



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 1, 2021 – 12:47 pm GMT

PDB ID : 7BEE
Title : Crystal structure of a Hsp47-collagen peptide complex
Authors : Abraham, E.T.; Gebauer, J.M.; Baumann, U.
Deposited on : 2020-12-23
Resolution : 1.94 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 ⓘ 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 : 2.23.2
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0267
CCP4 : 7.1.010 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.23.2

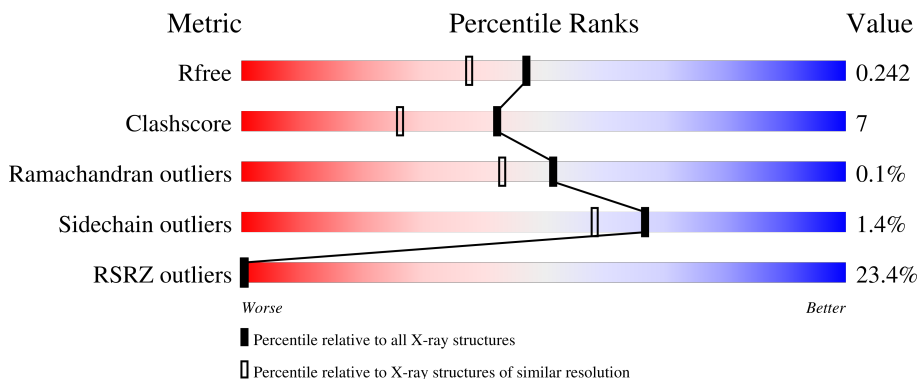
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.94 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.



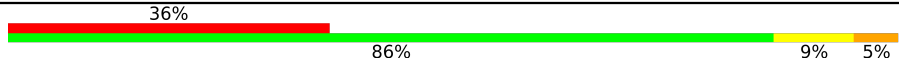


Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4310 (1.96-1.92)
Clashscore	141614	1023 (1.94-1.94)
Ramachandran outliers	138981	1007 (1.94-1.94)
Sidechain outliers	138945	1007 (1.94-1.94)
RSRZ outliers	127900	4250 (1.96-1.92)

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 ≥ 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	392	
1	B	392	
2	C	22	
2	D	22	
2	E	22	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	22	
2	G	22	
2	H	22	

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 7032 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 Collagen-binding protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	378	2982	1893	522	554	13	0	0	0
1	B	373	2938	1867	514	543	14	0	0	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	35	MET	-	initiating methionine	UNP E2RHY7
A	419	LEU	-	expression tag	UNP E2RHY7
A	420	GLU	-	expression tag	UNP E2RHY7
A	421	HIS	-	expression tag	UNP E2RHY7
A	422	HIS	-	expression tag	UNP E2RHY7
A	423	HIS	-	expression tag	UNP E2RHY7
A	424	HIS	-	expression tag	UNP E2RHY7
A	425	HIS	-	expression tag	UNP E2RHY7
A	426	HIS	-	expression tag	UNP E2RHY7
B	35	MET	-	initiating methionine	UNP E2RHY7
B	419	LEU	-	expression tag	UNP E2RHY7
B	420	GLU	-	expression tag	UNP E2RHY7
B	421	HIS	-	expression tag	UNP E2RHY7
B	422	HIS	-	expression tag	UNP E2RHY7
B	423	HIS	-	expression tag	UNP E2RHY7
B	424	HIS	-	expression tag	UNP E2RHY7
B	425	HIS	-	expression tag	UNP E2RHY7
B	426	HIS	-	expression tag	UNP E2RHY7

- Molecule 2 is a protein called 21er collagen model peptide.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	C	21	133	89	23	21	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	D	20	Total	C	N	O	0	0	0
			126	84	22	20			
2	E	20	Total	C	N	O	0	0	0
			126	84	22	20			
2	F	22	Total	C	N	O	0	0	0
			137	91	24	22			
2	G	20	Total	C	N	O	0	0	0
			126	84	22	20			
2	H	19	Total	C	N	O	0	0	0
			119	79	21	19			

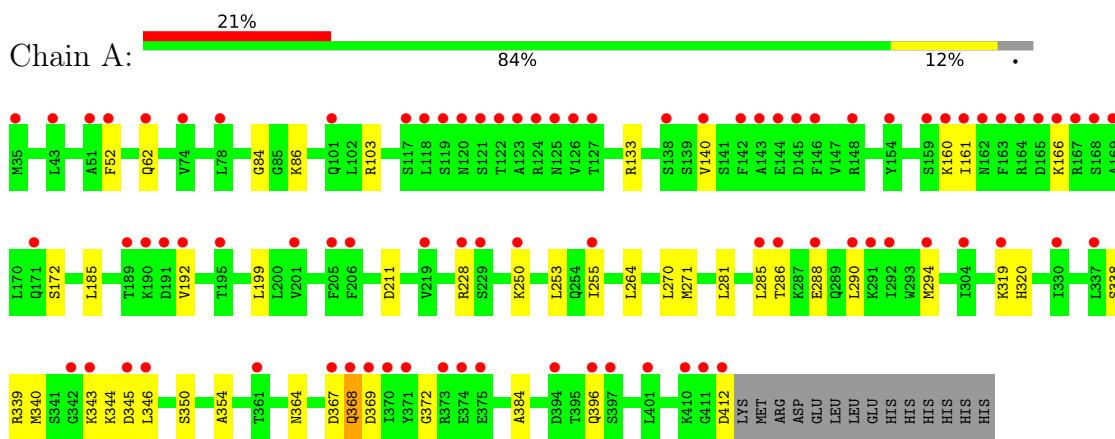
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	145	Total	O	0	0
			145	145		
3	B	143	Total	O	0	0
			143	143		
3	C	6	Total	O	0	0
			6	6		
3	D	17	Total	O	0	0
			17	17		
3	E	7	Total	O	0	0
			7	7		
3	F	9	Total	O	0	0
			9	9		
3	G	7	Total	O	0	0
			7	7		
3	H	11	Total	O	0	0
			11	11		

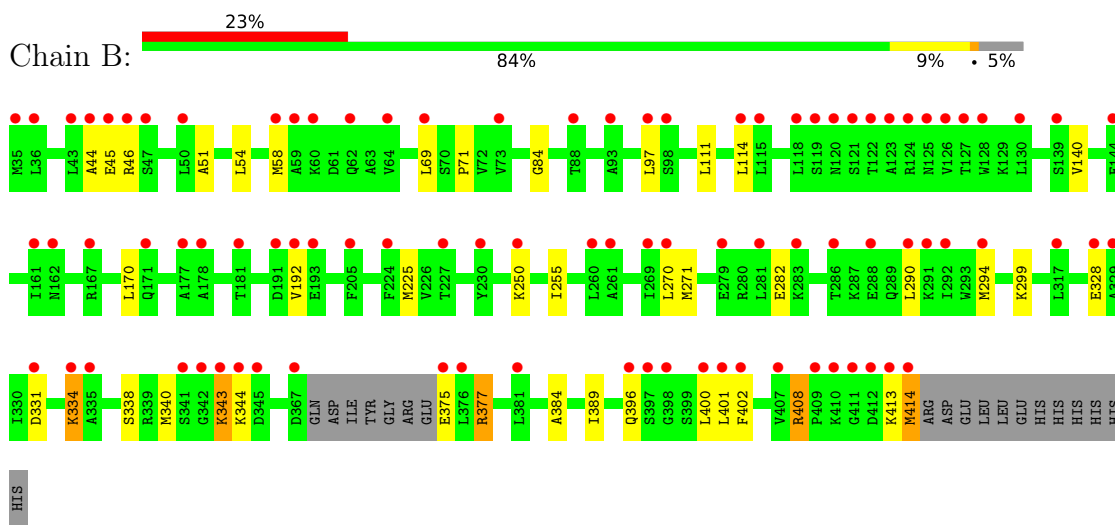
3 Residue-property plots [i](#)

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.

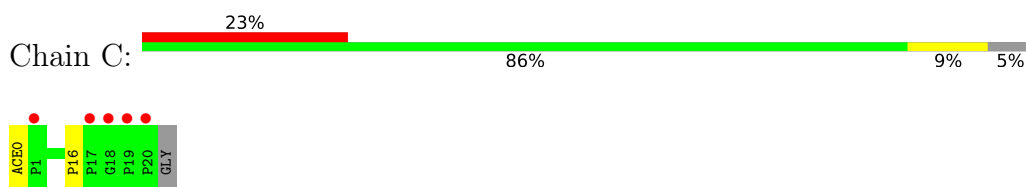
- Molecule 1: Collagen-binding protein



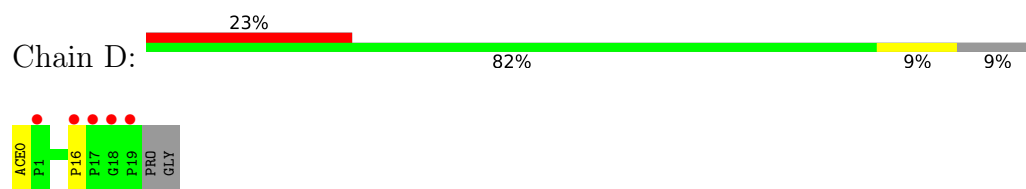
- Molecule 1: Collagen-binding protein



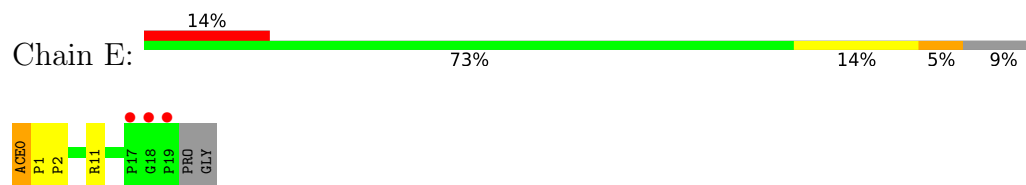
- Molecule 2: 21er collagen model peptide



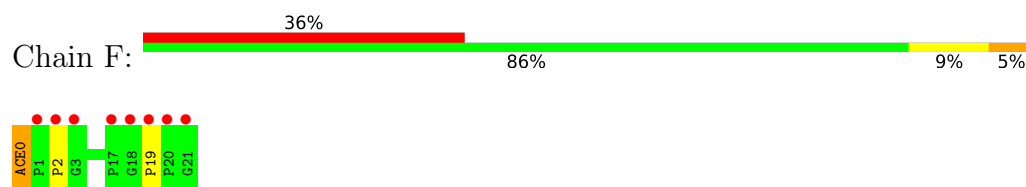
- Molecule 2: 21er collagen model peptide



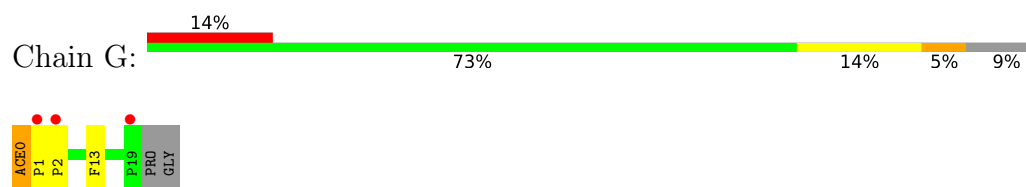
- Molecule 2: 21er collagen model peptide



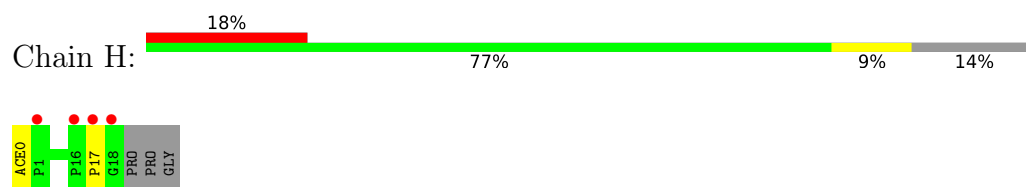
- Molecule 2: 21er collagen model peptide



- Molecule 2: 21er collagen model peptide



- Molecule 2: 21er collagen model peptide



4 Data and refinement statistics i

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α , β , γ	96.68Å 96.68Å 187.48Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	46.81 – 1.94 46.81 – 1.94	Depositor EDS
% Data completeness (in resolution range)	99.8 (46.81-1.94) 99.8 (46.81-1.94)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.37 (at 1.94Å)	Xtrriage
Refinement program	PHENIX 1.13rc2_2986	Depositor
R, R_{free}	0.207 , 0.242 0.207 , 0.242	Depositor DCC
R_{free} test set	3838 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å ²)	39.0	Xtrriage
Anisotropy	0.105	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	(Not available) , (Not available)	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.023 for -h,-k,l	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	7032	wwPDB-VP
Average B, all atoms (Å ²)	47.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.37% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: ACE

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	0.42	0/3043	0.61	0/4105
1	B	0.43	0/2997	0.60	2/4040 (0.0%)
2	C	0.35	0/143	0.70	1/203 (0.5%)
2	D	0.35	0/135	0.85	1/191 (0.5%)
2	E	0.37	0/135	0.65	1/191 (0.5%)
2	F	0.38	0/147	1.12	3/208 (1.4%)
2	G	0.39	0/135	0.80	1/191 (0.5%)
2	H	0.40	0/127	1.13	2/179 (1.1%)
All	All	0.42	0/6862	0.65	11/9308 (0.1%)

There are no bond length outliers.

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	F	0	ACE	O-C-N	11.22	142.42	121.10
2	H	0	ACE	O-C-N	-10.82	100.54	121.10
2	D	0	ACE	O-C-N	-9.60	102.86	121.10
2	F	0	ACE	C-N-CD	8.48	146.21	128.40
2	H	0	ACE	C-N-CD	-8.01	102.98	120.60

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

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	2982	0	3006	48	0
1	B	2938	0	2973	39	0
2	C	133	0	127	1	0
2	D	126	0	120	7	0
2	E	126	0	120	3	0
2	F	137	0	130	3	0
2	G	126	0	120	4	0
2	H	119	0	113	1	0
3	A	145	0	0	2	0
3	B	143	0	0	2	0
3	C	6	0	0	0	0
3	D	17	0	0	0	0
3	E	7	0	0	0	0
3	F	9	0	0	0	0
3	G	7	0	0	0	0
3	H	11	0	0	0	0
All	All	7032	0	6709	93	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 7.

The worst 5 of 93 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:331:ASP:OD2	1:B:334:LYS:HE3	1.36	1.22
1:B:331:ASP:CG	1:B:334:LYS:HE3	1.69	1.09
1:B:54:LEU:HG	1:B:58:MET:HE1	1.37	1.01
1:A:343:LYS:HZ3	1:A:345:ASP:H	1.15	0.92
1:A:343:LYS:NZ	1:A:345:ASP:H	1.69	0.90

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	376/392 (96%)	364 (97%)	11 (3%)	1 (0%)	41	32
1	B	369/392 (94%)	363 (98%)	6 (2%)	0	100	100
2	C	19/22 (86%)	19 (100%)	0	0	100	100
2	D	18/22 (82%)	18 (100%)	0	0	100	100
2	E	18/22 (82%)	18 (100%)	0	0	100	100
2	F	20/22 (91%)	20 (100%)	0	0	100	100
2	G	18/22 (82%)	18 (100%)	0	0	100	100
2	H	17/22 (77%)	17 (100%)	0	0	100	100
All	All	855/916 (93%)	837 (98%)	17 (2%)	1 (0%)	51	43

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	368	GLN

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	323/337 (96%)	320 (99%)	3 (1%)	78	75
1	B	319/337 (95%)	312 (98%)	7 (2%)	52	39
2	C	14/14 (100%)	14 (100%)	0	100	100
2	D	13/14 (93%)	13 (100%)	0	100	100
2	E	13/14 (93%)	13 (100%)	0	100	100
2	F	14/14 (100%)	14 (100%)	0	100	100
2	G	13/14 (93%)	13 (100%)	0	100	100
2	H	12/14 (86%)	12 (100%)	0	100	100
All	All	721/758 (95%)	711 (99%)	10 (1%)	67	58

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	377	ARG
1	B	408	ARG
1	B	414	MET
1	B	250	LYS
1	B	299	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	396	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](#)

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, 95th 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	378/392 (96%)	1.60	84 (22%) 0 0	27, 41, 92, 159	0
1	B	373/392 (95%)	1.48	91 (24%) 0 0	28, 43, 76, 113	0
2	C	20/22 (90%)	1.67	5 (25%) 0 0	29, 40, 100, 102	0
2	D	19/22 (86%)	1.53	5 (26%) 0 0	30, 38, 79, 101	0
2	E	19/22 (86%)	1.74	3 (15%) 2 2	32, 38, 79, 90	0
2	F	21/22 (95%)	1.92	8 (38%) 0 0	31, 48, 100, 114	0
2	G	19/22 (86%)	1.05	3 (15%) 2 2	32, 43, 83, 93	0
2	H	18/22 (81%)	1.43	4 (22%) 0 0	34, 41, 75, 99	0
All	All	867/916 (94%)	1.54	203 (23%) 0 0	27, 42, 85, 159	0

The worst 5 of 203 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	121	SER	20.5
1	A	123	ALA	17.1
1	A	369	ASP	10.5
1	A	374	GLU	10.1
2	E	19	PRO	9.9

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

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.