wwPDB X-ray Structure Validation Summary Report (i)

Sep 17, 2023 – 06:12 PM EDT

PDB ID	:	4XTK
Title	:	Structure of TM1797, a CAS1 protein from Thermotoga maritima
Authors	:	Petit, P.; Beloglazova, N.; Skarina, T.; Chang, C.; Edwards, A.; Joachimiak, A.; Savchenko, A.; Midwest Center for Structural Genomics (MCSG)
Deposited on	:	2015-01-23
Resolution	:	2.70 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 (1)) were used in the production of this report:

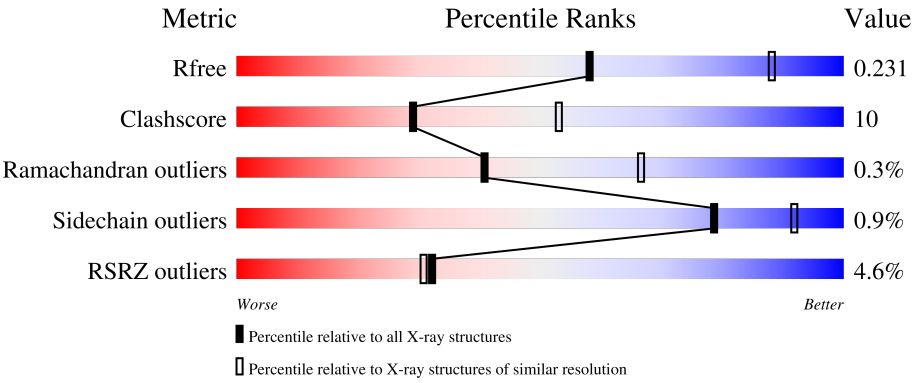
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 2.70 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.



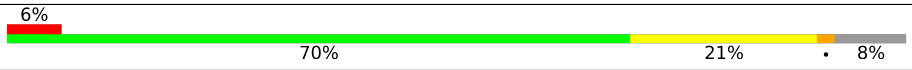
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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 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	326	5% 70% 23% • 5%
1	B	326	4% 76% 19% 5%
1	C	326	6% 68% 23% 9%
1	D	326	4% 79% 16% 5%
1	E	326	2% 81% 16% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	326	
1	G	326	
1	H	326	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 20014 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 CRISPR-associated endonuclease Cas1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	311	Total	C	N	O	S	0	0	0
			2476	1583	426	457	10			
1	B	309	Total	C	N	O	S	0	0	0
			2459	1575	415	459	10			
1	C	298	Total	C	N	O	S	0	0	0
			2382	1532	398	443	9			
1	D	311	Total	C	N	O	S	0	0	0
			2487	1595	423	460	9			
1	E	319	Total	C	N	O	S	0	0	0
			2609	1675	452	472	10			
1	F	314	Total	C	N	O	S	0	1	0
			2562	1651	443	458	10			
1	G	304	Total	C	N	O	S	0	1	0
			2465	1581	428	447	9			
1	H	300	Total	C	N	O	S	0	1	0
			2441	1569	421	442	9			

There are 56 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-6	ARG	-	expression tag	UNP Q9X2B7
A	-5	GLU	-	expression tag	UNP Q9X2B7
A	-4	LEU	-	expression tag	UNP Q9X2B7
A	-3	TYR	-	expression tag	UNP Q9X2B7
A	-2	PHE	-	expression tag	UNP Q9X2B7
A	-1	GLN	-	expression tag	UNP Q9X2B7
A	0	GLY	-	expression tag	UNP Q9X2B7
B	-6	ARG	-	expression tag	UNP Q9X2B7
B	-5	GLU	-	expression tag	UNP Q9X2B7
B	-4	LEU	-	expression tag	UNP Q9X2B7
B	-3	TYR	-	expression tag	UNP Q9X2B7
B	-2	PHE	-	expression tag	UNP Q9X2B7
B	-1	GLN	-	expression tag	UNP Q9X2B7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	GLY	-	expression tag	UNP Q9X2B7
C	-6	ARG	-	expression tag	UNP Q9X2B7
C	-5	GLU	-	expression tag	UNP Q9X2B7
C	-4	LEU	-	expression tag	UNP Q9X2B7
C	-3	TYR	-	expression tag	UNP Q9X2B7
C	-2	PHE	-	expression tag	UNP Q9X2B7
C	-1	GLN	-	expression tag	UNP Q9X2B7
C	0	GLY	-	expression tag	UNP Q9X2B7
D	-6	ARG	-	expression tag	UNP Q9X2B7
D	-5	GLU	-	expression tag	UNP Q9X2B7
D	-4	LEU	-	expression tag	UNP Q9X2B7
D	-3	TYR	-	expression tag	UNP Q9X2B7
D	-2	PHE	-	expression tag	UNP Q9X2B7
D	-1	GLN	-	expression tag	UNP Q9X2B7
D	0	GLY	-	expression tag	UNP Q9X2B7
E	-6	ARG	-	expression tag	UNP Q9X2B7
E	-5	GLU	-	expression tag	UNP Q9X2B7
E	-4	LEU	-	expression tag	UNP Q9X2B7
E	-3	TYR	-	expression tag	UNP Q9X2B7
E	-2	PHE	-	expression tag	UNP Q9X2B7
E	-1	GLN	-	expression tag	UNP Q9X2B7
E	0	GLY	-	expression tag	UNP Q9X2B7
F	-6	ARG	-	expression tag	UNP Q9X2B7
F	-5	GLU	-	expression tag	UNP Q9X2B7
F	-4	LEU	-	expression tag	UNP Q9X2B7
F	-3	TYR	-	expression tag	UNP Q9X2B7
F	-2	PHE	-	expression tag	UNP Q9X2B7
F	-1	GLN	-	expression tag	UNP Q9X2B7
F	0	GLY	-	expression tag	UNP Q9X2B7
G	-6	ARG	-	expression tag	UNP Q9X2B7
G	-5	GLU	-	expression tag	UNP Q9X2B7
G	-4	LEU	-	expression tag	UNP Q9X2B7
G	-3	TYR	-	expression tag	UNP Q9X2B7
G	-2	PHE	-	expression tag	UNP Q9X2B7
G	-1	GLN	-	expression tag	UNP Q9X2B7
G	0	GLY	-	expression tag	UNP Q9X2B7
H	-6	ARG	-	expression tag	UNP Q9X2B7
H	-5	GLU	-	expression tag	UNP Q9X2B7
H	-4	LEU	-	expression tag	UNP Q9X2B7
H	-3	TYR	-	expression tag	UNP Q9X2B7
H	-2	PHE	-	expression tag	UNP Q9X2B7
H	-1	GLN	-	expression tag	UNP Q9X2B7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
H	0	GLY	-	expression tag	UNP Q9X2B7

- Molecule 2 is water.

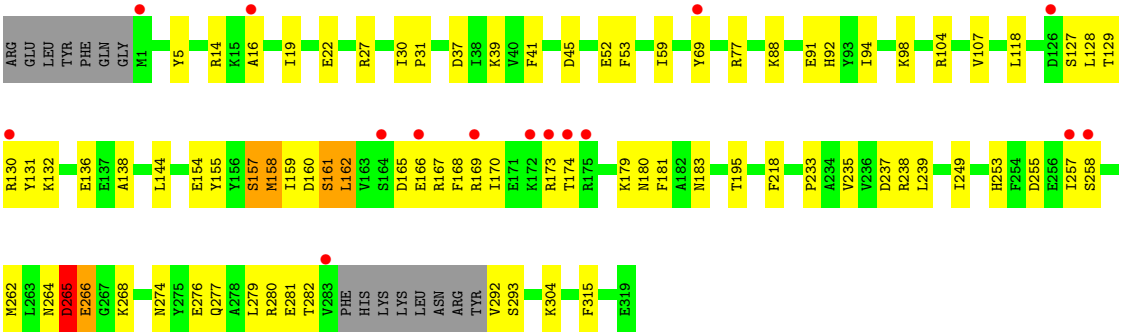
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	11	Total O 11 11	0	0
2	B	8	Total O 8 8	0	0
2	C	9	Total O 9 9	0	0
2	D	18	Total O 18 18	0	0
2	E	30	Total O 30 30	0	0
2	F	23	Total O 23 23	0	0
2	G	18	Total O 18 18	0	0
2	H	16	Total O 16 16	0	0

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.

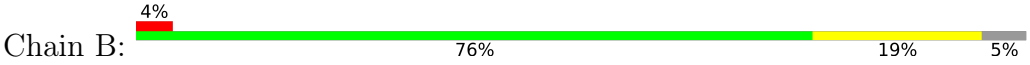
• Molecule 1: CRISPR-associated endonuclease Cas1

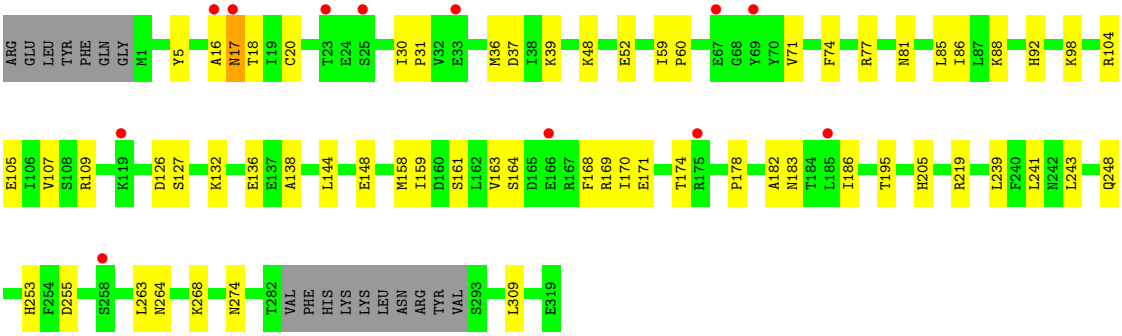
Chain A:



• Molecule 1: CRISPR-associated endonuclease Cas1

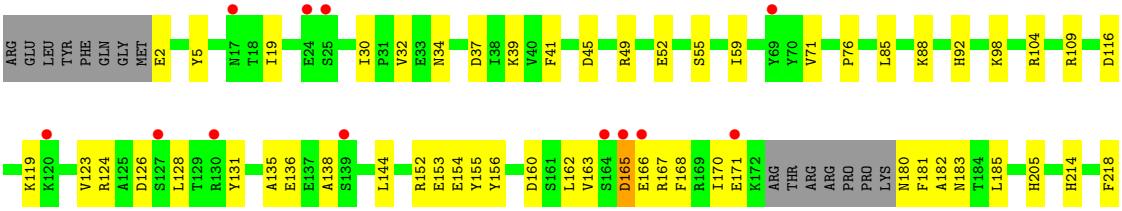
Chain B:

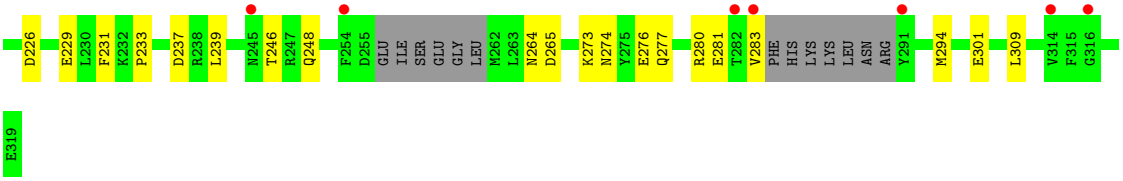




• Molecule 1: CRISPR-associated endonuclease Cas1

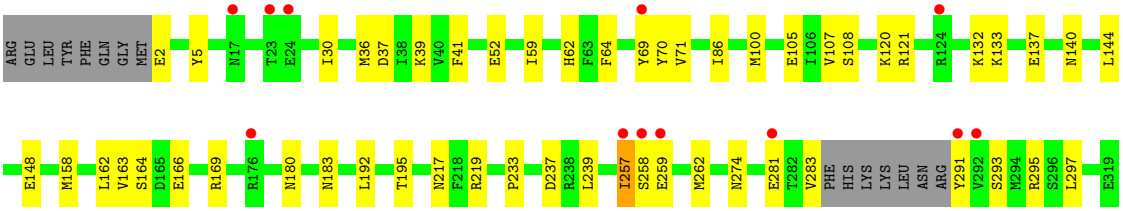
Chain C:





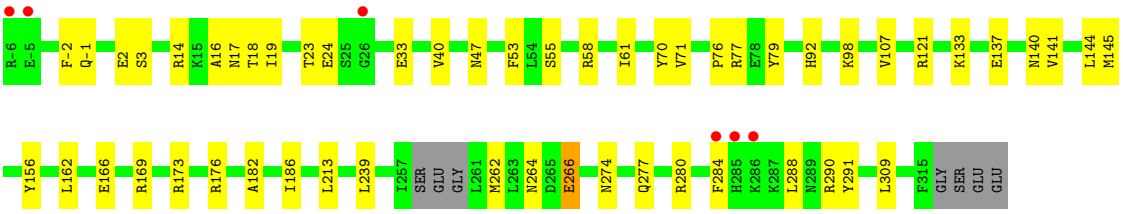
• Molecule 1: CRISPR-associated endonuclease Cas1

Chain D:



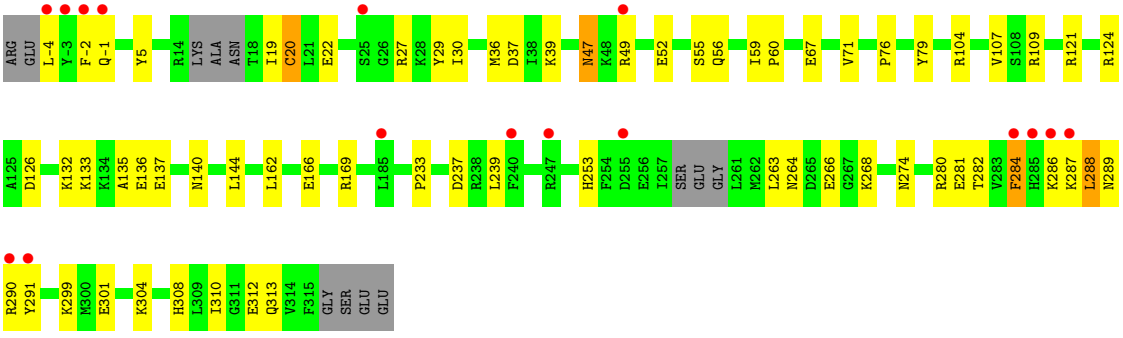
• Molecule 1: CRISPR-associated endonuclease Cas1

Chain E:



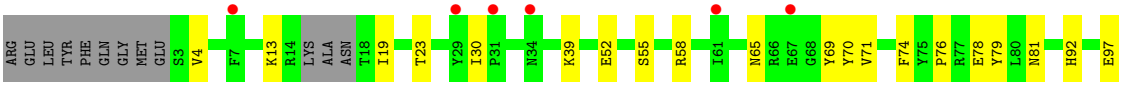
• Molecule 1: CRISPR-associated endonuclease Cas1

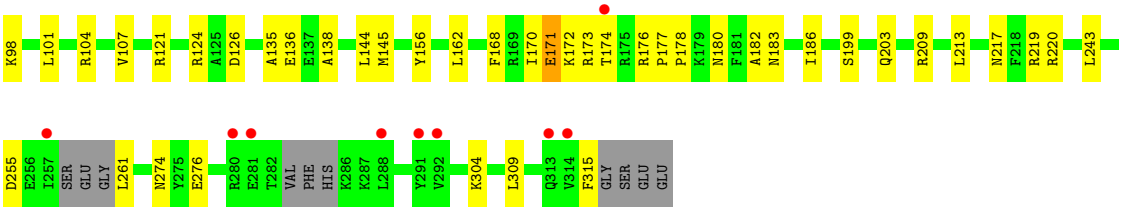
Chain F:



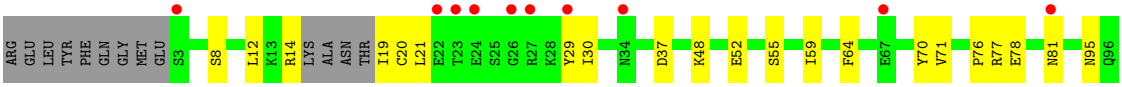
• Molecule 1: CRISPR-associated endonuclease Cas1

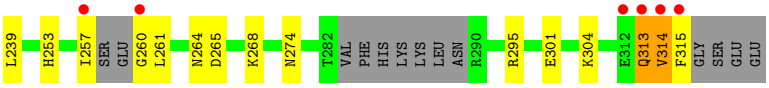
Chain G:





• Molecule 1: CRISPR-associated endonuclease Cas1





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	92.28Å 94.56Å 106.39Å 93.03° 115.06° 102.97°	Depositor
Resolution (Å)	19.85 – 2.70 19.85 – 2.70	Depositor EDS
% Data completeness (in resolution range)	97.6 (19.85-2.70) 97.6 (19.85-2.70)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	0.07	Depositor
$< I/\sigma(I) >$ [1]	3.09 (at 2.71Å)	Xtriage
Refinement program	PHENIX (phenix.refine: dev_1702)	Depositor
R, R_{free}	0.183 , 0.230 0.187 , 0.231	Depositor DCC
R_{free} test set	4202 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	56.9	Xtriage
Anisotropy	0.346	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 64.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	20014	wwPDB-VP
Average B, all atoms (Å^2)	76.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.74% 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	0.37	0/2522	0.57	0/3402
1	B	0.33	0/2505	0.54	0/3380
1	C	0.36	0/2425	0.59	0/3269
1	D	0.33	0/2534	0.50	0/3417
1	E	0.30	0/2660	0.48	0/3581
1	F	0.33	0/2616	0.54	1/3523 (0.0%)
1	G	0.33	0/2513	0.54	0/3385
1	H	0.33	0/2490	0.59	3/3351 (0.1%)
All	All	0.33	0/20265	0.54	4/27308 (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	A	0	2
1	B	0	1
1	C	0	2
1	D	0	1
1	G	0	1
1	H	0	2
All	All	0	9

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	H	144	LEU	CA-CB-CG	6.34	129.88	115.30
1	F	288	LEU	N-CA-C	-5.89	95.09	111.00
1	H	143	GLU	CA-CB-CG	5.43	125.35	113.40
1	H	140	ASN	N-CA-C	-5.01	97.47	111.00

There are no chirality outliers.

5 of 9 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	16	ALA	Peptide
1	A	265	ASP	Peptide
1	B	16	ALA	Peptide
1	C	126	ASP	Peptide
1	C	165	ASP	Peptide

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	2476	0	2405	68	0
1	B	2459	0	2386	43	0
1	C	2382	0	2310	53	0
1	D	2487	0	2426	38	0
1	E	2609	0	2567	41	0
1	F	2562	0	2528	48	1
1	G	2465	0	2421	51	1
1	H	2441	0	2404	66	0
2	A	11	0	0	0	0
2	B	8	0	0	0	0
2	C	9	0	0	1	0
2	D	18	0	0	0	0
2	E	30	0	0	1	0
2	F	23	0	0	2	0
2	G	18	0	0	0	0
2	H	16	0	0	3	0
All	All	20014	0	19447	376	1

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 10.

The worst 5 of 376 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:255:ASP:OD1	1:A:264:ASN:ND2	1.84	1.09
1:A:154:GLU:OE2	1:A:158:MET:CE	2.00	1.09
1:G:92:HIS:ND1	1:G:98:LYS:HD3	1.75	1.01
1:D:100:MET:HE1	1:D:140:ASN:HA	1.46	0.97
1:A:154:GLU:OE2	1:A:158:MET:HE1	1.66	0.94

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:F:280:ARG:NH1	1:G:23:THR:O[1_455]	2.13	0.07

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	307/326 (94%)	284 (92%)	18 (6%)	5 (2%)	9 24
1	B	305/326 (94%)	294 (96%)	10 (3%)	1 (0%)	41 66
1	C	290/326 (89%)	275 (95%)	14 (5%)	1 (0%)	41 66
1	D	307/326 (94%)	298 (97%)	9 (3%)	0	100 100
1	E	315/326 (97%)	305 (97%)	10 (3%)	0	100 100
1	F	309/326 (95%)	298 (96%)	11 (4%)	0	100 100
1	G	297/326 (91%)	287 (97%)	10 (3%)	0	100 100
1	H	293/326 (90%)	280 (96%)	12 (4%)	1 (0%)	41 66
All	All	2423/2608 (93%)	2321 (96%)	94 (4%)	8 (0%)	41 66

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	158	MET

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	166	GLU
1	A	157	SER
1	A	162	LEU
1	A	265	ASP

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	261/300 (87%)	260 (100%)	1 (0%)	91	97
1	B	261/300 (87%)	260 (100%)	1 (0%)	91	97
1	C	253/300 (84%)	251 (99%)	2 (1%)	81	93
1	D	264/300 (88%)	261 (99%)	3 (1%)	73	90
1	E	277/300 (92%)	275 (99%)	2 (1%)	84	94
1	F	273/300 (91%)	268 (98%)	5 (2%)	59	83
1	G	263/300 (88%)	260 (99%)	3 (1%)	73	90
1	H	261/300 (87%)	259 (99%)	2 (1%)	81	93
All	All	2113/2400 (88%)	2094 (99%)	19 (1%)	78	92

5 of 19 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	126	ASP
1	H	261	LEU
1	H	314	VAL
1	G	171	GLU
1	E	266	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	81	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	G	92	HIS
1	H	96	GLN
1	G	274	ASN
1	F	-1	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 monosaccharides 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	311/326 (95%)	-0.00	15 (4%)	30	28	32, 81, 127, 154	0
1	B	309/326 (94%)	-0.05	12 (3%)	39	38	37, 81, 126, 142	0
1	C	298/326 (91%)	0.10	19 (6%)	19	18	35, 85, 133, 162	0
1	D	311/326 (95%)	-0.04	12 (3%)	39	38	36, 79, 126, 141	0
1	E	319/326 (97%)	-0.37	6 (1%)	66	69	32, 59, 100, 124	0
1	F	314/326 (96%)	-0.18	16 (5%)	28	26	34, 66, 108, 135	0
1	G	304/326 (93%)	-0.08	15 (4%)	29	28	38, 72, 108, 135	0
1	H	300/326 (92%)	0.03	19 (6%)	20	19	42, 76, 109, 134	0
All	All	2466/2608 (94%)	-0.08	114 (4%)	32	31	32, 74, 122, 162	0

The worst 5 of 114 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	291	TYR	6.2
1	A	257	ILE	5.9
1	H	315	PHE	5.4
1	A	258	SER	5.3
1	H	34	ASN	5.2

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.