wwPDB X-ray Structure Validation Summary Report (i)

Oct 20, 2024 – 08:30 PM EDT

PDB ID	:	9CQF
Title	:	CRYSTAL STRUCTURE OF APO C-TERMINAL HIS-TAG DOG HSP47(36-418) IN A C 2 2 21 CRYSTAL FORM
Authors	:	Sheriff, S.
Deposited on	:	2024-07-19
Resolution	:	2.93 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:

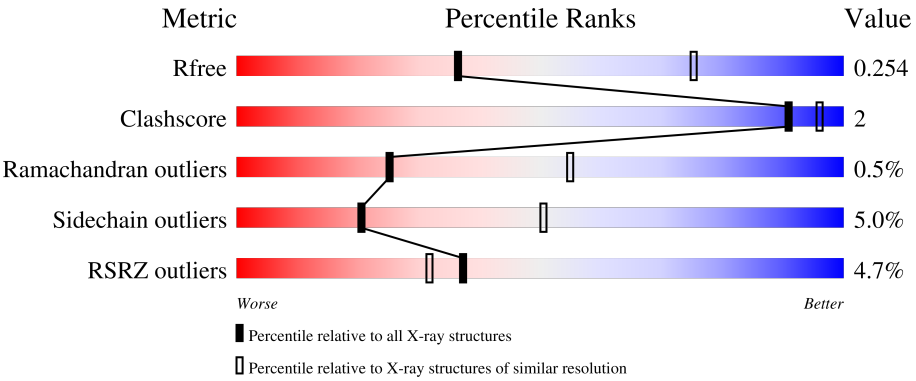
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

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 2.93 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.



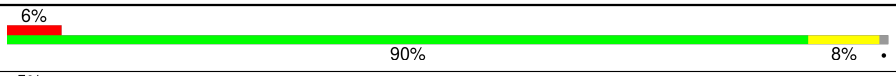
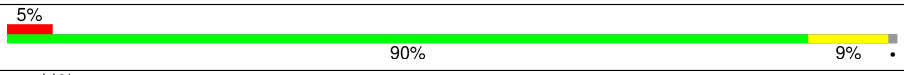
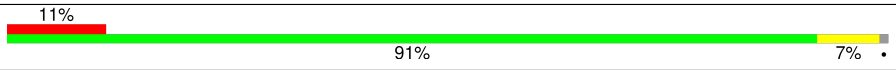
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	2797 (2.94-2.90)
Clashscore	180529	3049 (2.94-2.90)
Ramachandran outliers	177936	2981 (2.94-2.90)
Sidechain outliers	177891	2983 (2.94-2.90)
RSRZ outliers	164620	2799 (2.94-2.90)

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	392	3% 91% 7% ••
1	B	392	3% 91% 8% ••
1	D	392	3% 90% 9% •
1	E	392	3% 90% 6% ••
1	F	392	4% 90% 9% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	392	
1	H	392	
1	I	392	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 48235 atoms, of which 24093 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	387	Total	C	H	N	O	S	3002	0	0
			6011	1914	3002	525	556	14			
1	B	388	Total	C	H	N	O	S	3016	0	0
			6037	1922	3016	528	557	14			
1	D	390	Total	C	H	N	O	S	3028	0	0
			6069	1933	3028	534	560	14			
1	E	387	Total	C	H	N	O	S	3009	0	0
			6018	1912	3009	525	558	14			
1	F	388	Total	C	H	N	O	S	3020	0	0
			6051	1926	3020	530	561	14			
1	G	387	Total	C	H	N	O	S	3007	0	0
			6016	1913	3007	527	555	14			
1	H	387	Total	C	H	N	O	S	3002	0	0
			6006	1910	3002	525	555	14			
1	I	387	Total	C	H	N	O	S	3009	0	0
			6020	1916	3009	525	556	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.

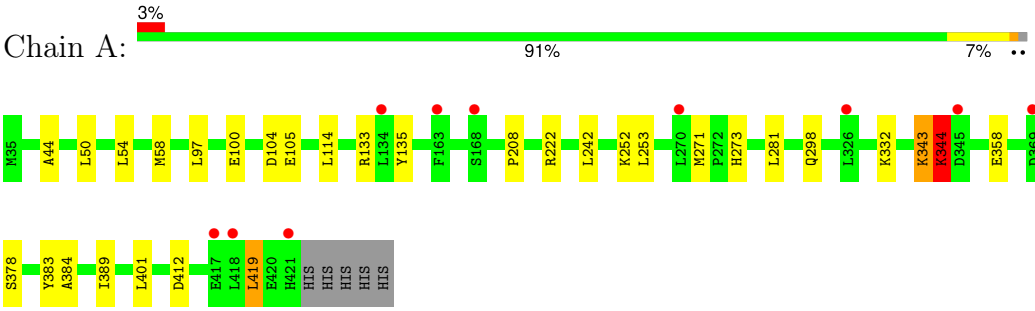
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	2	Total O 2 2	0	0
2	D	2	Total O 2 2	0	0
2	E	2	Total O 2 2	0	0
2	F	1	Total O 1 1	0	0

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.

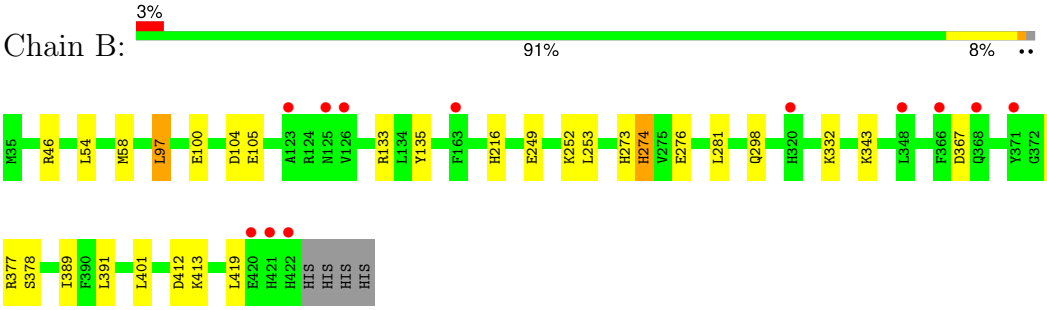
- Molecule 1: Serpin H1

Chain A:



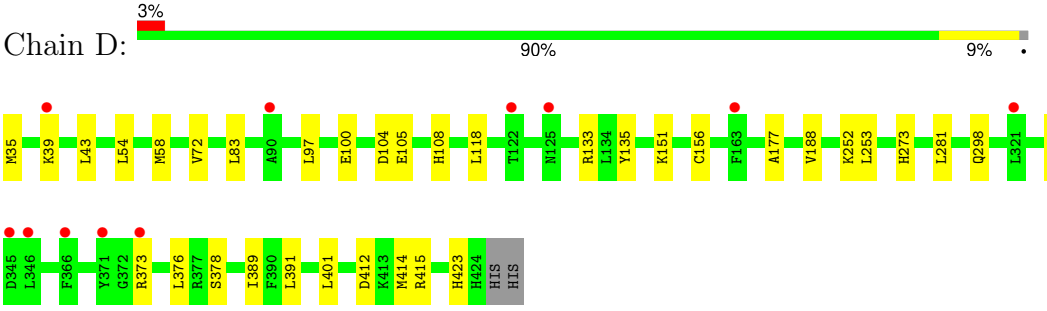
- Molecule 1: Serpin H1

Chain B:



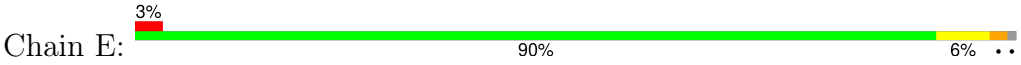
- Molecule 1: Serpin H1

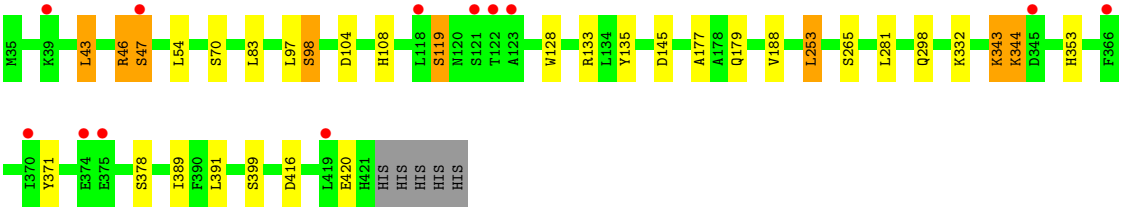
Chain D:



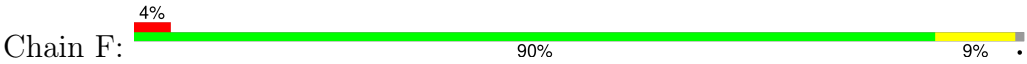
- Molecule 1: Serpin H1

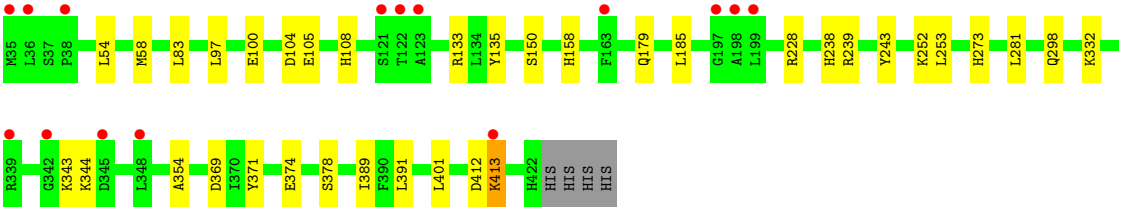
Chain E:



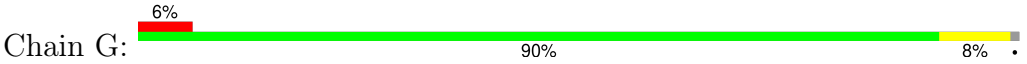


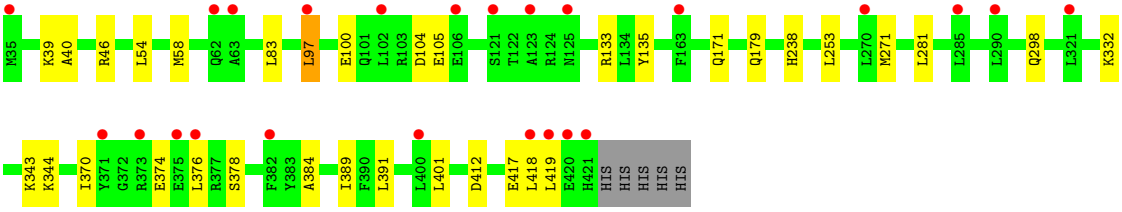
• Molecule 1: Serpin H1



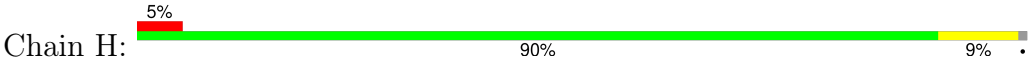


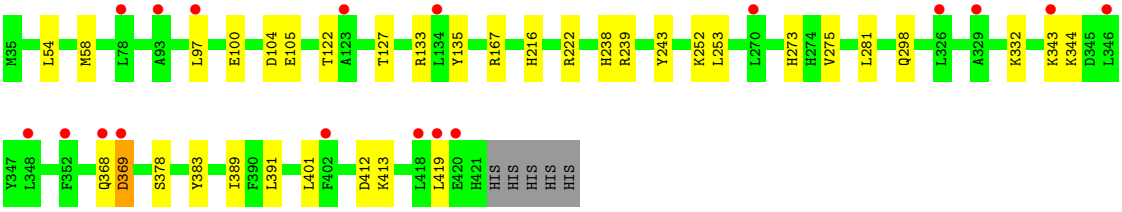
• Molecule 1: Serpin H1



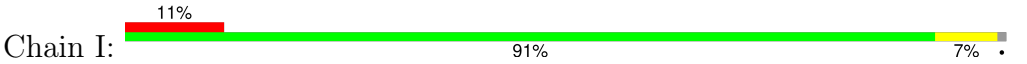


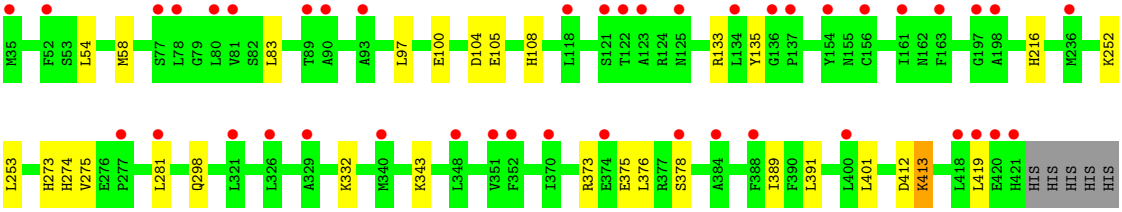
• Molecule 1: Serpin H1





• Molecule 1: Serpin H1





4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	117.78Å 129.94Å 501.93Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	250.97 – 2.93 250.97 – 2.93	Depositor EDS
% Data completeness (in resolution range)	76.0 (250.97-2.93) 76.0 (250.97-2.93)	Depositor EDS
R_{merge}	0.22	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.06 (at 2.91Å)	Xtriage
Refinement program	BUSTER 2.11.8	Depositor
R, R_{free}	0.228 , 0.252 0.231 , 0.254	Depositor DCC
R_{free} test set	3075 reflections (4.86%)	wwPDB-VP
Wilson B-factor (Å^2)	71.6	Xtriage
Anisotropy	0.058	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 83.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	48235	wwPDB-VP
Average B, all atoms (Å^2)	78.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.31% 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	0.48	0/3070	0.65	1/4144 (0.0%)
1	B	0.43	0/3083	0.61	0/4162
1	D	0.46	0/3105	0.63	0/4192
1	E	0.50	0/3069	0.65	0/4142
1	F	0.43	0/3094	0.61	0/4176
1	G	0.43	0/3070	0.62	0/4144
1	H	0.40	0/3064	0.60	0/4136
1	I	0.39	0/3072	0.60	0/4147
All	All	0.44	0/24627	0.62	1/33243 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	419	LEU	N-CA-C	-5.42	96.36	111.00

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	3009	3002	3002	11	0
1	B	3021	3016	3016	8	0
1	D	3041	3028	3028	11	0
1	E	3009	3009	3009	12	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	F	3031	3020	3020	16	0
1	G	3009	3007	3007	11	0
1	H	3004	3002	3002	12	0
1	I	3011	3009	3009	8	0
2	A	2	0	0	0	0
2	D	2	0	0	0	0
2	E	2	0	0	0	0
2	F	1	0	0	0	0
All	All	24142	24093	24093	79	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 79 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:D:83:LEU:HD13	1:D:108:HIS:CD2	2.36	0.60
1:A:273:HIS:NE2	1:F:228:ARG:NH1	2.52	0.57
1:H:167:ARG:NH2	1:I:413:LYS:HZ1	2.03	0.57
1:F:83:LEU:HD13	1:F:108:HIS:CD2	2.43	0.53
1:F:252:LYS:HA	1:F:273:HIS:CE1	2.44	0.53

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	385/392 (98%)	363 (94%)	21 (6%)	1 (0%)	37 65
1	B	386/392 (98%)	357 (92%)	25 (6%)	4 (1%)	13 38
1	D	388/392 (99%)	362 (93%)	25 (6%)	1 (0%)	37 65

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	E	385/392 (98%)	354 (92%)	29 (8%)	2 (0%)	25	55
1	F	386/392 (98%)	361 (94%)	24 (6%)	1 (0%)	37	65
1	G	385/392 (98%)	355 (92%)	28 (7%)	2 (0%)	25	55
1	H	385/392 (98%)	361 (94%)	22 (6%)	2 (0%)	25	55
1	I	385/392 (98%)	353 (92%)	29 (8%)	3 (1%)	16	44
All	All	3085/3136 (98%)	2866 (93%)	203 (7%)	16 (0%)	25	55

5 of 16 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	371	TYR
1	A	344	LYS
1	B	367	ASP
1	B	373	ARG
1	E	420	GLU

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	318/337 (94%)	303 (95%)	15 (5%)	22	53
1	B	320/337 (95%)	303 (95%)	17 (5%)	19	48
1	D	322/337 (96%)	303 (94%)	19 (6%)	16	43
1	E	320/337 (95%)	301 (94%)	19 (6%)	16	43
1	F	322/337 (96%)	309 (96%)	13 (4%)	27	59
1	G	319/337 (95%)	303 (95%)	16 (5%)	20	50
1	H	318/337 (94%)	304 (96%)	14 (4%)	24	55
1	I	319/337 (95%)	304 (95%)	15 (5%)	22	53
All	All	2558/2696 (95%)	2430 (95%)	128 (5%)	20	50

5 of 128 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	I	104	ASP
1	I	298	GLN
1	D	412	ASP
1	D	401	LEU
1	I	343	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	H	353	HIS
1	I	353	HIS
1	D	423	HIS
1	E	171	GLN
1	E	353	HIS

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 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, 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	387/392 (98%)	0.04	12 (3%)	51	46	16, 31, 56, 70	0
1	B	388/392 (98%)	0.43	12 (3%)	51	46	18, 40, 65, 77	0
1	D	390/392 (99%)	0.18	11 (2%)	55	49	15, 34, 58, 72	0
1	E	387/392 (98%)	-0.01	12 (3%)	51	46	16, 29, 51, 75	0
1	F	388/392 (98%)	0.28	15 (3%)	44	38	21, 38, 66, 80	0
1	G	387/392 (98%)	0.57	24 (6%)	28	24	19, 40, 64, 94	0
1	H	387/392 (98%)	0.62	18 (4%)	37	32	24, 46, 73, 87	0
1	I	387/392 (98%)	0.92	43 (11%)	12	11	25, 53, 80, 104	0
All	All	3101/3136 (98%)	0.38	147 (4%)	37	32	15, 38, 69, 104	0

The worst 5 of 147 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	125	ASN	5.3
1	I	421	HIS	5.1
1	I	121	SER	4.7
1	F	345	ASP	4.5
1	I	78	LEU	4.2

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.