wwPDB X-ray Structure Validation Summary Report (i)

Feb 22, 2024 – 03:05 AM JST

PDB ID	:	8IJ8
Title	:	Crystal structure of alcohol dehydrogenase M4 mutant from Burkholderia gladioli
Authors	:	Han, X.; Mei, Z.L.; Liu, W.D.; Sun, Z.T.; Ma, J.A.
Deposited on	:	2023-02-26
Resolution	:	2.38 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

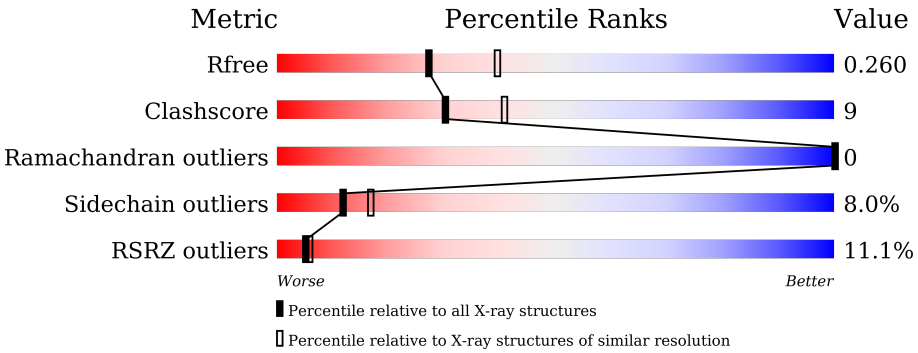
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.38 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.



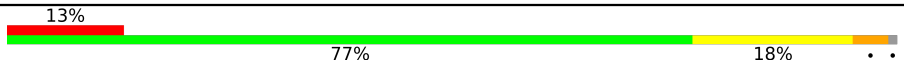
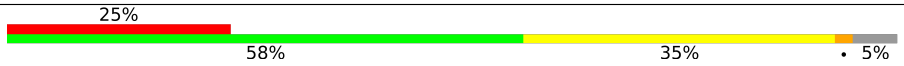
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	5509 (2.40-2.36)
Clashscore	141614	6082 (2.40-2.36)
Ramachandran outliers	138981	5973 (2.40-2.36)
Sidechain outliers	138945	5975 (2.40-2.36)
RSRZ outliers	127900	5397 (2.40-2.36)

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 <=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	251	8% 85% 12% ·
1	B	251	6% 84% 14% ·
1	C	251	11% 79% 18% ··
1	D	251	8% 82% 16% ··
1	E	251	8% 78% 20% ··
1	F	251	9% 78% 20% ··

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	251	
1	H	251	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 14566 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 Putative short-chain dehydrogenases/reductase family protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	250	Total	C	N	O	S	0	0	0
			1806	1121	327	356	2			
1	B	250	Total	C	N	O	S	0	0	0
			1802	1118	326	356	2			
1	C	249	Total	C	N	O	S	0	0	0
			1786	1111	316	357	2			
1	D	249	Total	C	N	O	S	0	0	0
			1799	1118	323	356	2			
1	E	249	Total	C	N	O	S	0	0	0
			1790	1113	317	358	2			
1	F	249	Total	C	N	O	S	0	0	0
			1776	1105	317	352	2			
1	G	248	Total	C	N	O	S	0	0	0
			1754	1093	306	353	2			
1	H	239	Total	C	N	O	S	0	0	0
			1668	1046	286	334	2			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	84	ILE	VAL	engineered mutation	UNP F2LIG4
A	92	ALA	GLY	engineered mutation	UNP F2LIG4
A	140	LYS	ALA	engineered mutation	UNP F2LIG4
A	203	THR	LEU	engineered mutation	UNP F2LIG4
B	84	ILE	VAL	engineered mutation	UNP F2LIG4
B	92	ALA	GLY	engineered mutation	UNP F2LIG4
B	140	LYS	ALA	engineered mutation	UNP F2LIG4
B	203	THR	LEU	engineered mutation	UNP F2LIG4
C	84	ILE	VAL	engineered mutation	UNP F2LIG4
C	92	ALA	GLY	engineered mutation	UNP F2LIG4
C	140	LYS	ALA	engineered mutation	UNP F2LIG4
C	203	THR	LEU	engineered mutation	UNP F2LIG4
D	84	ILE	VAL	engineered mutation	UNP F2LIG4

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	92	ALA	GLY	engineered mutation	UNP F2LIG4
D	140	LYS	ALA	engineered mutation	UNP F2LIG4
D	203	THR	LEU	engineered mutation	UNP F2LIG4
E	84	ILE	VAL	engineered mutation	UNP F2LIG4
E	92	ALA	GLY	engineered mutation	UNP F2LIG4
E	140	LYS	ALA	engineered mutation	UNP F2LIG4
E	203	THR	LEU	engineered mutation	UNP F2LIG4
F	84	ILE	VAL	engineered mutation	UNP F2LIG4
F	92	ALA	GLY	engineered mutation	UNP F2LIG4
F	140	LYS	ALA	engineered mutation	UNP F2LIG4
F	203	THR	LEU	engineered mutation	UNP F2LIG4
G	84	ILE	VAL	engineered mutation	UNP F2LIG4
G	92	ALA	GLY	engineered mutation	UNP F2LIG4
G	140	LYS	ALA	engineered mutation	UNP F2LIG4
G	203	THR	LEU	engineered mutation	UNP F2LIG4
H	84	ILE	VAL	engineered mutation	UNP F2LIG4
H	92	ALA	GLY	engineered mutation	UNP F2LIG4
H	140	LYS	ALA	engineered mutation	UNP F2LIG4
H	203	THR	LEU	engineered mutation	UNP F2LIG4

- Molecule 2 is water.

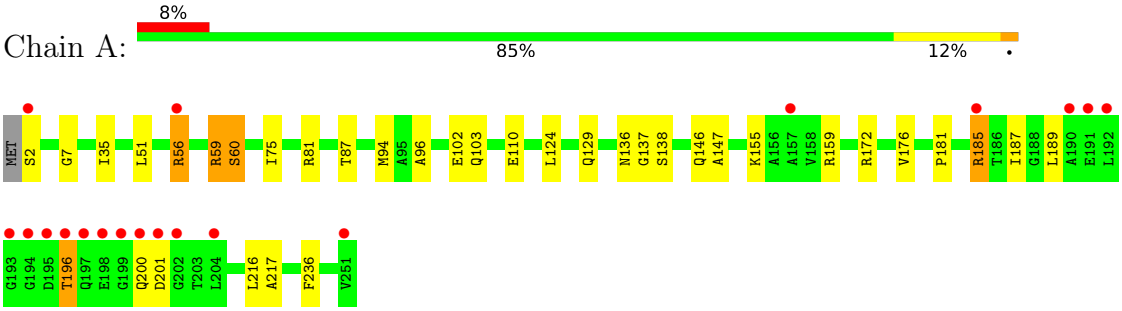
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	79	Total O 79 79	0	0
2	B	76	Total O 76 76	0	0
2	C	48	Total O 48 48	0	0
2	D	46	Total O 46 46	0	0
2	E	48	Total O 48 48	0	0
2	F	39	Total O 39 39	0	0
2	G	33	Total O 33 33	0	0
2	H	16	Total O 16 16	0	0

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.

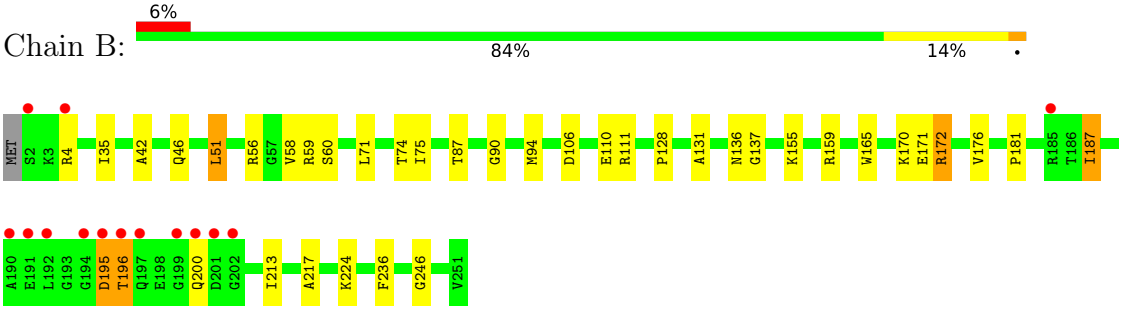
• Molecule 1: Putative short-chain dehydrogenases/reductase family protein

Chain A:



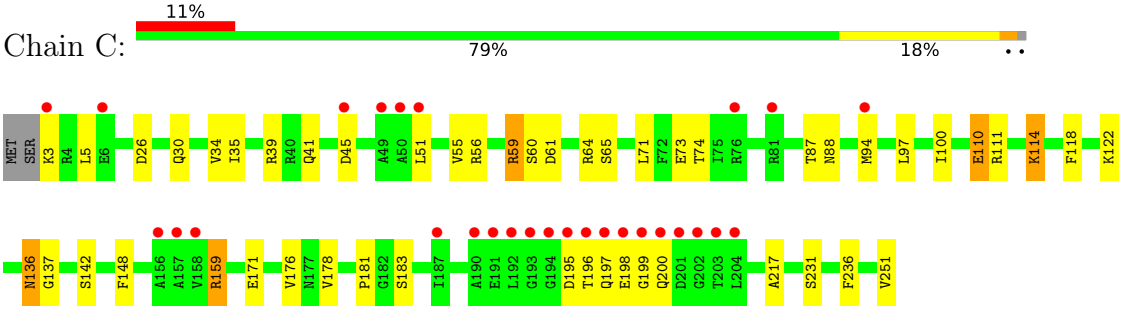
• Molecule 1: Putative short-chain dehydrogenases/reductase family protein

Chain B:



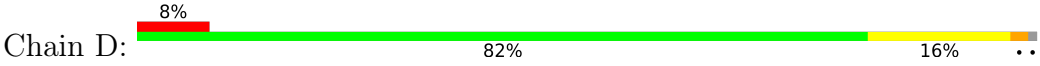
• Molecule 1: Putative short-chain dehydrogenases/reductase family protein

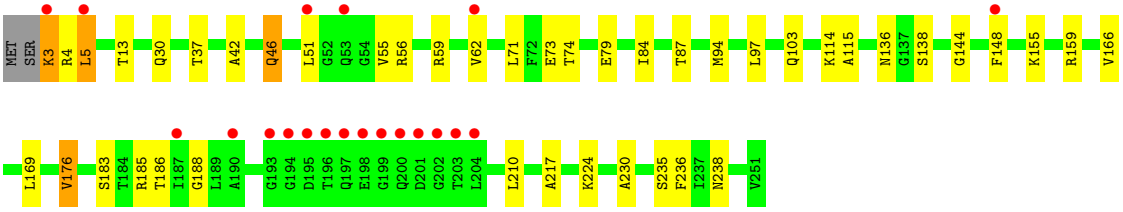
Chain C:



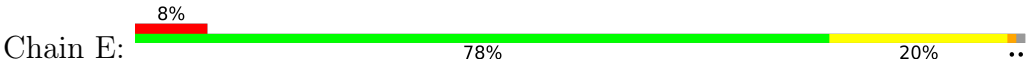
• Molecule 1: Putative short-chain dehydrogenases/reductase family protein

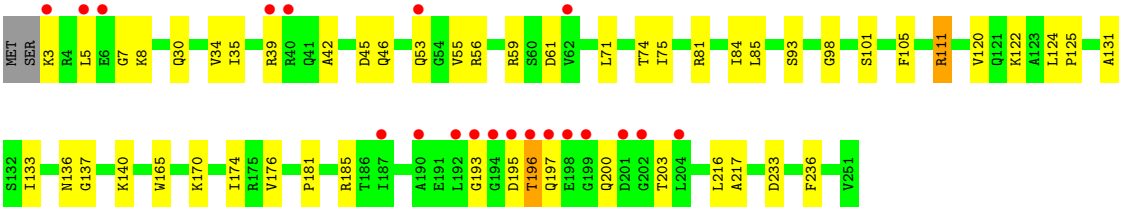
Chain D:



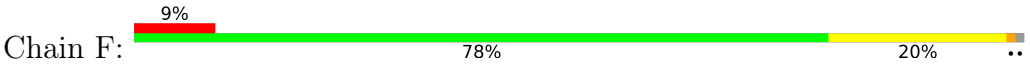


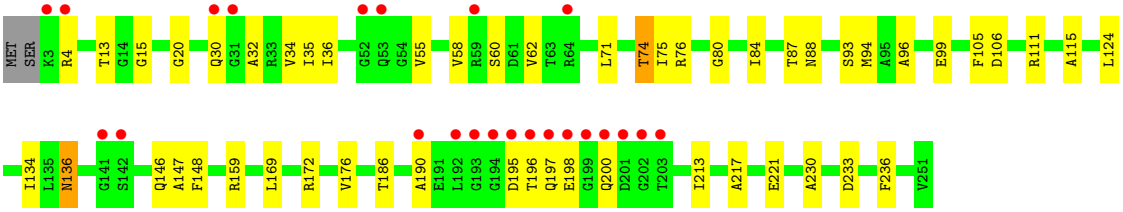
• Molecule 1: Putative short-chain dehydrogenases/reductase family protein



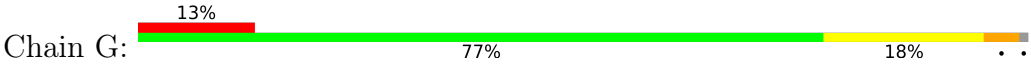


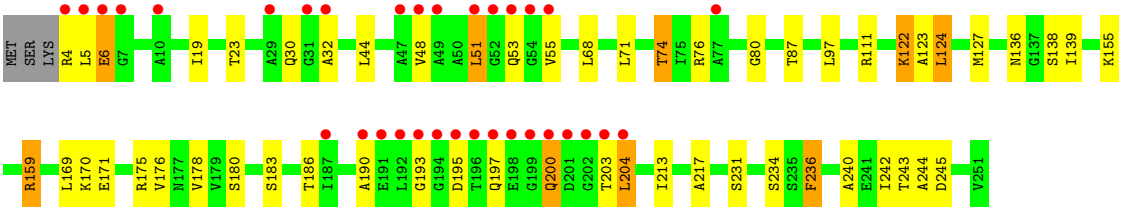
• Molecule 1: Putative short-chain dehydrogenases/reductase family protein



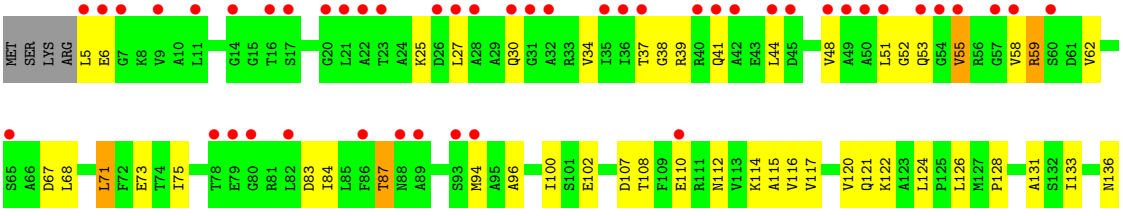


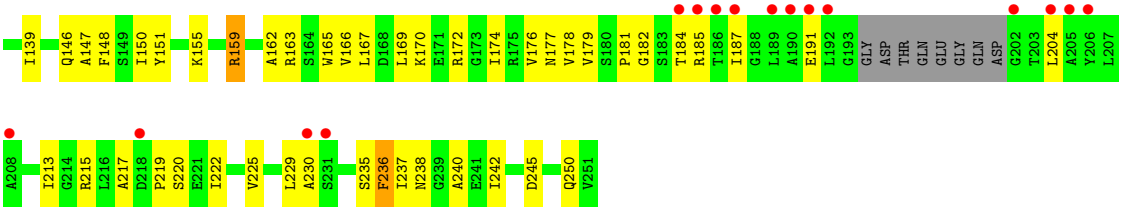
• Molecule 1: Putative short-chain dehydrogenases/reductase family protein





• Molecule 1: Putative short-chain dehydrogenases/reductase family protein




I139
Q146
A147
F148
S149
I150
Y151
K155
R159
A162
R163
S164
W165
V166
L167
D168
L169
K170
E171
R172
G173
I174
R175
V176
N177
V178
V179
S180
P181
G182
S183
T184
R185
T186
I187
G188
L189
A190
E191
L192
G193
GLY
ASP
THR
GLN
GLU
GLY
GLN
ASP
G202
T203
L204
A205
Y206
L207
A208
I213
G214
R215
L216
A217
D218
P219
S220
E221
I222
V225
L229
A230
S231
S235
F236
I237
N238
G239
A240
E241
I242
D245
Q250
V251

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 4	Depositor
Cell constants a, b, c, α, β, γ	141.31Å 141.31Å 164.76Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.96 – 2.38 49.96 – 2.38	Depositor EDS
% Data completeness (in resolution range)	98.0 (49.96-2.38) 98.0 (49.96-2.38)	Depositor EDS
R_{merge}	0.17	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.48 (at 2.37Å)	Xtriage
Refinement program	PHENIX 1.1932-4158	Depositor
R, R_{free}	0.225 , 0.258 0.231 , 0.260	Depositor DCC
R_{free} test set	6312 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	41.4	Xtriage
Anisotropy	0.487	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 51.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.048 for h,-k,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	14566	wwPDB-VP
Average B, all atoms (Å^2)	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.35% of the height of the origin peak. No significant pseudotranslation is detected.*

[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	0.40	0/1823	0.64	0/2465
1	B	0.38	0/1819	0.63	0/2461
1	C	0.43	0/1803	0.63	0/2441
1	D	0.40	0/1816	0.61	0/2456
1	E	0.38	0/1807	0.60	0/2446
1	F	0.43	0/1793	0.63	0/2429
1	G	0.42	0/1771	0.63	0/2402
1	H	0.41	0/1684	0.70	0/2289
All	All	0.41	0/14316	0.63	0/19389

There are no bond length outliers.

There are no bond angle outliers.

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	1806	0	1838	36	0
1	B	1802	0	1827	24	0
1	C	1786	0	1802	33	0
1	D	1799	0	1828	33	0
1	E	1790	0	1808	38	0
1	F	1776	0	1788	37	0
1	G	1754	0	1749	37	0
1	H	1668	0	1658	65	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	79	0	0	3	0
2	B	76	0	0	1	0
2	C	48	0	0	1	0
2	D	46	0	0	1	0
2	E	48	0	0	2	0
2	F	39	0	0	4	0
2	G	33	0	0	1	0
2	H	16	0	0	3	0
All	All	14566	0	14298	265	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 9.

The worst 5 of 265 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:59:ARG:HG2	1:A:59:ARG:HH11	1.13	1.10
1:E:39:ARG:HD3	1:E:111:ARG:NH2	1.70	1.05
1:A:185:ARG:HG3	1:A:185:ARG:NH1	1.67	1.03
1:A:185:ARG:HG3	1:A:185:ARG:HH11	0.86	1.01
1:A:185:ARG:HH11	1:A:185:ARG:CG	1.77	0.93

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	248/251 (99%)	244 (98%)	4 (2%)	0	100	100
1	B	248/251 (99%)	246 (99%)	2 (1%)	0	100	100
1	C	247/251 (98%)	245 (99%)	2 (1%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	247/251 (98%)	245 (99%)	2 (1%)	0	100	100
1	E	247/251 (98%)	243 (98%)	4 (2%)	0	100	100
1	F	247/251 (98%)	244 (99%)	3 (1%)	0	100	100
1	G	246/251 (98%)	244 (99%)	2 (1%)	0	100	100
1	H	235/251 (94%)	232 (99%)	3 (1%)	0	100	100
All	All	1965/2008 (98%)	1943 (99%)	22 (1%)	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	183/186 (98%)	168 (92%)	15 (8%)	11	15
1	B	182/186 (98%)	170 (93%)	12 (7%)	16	24
1	C	180/186 (97%)	165 (92%)	15 (8%)	11	15
1	D	182/186 (98%)	170 (93%)	12 (7%)	16	24
1	E	181/186 (97%)	171 (94%)	10 (6%)	21	32
1	F	177/186 (95%)	164 (93%)	13 (7%)	14	20
1	G	174/186 (94%)	154 (88%)	20 (12%)	5	6
1	H	164/186 (88%)	147 (90%)	17 (10%)	7	9
All	All	1423/1488 (96%)	1309 (92%)	114 (8%)	12	17

5 of 114 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	93	SER
1	H	204	LEU
1	F	159	ARG
1	H	176	VAL
1	H	59	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	H	177	ASN
1	H	250	GLN
1	F	88	ASN
1	H	53	GLN
1	H	88	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 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

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	250/251 (99%)	0.67	19 (7%) 13 15	27, 37, 72, 121	0
1	B	250/251 (99%)	0.54	14 (5%) 24 27	31, 41, 74, 114	0
1	C	249/251 (99%)	0.74	28 (11%) 5 6	31, 48, 79, 112	0
1	D	249/251 (99%)	0.69	20 (8%) 12 13	31, 50, 81, 110	0
1	E	249/251 (99%)	0.68	20 (8%) 12 13	36, 54, 81, 118	0
1	F	249/251 (99%)	0.76	23 (9%) 9 10	33, 50, 83, 125	0
1	G	248/251 (98%)	0.99	33 (13%) 3 3	34, 59, 93, 131	0
1	H	239/251 (95%)	1.48	63 (26%) 0 0	42, 70, 110, 154	0
All	All	1983/2008 (98%)	0.81	220 (11%) 5 6	27, 50, 90, 154	0

The worst 5 of 220 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	2	SER	9.1
1	G	195	ASP	8.8
1	B	194	GLY	8.1
1	D	196	THR	7.9
1	G	197	GLN	7.9

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 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.