



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 18, 2026 – 12:04 AM UTC

PDB ID : 5HGB / pdb_00005hgb
Title : HLA*A2402 complexed with HIV nef138 8mer epitope
Authors : Shi, Y.; Qi, J.; Gao, G.F.
Deposited on : 2016-01-08
Resolution : 2.40 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 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	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

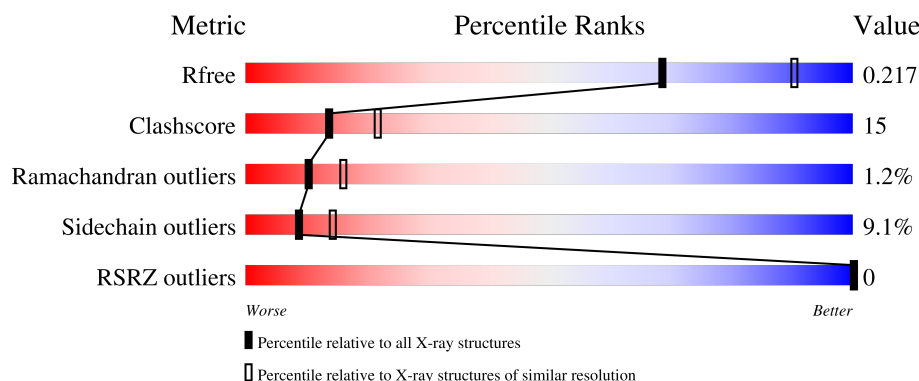
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.40 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.



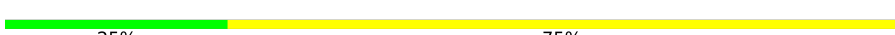
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	180053	4912 (2.40-2.40)
Clashscore	190562	5391 (2.40-2.40)
Ramachandran outliers	187476	5320 (2.40-2.40)
Sidechain outliers	187428	5321 (2.40-2.40)
RSRZ outliers	180081	4916 (2.40-2.40)

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	275	68% 29% .
1	D	275	64% 31% .
1	G	275	67% 29% .
1	J	275	63% 33% .
2	B	100	69% 26% . .

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	E	100	 63%34%..
2	H	100	 64%31%..
2	K	100	 71%23%5%.
3	C	8	 50%50%
3	F	8	 38%50%12%
3	I	8	 50%50%
3	L	8	 25%75%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 13137 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 HLA class I histocompatibility antigen, A-24 alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	274	Total	C	N	O	S	0	0	0
			2221	1382	403	426	10			
1	D	274	Total	C	N	O	S	0	0	0
			2221	1382	403	426	10			
1	G	274	Total	C	N	O	S	0	0	0
			2221	1382	403	426	10			
1	J	274	Total	C	N	O	S	0	0	0
			2221	1382	403	426	10			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	initiating methionine	UNP P05534
D	0	MET	-	initiating methionine	UNP P05534
G	0	MET	-	initiating methionine	UNP P05534
J	0	MET	-	initiating methionine	UNP P05534

- Molecule 2 is a protein called Beta-2-microglobulin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	100	Total	C	N	O	S	0	0	0
			836	533	141	158	4			
2	E	100	Total	C	N	O	S	0	0	0
			836	533	141	158	4			
2	H	100	Total	C	N	O	S	0	0	0
			836	533	141	158	4			
2	K	100	Total	C	N	O	S	0	0	0
			836	533	141	158	4			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	MET	-	initiating methionine	UNP P61769
E	0	MET	-	initiating methionine	UNP P61769
H	0	MET	-	initiating methionine	UNP P61769
K	0	MET	-	initiating methionine	UNP P61769

- Molecule 3 is a protein called Protein Nef.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	8	Total	C	N	O	0	0	0
			75	52	12	11			
3	F	8	Total	C	N	O	0	0	0
			75	52	12	11			
3	I	8	Total	C	N	O	0	0	0
			75	52	12	11			
3	L	8	Total	C	N	O	0	0	0
			75	52	12	11			

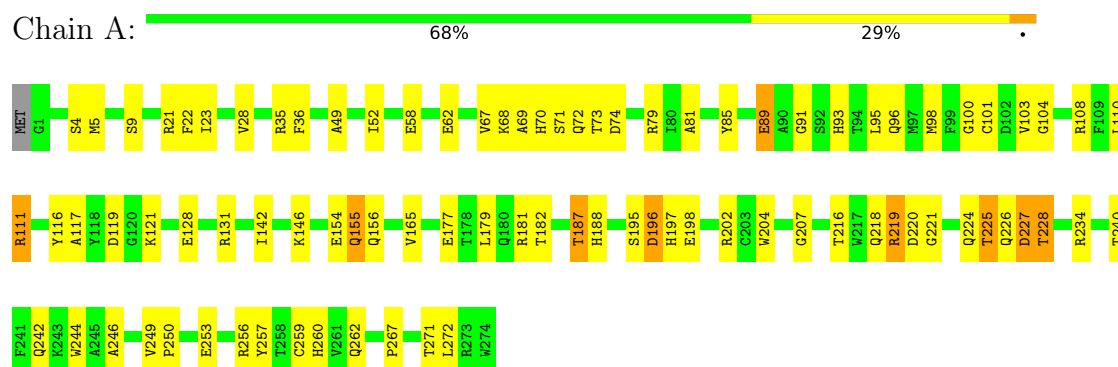
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	114	Total	O	0	0
			114	114		
4	B	35	Total	O	0	0
			35	35		
4	C	5	Total	O	0	0
			5	5		
4	D	102	Total	O	0	0
			102	102		
4	E	28	Total	O	0	0
			28	28		
4	F	4	Total	O	0	0
			4	4		
4	G	140	Total	O	0	0
			140	140		
4	H	46	Total	O	0	0
			46	46		
4	I	5	Total	O	0	0
			5	5		
4	J	99	Total	O	0	0
			99	99		
4	K	23	Total	O	0	0
			23	23		
4	L	8	Total	O	0	0
			8	8		

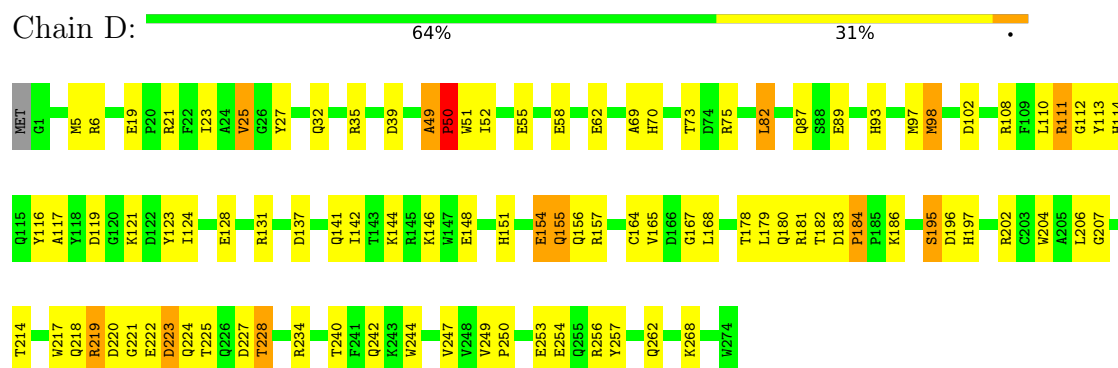
3 Residue-property plots [i](#)

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.

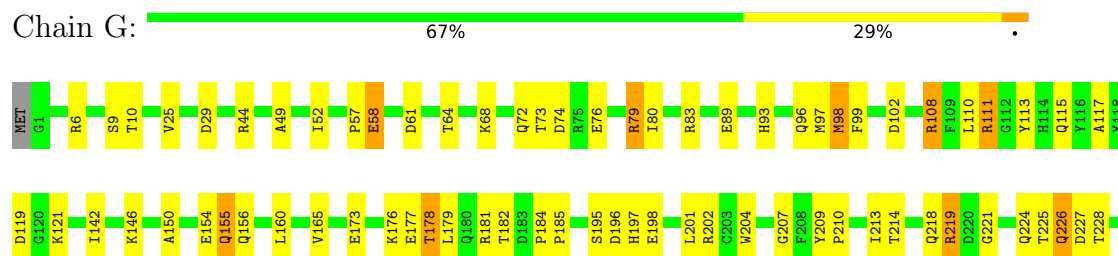
- Molecule 1: HLA class I histocompatibility antigen, A-24 alpha chain

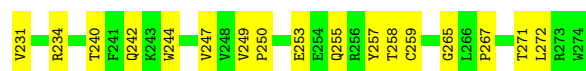


- Molecule 1: HLA class I histocompatibility antigen, A-24 alpha chain



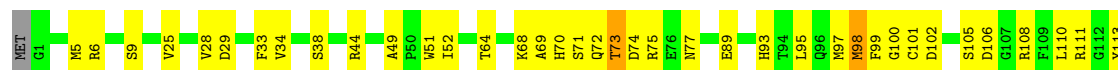
- Molecule 1: HLA class I histocompatibility antigen, A-24 alpha chain





- Molecule 1: HLA class I histocompatibility antigen, A-24 alpha chain

Chain J: 63% 33% .



- Molecule 2: Beta-2-microglobulin

Chain B: 69% 26% . .



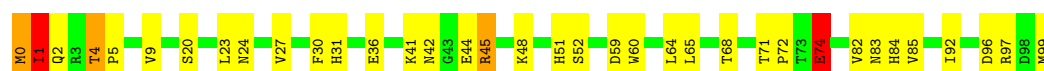
- Molecule 2: Beta-2-microglobulin

Chain E: 63% 34% . .



- Molecule 2: Beta-2-microglobulin

Chain H: 64% 31% . .



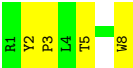
- Molecule 2: Beta-2-microglobulin

Chain K: 71% 23% 5% .

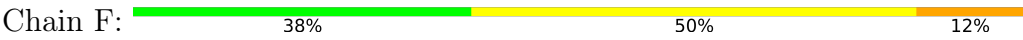


- Molecule 3: Protein Nef

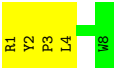
Chain C: 50% 50%



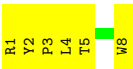
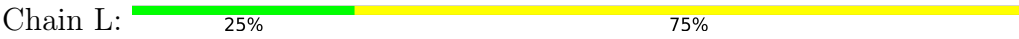
● Molecule 3: Protein Nef



● Molecule 3: Protein Nef



● Molecule 3: Protein Nef



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	71.95Å 152.51Å 90.17Å 90.00° 90.12° 90.00°	Depositor
Resolution (Å)	35.98 – 2.40 35.98 – 2.40	Depositor EDS
% Data completeness (in resolution range)	92.8 (35.98-2.40) 98.8 (35.98-2.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.37 (at 2.39Å)	Xtriage
Refinement program	PHENIX 1.5_2	Depositor
R, R_{free}	0.195 , 0.218 0.201 , 0.217	Depositor DCC
R_{free} test set	3788 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å ²)	39.8	Xtriage
Anisotropy	0.553	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 31.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.478 for h,-k,-l	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	13137	wwPDB-VP
Average B, all atoms (Å ²)	46.0	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 44.70 % 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 1.4694e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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 [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	0.33	0/2281	0.73	1/3092 (0.0%)
1	D	0.34	0/2281	0.73	4/3092 (0.1%)
1	G	0.36	0/2281	0.72	2/3092 (0.1%)
1	J	0.33	0/2281	0.72	2/3092 (0.1%)
2	B	0.31	0/859	0.71	0/1162
2	E	0.32	0/859	0.75	0/1162
2	H	0.32	0/859	0.76	0/1162
2	K	0.29	0/859	0.68	0/1162
3	C	0.27	0/79	0.86	0/106
3	F	0.25	0/79	0.74	0/106
3	I	0.26	0/79	0.82	0/106
3	L	0.28	0/79	0.66	0/106
All	All	0.33	0/12876	0.73	9/17440 (0.1%)

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	227	ASP	N-CA-C	-7.96	103.56	113.28
1	J	227	ASP	N-CA-C	-6.07	105.68	114.12
1	G	227	ASP	N-CA-C	-5.68	106.33	113.72
1	J	184	PRO	O-C-N	5.40	123.69	121.15
1	D	227	ASP	N-CA-C	-5.38	106.09	113.30

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	2221	0	2082	60	0
1	D	2221	0	2082	70	0
1	G	2221	0	2082	66	0
1	J	2221	0	2082	62	0
2	B	836	0	803	24	0
2	E	836	0	803	29	0
2	H	836	0	803	40	0
2	K	836	0	803	29	0
3	C	75	0	71	4	0
3	F	75	0	71	4	0
3	I	75	0	71	2	0
3	L	75	0	71	6	0
4	A	114	0	0	6	0
4	B	35	0	0	5	0
4	C	5	0	0	0	0
4	D	102	0	0	2	0
4	E	28	0	0	7	0
4	F	4	0	0	0	0
4	G	140	0	0	5	0
4	H	46	0	0	11	0
4	I	5	0	0	0	0
4	J	99	0	0	4	0
4	K	23	0	0	3	0
4	L	8	0	0	0	0
All	All	13137	0	11824	354	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 15.

The worst 5 of 354 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)
2:K:1:ILE:HG23	2:K:2:GLN:N	1.40	1.17
2:K:1:ILE:CG2	2:K:2:GLN:H	1.50	1.15
2:K:1:ILE:CG2	2:K:2:GLN:N	2.12	1.06
1:J:234:ARG:HH22	2:K:99:MET:HE1	1.20	1.02
2:H:71:THR:HA	4:H:104:HOH:O	1.66	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	272/275 (99%)	257 (94%)	13 (5%)	2 (1%)	18	28
1	D	272/275 (99%)	252 (93%)	16 (6%)	4 (2%)	8	12
1	G	272/275 (99%)	252 (93%)	18 (7%)	2 (1%)	18	28
1	J	272/275 (99%)	257 (94%)	12 (4%)	3 (1%)	11	18
2	B	98/100 (98%)	91 (93%)	6 (6%)	1 (1%)	12	20
2	E	98/100 (98%)	92 (94%)	5 (5%)	1 (1%)	12	20
2	H	98/100 (98%)	93 (95%)	2 (2%)	3 (3%)	3	3
2	K	98/100 (98%)	91 (93%)	5 (5%)	2 (2%)	6	7
3	C	6/8 (75%)	6 (100%)	0	0	100	100
3	F	6/8 (75%)	6 (100%)	0	0	100	100
3	I	6/8 (75%)	6 (100%)	0	0	100	100
3	L	6/8 (75%)	6 (100%)	0	0	100	100
All	All	1504/1532 (98%)	1409 (94%)	77 (5%)	18 (1%)	10	16

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	1	ILE
1	D	196	ASP
2	E	1	ILE
1	G	196	ASP
1	J	176	LYS

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	230/231 (100%)	210 (91%)	20 (9%)	9	16
1	D	230/231 (100%)	212 (92%)	18 (8%)	11	20
1	G	230/231 (100%)	210 (91%)	20 (9%)	9	16
1	J	230/231 (100%)	211 (92%)	19 (8%)	10	17
2	B	95/95 (100%)	86 (90%)	9 (10%)	8	13
2	E	95/95 (100%)	83 (87%)	12 (13%)	4	6
2	H	95/95 (100%)	87 (92%)	8 (8%)	10	17
2	K	95/95 (100%)	83 (87%)	12 (13%)	4	6
3	C	7/7 (100%)	7 (100%)	0	100	100
3	F	7/7 (100%)	6 (86%)	1 (14%)	3	4
3	I	7/7 (100%)	6 (86%)	1 (14%)	3	4
3	L	7/7 (100%)	6 (86%)	1 (14%)	3	4
All	All	1328/1332 (100%)	1207 (91%)	121 (9%)	9	14

5 of 121 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	E	97	ARG
2	K	9	VAL
1	G	155	GLN
2	K	1	ILE
2	K	97	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 53 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	174	ASN
2	H	51	HIS
2	K	24	ASN
1	G	180	GLN
1	G	242	GLN

5.3.3 RNA ⓘ

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 oligosaccharides 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	274/275 (99%)	-1.57	0 100 100	25, 40, 83, 130	0
1	D	274/275 (99%)	-1.59	0 100 100	24, 42, 69, 136	0
1	G	274/275 (99%)	-1.59	0 100 100	21, 39, 72, 142	0
1	J	274/275 (99%)	-1.59	0 100 100	25, 42, 71, 154	0
2	B	100/100 (100%)	-1.50	0 100 100	24, 45, 83, 129	0
2	E	100/100 (100%)	-1.46	0 100 100	28, 54, 85, 115	0
2	H	100/100 (100%)	-1.55	0 100 100	24, 47, 80, 107	0
2	K	100/100 (100%)	-1.53	0 100 100	26, 49, 85, 144	0
3	C	8/8 (100%)	-1.63	0 100 100	25, 29, 34, 37	0
3	F	8/8 (100%)	-1.64	0 100 100	29, 34, 42, 43	0
3	I	8/8 (100%)	-1.70	0 100 100	30, 34, 36, 38	0
3	L	8/8 (100%)	-1.64	0 100 100	24, 35, 37, 43	0
All	All	1528/1532 (99%)	-1.57	0 100 100	21, 42, 78, 154	0

There are no RSRZ outliers to report.

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 oligosaccharides 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.