72.67Å 173.91Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	72.68 – 2.50 72.67 – 2.50	Depositor EDS
% Data completeness (in resolution range)	99.5 (72.68-2.50) 99.5 (72.67-2.50)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	5.45 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.195 , 0.256 0.202 , 0.232	Depositor DCC
R_{free} test set	1595 reflections (5.17%)	wwPDB-VP
Wilson B-factor (Å^2)	50.5	Xtriage
Anisotropy	0.221	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 58.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.52$, $< L^2 > = 0.35$	Xtriage
Estimated twinning fraction	0.277 for h,-k,-l	Xtriage
Reported twinning fraction	0.510 for H, K, L 0.490 for -H, K, -L	Depositor
Outliers	0 of 30856 reflections	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	5379	wwPDB-VP
Average B, all atoms (Å^2)	58.0	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 24.38 % 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 3.8776e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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

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	0.99	1/627 (0.2%)	1.07	4/853 (0.5%)
1	B	0.88	0/677	1.02	4/920 (0.4%)
1	C	1.06	1/703 (0.1%)	0.96	1/957 (0.1%)
1	D	1.04	1/682 (0.1%)	0.97	2/927 (0.2%)
1	E	0.90	0/685	0.95	2/933 (0.2%)
1	F	0.98	1/627 (0.2%)	1.08	4/853 (0.5%)
1	G	0.97	1/700 (0.1%)	1.01	2/952 (0.2%)
1	H	0.95	1/698 (0.1%)	0.95	1/950 (0.1%)
All	All	0.97	6/5399 (0.1%)	1.00	20/7345 (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	C	0	1
1	D	0	1
1	G	0	2
All	All	0	4

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	117	PRO	N-CD	10.10	1.61	1.47
1	C	61	ASP	CB-CG	-7.34	1.36	1.51
1	G	114	CYS	CB-SG	-5.95	1.72	1.81
1	H	130	VAL	CB-CG1	5.84	1.65	1.52
1	A	44	GLU	CG-CD	5.41	1.60	1.51

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	F	112	THR	CB-CA-C	10.16	139.05	111.60
1	A	112	THR	CB-CA-C	9.86	138.22	111.60
1	G	108	THR	C-N-CA	9.46	145.34	121.70
1	A	112	THR	N-CA-C	-9.04	86.58	111.00
1	F	112	THR	N-CA-C	-8.58	87.84	111.00

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	116	LEU	Peptide
1	D	116	LEU	Peptide
1	G	108	THR	Peptide
1	G	39	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	618	0	614	30	1
1	B	668	0	667	31	0
1	C	693	0	685	51	1
1	D	673	0	672	30	0
1	E	675	0	669	17	0
1	F	618	0	614	19	0
1	G	690	0	684	29	1
1	H	688	0	683	34	0
2	A	9	0	0	3	2
2	B	5	0	0	1	0
2	C	3	0	0	0	0
2	D	9	0	0	4	1
2	E	5	0	0	1	0
2	F	9	0	0	1	2
2	G	6	0	0	4	3
2	H	10	0	0	6	1
All	All	5379	0	5288	214	6

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 20.

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:G:118:ILE:HD12	2:G:136:HOH:O	1.07	1.23
1:A:84:ALA:HA	1:A:87:LYS:CD	1.69	1.22
1:C:87:LYS:NZ	1:D:131:GLN:HE22	1.36	1.21
1:C:111:SER:O	1:C:115:SER:HA	1.37	1.19
1:H:80:PRO:HD3	2:H:139:HOH:O	1.44	1.17

The worst 5 of 6 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:D:140:HOH:O	2:F:143:HOH:O[1_545]	1.14	1.06
2:G:140:HOH:O	2:H:141:HOH:O[3_655]	1.43	0.77
2:A:137:HOH:O	2:G:139:HOH:O[4_564]	1.47	0.73
1:A:63:HIS:CE1	2:G:140:HOH:O[4_564]	1.94	0.26
1:G:65:GLU:CG	2:A:137:HOH:O[3_655]	2.09	0.11

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	84/101 (83%)	74 (88%)	7 (8%)	3 (4%)	3 4
1	B	91/101 (90%)	82 (90%)	6 (7%)	3 (3%)	4 5
1	C	95/101 (94%)	86 (90%)	7 (7%)	2 (2%)	7 11
1	D	92/101 (91%)	82 (89%)	6 (6%)	4 (4%)	2 3
1	E	92/101 (91%)	87 (95%)	4 (4%)	1 (1%)	14 26
1	F	84/101 (83%)	77 (92%)	6 (7%)	1 (1%)	13 24

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	G	95/101 (94%)	85 (90%)	10 (10%)	0	100 100
1	H	94/101 (93%)	83 (88%)	6 (6%)	5 (5%)	2 2
All	All	727/808 (90%)	656 (90%)	52 (7%)	19 (3%)	5 8

5 of 19 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	117	PRO
1	B	117	PRO
1	C	117	PRO
1	D	80	PRO
1	D	117	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	69/79 (87%)	60 (87%)	9 (13%)	4 7
1	B	74/79 (94%)	65 (88%)	9 (12%)	5 9
1	C	76/79 (96%)	64 (84%)	12 (16%)	2 4
1	D	74/79 (94%)	66 (89%)	8 (11%)	6 12
1	E	75/79 (95%)	70 (93%)	5 (7%)	16 31
1	F	69/79 (87%)	56 (81%)	13 (19%)	1 2
1	G	75/79 (95%)	60 (80%)	15 (20%)	1 2
1	H	76/79 (96%)	72 (95%)	4 (5%)	22 43
All	All	588/632 (93%)	513 (87%)	75 (13%)	4 8

5 of 75 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	87	LYS
1	E	98	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	F	122	GLN
1	D	96	VAL
1	E	45	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 26 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	128	GLN
1	E	89	HIS
1	H	77	GLN
1	D	131	GLN
1	E	76	GLN

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

There are no ligands in this entry.

5.7 Other polymers

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

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^{th} 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^2)	Q<0.9
1	A	86/101 (85%)	0.45	5 (5%)	23	24	36, 47, 115, 119	0
1	B	93/101 (92%)	0.28	2 (2%)	62	65	35, 52, 90, 101	0
1	C	97/101 (96%)	0.34	2 (2%)	63	66	31, 56, 81, 84	0
1	D	94/101 (93%)	0.53	9 (9%)	8	7	37, 56, 89, 95	0
1	E	94/101 (93%)	0.43	4 (4%)	35	38	36, 53, 84, 92	0
1	F	86/101 (85%)	0.69	8 (9%)	8	8	34, 47, 114, 118	0
1	G	97/101 (96%)	0.32	2 (2%)	63	66	35, 61, 81, 87	0
1	H	96/101 (95%)	0.65	6 (6%)	20	21	38, 58, 91, 97	0
All	All	743/808 (91%)	0.46	38 (5%)	28	29	31, 55, 91, 119	0

The worst 5 of 38 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	125	GLY	13.6
1	F	121	LEU	8.3
1	H	117	PRO	7.0
1	D	130	VAL	6.4
1	A	125	GLY	6.2

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

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.