



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 1, 2026 – 11:51 AM UTC

PDB ID : 9CQG / pdb_00009cQG
Title : CRYSTAL STRUCTURE OF APO C-TERMINAL HIS-TAG DOG
HSP47(36-418) IN A P 1 CRYSTAL FORM
Authors : Sheriff, S.
Deposited on : 2024-07-19
Resolution : 2.47 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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

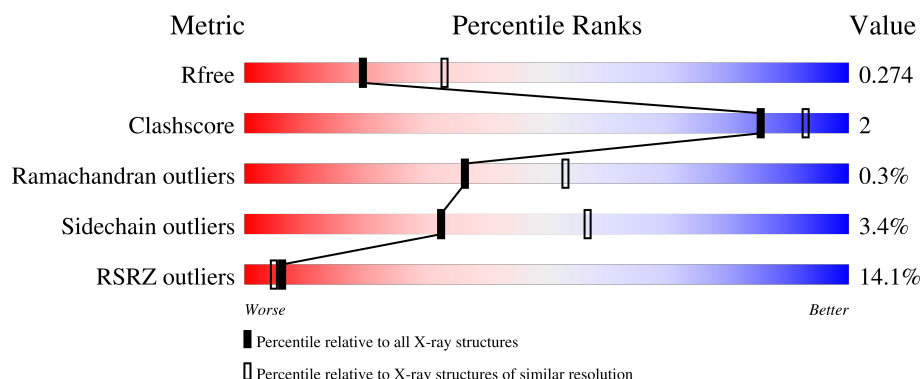
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.47 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}	180053	7589 (2.50-2.46)
Clashscore	190562	8295 (2.50-2.46)
Ramachandran outliers	187476	8164 (2.50-2.46)
Sidechain outliers	187428	8166 (2.50-2.46)
RSRZ outliers	180081	7593 (2.50-2.46)

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 $\leq 5\%$. 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	392	10% 93% 6% ..
1	B	392	10% 92% 8% .
1	D	392	8% 90% 8% .
1	E	392	26% 91% 9%
1	F	392	8% 93% 6% ..

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	392	22% 91% 8%
1	H	392	10% 92% 7%
1	I	392	17% 90% 9%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 48565 atoms, of which 24169 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 Serpin H1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	390	Total	C	H	N	O	S	3013	0	0
			6052	1932	3013	531	562	14			
1	B	392	Total	C	H	N	O	S	3047	0	0
			6115	1949	3047	541	564	14			
1	D	387	Total	C	H	N	O	S	2999	0	0
			6008	1915	2999	523	557	14			
1	E	392	Total	C	H	N	O	S	3037	0	0
			6098	1945	3037	538	564	14			
1	F	390	Total	C	H	N	O	S	3035	0	0
			6081	1937	3035	536	559	14			
1	G	391	Total	C	H	N	O	S	2991	0	0
			6027	1925	2991	533	564	14			
1	H	390	Total	C	H	N	O	S	3020	0	0
			6061	1934	3020	530	563	14			
1	I	391	Total	C	H	N	O	S	3027	0	0
			6078	1938	3027	538	561	14			

There are 72 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	35	MET	-	initiating methionine	UNP C7C419
A	419	LEU	-	expression tag	UNP C7C419
A	420	GLU	-	expression tag	UNP C7C419
A	421	HIS	-	expression tag	UNP C7C419
A	422	HIS	-	expression tag	UNP C7C419
A	423	HIS	-	expression tag	UNP C7C419
A	424	HIS	-	expression tag	UNP C7C419
A	425	HIS	-	expression tag	UNP C7C419
A	426	HIS	-	expression tag	UNP C7C419
B	35	MET	-	initiating methionine	UNP C7C419
B	419	LEU	-	expression tag	UNP C7C419
B	420	GLU	-	expression tag	UNP C7C419
B	421	HIS	-	expression tag	UNP C7C419

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	422	HIS	-	expression tag	UNP C7C419
B	423	HIS	-	expression tag	UNP C7C419
B	424	HIS	-	expression tag	UNP C7C419
B	425	HIS	-	expression tag	UNP C7C419
B	426	HIS	-	expression tag	UNP C7C419
D	35	MET	-	initiating methionine	UNP C7C419
D	419	LEU	-	expression tag	UNP C7C419
D	420	GLU	-	expression tag	UNP C7C419
D	421	HIS	-	expression tag	UNP C7C419
D	422	HIS	-	expression tag	UNP C7C419
D	423	HIS	-	expression tag	UNP C7C419
D	424	HIS	-	expression tag	UNP C7C419
D	425	HIS	-	expression tag	UNP C7C419
D	426	HIS	-	expression tag	UNP C7C419
E	35	MET	-	initiating methionine	UNP C7C419
E	419	LEU	-	expression tag	UNP C7C419
E	420	GLU	-	expression tag	UNP C7C419
E	421	HIS	-	expression tag	UNP C7C419
E	422	HIS	-	expression tag	UNP C7C419
E	423	HIS	-	expression tag	UNP C7C419
E	424	HIS	-	expression tag	UNP C7C419
E	425	HIS	-	expression tag	UNP C7C419
E	426	HIS	-	expression tag	UNP C7C419
F	35	MET	-	initiating methionine	UNP C7C419
F	419	LEU	-	expression tag	UNP C7C419
F	420	GLU	-	expression tag	UNP C7C419
F	421	HIS	-	expression tag	UNP C7C419
F	422	HIS	-	expression tag	UNP C7C419
F	423	HIS	-	expression tag	UNP C7C419
F	424	HIS	-	expression tag	UNP C7C419
F	425	HIS	-	expression tag	UNP C7C419
F	426	HIS	-	expression tag	UNP C7C419
G	35	MET	-	initiating methionine	UNP C7C419
G	419	LEU	-	expression tag	UNP C7C419
G	420	GLU	-	expression tag	UNP C7C419
G	421	HIS	-	expression tag	UNP C7C419
G	422	HIS	-	expression tag	UNP C7C419
G	423	HIS	-	expression tag	UNP C7C419
G	424	HIS	-	expression tag	UNP C7C419
G	425	HIS	-	expression tag	UNP C7C419
G	426	HIS	-	expression tag	UNP C7C419
H	35	MET	-	initiating methionine	UNP C7C419

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
H	419	LEU	-	expression tag	UNP C7C419
H	420	GLU	-	expression tag	UNP C7C419
H	421	HIS	-	expression tag	UNP C7C419
H	422	HIS	-	expression tag	UNP C7C419
H	423	HIS	-	expression tag	UNP C7C419
H	424	HIS	-	expression tag	UNP C7C419
H	425	HIS	-	expression tag	UNP C7C419
H	426	HIS	-	expression tag	UNP C7C419
I	35	MET	-	initiating methionine	UNP C7C419
I	419	LEU	-	expression tag	UNP C7C419
I	420	GLU	-	expression tag	UNP C7C419
I	421	HIS	-	expression tag	UNP C7C419
I	422	HIS	-	expression tag	UNP C7C419
I	423	HIS	-	expression tag	UNP C7C419
I	424	HIS	-	expression tag	UNP C7C419
I	425	HIS	-	expression tag	UNP C7C419
I	426	HIS	-	expression tag	UNP C7C419

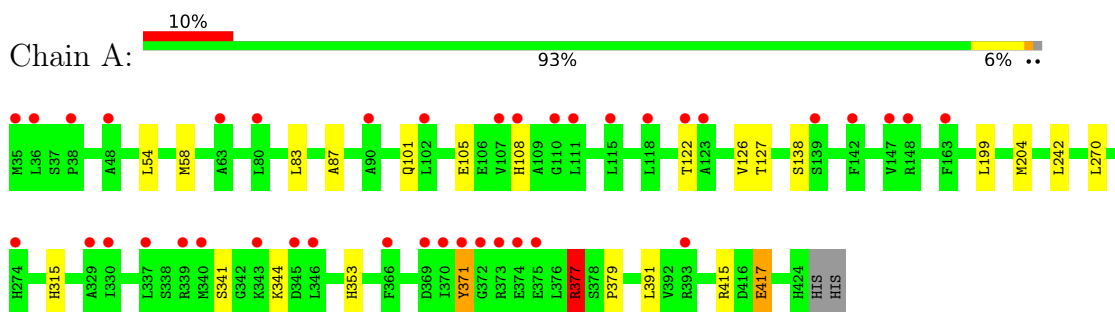
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	12	Total O 12 12	0	0
2	B	7	Total O 7 7	0	0
2	D	4	Total O 4 4	0	0
2	E	8	Total O 8 8	0	0
2	F	4	Total O 4 4	0	0
2	G	5	Total O 5 5	0	0
2	H	1	Total O 1 1	0	0
2	I	4	Total O 4 4	0	0

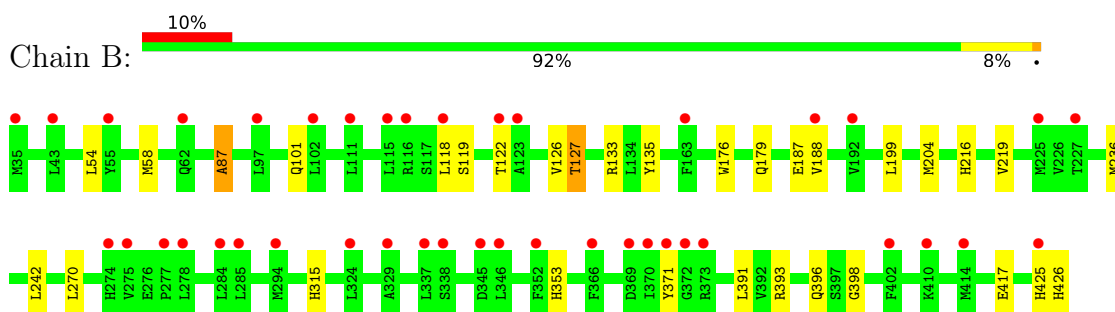
3 Residue-property plots [i](#)

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.

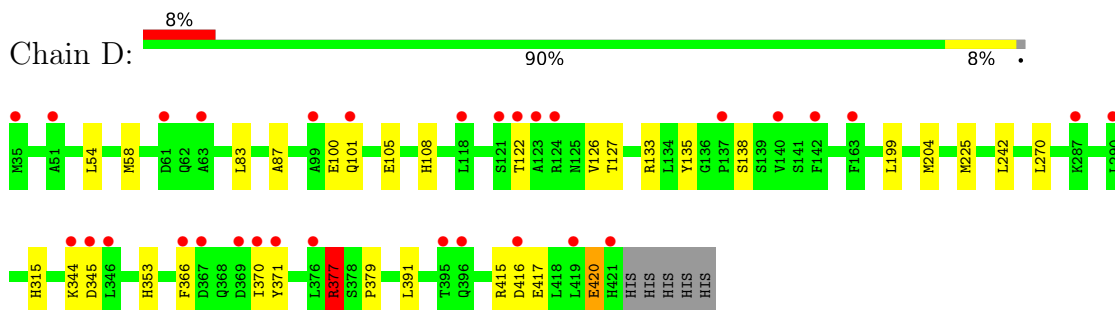
• Molecule 1: Serpin H1



• Molecule 1: Serpin H1

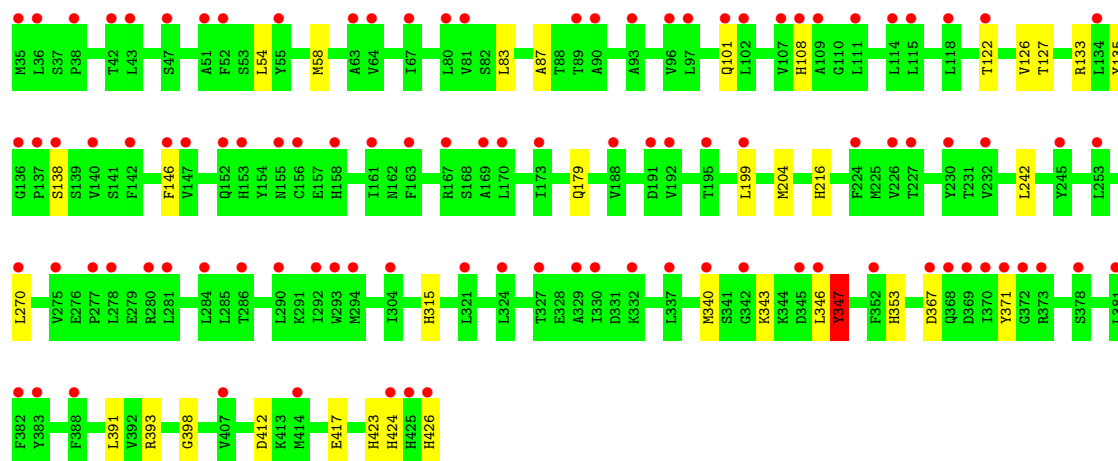


• Molecule 1: Serpin H1

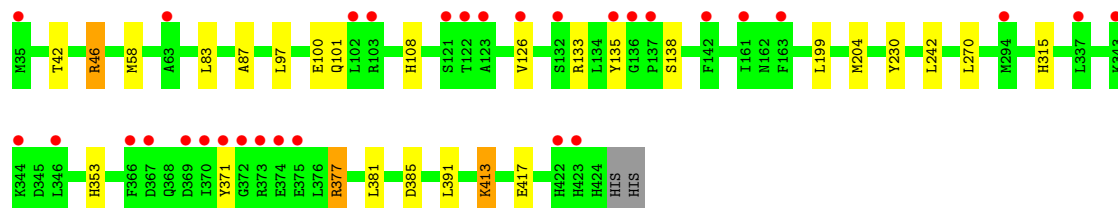
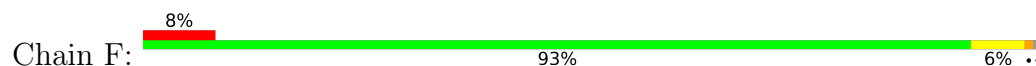


• Molecule 1: Serpin H1

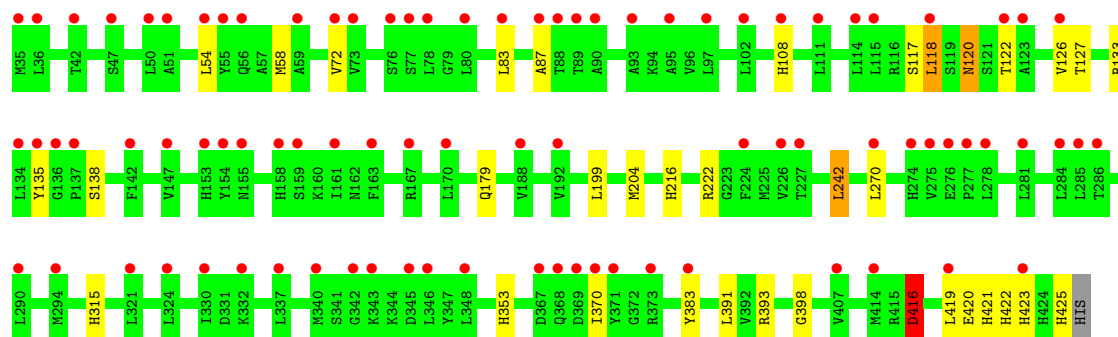
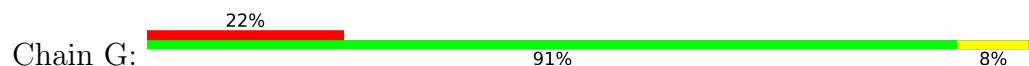




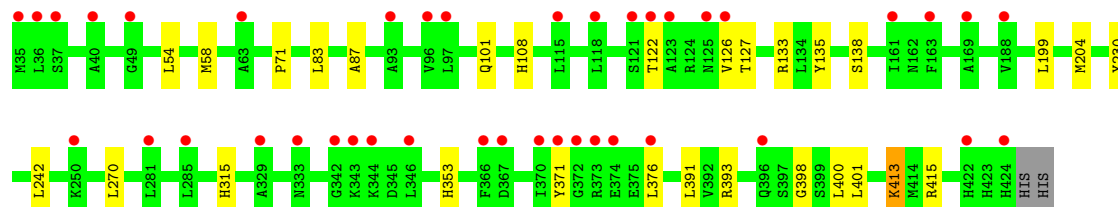
• Molecule 1: Serpin H1



• Molecule 1: Serpin H1

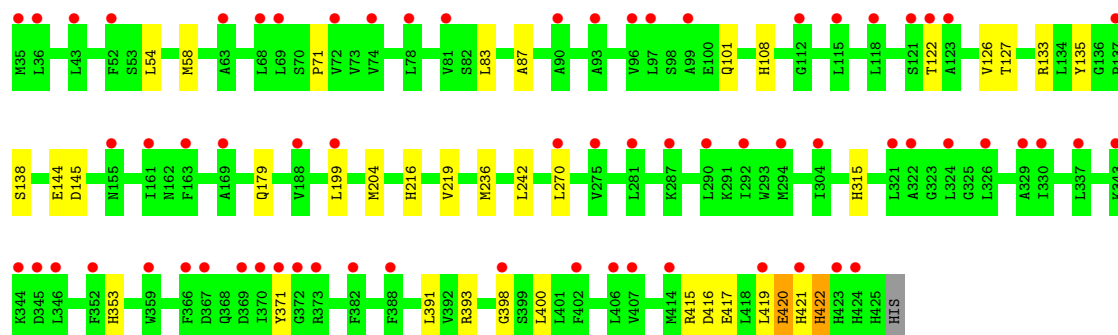


• Molecule 1: Serpin H1



● Molecule 1: Serpin H1

Chain I:  17% 90% 9%



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	87.30Å 92.00Å 123.58Å 85.23° 70.69° 83.23°	Depositor
Resolution (Å)	116.50 – 2.47 116.50 – 2.47	Depositor EDS
% Data completeness (in resolution range)	69.4 (116.50-2.47) 69.4 (116.50-2.47)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.81 (at 2.48Å)	Xtriage
Refinement program	BUSTER 2.11.8	Depositor
R, R_{free}	0.243 , 0.264 (Not available) , 0.274	Depositor DCC
R_{free} test set	4465 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å ²)	62.3	Xtriage
Anisotropy	0.022	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 48.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	48565	wwPDB-VP
Average B, all atoms (Å ²)	73.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.78% 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	0.79	3/3104 (0.1%)	1.03	1/4193 (0.0%)
1	B	0.81	2/3135 (0.1%)	1.04	1/4233 (0.0%)
1	D	0.79	2/3071 (0.1%)	1.03	0/4147
1	E	0.77	2/3128 (0.1%)	1.02	0/4224
1	F	0.78	2/3111 (0.1%)	1.02	0/4200
1	G	0.75	1/3101 (0.0%)	1.04	3/4190 (0.1%)
1	H	0.75	1/3105 (0.0%)	1.00	0/4193
1	I	0.76	2/3117 (0.1%)	1.04	6/4209 (0.1%)
All	All	0.78	15/24872 (0.1%)	1.03	11/33589 (0.0%)

The worst 5 of 15 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	87	ALA	CA-C	7.05	1.62	1.52
1	D	87	ALA	CA-C	6.73	1.61	1.52
1	I	87	ALA	CA-C	6.57	1.61	1.52
1	E	87	ALA	CA-C	6.29	1.61	1.52
1	G	87	ALA	CA-C	6.28	1.61	1.52

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	G	120	ASN	CA-CB-CG	8.18	120.78	112.60
1	G	416	ASP	CA-CB-CG	7.00	119.60	112.60
1	I	419	LEU	CA-C-N	6.87	132.46	122.77
1	I	419	LEU	C-N-CA	6.87	132.46	122.77
1	I	145	ASP	CA-CB-CG	6.64	119.24	112.60

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	3039	3013	3013	9	0
1	B	3068	3047	3047	15	0
1	D	3009	2999	2999	12	0
1	E	3061	3037	3037	16	0
1	F	3046	3035	3035	12	0
1	G	3036	2991	2991	20	0
1	H	3041	3020	3020	13	0
1	I	3051	3027	3027	16	0
2	A	12	0	0	0	0
2	B	7	0	0	0	0
2	D	4	0	0	0	0
2	E	8	0	0	0	0
2	F	4	0	0	0	0
2	G	5	0	0	0	0
2	H	1	0	0	0	0
2	I	4	0	0	0	0
All	All	24396	24169	24169	105	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 2.

The worst 5 of 105 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:E:58:MET:HE1	1:E:315:HIS:ND1	2.06	0.71
1:H:58:MET:HE1	1:H:315:HIS:ND1	2.06	0.70
1:I:58:MET:HE1	1:I:315:HIS:ND1	2.07	0.70
1:A:58:MET:HE1	1:A:315:HIS:ND1	2.07	0.69
1:D:58:MET:HE1	1:D:315:HIS:ND1	2.07	0.69

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	388/392 (99%)	360 (93%)	26 (7%)	2 (0%)	24	40
1	B	390/392 (100%)	364 (93%)	25 (6%)	1 (0%)	36	53
1	D	385/392 (98%)	350 (91%)	32 (8%)	3 (1%)	16	28
1	E	390/392 (100%)	359 (92%)	29 (7%)	2 (0%)	24	40
1	F	388/392 (99%)	362 (93%)	25 (6%)	1 (0%)	36	53
1	G	389/392 (99%)	363 (93%)	26 (7%)	0	100	100
1	H	388/392 (99%)	362 (93%)	26 (7%)	0	100	100
1	I	389/392 (99%)	364 (94%)	24 (6%)	1 (0%)	36	53
All	All	3107/3136 (99%)	2884 (93%)	213 (7%)	10 (0%)	36	53

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	347	TYR
1	I	422	HIS
1	A	377	ARG
1	B	87	ALA
1	D	377	ARG

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	322/337 (96%)	311 (97%)	11 (3%)	32	57

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	326/337 (97%)	314 (96%)	12 (4%)	30	54
1	D	319/337 (95%)	305 (96%)	14 (4%)	25	47
1	E	325/337 (96%)	314 (97%)	11 (3%)	32	57
1	F	323/337 (96%)	312 (97%)	11 (3%)	32	57
1	G	320/337 (95%)	312 (98%)	8 (2%)	42	66
1	H	322/337 (96%)	311 (97%)	11 (3%)	32	57
1	I	323/337 (96%)	312 (97%)	11 (3%)	32	57
All	All	2580/2696 (96%)	2491 (97%)	89 (3%)	32	57

5 of 89 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	413	LYS
1	H	242	LEU
1	G	118	LEU
1	G	416	ASP
1	H	400	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	353	HIS
1	H	171	GLN
1	G	425	HIS
1	H	353	HIS
1	B	426	HIS

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 [i](#)

There are no oligosaccharides 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, 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	390/392 (99%)	0.65	39 (10%) 12 10	15, 31, 59, 72	0
1	B	392/392 (100%)	0.88	41 (10%) 11 9	15, 36, 56, 72	0
1	D	387/392 (98%)	0.69	31 (8%) 18 16	16, 32, 56, 72	0
1	E	392/392 (100%)	1.35	102 (26%) 1 1	15, 40, 63, 82	0
1	F	390/392 (99%)	0.66	31 (7%) 18 16	17, 32, 62, 75	0
1	G	391/392 (99%)	1.22	87 (22%) 2 2	15, 41, 73, 91	0
1	H	390/392 (99%)	0.87	40 (10%) 12 10	19, 37, 62, 77	0
1	I	391/392 (99%)	1.18	68 (17%) 4 3	20, 41, 63, 76	0
All	All	3123/3136 (99%)	0.94	439 (14%) 6 5	15, 36, 63, 91	0

The worst 5 of 439 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	118	LEU	5.2
1	E	163	PHE	5.0
1	I	345	ASP	4.9
1	B	371	TYR	4.7
1	I	371	TYR	4.6

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 oligosaccharides in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.