



wwPDB X-ray Structure Validation Summary Report i

Feb 27, 2014 – 08:56 AM GMT

PDB ID : 3S5U
Title : Crystal structure of CRISPR associated protein
Authors : Ke, A.; Nam, K.H.
Deposited on : 2011-05-23
Resolution : 2.70 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

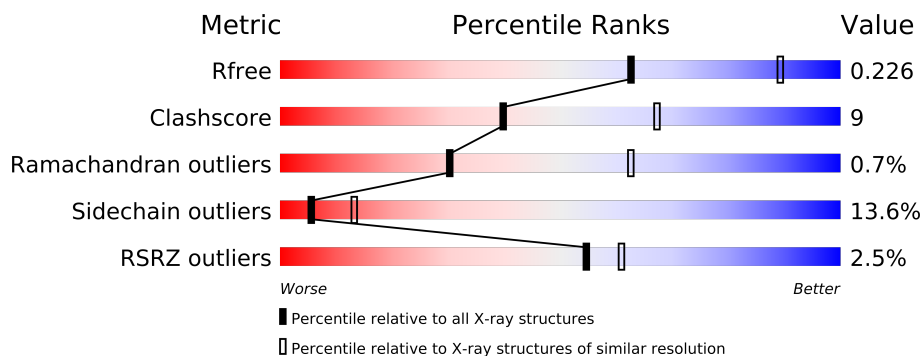
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

1 Overall quality at a glance

The reported resolution of this entry is 2.70 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}	66092	1557 (2.70-2.70)
Clashscore	79885	1939 (2.70-2.70)
Ramachandran outliers	78287	1905 (2.70-2.70)
Sidechain outliers	78261	1905 (2.70-2.70)
RSRZ outliers	66119	1559 (2.70-2.70)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	220	
1	B	220	
1	C	220	
1	D	220	
1	E	220	
1	F	220	
1	G	220	
1	H	220	

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 13839 atoms, of which 0 are hydrogen and 0 are deuterium.

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 uncharacterized protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	212	Total	C	N	O	S	0	0	0
			1736	1126	268	338	4			
1	B	199	Total	C	N	O	S	0	0	0
			1624	1052	251	317	4			
1	C	210	Total	C	N	O	S	0	0	0
			1712	1110	264	334	4			
1	D	212	Total	C	N	O	S	0	0	0
			1737	1128	266	339	4			
1	E	213	Total	C	N	O	S	0	0	0
			1744	1130	269	341	4			
1	F	210	Total	C	N	O	S	0	0	0
			1717	1115	265	333	4			
1	G	209	Total	C	N	O	S	0	0	0
			1712	1112	263	333	4			
1	H	213	Total	C	N	O	S	0	0	0
			1729	1125	264	336	4			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	SER	-	EXPRESSION TAG	UNP C7UDU4
B	0	SER	-	EXPRESSION TAG	UNP C7UDU4
C	0	SER	-	EXPRESSION TAG	UNP C7UDU4
D	0	SER	-	EXPRESSION TAG	UNP C7UDU4
E	0	SER	-	EXPRESSION TAG	UNP C7UDU4
F	0	SER	-	EXPRESSION TAG	UNP C7UDU4
G	0	SER	-	EXPRESSION TAG	UNP C7UDU4
H	0	SER	-	EXPRESSION TAG	UNP C7UDU4

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	G	2	Total Ca 2 2	0	0
2	D	2	Total Ca 2 2	0	0
2	E	2	Total Ca 2 2	0	0
2	H	2	Total Ca 2 2	0	0
2	B	2	Total Ca 2 2	0	0
2	C	2	Total Ca 2 2	0	0
2	A	2	Total Ca 2 2	0	0
2	F	2	Total Ca 2 2	0	0

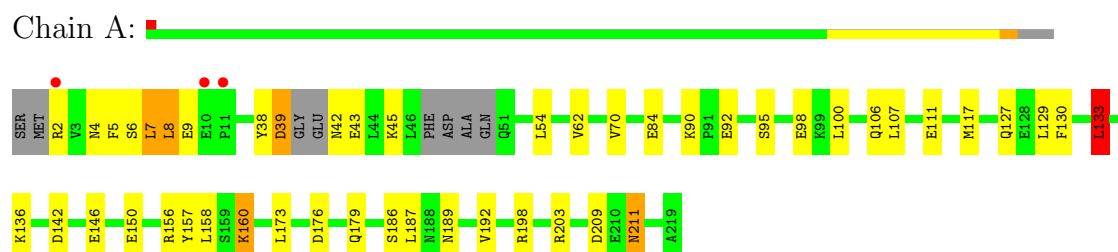
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	26	Total O 26 26	0	0
3	B	12	Total O 12 12	0	0
3	C	6	Total O 6 6	0	0
3	D	11	Total O 11 11	0	0
3	E	16	Total O 16 16	0	0
3	F	13	Total O 13 13	0	0
3	G	12	Total O 12 12	0	0
3	H	16	Total O 16 16	0	0

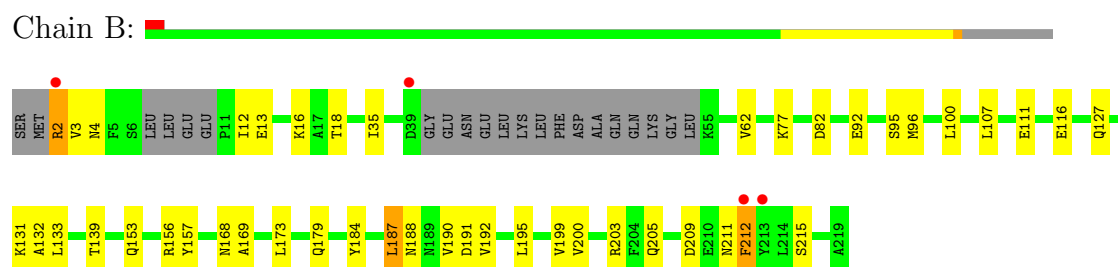
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

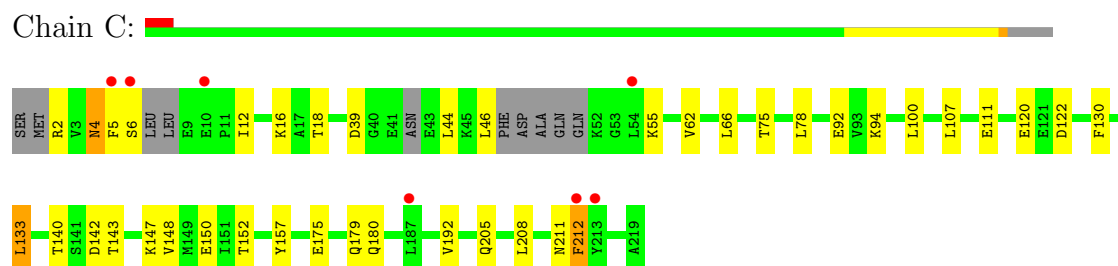
- Molecule 1: Putative uncharacterized protein



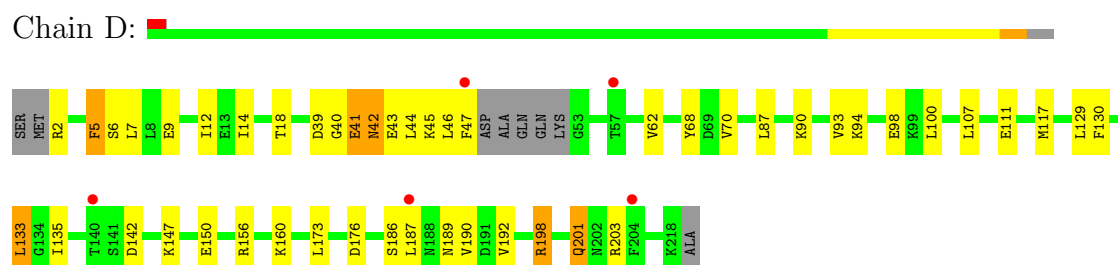
- Molecule 1: Putative uncharacterized protein



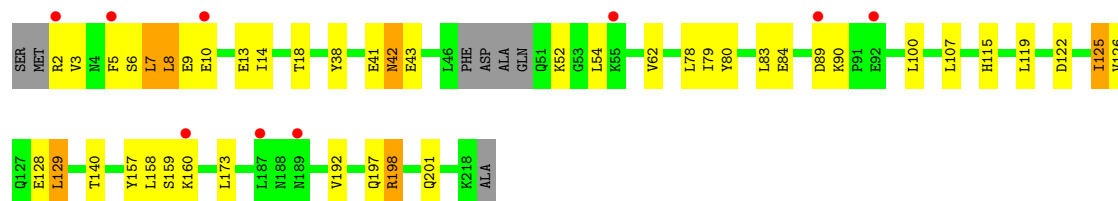
- Molecule 1: Putative uncharacterized protein



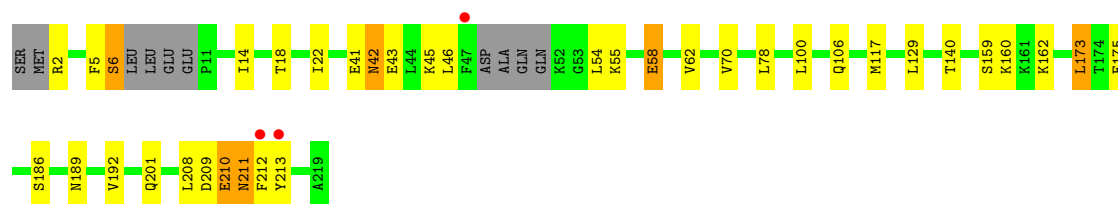
- Molecule 1: Putative uncharacterized protein



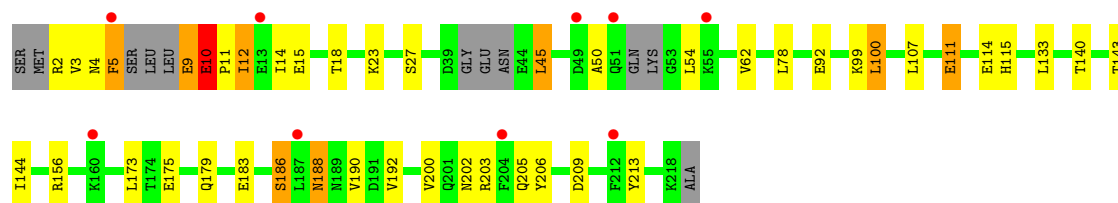
- Molecule 1: Putative uncharacterized protein

Chain E: 

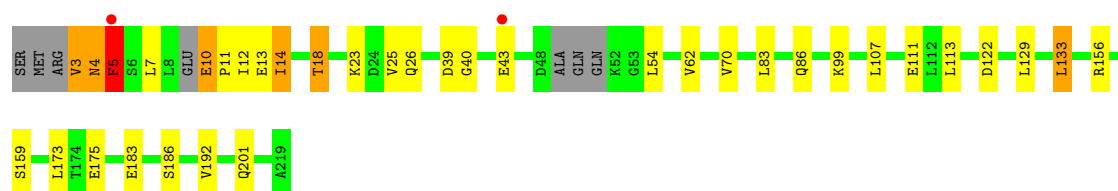
- Molecule 1: Putative uncharacterized protein

Chain F: 

- Molecule 1: Putative uncharacterized protein

Chain G: 

- Molecule 1: Putative uncharacterized protein

Chain H: 

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	104.11Å 140.09Å 148.43Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.70 19.85 – 2.70	Depositor EDS
% Data completeness (in resolution range)	98.8 (20.00-2.70) 98.8 (19.85-2.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.66 (at 2.71Å)	Xtriage
Refinement program	phenix.refinement	Depositor
R, R_{free}	0.219 , 0.275 0.221 , 0.226	Depositor DCC
R_{free} test set	3003 reflections (5.06%)	DCC
Wilson B-factor (Å ²)	51.6	Xtriage
Anisotropy	0.619	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 27.0	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 59542 reflections	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	13839	wwPDB-VP
Average B, all atoms (Å ²)	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.10% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

5 Model quality ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: CA

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	0.79	0/1760	0.82	1/2376 (0.0%)
1	B	0.80	0/1647	0.79	0/2224
1	C	0.91	0/1735	0.81	0/2340
1	D	0.78	0/1763	0.82	0/2382
1	E	0.80	0/1769	0.79	1/2389 (0.0%)
1	F	0.85	0/1742	0.80	1/2350 (0.0%)
1	G	0.84	3/1736 (0.2%)	0.87	5/2343 (0.2%)
1	H	0.82	1/1754 (0.1%)	0.83	1/2367 (0.0%)
All	All	0.82	4/13906 (0.0%)	0.82	9/18771 (0.0%)

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

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	G	5	PHE	CD2-CE2	6.32	1.51	1.39
1	G	5	PHE	CD1-CE1	5.40	1.50	1.39
1	G	5	PHE	CE2-CZ	5.22	1.47	1.37
1	H	5	PHE	CE2-CZ	5.11	1.47	1.37

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	7	LEU	CA-CB-CG	-6.33	100.73	115.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	G	100	LEU	CA-CB-CG	6.30	129.78	115.30
1	A	133	LEU	CA-CB-CG	6.19	129.54	115.30
1	F	173	LEU	CA-CB-CG	6.07	129.27	115.30
1	G	133	LEU	N-CA-C	-5.61	95.84	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	39	ASP	Peptide

5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1736	0	0	25	0
1	B	1624	0	0	19	0
1	C	1712	0	0	10	0
1	D	1737	0	5	19	0
1	E	1744	0	0	21	0
1	F	1717	0	0	17	0
1	G	1712	0	0	15	0
1	H	1729	0	0	10	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
2	C	2	0	0	0	0
2	D	2	0	0	0	0
2	E	2	0	0	0	0
2	F	2	0	0	0	0
2	G	2	0	0	0	0
2	H	2	0	0	0	0
3	A	26	0	0	3	0
3	B	12	0	0	2	0
3	C	6	0	0	0	0
3	D	11	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	E	16	0	0	0	0
3	F	13	0	0	1	0
3	G	12	0	0	0	0
3	H	16	0	0	0	0
All	All	13839	0	5	130	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 9.

The worst 5 of 130 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:D:5:PHE:CZ	1:D:9:GLU:O	2.01	1.13
1:A:5:PHE:CD1	1:A:8:LEU:CD1	2.33	1.11
1:A:5:PHE:CE1	1:A:9:GLU:N	2.26	1.04
1:A:7:LEU:CD2	1:A:7:LEU:C	2.30	1.00
1:B:187:LEU:CD1	1:B:187:LEU:C	2.30	0.99

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone ⓘ

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	206/220 (94%)	194 (94%)	12 (6%)	0	100	100
1	B	193/220 (88%)	181 (94%)	11 (6%)	1 (0%)	38	70
1	C	202/220 (92%)	191 (95%)	10 (5%)	1 (0%)	38	70
1	D	208/220 (94%)	193 (93%)	13 (6%)	2 (1%)	22	51
1	E	209/220 (95%)	186 (89%)	21 (10%)	2 (1%)	22	51
1	F	204/220 (93%)	191 (94%)	12 (6%)	1 (0%)	38	70
1	G	201/220 (91%)	190 (94%)	9 (4%)	2 (1%)	22	51
1	H	207/220 (94%)	196 (95%)	9 (4%)	2 (1%)	22	51
All	All	1630/1760 (93%)	1522 (93%)	97 (6%)	11 (1%)	30	62

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	212	PHE
1	D	41	GLU
1	G	10	GLU
1	D	42	ASN
1	F	211	ASN

5.3.2 Protein sidechains ⓘ

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	197/203 (97%)	167 (85%)	30 (15%)	4	10
1	B	183/203 (90%)	161 (88%)	22 (12%)	7	17
1	C	193/203 (95%)	169 (88%)	24 (12%)	7	16
1	D	197/203 (97%)	168 (85%)	29 (15%)	4	11
1	E	198/203 (98%)	176 (89%)	22 (11%)	9	20
1	F	194/203 (96%)	167 (86%)	27 (14%)	5	13
1	G	193/203 (95%)	166 (86%)	27 (14%)	5	12
1	H	194/203 (96%)	165 (85%)	29 (15%)	4	11
All	All	1549/1624 (95%)	1339 (86%)	210 (14%)	5	13

5 of 210 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	133	LEU
1	E	119	LEU
1	H	62	VAL
1	D	173	LEU
1	E	8	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA ⓘ

There are no RNA chains 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 ⓘ

Of 16 ligands modelled in this entry, 16 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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, 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	212/220 (96%)	-0.11	3 (1%) 72 77	29, 47, 64, 75	0
1	B	199/220 (90%)	-0.05	4 (2%) 62 68	37, 50, 69, 79	0
1	C	210/220 (95%)	0.08	7 (3%) 44 49	34, 51, 66, 79	0
1	D	212/220 (96%)	-0.04	5 (2%) 56 62	36, 52, 69, 74	0
1	E	213/220 (96%)	0.00	9 (4%) 35 39	37, 51, 77, 81	0
1	F	210/220 (95%)	-0.11	3 (1%) 72 77	36, 50, 69, 76	0
1	G	209/220 (95%)	-0.08	9 (4%) 34 38	33, 50, 68, 76	0
1	H	213/220 (96%)	-0.11	2 (0%) 81 85	37, 50, 65, 73	0
All	All	1678/1760 (95%)	-0.05	42 (2%) 54 61	29, 50, 69, 81	0

The worst 5 of 42 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	2	ARG	5.7
1	D	47	PHE	5.1
1	H	5	PHE	4.4
1	B	2	ARG	4.3
1	C	5	PHE	3.7

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

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors(Å ²)	Q<0.9
2	CA	A	220	1/1	0.11	-0.42	47,47,47,47	0
2	CA	D	220	1/1	0.12	-0.88	53,53,53,53	0
2	CA	G	221	1/1	0.10	-0.90	50,50,50,50	0
2	CA	H	221	1/1	0.10	-1.23	47,47,47,47	0
2	CA	F	221	1/1	0.09	-1.35	51,51,51,51	0
2	CA	B	221	1/1	0.10	-1.36	59,59,59,59	0
2	CA	F	220	1/1	0.09	-1.36	44,44,44,44	0
2	CA	A	221	1/1	0.08	-1.80	37,37,37,37	0
2	CA	E	220	1/1	0.04	-2.16	56,56,56,56	0
2	CA	C	220	1/1	0.06	-2.24	49,49,49,49	0
2	CA	H	220	1/1	0.05	-2.44	42,42,42,42	0
2	CA	G	220	1/1	0.08	-2.51	40,40,40,40	0
2	CA	B	220	1/1	0.07	-2.64	48,48,48,48	0
2	CA	C	221	1/1	0.06	-2.72	65,65,65,65	0
2	CA	D	221	1/1	0.04	-3.04	49,49,49,49	0
2	CA	E	221	1/1	0.04	-3.41	54,54,54,54	0

6.5 Other polymers

There are no such residues in this entry.