wwPDB/EMDataBank EM Map/Model Validation Summary Report (i)

Feb 20, 2018 – 09:49 pm GMT

PDB ID	:	1Z8Y
EMDB ID:	:	EMD-1121
Title	:	Mapping the E2 Glycoprotein of Alphaviruses
Authors	:	Mukhopadhyay, S.; Zhang, W.; Gabler, S.; Chipman, P.R.; Strauss, E.G.; Strauss, J.H.; Baker, T.S.; Kuhn, R.J.; Rossmann, M.G.
Deposited on	:	2005-03-31
Resolution	:	9.00 Å(reported)
Based on PDB ID	:	1WYK, 1I9W

This is a wwPDB/EMDataBank EM Map/Model Validation Summary Report for a publicly released PDB/EMDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the (i) symbol.

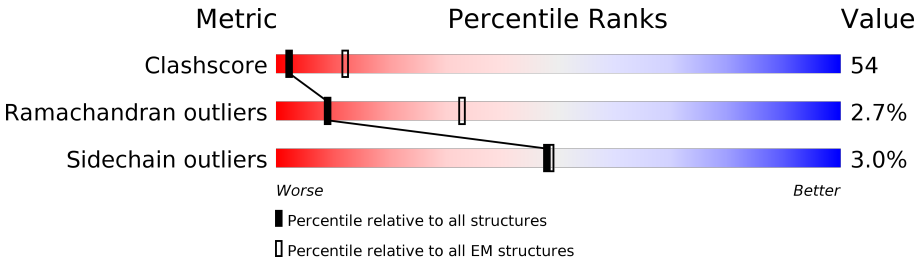
MolProbity	:	4.02b-467
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk30686

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 9.00 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)	EM structures (#Entries)
Clashscore	136279	1886
Ramachandran outliers	132675	1663
Sidechain outliers	132484	1531

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the 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%

Mol	Chain	Length	Quality of chain
1	A	290	34% 62% .
1	C	290	34% 61% .
1	E	290	34% 61% .
1	G	290	35% 60% .
2	B	89	24% 65% 10% .
2	D	89	25% 64% 10% .
2	F	89	24% 65% 10% .
2	H	89	22% 66% 10% .
3	I	31	87% 13%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	K	31	87% 13%
3	M	31	87% 13%
3	O	31	87% 13%
4	J	36	69% 31%
4	L	36	75% 25%
4	N	36	75% 25%
4	P	36	75% 25%
5	Q	151	80% 17% •
5	R	151	83% 15% •
5	S	151	83% 15% •
5	T	151	84% 14% •

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 18071 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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 Spike glycoprotein E1.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	290	Total	C	N	O	S	0	0
			2204	1405	364	419	16		
1	C	290	Total	C	N	O	S	0	0
			2205	1405	365	419	16		
1	E	290	Total	C	N	O	S	0	0
			2205	1405	365	419	16		
1	G	290	Total	C	N	O	S	0	0
			2205	1405	365	419	16		

- Molecule 2 is a protein called Spike glycoprotein E1.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	89	Total	C	N	O	S	0	0
			657	404	111	135	7		
2	D	89	Total	C	N	O	S	0	0
			657	404	111	135	7		
2	F	89	Total	C	N	O	S	0	0
			657	404	111	135	7		
2	H	89	Total	C	N	O	S	0	0
			657	404	111	135	7		

- Molecule 3 is a protein called Spike glycoprotein E1.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	I	31	Total	C	N	O	S	0	0
			237	157	38	38	4		
3	K	31	Total	C	N	O	S	0	0
			237	157	38	38	4		
3	M	31	Total	C	N	O	S	0	0
			237	157	38	38	4		
3	O	31	Total	C	N	O	S	0	0
			237	157	38	38	4		

- Molecule 4 is a protein called Spike glycoprotein E2.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	J	36	Total	C	N	O	S	0	0
			257	164	44	44	5		
4	L	36	Total	C	N	O	S	0	0
			257	164	44	44	5		
4	N	36	Total	C	N	O	S	0	0
			257	164	44	44	5		
4	P	36	Total	C	N	O	S	0	0
			257	164	44	44	5		

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
J	29	LEU	LYS	SEE REMARK 999	UNP P11259
L	29	LEU	LYS	SEE REMARK 999	UNP P11259
N	29	LEU	LYS	SEE REMARK 999	UNP P11259
P	29	LEU	LYS	SEE REMARK 999	UNP P11259

- Molecule 5 is a protein called Capsid protein C.

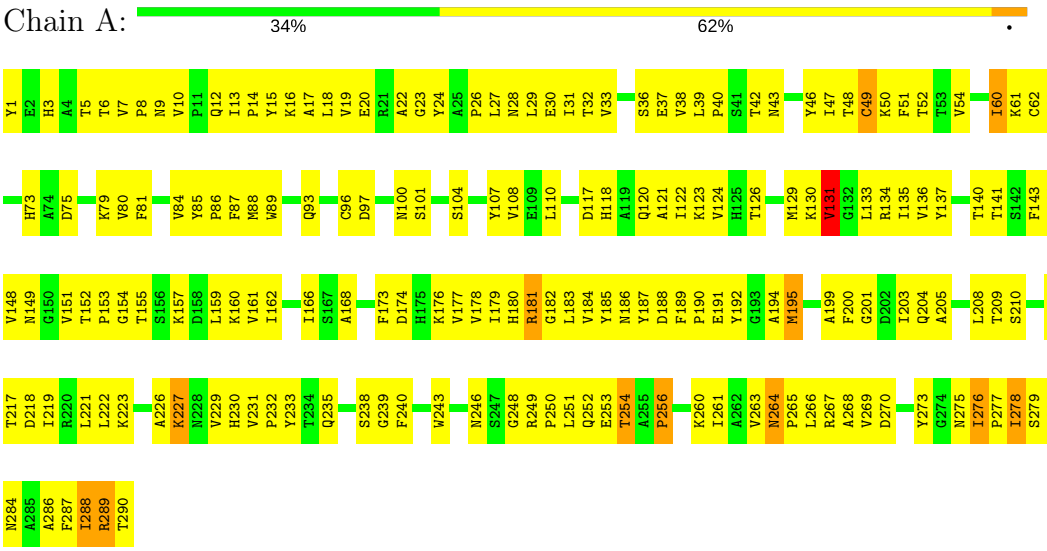
Mol	Chain	Residues	Atoms					AltConf	Trace
5	Q	151	Total	C	N	O	S	0	0
			1162	731	207	219	5		
5	R	151	Total	C	N	O	S	0	0
			1162	731	207	219	5		
5	S	151	Total	C	N	O	S	0	0
			1162	731	207	219	5		
5	T	151	Total	C	N	O	S	0	0
			1162	731	207	219	5		

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. 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. 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.

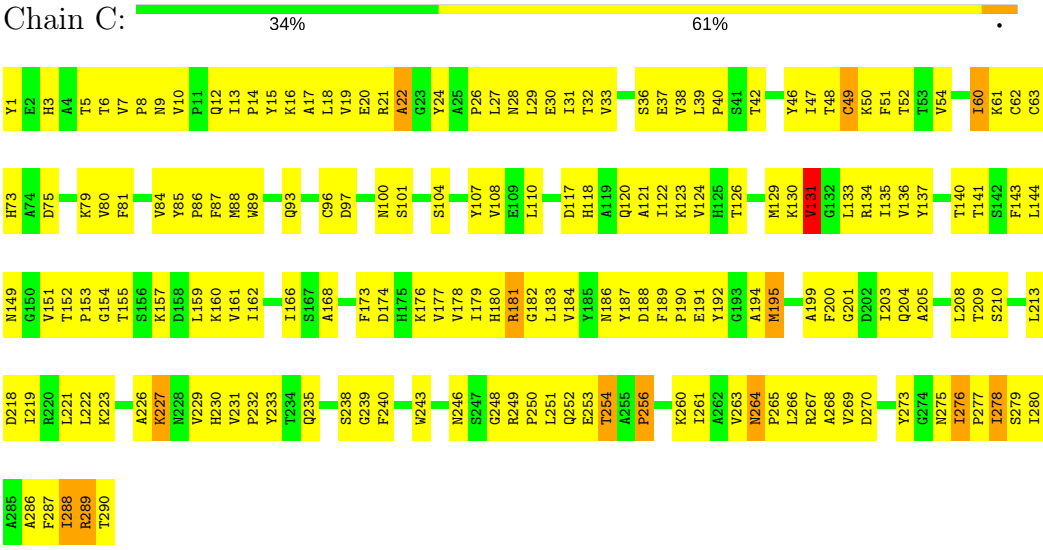
• Molecule 1: Spike glycoprotein E1

Chain A:



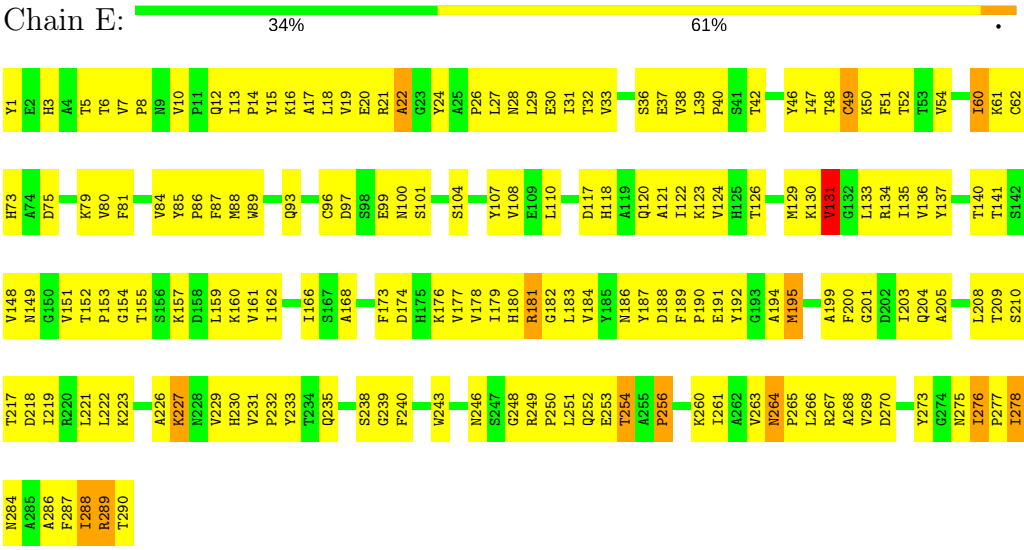
• Molecule 1: Spike glycoprotein E1

Chain C:



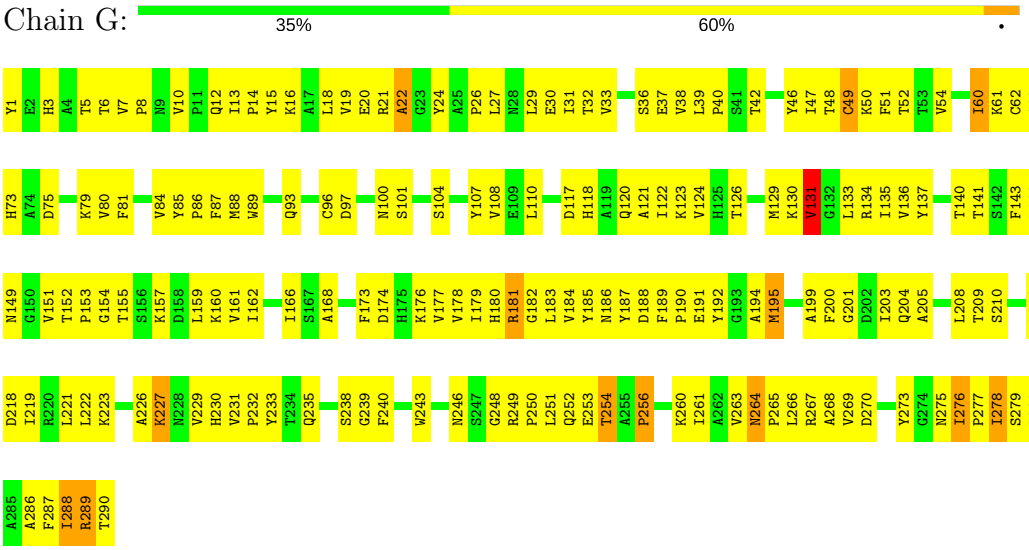
• Molecule 1: Spike glycoprotein E1

Chain E:



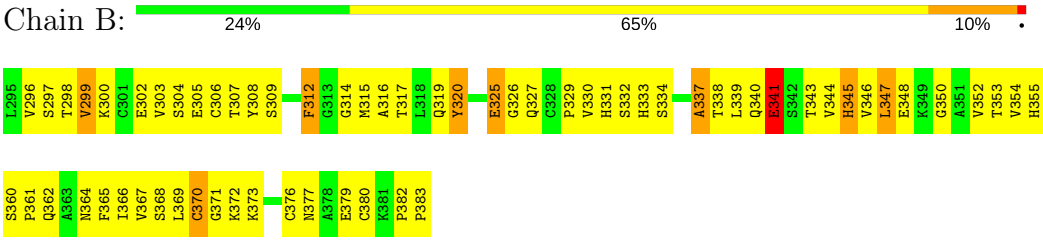
• Molecule 1: Spike glycoprotein E1

Chain G:



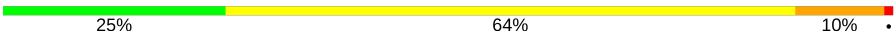
• Molecule 2: Spike glycoprotein E1

Chain B:



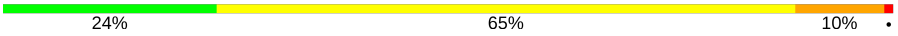
• Molecule 2: Spike glycoprotein E1

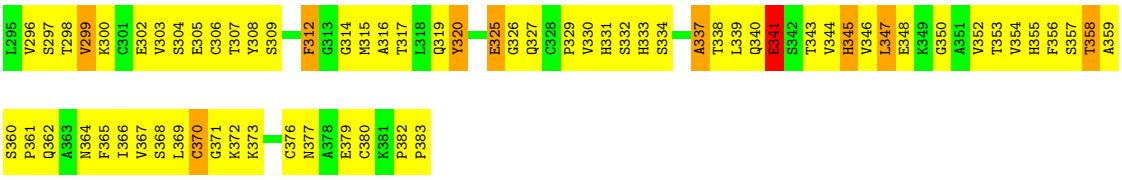
Chain D:



• Molecule 2: Spike glycoprotein E1

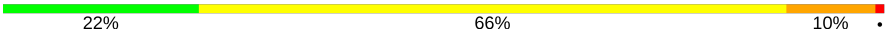
Chain F:





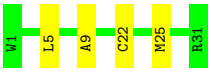
• Molecule 2: Spike glycoprotein E1

Chain H:



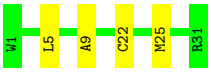
• Molecule 3: Spike glycoprotein E1

Chain I:



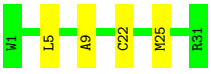
• Molecule 3: Spike glycoprotein E1

Chain K:



• Molecule 3: Spike glycoprotein E1

Chain M:



• Molecule 3: Spike glycoprotein E1

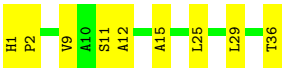
Chain O:

• Molecule 4: Spike glycoprotein E2

Chain J:

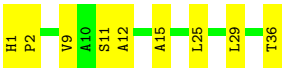
• Molecule 4: Spike glycoprotein E2

Chain L:



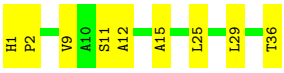
• Molecule 4: Spike glycoprotein E2

Chain N:



• Molecule 4: Spike glycoprotein E2

Chain P:



• Molecule 5: Capsid protein C

Chain Q:

• Molecule 5: Capsid protein C

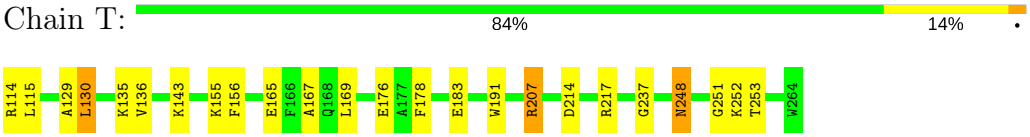
Chain R:

• Molecule 5: Capsid protein C

Chain S:

• Molecule 5: Capsid protein C

Chain T:



4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, I	Depositor
Number of particles used	7085	Depositor
Resolution determination method	Not provided	Depositor
CTF correction method	Fourier transform of each image was modified	Depositor
Microscope	FEI/PHILIPS CM200FEG	Depositor
Voltage (kV)	200	Depositor
Electron dose ($e^-/Å^2$)	18	Depositor
Minimum defocus (nm)	1100	Depositor
Maximum defocus (nm)	2580	Depositor
Magnification	38000	Depositor
Image detector	KODAK SO-163 FILM	Depositor

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	#$\|Z\|>2$	RMSZ	#$\|Z\|>2$
1	A	0.91	0/2261	1.12	3/3083 (0.1%)
1	C	0.91	0/2262	1.12	3/3085 (0.1%)
1	E	0.91	0/2262	1.12	3/3085 (0.1%)
1	G	0.91	0/2262	1.12	3/3085 (0.1%)
2	B	0.83	0/669	1.16	2/908 (0.2%)
2	D	0.83	0/669	1.16	2/908 (0.2%)
2	F	0.83	0/669	1.16	2/908 (0.2%)
2	H	0.83	0/669	1.16	2/908 (0.2%)
3	I	0.37	0/241	0.46	0/321
3	K	0.37	0/241	0.46	0/321
3	M	0.38	0/241	0.46	0/321
3	O	0.37	0/241	0.46	0/321
4	J	0.28	0/259	0.43	0/354
4	L	0.28	0/259	0.43	0/354
4	N	0.28	0/259	0.43	0/354
4	P	0.28	0/259	0.43	0/354
5	Q	0.52	0/1190	0.81	1/1608 (0.1%)
5	R	0.52	0/1190	0.81	1/1608 (0.1%)
5	S	0.52	0/1190	0.81	1/1608 (0.1%)
5	T	0.52	0/1190	0.81	1/1608 (0.1%)
All	All	0.77	0/18483	1.00	24/25102 (0.1%)

There are no bond length outliers.

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	G	254	THR	C-N-CA	5.47	135.38	121.70
1	A	254	THR	C-N-CA	5.47	135.38	121.70
1	E	254	THR	C-N-CA	5.47	135.37	121.70
1	C	254	THR	C-N-CA	5.46	135.35	121.70
1	C	131	VAL	C-N-CA	5.45	133.74	122.30

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	2204	0	2157	368	0
1	C	2205	0	2159	369	0
1	E	2205	0	2161	359	0
1	G	2205	0	2161	328	0
2	B	657	0	630	152	0
2	D	657	0	630	149	0
2	F	657	0	630	144	0
2	H	657	0	630	116	0
3	I	237	0	254	2	0
3	K	237	0	254	2	0
3	M	237	0	254	2	0
3	O	237	0	254	2	0
4	J	257	0	281	18	0
4	L	257	0	281	8	0
4	N	257	0	281	6	0
4	P	257	0	281	7	0
5	Q	1162	0	1131	28	0
5	R	1162	0	1131	14	0
5	S	1162	0	1131	13	0
5	T	1162	0	1131	13	0
All	All	18071	0	17822	1929	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 54.

The worst 5 of 1929 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:A:18:LEU:HD23	2:B:331:HIS:CD2	1.49	1.46
1:C:18:LEU:HB2	2:D:331:HIS:NE2	1.35	1.41
1:C:28:ASN:CG	2:D:331:HIS:ND1	1.76	1.37
1:A:16:LYS:HZ2	2:B:338:THR:CG2	1.38	1.34
1:A:18:LEU:HB2	2:B:331:HIS:NE2	1.42	1.33

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 PDB entries followed by that with respect to all EM entries.

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	288/290 (99%)	250 (87%)	30 (10%)	8 (3%)	5	37
1	C	288/290 (99%)	250 (87%)	30 (10%)	8 (3%)	5	37
1	E	288/290 (99%)	250 (87%)	30 (10%)	8 (3%)	5	37
1	G	288/290 (99%)	250 (87%)	30 (10%)	8 (3%)	5	37
2	B	87/89 (98%)	63 (72%)	16 (18%)	8 (9%)	1	14
2	D	87/89 (98%)	63 (72%)	16 (18%)	8 (9%)	1	14
2	F	87/89 (98%)	63 (72%)	16 (18%)	8 (9%)	1	14
2	H	87/89 (98%)	63 (72%)	16 (18%)	8 (9%)	1	14
3	I	29/31 (94%)	28 (97%)	1 (3%)	0	100	100
3	K	29/31 (94%)	28 (97%)	1 (3%)	0	100	100
3	M	29/31 (94%)	28 (97%)	1 (3%)	0	100	100
3	O	29/31 (94%)	28 (97%)	1 (3%)	0	100	100
4	J	34/36 (94%)	33 (97%)	1 (3%)	0	100	100
4	L	34/36 (94%)	33 (97%)	1 (3%)	0	100	100
4	N	34/36 (94%)	33 (97%)	1 (3%)	0	100	100
4	P	34/36 (94%)	33 (97%)	1 (3%)	0	100	100
5	Q	149/151 (99%)	144 (97%)	5 (3%)	0	100	100
5	R	149/151 (99%)	144 (97%)	5 (3%)	0	100	100
5	S	149/151 (99%)	144 (97%)	5 (3%)	0	100	100
5	T	149/151 (99%)	144 (97%)	5 (3%)	0	100	100
All	All	2348/2388 (98%)	2072 (88%)	212 (9%)	64 (3%)	9	38

5 of 64 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	131	VAL
1	A	195	MET
1	A	227	LYS
1	A	288	ILE
2	B	299	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 PDB entries followed by that with respect to all EM entries.

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	242/243 (100%)	237 (98%)	5 (2%)	56	78
1	C	243/243 (100%)	238 (98%)	5 (2%)	56	78
1	E	243/243 (100%)	238 (98%)	5 (2%)	56	78
1	G	243/243 (100%)	238 (98%)	5 (2%)	56	78
2	B	77/77 (100%)	74 (96%)	3 (4%)	35	63
2	D	77/77 (100%)	74 (96%)	3 (4%)	35	63
2	F	77/77 (100%)	74 (96%)	3 (4%)	35	63
2	H	77/77 (100%)	74 (96%)	3 (4%)	35	63
3	I	25/25 (100%)	25 (100%)	0	100	100
3	K	25/25 (100%)	25 (100%)	0	100	100
3	M	25/25 (100%)	25 (100%)	0	100	100
3	O	25/25 (100%)	25 (100%)	0	100	100
4	J	28/28 (100%)	28 (100%)	0	100	100
4	L	28/28 (100%)	28 (100%)	0	100	100
4	N	28/28 (100%)	28 (100%)	0	100	100
4	P	28/28 (100%)	28 (100%)	0	100	100
5	Q	122/123 (99%)	115 (94%)	7 (6%)	23	53
5	R	122/123 (99%)	115 (94%)	7 (6%)	23	53
5	S	122/123 (99%)	115 (94%)	7 (6%)	23	53
5	T	122/123 (99%)	115 (94%)	7 (6%)	23	53

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1979/1984 (100%)	1919 (97%)	60 (3%)	48 69

5 of 60 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	278	ILE
5	Q	155	LYS
5	T	183	GLU
2	H	347	LEU
5	Q	191	TRP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 62 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	100	ASN
2	F	362	GLN
5	S	197	GLN
1	E	125	HIS
1	G	43	ASN

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.