wwPDB X-ray Structure Validation Summary Report (i)

Nov 7, 2023 – 01:57 PM JST

PDB ID	:	5B2G
Title	:	Crystal structure of human claudin-4 in complex with C-terminal fragment of Clostridium perfringens enterotoxin
Authors	:	Shinoda, T.; Kimura-Someya, T.; Shirouzu, M.; Yokoyama, S.
Deposited on	:	2016-01-15
Resolution	:	3.50 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:

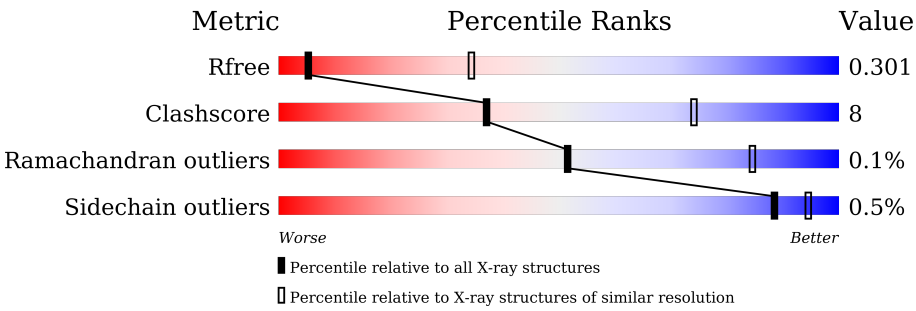
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 3.50 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	1659 (3.60-3.40)
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)

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%

Mol	Chain	Length	Quality of chain
1	A	364	73% 18% 9%
1	C	364	80% 15% • •
1	E	364	71% 21% 8%
1	G	364	78% 13% • 8%
2	B	142	70% 12% 18%
2	D	142	78% 9% 13%
2	F	142	65% 18% •• 15%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	H	142	78% 9% 13%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 14200 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 Endolysin,Claudin-4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	332	2540	1628	438	456	18	0	0	0
1	C	350	2657	1695	459	481	22	0	0	0
1	E	335	2563	1642	441	460	20	0	0	0
1	G	336	2558	1637	443	459	19	0	0	0

There are 96 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	995	GLY	-	expression tag	UNP P00720
A	996	SER	-	expression tag	UNP P00720
A	997	SER	-	expression tag	UNP P00720
A	998	GLY	-	expression tag	UNP P00720
A	999	SER	-	expression tag	UNP P00720
A	1000	SER	-	expression tag	UNP P00720
A	1001	GLY	-	expression tag	UNP P00720
A	1012	GLY	ARG	engineered mutation	UNP P00720
A	1054	THR	CYS	engineered mutation	UNP P00720
A	1097	ALA	CYS	engineered mutation	UNP P00720
A	1137	ARG	ILE	engineered mutation	UNP P00720
A	1163	GLY	-	linker	UNP P00720
A	1347	SER	-	expression tag	UNP O14493
A	1348	GLY	-	expression tag	UNP O14493
A	1349	PRO	-	expression tag	UNP O14493
A	1350	SER	-	expression tag	UNP O14493
A	1351	SER	-	expression tag	UNP O14493
A	1352	GLY	-	expression tag	UNP O14493
A	1353	GLU	-	expression tag	UNP O14493
A	1354	ASN	-	expression tag	UNP O14493
A	1355	LEU	-	expression tag	UNP O14493

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	1356	TYR	-	expression tag	UNP O14493
A	1357	PHE	-	expression tag	UNP O14493
A	1358	GLN	-	expression tag	UNP O14493
C	995	GLY	-	expression tag	UNP P00720
C	996	SER	-	expression tag	UNP P00720
C	997	SER	-	expression tag	UNP P00720
C	998	GLY	-	expression tag	UNP P00720
C	999	SER	-	expression tag	UNP P00720
C	1000	SER	-	expression tag	UNP P00720
C	1001	GLY	-	expression tag	UNP P00720
C	1012	GLY	ARG	engineered mutation	UNP P00720
C	1054	THR	CYS	engineered mutation	UNP P00720
C	1097	ALA	CYS	engineered mutation	UNP P00720
C	1137	ARG	ILE	engineered mutation	UNP P00720
C	1163	GLY	-	linker	UNP P00720
C	184	SER	-	expression tag	UNP O14493
C	185	GLY	-	expression tag	UNP O14493
C	186	PRO	-	expression tag	UNP O14493
C	187	SER	-	expression tag	UNP O14493
C	188	SER	-	expression tag	UNP O14493
C	189	GLY	-	expression tag	UNP O14493
C	190	GLU	-	expression tag	UNP O14493
C	191	ASN	-	expression tag	UNP O14493
C	192	LEU	-	expression tag	UNP O14493
C	193	TYR	-	expression tag	UNP O14493
C	194	PHE	-	expression tag	UNP O14493
C	195	GLN	-	expression tag	UNP O14493
E	995	GLY	-	expression tag	UNP P00720
E	996	SER	-	expression tag	UNP P00720
E	997	SER	-	expression tag	UNP P00720
E	998	GLY	-	expression tag	UNP P00720
E	999	SER	-	expression tag	UNP P00720
E	1000	SER	-	expression tag	UNP P00720
E	1001	GLY	-	expression tag	UNP P00720
E	1012	GLY	ARG	engineered mutation	UNP P00720
E	1054	THR	CYS	engineered mutation	UNP P00720
E	1097	ALA	CYS	engineered mutation	UNP P00720
E	1137	ARG	ILE	engineered mutation	UNP P00720
E	1163	GLY	-	linker	UNP P00720
E	1347	SER	-	expression tag	UNP O14493
E	1348	GLY	-	expression tag	UNP O14493
E	1349	PRO	-	expression tag	UNP O14493

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	1350	SER	-	expression tag	UNP O14493
E	1351	SER	-	expression tag	UNP O14493
E	1352	GLY	-	expression tag	UNP O14493
E	1353	GLU	-	expression tag	UNP O14493
E	1354	ASN	-	expression tag	UNP O14493
E	1355	LEU	-	expression tag	UNP O14493
E	1356	TYR	-	expression tag	UNP O14493
E	1357	PHE	-	expression tag	UNP O14493
E	1358	GLN	-	expression tag	UNP O14493
G	995	GLY	-	expression tag	UNP P00720
G	996	SER	-	expression tag	UNP P00720
G	997	SER	-	expression tag	UNP P00720
G	998	GLY	-	expression tag	UNP P00720
G	999	SER	-	expression tag	UNP P00720
G	1000	SER	-	expression tag	UNP P00720
G	1001	GLY	-	expression tag	UNP P00720
G	1012	GLY	ARG	engineered mutation	UNP P00720
G	1054	THR	CYS	engineered mutation	UNP P00720
G	1097	ALA	CYS	engineered mutation	UNP P00720
G	1137	ARG	ILE	engineered mutation	UNP P00720
G	1163	GLY	-	linker	UNP P00720
G	1347	SER	-	expression tag	UNP O14493
G	1348	GLY	-	expression tag	UNP O14493
G	1349	PRO	-	expression tag	UNP O14493
G	1350	SER	-	expression tag	UNP O14493
G	1351	SER	-	expression tag	UNP O14493
G	1352	GLY	-	expression tag	UNP O14493
G	1353	GLU	-	expression tag	UNP O14493
G	1354	ASN	-	expression tag	UNP O14493
G	1355	LEU	-	expression tag	UNP O14493
G	1356	TYR	-	expression tag	UNP O14493
G	1357	PHE	-	expression tag	UNP O14493
G	1358	GLN	-	expression tag	UNP O14493

- Molecule 2 is a protein called Heat-labile enterotoxin B chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
2	B	117	933	595	157	180	1	0	0	0
2	D	124	992	633	165	193	1	0	0	0
2	F	121	965	617	161	186	1	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
2	H	124	992	633	165	193	1	0	0	0

There are 36 discrepancies between the modelled and reference sequences:

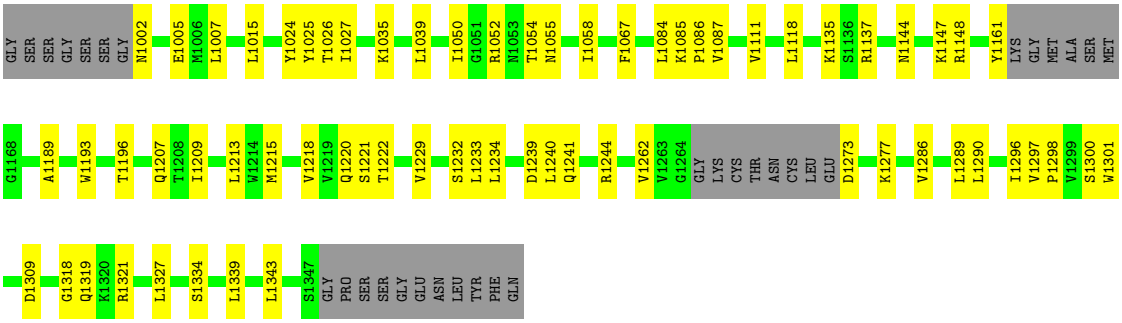
Chain	Residue	Modelled	Actual	Comment	Reference
B	178	GLY	-	expression tag	UNP P01558
B	179	SER	-	expression tag	UNP P01558
B	180	SER	-	expression tag	UNP P01558
B	181	GLY	-	expression tag	UNP P01558
B	182	SER	-	expression tag	UNP P01558
B	183	SER	-	expression tag	UNP P01558
B	184	GLY	-	expression tag	UNP P01558
B	185	ARG	-	expression tag	UNP P01558
B	186	TRP	-	expression tag	UNP P01558
D	178	GLY	-	expression tag	UNP P01558
D	179	SER	-	expression tag	UNP P01558
D	180	SER	-	expression tag	UNP P01558
D	181	GLY	-	expression tag	UNP P01558
D	182	SER	-	expression tag	UNP P01558
D	183	SER	-	expression tag	UNP P01558
D	184	GLY	-	expression tag	UNP P01558
D	185	ARG	-	expression tag	UNP P01558
D	186	TRP	-	expression tag	UNP P01558
F	178	GLY	-	expression tag	UNP P01558
F	179	SER	-	expression tag	UNP P01558
F	180	SER	-	expression tag	UNP P01558
F	181	GLY	-	expression tag	UNP P01558
F	182	SER	-	expression tag	UNP P01558
F	183	SER	-	expression tag	UNP P01558
F	184	GLY	-	expression tag	UNP P01558
F	185	ARG	-	expression tag	UNP P01558
F	186	TRP	-	expression tag	UNP P01558
H	178	GLY	-	expression tag	UNP P01558
H	179	SER	-	expression tag	UNP P01558
H	180	SER	-	expression tag	UNP P01558
H	181	GLY	-	expression tag	UNP P01558
H	182	SER	-	expression tag	UNP P01558
H	183	SER	-	expression tag	UNP P01558
H	184	GLY	-	expression tag	UNP P01558
H	185	ARG	-	expression tag	UNP P01558
H	186	TRP	-	expression tag	UNP P01558

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. 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.

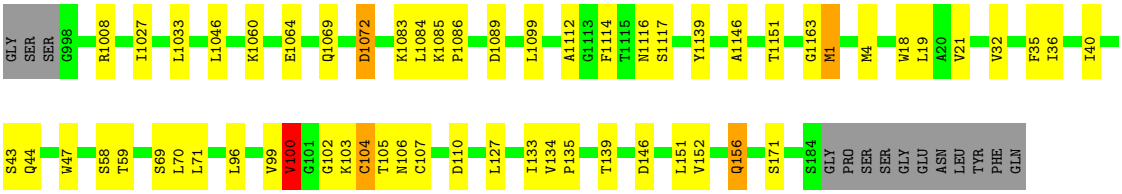
- Molecule 1: Endolysin,Claudin-4

Chain A:



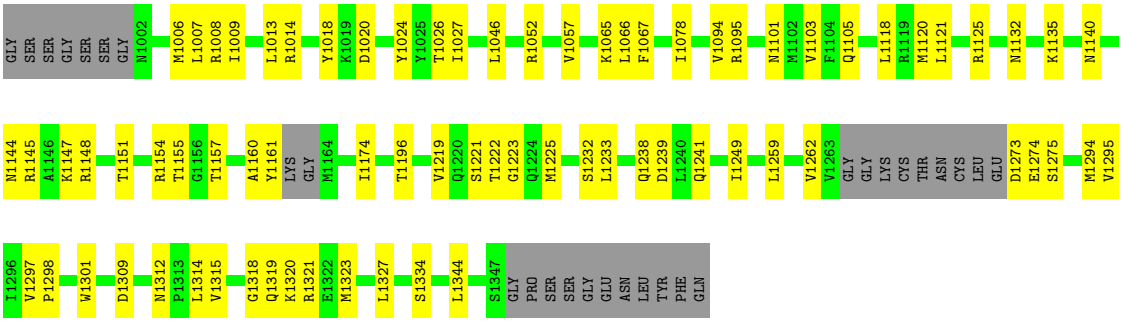
- Molecule 1: Endolysin,Claudin-4

Chain C:



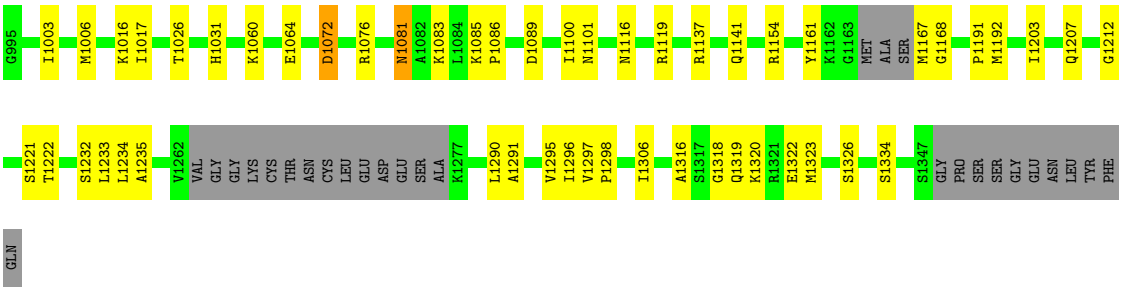
- Molecule 1: Endolysin,Claudin-4

Chain E:



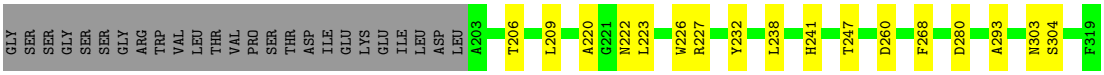
- Molecule 1: Endolysin,Claudin-4

Chain G:



- Molecule 2: Heat-labile enterotoxin B chain

Chain B:

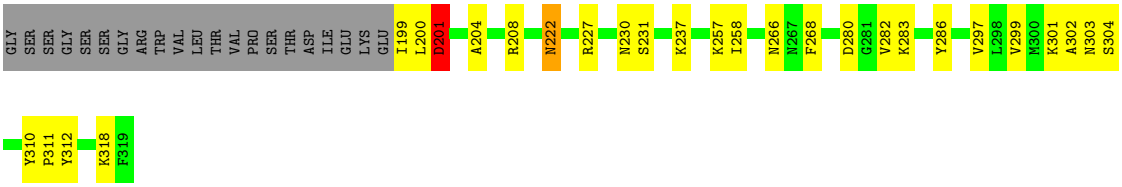


- Molecule 2: Heat-labile enterotoxin B chain

Chain D:

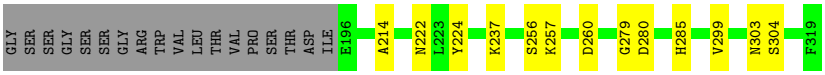
- Molecule 2: Heat-labile enterotoxin B chain

Chain F:



- Molecule 2: Heat-labile enterotoxin B chain

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43	Depositor
Cell constants a, b, c, α, β, γ	105.92Å 105.92Å 244.28Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.59 – 3.50 48.59 – 3.35	Depositor EDS
% Data completeness (in resolution range)	99.8 (48.59-3.50) 99.8 (48.59-3.35)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.41 (at 3.33Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.288 , 0.309 0.285 , 0.301	Depositor DCC
R_{free} test set	1925 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	129.9	Xtriage
Anisotropy	0.306	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 61.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.36$, $< L^2 > = 0.19$	Xtriage
Estimated twinning fraction	0.387 for h,-k,-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	14200	wwPDB-VP
Average B, all atoms (Å^2)	99.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.52% 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 (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	0.23	0/2581	0.44	0/3498
1	C	0.32	0/2700	0.53	3/3657 (0.1%)
1	E	0.24	0/2604	0.49	0/3528
1	G	0.22	0/2599	0.46	2/3518 (0.1%)
2	B	0.21	0/953	0.41	0/1294
2	D	0.26	0/1012	0.40	0/1373
2	F	0.29	0/985	0.45	1/1338 (0.1%)
2	H	0.22	0/1012	0.40	0/1373
All	All	0.25	0/14446	0.46	6/19579 (0.0%)

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	156	GLN	CB-CA-C	-9.54	91.33	110.40
1	C	104	CYS	N-CA-C	-7.56	90.59	111.00
1	G	1081	ASN	CB-CA-C	5.87	122.14	110.40
2	F	201	ASP	CB-CA-C	5.78	121.96	110.40
1	G	1081	ASN	N-CA-C	-5.78	95.40	111.00

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	2540	0	2622	54	0
1	C	2657	0	2740	45	0
1	E	2563	0	2647	61	0
1	G	2558	0	2644	36	0
2	B	933	0	907	13	0
2	D	992	0	969	11	0
2	F	965	0	944	23	0
2	H	992	0	969	8	0
All	All	14200	0	14442	223	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 8.

The worst 5 of 223 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:1213:LEU:CD1	1:A:1240:LEU:HD21	1.83	1.09
1:A:1213:LEU:HD12	1:A:1240:LEU:CD2	1.82	1.08
1:A:1213:LEU:HD12	1:A:1240:LEU:HD21	1.09	1.06
1:C:103:LYS:HG3	1:C:104:CYS:O	1.58	1.00
1:A:1213:LEU:CD1	1:A:1240:LEU:CD2	2.42	0.96

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	326/364 (90%)	304 (93%)	22 (7%)	0	100 100
1	C	348/364 (96%)	313 (90%)	33 (10%)	2 (1%)	25 64
1	E	329/364 (90%)	303 (92%)	26 (8%)	0	100 100
1	G	330/364 (91%)	304 (92%)	26 (8%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	115/142 (81%)	109 (95%)	6 (5%)	0	100	100
2	D	122/142 (86%)	117 (96%)	5 (4%)	0	100	100
2	F	119/142 (84%)	111 (93%)	8 (7%)	0	100	100
2	H	122/142 (86%)	116 (95%)	6 (5%)	0	100	100
All	All	1811/2024 (90%)	1677 (93%)	132 (7%)	2 (0%)	51	84

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	1	MET
1	C	100	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 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	268/292 (92%)	268 (100%)	0	100	100
1	C	281/292 (96%)	278 (99%)	3 (1%)	73	88
1	E	271/292 (93%)	271 (100%)	0	100	100
1	G	270/292 (92%)	269 (100%)	1 (0%)	91	96
2	B	102/123 (83%)	101 (99%)	1 (1%)	76	88
2	D	109/123 (89%)	109 (100%)	0	100	100
2	F	106/123 (86%)	104 (98%)	2 (2%)	57	80
2	H	109/123 (89%)	108 (99%)	1 (1%)	78	90
All	All	1516/1660 (91%)	1508 (100%)	8 (0%)	88	94

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	H	260	ASP
1	G	1072	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	F	201	ASP
1	C	4	MET
2	F	222	ASN

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

Mol	Chain	Res	Type
1	A	1220	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 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 (i)

6.1 Protein, DNA and RNA chains (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.