



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 5, 2017 – 11:56 PM EDT

PDB ID : 3GR1
Title : Periplasmic domain of the T3SS inner membrane protein PrgH from *S.typhimurium* (fragment 170-392)
Authors : Yip, C.K.; Vockovic, M.; Yu, A.C.; Strynadka, N.C.J.
Deposited on : unknown
Resolution : 2.80 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

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

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

MolProbity : 4.02b-467
Xtriage (Phenix) : 1.9-1692
EDS : rb-20029824
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
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) : rb-20029824

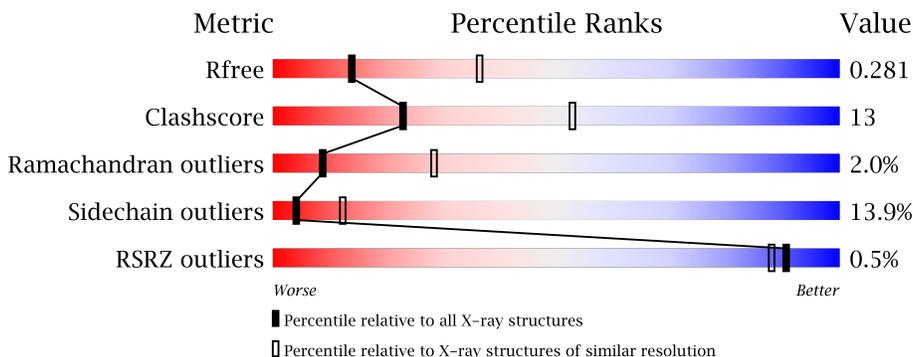
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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}	100719	2583 (2.80-2.80)
Clashscore	112137	3033 (2.80-2.80)
Ramachandran outliers	110173	2983 (2.80-2.80)
Sidechain outliers	110143	2985 (2.80-2.80)
RSRZ outliers	101464	2610 (2.80-2.80)

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 $\leq 5\%$. 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	227	
1	B	227	
1	C	227	
1	D	227	
1	E	227	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	227	 57% 22% 17%
1	G	227	 54% 24% 17%
1	H	227	 53% 25% 7% 16%

2 Entry composition [i](#)

There is only 1 type of molecule in this entry. The entry contains 12613 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 Protein prgH.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	195	Total 1602	C 1014	N 289	O 294	S 5	0	0	0
1	B	193	Total 1589	C 1006	N 287	O 292	S 4	0	0	0
1	C	190	Total 1570	C 995	N 284	O 287	S 4	0	0	0
1	D	193	Total 1589	C 1006	N 287	O 292	S 4	0	0	0
1	E	192	Total 1584	C 1003	N 286	O 291	S 4	0	0	0
1	F	189	Total 1554	C 986	N 281	O 283	S 4	0	0	0
1	G	189	Total 1554	C 986	N 281	O 283	S 4	0	0	0
1	H	191	Total 1571	C 996	N 285	O 285	S 5	0	0	0

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	166	GLY	-	EXPRESSION TAG	UNP P41783
A	167	SER	-	EXPRESSION TAG	UNP P41783
A	168	HIS	-	EXPRESSION TAG	UNP P41783
A	169	MET	-	EXPRESSION TAG	UNP P41783
B	166	GLY	-	EXPRESSION TAG	UNP P41783
B	167	SER	-	EXPRESSION TAG	UNP P41783
B	168	HIS	-	EXPRESSION TAG	UNP P41783
B	169	MET	-	EXPRESSION TAG	UNP P41783
C	166	GLY	-	EXPRESSION TAG	UNP P41783
C	167	SER	-	EXPRESSION TAG	UNP P41783
C	168	HIS	-	EXPRESSION TAG	UNP P41783
C	169	MET	-	EXPRESSION TAG	UNP P41783
D	166	GLY	-	EXPRESSION TAG	UNP P41783

Continued on next page...

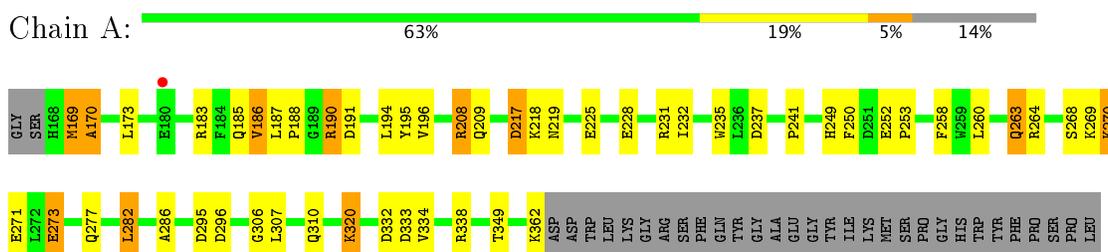
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	167	SER	-	EXPRESSION TAG	UNP P41783
D	168	HIS	-	EXPRESSION TAG	UNP P41783
D	169	MET	-	EXPRESSION TAG	UNP P41783
E	166	GLY	-	EXPRESSION TAG	UNP P41783
E	167	SER	-	EXPRESSION TAG	UNP P41783
E	168	HIS	-	EXPRESSION TAG	UNP P41783
E	169	MET	-	EXPRESSION TAG	UNP P41783
F	166	GLY	-	EXPRESSION TAG	UNP P41783
F	167	SER	-	EXPRESSION TAG	UNP P41783
F	168	HIS	-	EXPRESSION TAG	UNP P41783
F	169	MET	-	EXPRESSION TAG	UNP P41783
G	166	GLY	-	EXPRESSION TAG	UNP P41783
G	167	SER	-	EXPRESSION TAG	UNP P41783
G	168	HIS	-	EXPRESSION TAG	UNP P41783
G	169	MET	-	EXPRESSION TAG	UNP P41783
H	166	GLY	-	EXPRESSION TAG	UNP P41783
H	167	SER	-	EXPRESSION TAG	UNP P41783
H	168	HIS	-	EXPRESSION TAG	UNP P41783
H	169	MET	-	EXPRESSION TAG	UNP P41783

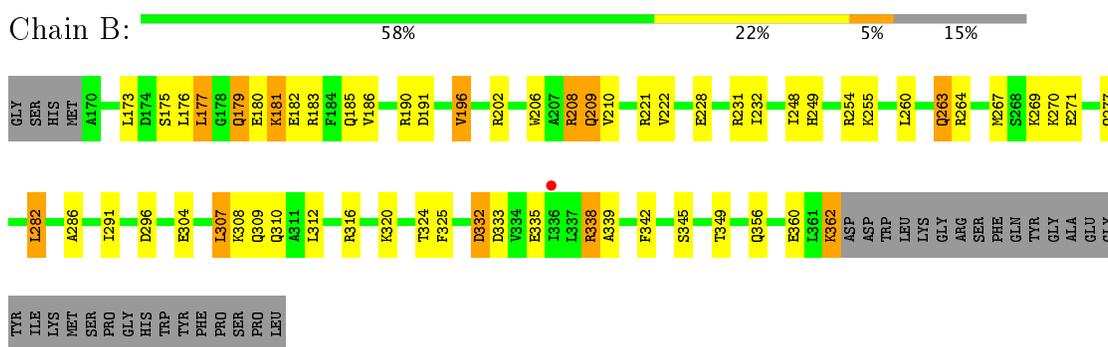
3 Residue-property plots [i](#)

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.

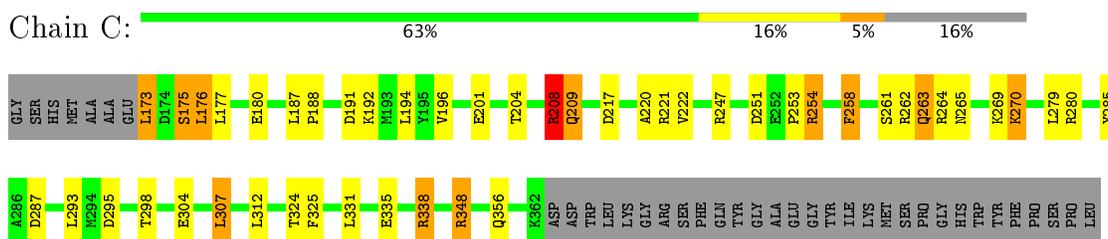
- Molecule 1: Protein prgH



- Molecule 1: Protein prgH

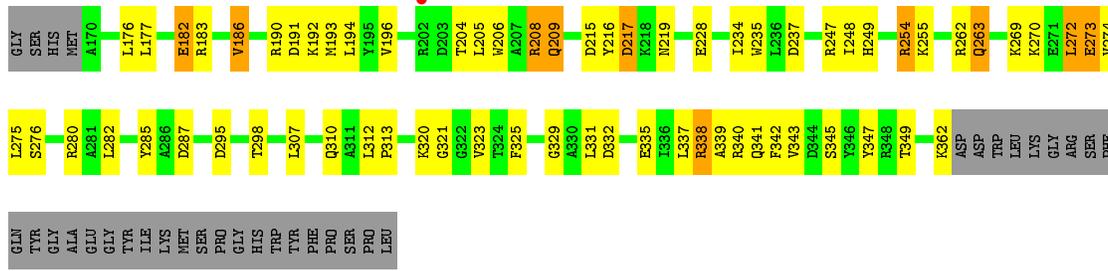


- Molecule 1: Protein prgH

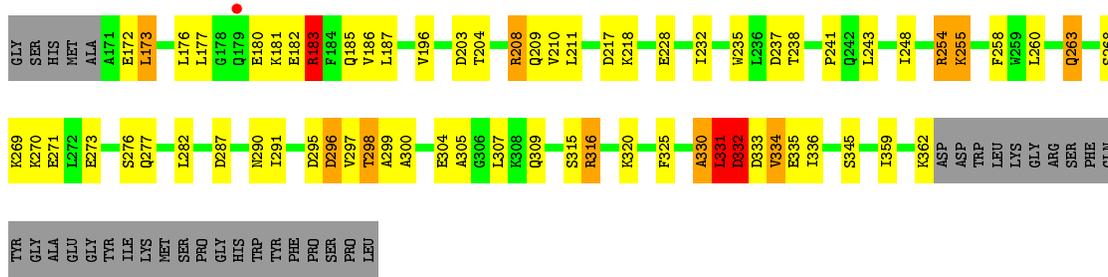


- Molecule 1: Protein prgH

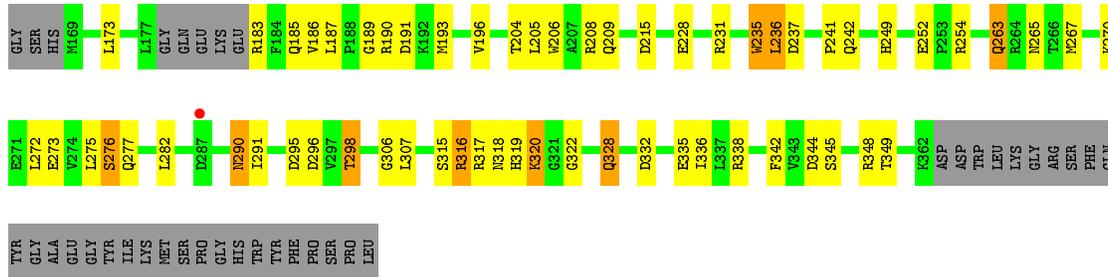




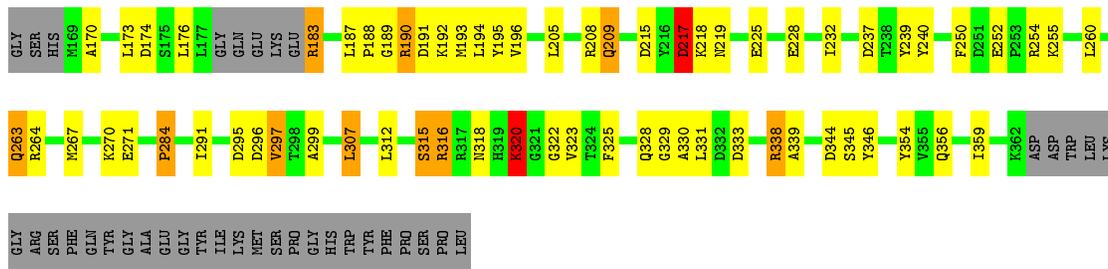
• Molecule 1: Protein prgH



• Molecule 1: Protein prgH



• Molecule 1: Protein prgH



• Molecule 1: Protein prgH



GLY	S268	S269	K270	E271	L272	E273	L173	D174	S175	L176	L177	G178	GLN	GLU	LYS	GLU	R183	V186	L187	R190	D191	K192	M193	L194	Y195	L205	R208	Q209	Y216	D217	K218	Y222	S233	L236	Y240	I248	H249	F250	D251	F262	F263	R264	K265	K266	P266	L260	S261	R262	Q263	R264								
S268	K270	E271	L272	E273	K278	L279	R280	A281	L282	M283	F284	Y285	A286	D287	M290	L291	T292	L293	M294	D295	D296	V297	T298	Q302	A303	E304	A305	G306	L307	K308	Q309	Q310	R316	R317	N318	H319	K320	G321	G322	F325	G329	A330	L331	D332	D333	V334	R338	A339	T349									
Q356	K362	ASP	TRP	LEU	LYS	GLY	ARG	SER	PHE	GLN	TYR	GLY	ALA	GLU	GLY	TYR	ILE	LYS	MET	SER	PRO	GLY	HIS	TRP	TYR	PHE	PRO	SER	PRO	LEU	ASP	TRP	LEU	LYS	ARG	SER	PHE	GLN	TYR	GLY	ALA	GLU	GLY	TYR	ILE	LYS	MET	SER	PRO	GLY	HIS	TRP	TYR	PHE	PRO	SER	PRO	LEU

4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	79.61Å 188.75Å 305.56Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	152.50 – 2.80 152.78 – 2.80	Depositor EDS
% Data completeness (in resolution range)	96.5 (152.50-2.80) 96.5 (152.78-2.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.65 (at 2.82Å)	Xtrriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.247 , 0.295 0.237 , 0.281	Depositor DCC
R_{free} test set	2812 reflections (5.37%)	DCC
Wilson B-factor (Å ²)	55.2	Xtrriage
Anisotropy	0.429	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 43.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	12613	wwPDB-VP
Average B, all atoms (Å ²)	48.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 14.51% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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	0.77	0/1634	0.81	1/2206 (0.0%)
1	B	0.77	0/1621	0.82	1/2189 (0.0%)
1	C	0.70	0/1602	0.82	2/2163 (0.1%)
1	D	0.68	0/1621	0.79	2/2189 (0.1%)
1	E	0.70	0/1616	0.81	2/2182 (0.1%)
1	F	0.70	0/1585	0.76	0/2141
1	G	0.70	1/1585 (0.1%)	0.76	0/2141
1	H	0.65	0/1603	0.76	1/2164 (0.0%)
All	All	0.71	1/12867 (0.0%)	0.79	9/17375 (0.1%)

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	D	0	1
1	E	0	1
All	All	0	2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	G	330	ALA	CA-CB	5.22	1.63	1.52

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	208	ARG	NE-CZ-NH2	7.46	124.03	120.30
1	D	321	GLY	N-CA-C	-6.54	96.75	113.10
1	B	208	ARG	NE-CZ-NH2	6.46	123.53	120.30
1	D	208	ARG	NE-CZ-NH2	5.81	123.20	120.30
1	E	330	ALA	N-CA-C	-5.59	95.92	111.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	320	LYS	Peptide
1	E	331	LEU	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	1602	0	1588	34	0
1	B	1589	0	1577	47	0
1	C	1570	0	1561	35	0
1	D	1589	0	1577	47	0
1	E	1584	0	1572	50	0
1	F	1554	0	1542	41	0
1	G	1554	0	1542	39	0
1	H	1571	0	1559	52	0
All	All	12613	0	12518	323	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 13.

The worst 5 of 323 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:277:GLN:HG2	1:B:191:ASP:HA	1.36	1.07
1:G:316:ARG:HD3	1:G:318:ASN:HD21	1.21	1.05
1:E:330:ALA:CA	1:E:331:LEU:HB2	1.89	1.02
1:E:330:ALA:HA	1:E:331:LEU:HB2	1.02	1.02
1:E:330:ALA:HA	1:E:331:LEU:CB	1.90	1.01

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	193/227 (85%)	182 (94%)	9 (5%)	2 (1%)	18	50
1	B	191/227 (84%)	178 (93%)	9 (5%)	4 (2%)	8	27
1	C	188/227 (83%)	170 (90%)	15 (8%)	3 (2%)	11	36
1	D	191/227 (84%)	166 (87%)	22 (12%)	3 (2%)	11	36
1	E	190/227 (84%)	170 (90%)	15 (8%)	5 (3%)	6	21
1	F	185/227 (82%)	157 (85%)	23 (12%)	5 (3%)	6	20
1	G	185/227 (82%)	155 (84%)	24 (13%)	6 (3%)	5	16
1	H	187/227 (82%)	164 (88%)	21 (11%)	2 (1%)	17	47
All	All	1510/1816 (83%)	1342 (89%)	138 (9%)	30 (2%)	9	28

5 of 30 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	175	SER
1	C	176	LEU
1	D	272	LEU
1	D	273	GLU
1	E	331	LEU

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	166/193 (86%)	145 (87%)	21 (13%)	5	15

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	165/193 (86%)	143 (87%)	22 (13%)	4	13
1	C	164/193 (85%)	145 (88%)	19 (12%)	6	19
1	D	165/193 (86%)	146 (88%)	19 (12%)	6	20
1	E	165/193 (86%)	134 (81%)	31 (19%)	2	5
1	F	161/193 (83%)	142 (88%)	19 (12%)	6	18
1	G	161/193 (83%)	135 (84%)	26 (16%)	3	8
1	H	163/193 (84%)	138 (85%)	25 (15%)	3	9
All	All	1310/1544 (85%)	1128 (86%)	182 (14%)	4	12

5 of 182 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	173	LEU
1	E	307	LEU
1	H	263	GLN
1	E	177	LEU
1	E	218	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 32 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	249	HIS
1	E	185	GLN
1	H	263	GLN
1	D	341	GLN
1	E	263	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	195/227 (85%)	-0.14	1 (0%) 90 88	17, 40, 62, 73	0
1	B	193/227 (85%)	-0.13	1 (0%) 90 88	18, 40, 64, 70	0
1	C	190/227 (83%)	-0.06	0 100 100	20, 45, 69, 78	0
1	D	193/227 (85%)	0.01	1 (0%) 90 88	25, 49, 74, 83	0
1	E	192/227 (84%)	-0.06	1 (0%) 90 88	23, 47, 74, 89	0
1	F	189/227 (83%)	-0.01	1 (0%) 90 88	20, 47, 81, 94	0
1	G	189/227 (83%)	0.07	0 100 100	31, 53, 76, 80	0
1	H	191/227 (84%)	-0.00	2 (1%) 82 77	32, 51, 76, 83	0
All	All	1532/1816 (84%)	-0.04	7 (0%) 90 88	17, 47, 74, 94	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	H	330	ALA	2.8
1	B	336	ILE	2.8
1	F	287	ASP	2.6
1	A	180	GLU	2.3
1	E	179	GLN	2.1

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

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.