wwPDB X-ray Structure Validation Summary Report (i)

May 26, 2020 – 07:35 pm BST

PDB ID	:	6GLX
Title	:	Structure of galectin-10 in complex with the Fab fragment of a Charcot-Leyden crystal solubilizing antibody, 4E8
Authors	:	Verstraete, K.; Verschueren, K.; Savvides, S.N.
Deposited on	:	2018-05-23
Resolution	:	3.40 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 (i) symbol.

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

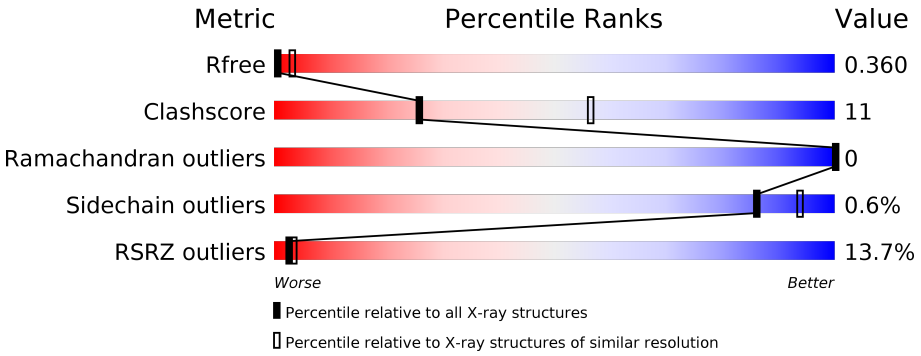
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 3.40 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}	130704	1026 (3.48-3.32)
Clashscore	141614	1055 (3.48-3.32)
Ramachandran outliers	138981	1038 (3.48-3.32)
Sidechain outliers	138945	1038 (3.48-3.32)
RSRZ outliers	127900	2173 (3.50-3.30)

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	179	2% 63% 14% • 23%
1	B	179	4% 59% 17% • 23%
2	C	232	15% 66% 31% •
2	H	232	13% 64% 28% 8%
3	D	213	11% 76% 24%
3	L	213	27% 78% 22%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 8685 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 Galectin-10.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	138	Total	C	N	O	S	0	0	0
			1120	717	186	210	7			
1	B	137	Total	C	N	O	S	0	0	0
			1108	708	185	208	7			

There are 74 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-36	MET	-	initiating methionine	UNP Q05315
A	-35	ALA	-	expression tag	UNP Q05315
A	-34	SER	-	expression tag	UNP Q05315
A	-33	THR	-	expression tag	UNP Q05315
A	-32	THR	-	expression tag	UNP Q05315
A	-31	HIS	-	expression tag	UNP Q05315
A	-30	HIS	-	expression tag	UNP Q05315
A	-29	HIS	-	expression tag	UNP Q05315
A	-28	HIS	-	expression tag	UNP Q05315
A	-27	HIS	-	expression tag	UNP Q05315
A	-26	HIS	-	expression tag	UNP Q05315
A	-25	ASP	-	expression tag	UNP Q05315
A	-24	THR	-	expression tag	UNP Q05315
A	-23	ASP	-	expression tag	UNP Q05315
A	-22	ILE	-	expression tag	UNP Q05315
A	-21	PRO	-	expression tag	UNP Q05315
A	-20	THR	-	expression tag	UNP Q05315
A	-19	THR	-	expression tag	UNP Q05315
A	-18	GLY	-	expression tag	UNP Q05315
A	-17	GLY	-	expression tag	UNP Q05315
A	-16	GLY	-	expression tag	UNP Q05315
A	-15	SER	-	expression tag	UNP Q05315
A	-14	ARG	-	expression tag	UNP Q05315
A	-13	PRO	-	expression tag	UNP Q05315
A	-12	ASP	-	expression tag	UNP Q05315

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	-11	ASP	-	expression tag	UNP Q05315
A	-10	ASP	-	expression tag	UNP Q05315
A	-9	ASP	-	expression tag	UNP Q05315
A	-8	LYS	-	expression tag	UNP Q05315
A	-7	GLU	-	expression tag	UNP Q05315
A	-6	ASN	-	expression tag	UNP Q05315
A	-5	LEU	-	expression tag	UNP Q05315
A	-4	TYR	-	expression tag	UNP Q05315
A	-3	PHE	-	expression tag	UNP Q05315
A	-2	GLN	-	expression tag	UNP Q05315
A	-1	GLY	-	expression tag	UNP Q05315
A	0	HIS	-	expression tag	UNP Q05315
B	-36	MET	-	initiating methionine	UNP Q05315
B	-35	ALA	-	expression tag	UNP Q05315
B	-34	SER	-	expression tag	UNP Q05315
B	-33	THR	-	expression tag	UNP Q05315
B	-32	THR	-	expression tag	UNP Q05315
B	-31	HIS	-	expression tag	UNP Q05315
B	-30	HIS	-	expression tag	UNP Q05315
B	-29	HIS	-	expression tag	UNP Q05315
B	-28	HIS	-	expression tag	UNP Q05315
B	-27	HIS	-	expression tag	UNP Q05315
B	-26	HIS	-	expression tag	UNP Q05315
B	-25	ASP	-	expression tag	UNP Q05315
B	-24	THR	-	expression tag	UNP Q05315
B	-23	ASP	-	expression tag	UNP Q05315
B	-22	ILE	-	expression tag	UNP Q05315
B	-21	PRO	-	expression tag	UNP Q05315
B	-20	THR	-	expression tag	UNP Q05315
B	-19	THR	-	expression tag	UNP Q05315
B	-18	GLY	-	expression tag	UNP Q05315
B	-17	GLY	-	expression tag	UNP Q05315
B	-16	GLY	-	expression tag	UNP Q05315
B	-15	SER	-	expression tag	UNP Q05315
B	-14	ARG	-	expression tag	UNP Q05315
B	-13	PRO	-	expression tag	UNP Q05315
B	-12	ASP	-	expression tag	UNP Q05315
B	-11	ASP	-	expression tag	UNP Q05315
B	-10	ASP	-	expression tag	UNP Q05315
B	-9	ASP	-	expression tag	UNP Q05315
B	-8	LYS	-	expression tag	UNP Q05315
B	-7	GLU	-	expression tag	UNP Q05315

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-6	ASN	-	expression tag	UNP Q05315
B	-5	LEU	-	expression tag	UNP Q05315
B	-4	TYR	-	expression tag	UNP Q05315
B	-3	PHE	-	expression tag	UNP Q05315
B	-2	GLN	-	expression tag	UNP Q05315
B	-1	GLY	-	expression tag	UNP Q05315
B	0	HIS	-	expression tag	UNP Q05315

- Molecule 2 is a protein called Fab 4E8 - Heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	224	Total	C	N	O	S	0	0	0
			1671	1055	273	337	6			
2	H	214	Total	C	N	O	S	0	0	0
			1600	1013	260	321	6			

- Molecule 3 is a protein called Fab 4E8 - Light chain.

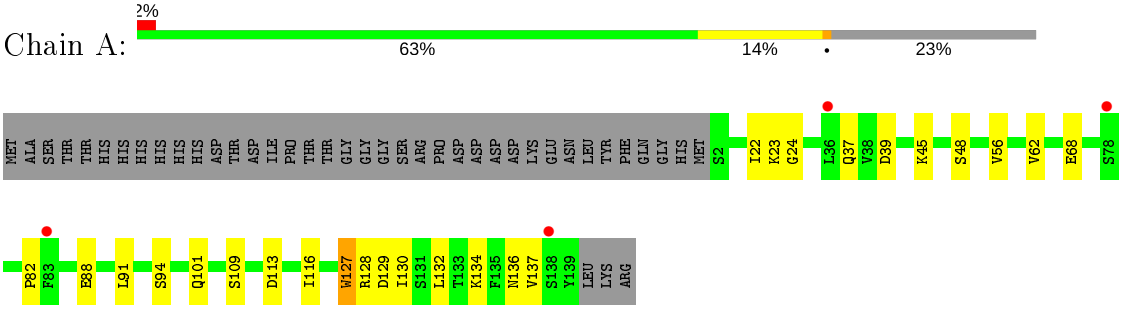
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	213	Total	C	N	O	S	0	0	0
			1593	992	267	329	5			
3	L	213	Total	C	N	O	S	0	0	0
			1593	992	267	329	5			

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 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.

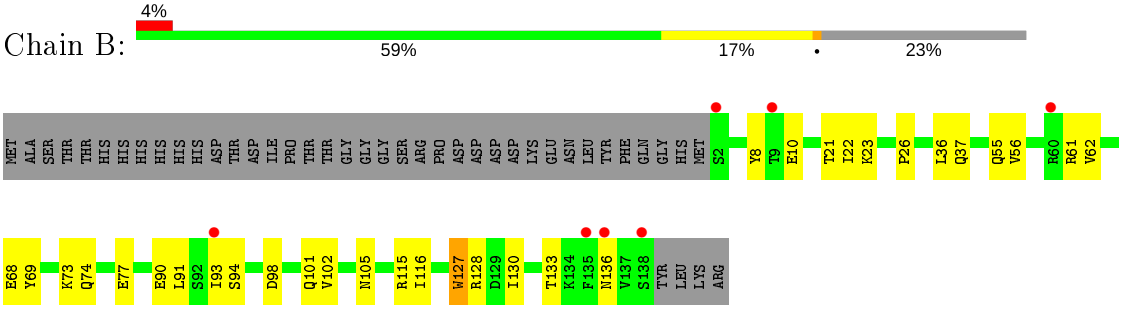
- Molecule 1: Galectin-10

Chain A:



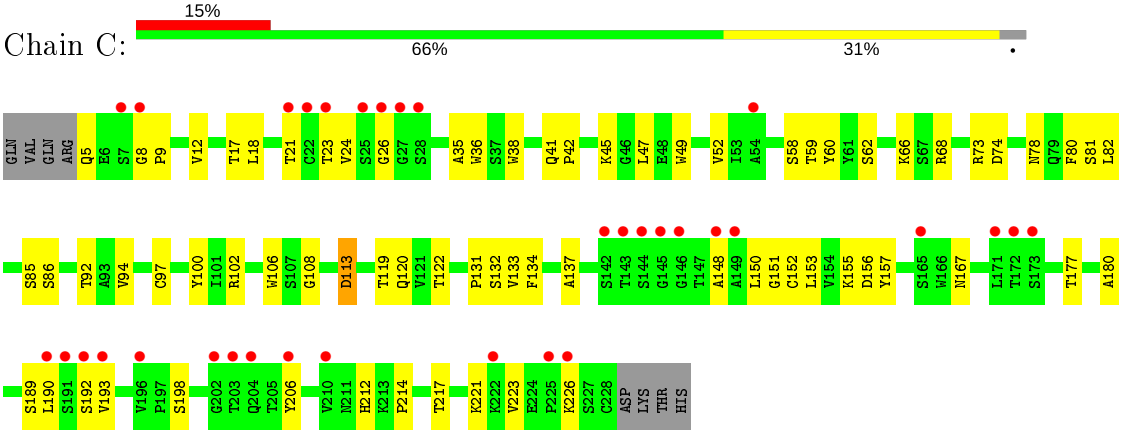
- Molecule 1: Galectin-10

Chain B:



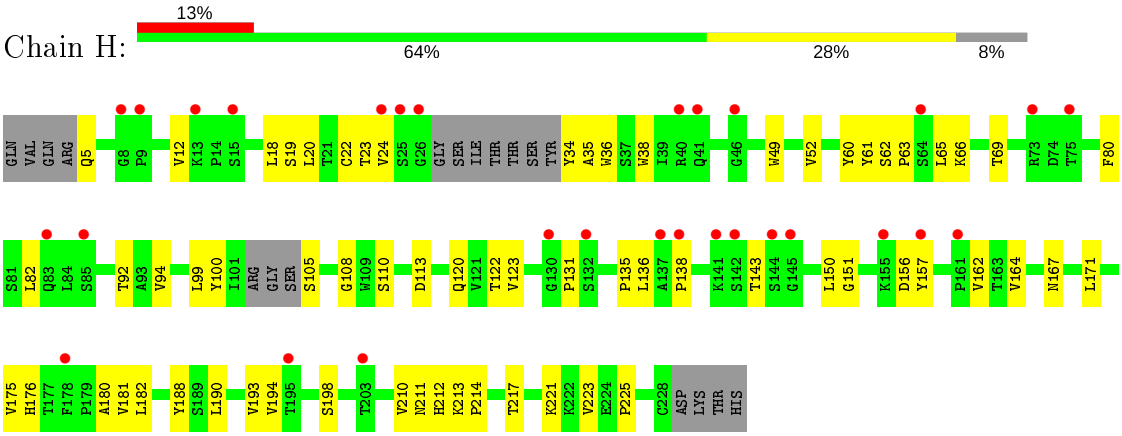
- Molecule 2: Fab 4E8 - Heavy chain

Chain C:



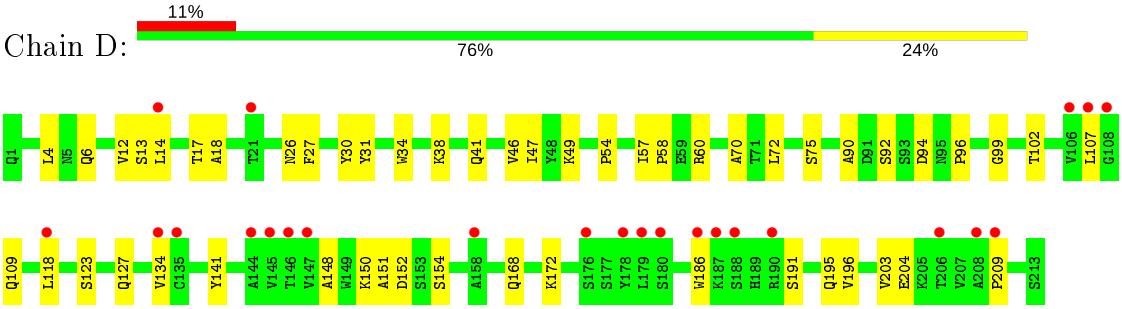
- Molecule 2: Fab 4E8 - Heavy chain

Chain H:



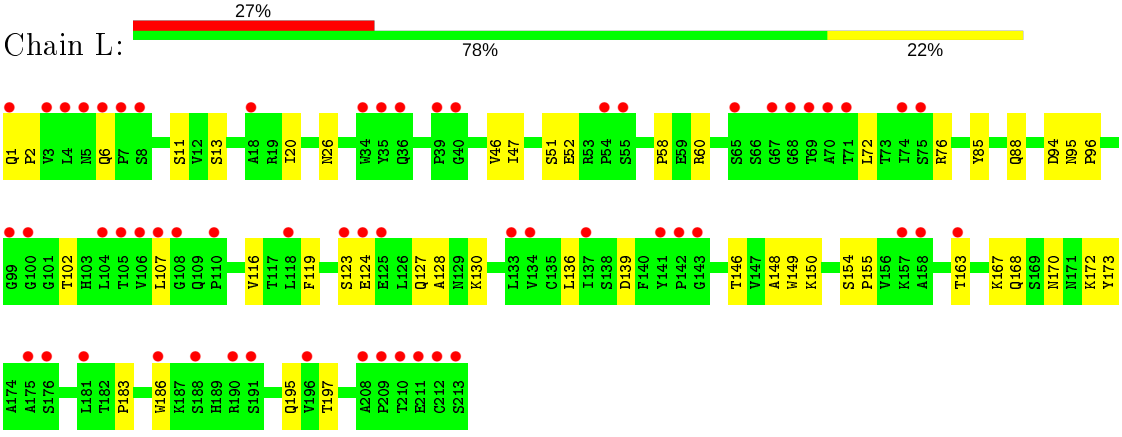
- Molecule 3: Fab 4E8 - Light chain

Chain D:



- Molecule 3: Fab 4E8 - Light chain

Chain L:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	41.33Å 150.45Å 94.50Å 90.00° 96.31° 90.00°	Depositor
Resolution (Å)	46.97 – 3.40 46.97 – 3.40	Depositor EDS
% Data completeness (in resolution range)	98.4 (46.97-3.40) 98.4 (46.97-3.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.33 (at 3.40Å)	Xtriage
Refinement program	PHENIX (1.12_2829: ???)	Depositor
R, R_{free}	0.303 , 0.360 0.303 , 0.360	Depositor DCC
R_{free} test set	1050 reflections (6.72%)	wwPDB-VP
Wilson B-factor (Å^2)	95.6	Xtriage
Anisotropy	0.122	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 86.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.25$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.84	EDS
Total number of atoms	8685	wwPDB-VP
Average B, all atoms (Å^2)	124.0	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 27.59 % 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 2.1409e-03.*

[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	0.25	0/1148	0.44	0/1554
1	B	0.25	0/1135	0.45	0/1536
2	C	0.26	0/1717	0.49	0/2348
2	H	0.25	0/1643	0.49	0/2245
3	D	0.26	0/1633	0.48	0/2229
3	L	0.25	0/1633	0.47	0/2229
All	All	0.26	0/8909	0.47	0/12141

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	A	0	1
1	B	0	1
All	All	0	2

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	127	TRP	Peptide
1	B	127	TRP	Peptide

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	1120	0	1093	17	0
1	B	1108	0	1084	23	0
2	C	1671	0	1623	43	0
2	H	1600	0	1553	47	0
3	D	1593	0	1520	31	0
3	L	1593	0	1520	35	0
All	All	8685	0	8393	180	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 11.

The worst 5 of 180 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)
3:L:139:ASP:H	3:L:168:GLN:HE22	1.18	0.88
2:H:5:GLN:HB2	2:H:23:THR:HB	1.67	0.76
1:B:68:GLU:HG2	1:B:116:ILE:HD11	1.70	0.74
1:A:68:GLU:HG2	1:A:116:ILE:HD11	1.68	0.73
2:C:8:GLY:H	2:C:9:PRO:HD3	1.53	0.73

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	136/179 (76%)	131 (96%)	5 (4%)	0	100	100
1	B	135/179 (75%)	130 (96%)	5 (4%)	0	100	100
2	C	222/232 (96%)	211 (95%)	11 (5%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	H	208/232 (90%)	202 (97%)	6 (3%)	0	100	100
3	D	211/213 (99%)	203 (96%)	8 (4%)	0	100	100
3	L	211/213 (99%)	202 (96%)	9 (4%)	0	100	100
All	All	1123/1248 (90%)	1079 (96%)	44 (4%)	0	100	100

There are no Ramachandran outliers to report.

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	127/163 (78%)	127 (100%)	0	100	100
1	B	126/163 (77%)	126 (100%)	0	100	100
2	C	193/201 (96%)	190 (98%)	3 (2%)	62	81
2	H	185/201 (92%)	185 (100%)	0	100	100
3	D	177/177 (100%)	176 (99%)	1 (1%)	86	94
3	L	177/177 (100%)	175 (99%)	2 (1%)	73	86
All	All	985/1082 (91%)	979 (99%)	6 (1%)	86	94

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	152	CYS
3	L	88	GLN
3	D	26	ASN
2	C	113	ASP
3	L	26	ASN

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

Mol	Chain	Res	Type
3	L	168	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	138/179 (77%)	0.28	4 (2%) 51 50	43, 87, 143, 177	0
1	B	137/179 (76%)	0.31	7 (5%) 28 28	31, 88, 166, 265	0
2	C	224/232 (96%)	0.88	34 (15%) 2 2	39, 119, 236, 340	0
2	H	214/232 (92%)	0.82	29 (13%) 3 3	45, 127, 216, 260	0
3	D	213/213 (100%)	0.75	24 (11%) 5 6	40, 121, 226, 268	0
3	L	213/213 (100%)	1.39	58 (27%) 0 0	58, 160, 274, 382	0
All	All	1139/1248 (91%)	0.80	156 (13%) 3 3	31, 120, 231, 382	0

The worst 5 of 156 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	L	213	SER	11.1
2	H	144	SER	8.5
3	D	108	GLY	8.3
3	L	107	LEU	7.8
2	C	28	SER	6.8

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 (i)

There are no such residues in this entry.