WORLDWIDE

wwPDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 02:55 AM GMT

PDB ID	:	2J8U
Title	:	LARGE CDR3A LOOP ALTERATION AS A FUNCTION OF MHC MUTATION.
Authors	:	Miller, P.J.; Benhar, Y.P.; Biddison, W.; Collins, E.J.
Deposited on	:	2006-10-27
Resolution	:	2.88 Å(reported)

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

|		
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 2.88 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.
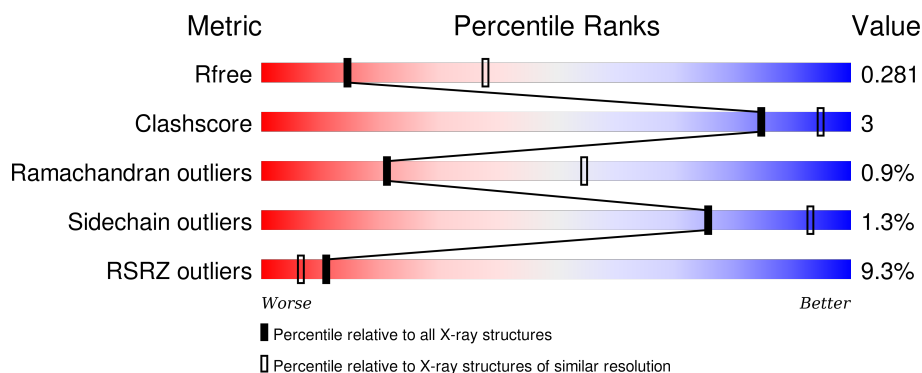
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1945 (2.90-2.86)
Clashscore	102246	2202 (2.90-2.86)
Ramachandran outliers	100387	2149 (2.90-2.86)
Sidechain outliers	100360	2152 (2.90-2.86)
RSRZ outliers	91569	1950 (2.90-2.86)

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. 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	275	7% / 92% / 7% ·
1	H	275	6% / 91% / 9%
2	B	100	6% / 98% ·
2	I	100	7% / 96% ·
3	C	9	89% / 11%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	J	9	100%
4	E	194	13% / 84% / 15%
4	L	194	15% / 80% / 16%
5	F	238	11% / 88% / 11%
5	M	238	9% / 93% / 6%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 13128 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-2 ALPHA CHAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	275	Total	C	N	O	S	0	0	0
			2243	1400	408	426	9			
1	H	275	Total	C	N	O	S	0	0	0
			2243	1400	408	426	9			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	66	ALA	LYS	ENGINEERED MUTATION	UNP P01892
H	66	ALA	LYS	ENGINEERED MUTATION	UNP P01892

- 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
			837	533	141	159	4			
2	I	100	Total	C	N	O	S	0	0	0
			837	533	141	159	4			

- Molecule 3 is a protein called SELF-PEPTIDE P1049.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	9	Total	C	N	O	0	0	0
			76	56	10	10			
3	J	9	Total	C	N	O	0	0	0
			76	56	10	10			

- Molecule 4 is a protein called AHIII TCR ALPHA CHAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	194	Total 1521	C 965	N 245	O 302	S 9	86	0	0
4	L	194	Total 1521	C 965	N 245	O 302	S 9	86	0	0

- Molecule 5 is a protein called AHIII TCR BETA CHAIN.

| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	237	Total 1887	C 1192	N 331	O 359	S 5	0	0	0
5	M	237	Total 1887	C 1192	N 331	O 359	S 5	0	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 errors 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.
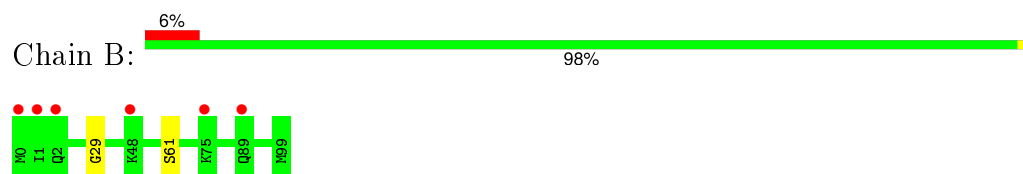
- Molecule 1: HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, A-2 ALPHA CHAIN
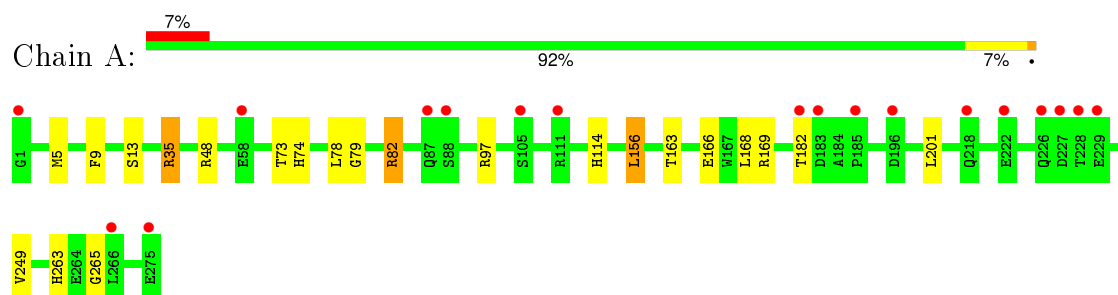
- Molecule 1: HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, A-2 ALPHA CHAIN
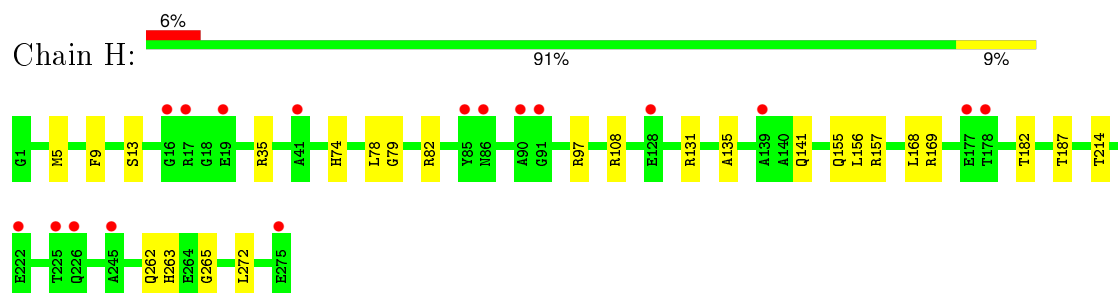
- Molecule 2: BETA-2-MICROGLOBULIN

- Molecule 2: BETA-2-MICROGLOBULIN
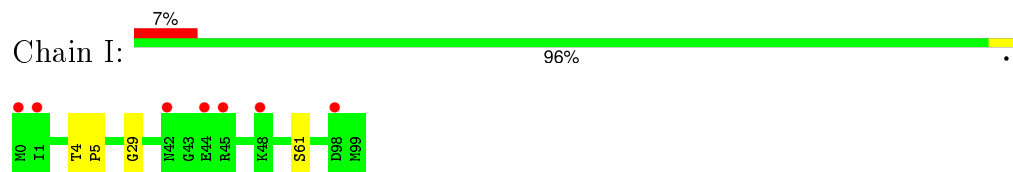
- Molecule 3: SELF-PEPTIDE P1049

• **Molecule 3: SELF-PEPTIDE P1049**

Chain J:

There are no outlier residues recorded for this chain.

• **Molecule 4: AHIII TCR ALPHA CHAIN**

Chain E:

• **Molecule 4: AHIII TCR ALPHA CHAIN**

Chain L:

• **Molecule 5: AHIII TCR BETA CHAIN**

Chain F:

• **Molecule 5: AHIII TCR BETA CHAIN**

Chain M:

4 Data and refinement statistics (i)

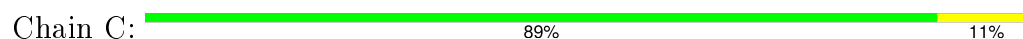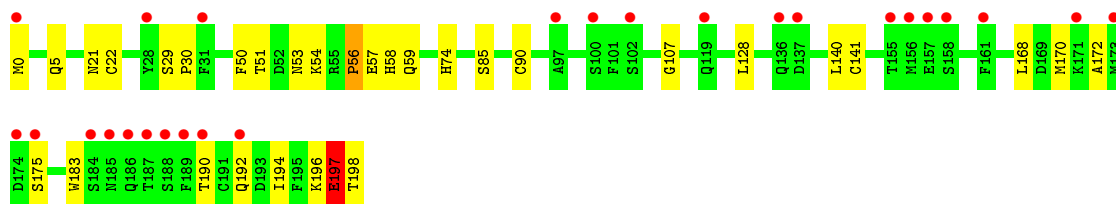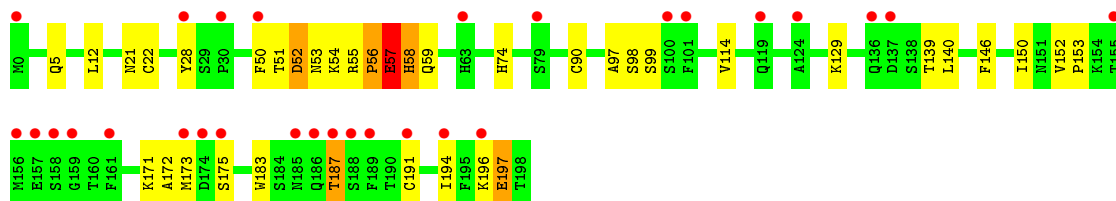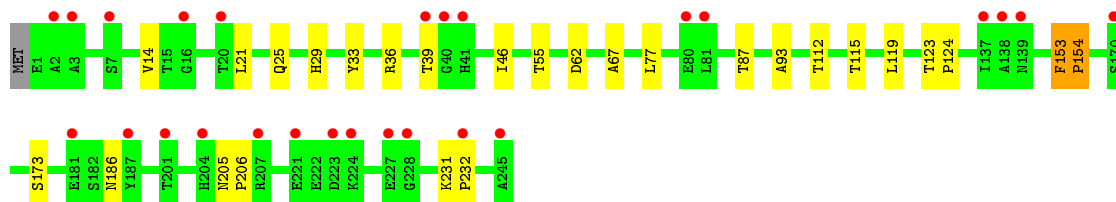| Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	93.42Å 83.89Å 122.27Å 90.00° 92.21° 90.00°	Depositor
Resolution (Å)	122.17 – 2.88 39.67 – 2.88	Depositor EDS
% Data completeness (in resolution range)	94.8 (122.17-2.88) 94.9 (39.67-2.88)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.60 (at 2.90Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.268 , 0.293 0.259 , 0.281	Depositor DCC
R_{free} test set	2007 reflections (5.19%)	DCC
Wilson B-factor (Å^2)	57.9	Xtriage
Anisotropy	0.433	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 37.3	EDS
Estimated twinning fraction	0.000 for h,-k,-l	Xtriage
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Outliers	3 of 40690 reflections (0.007%)	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	13128	wwPDB-VP
Average B, all atoms (Å^2)	47.0	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 45.67 % 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 1.2779e-04. 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.29	0/2308	0.43	0/3133
1	H	0.29	0/2308	0.43	0/3133
2	B	0.31	0/860	0.41	0/1162
2	I	0.30	0/860	0.41	0/1162
3	C	0.43	0/80	0.41	0/108
3	J	0.43	0/80	0.41	0/108
4	E	0.58	2/1556 (0.1%)	0.72	5/2109 (0.2%)
4	L	1.17	12/1556 (0.8%)	1.27	24/2109 (1.1%)
5	F	0.30	0/1943	0.43	0/2644
5	M	0.30	0/1942	0.43	0/2641
All	All	0.52	14/13493 (0.1%)	0.62	29/18309 (0.2%)

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
4	E	0	1
4	L	1	1
5	F	0	1
All	All	1	3

The worst 5 of 14 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	L	58	HIS	C-N	22.04	1.84	1.34
4	L	59	GLN	CA-CB	18.42	1.94	1.53
4	L	50	PHE	C-N	15.50	1.69	1.34
4	L	197	GLU	C-N	15.12	1.68	1.34
4	L	56	PRO	CG-CD	-14.70	1.02	1.50

The worst 5 of 29 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	L	50	PHE	CA-C-N	-21.21	70.53	117.20
4	E	50	PHE	CA-C-N	-15.65	82.76	117.20
4	L	56	PRO	N-CA-CB	-14.70	85.66	103.30
4	L	57	GLU	O-C-N	-14.28	99.85	122.70
4	L	59	GLN	N-CA-CB	14.06	135.91	110.60

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
4	L	57	GLU	CA

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	E	197	GLU	Mainchain
5	F	153	PHE	Peptide
4	L	57	GLU	Mainchain

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	2243	0	2088	13	0
1	H	2243	0	2088	11	2
2	B	837	0	803	1	0
2	I	837	0	803	2	0
3	C	76	0	76	1	0
3	J	76	0	76	0	0
4	E	1521	0	1474	13	3
4	L	1521	0	1473	14	0
5	F	1887	0	1790	16	0
5	M	1887	0	1789	6	1
All	All	13128	0	12460	73	3

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 3.

The worst 5 of 73 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)
4:L:22:CYS:H	4:L:74:HIS:HD2	1.33	0.74
4:E:22:CYS:H	4:E:74:HIS:HD2	1.36	0.73
1:A:97:ARG:HH21	1:A:114:HIS:HE1	1.45	0.65
5:F:25:GLN:HE22	5:F:29:HIS:H	1.43	0.63
2:B:29:GLY:HA2	2:B:61:SER:HB2	1.83	0.61

All (3) 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)
4:E:198:THR:O	1:H:169:ARG:NH2[2_645]	1.71	0.49
4:E:59:GLN:NE2	5:M:84:LEU:CD1[1_545]	1.79	0.41
4:E:198:THR:OG1	1:H:108:ARG:NE[2_645]	2.12	0.08

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	273/275 (99%)	268 (98%)	5 (2%)	0	100	100
1	H	273/275 (99%)	267 (98%)	6 (2%)	0	100	100
2	B	98/100 (98%)	96 (98%)	2 (2%)	0	100	100
2	I	98/100 (98%)	96 (98%)	2 (2%)	0	100	100
3	C	7/9 (78%)	7 (100%)	0	0	100	100
3	J	7/9 (78%)	7 (100%)	0	0	100	100
4	E	190/194 (98%)	174 (92%)	9 (5%)	7 (4%)	4	15
4	L	190/194 (98%)	169 (89%)	14 (7%)	7 (4%)	4	15
5	F	235/238 (99%)	227 (97%)	7 (3%)	1 (0%)	39	73
5	M	233/238 (98%)	228 (98%)	5 (2%)	0	100	100
All	All	1604/1632 (98%)	1539 (96%)	50 (3%)	15 (1%)	21	55

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	E	51	THR
4	E	53	ASN
4	E	56	PRO
4	E	57	GLU
5	F	154	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	230/230 (100%)	225 (98%)	5 (2%)	60	87
1	H	230/230 (100%)	227 (99%)	3 (1%)	76	93
2	B	95/95 (100%)	95 (100%)	0	100	100
2	I	95/95 (100%)	95 (100%)	0	100	100
3	C	7/7 (100%)	7 (100%)	0	100	100
3	J	7/7 (100%)	7 (100%)	0	100	100
4	E	177/177 (100%)	176 (99%)	1 (1%)	90	97
4	L	177/177 (100%)	173 (98%)	4 (2%)	58	86
5	F	204/206 (99%)	202 (99%)	2 (1%)	82	95
5	M	204/206 (99%)	201 (98%)	3 (2%)	72	91
All	All	1426/1430 (100%)	1408 (99%)	18 (1%)	76	93

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	H	35	ARG
1	H	156	LEU
4	L	187	THR
5	F	39	THR
5	F	62	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 23 such

sidechains are listed below:

Mol	Chain	Res	Type
1	H	87	GLN
1	H	141	GLN
5	M	156	HIS
1	H	93	HIS
1	H	155	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 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, 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	275/275 (100%)	0.61	18 (6%)	22	16	44, 49, 51, 52	0
1	H	275/275 (100%)	0.62	17 (6%)	24	18	42, 47, 49, 50	0
2	B	100/100 (100%)	0.69	6 (6%)	25	19	44, 45, 48, 49	0
2	I	100/100 (100%)	0.66	7 (7%)	19	14	44, 45, 49, 50	0
3	C	9/9 (100%)	-0.22	0	100	100	35, 36, 37, 37	0
3	J	9/9 (100%)	-0.07	0	100	100	38, 38, 40, 40	0
4	E	184/194 (94%)	0.94	26 (14%)	4	2	39, 47, 51, 52	0
4	L	184/194 (94%)	1.06	29 (15%)	3	1	42, 47, 50, 52	0
5	F	237/238 (99%)	0.85	26 (10%)	7	4	44, 49, 52, 52	0
5	M	237/238 (99%)	0.77	21 (8%)	12	7	43, 47, 55, 56	0
All	All	1610/1632 (98%)	0.76	150 (9%)	11	7	35, 47, 51, 56	0

The worst 5 of 150 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	M	245	ALA	6.8
5	F	245	ALA	6.8
4	L	173	MET	6.5
4	L	187	THR	6.3
4	L	189	PHE	6.3

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 (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.