wwPDB X-ray Structure Validation Summary Report (i)

Oct 19, 2024 – 06:39 PM EDT

PDB ID	:	9CQH
Title	:	CRYSTAL STRUCTURE OF APO CLEAVED N-TERNMINAL HIS-TAG GAGA-DOG HSP47(36-418) IN A C 2 CRYSTAL FORM
Authors	:	Sheriff, S.
Deposited on	:	2024-07-19
Resolution	:	2.01 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 (i)) were used in the production of this report:

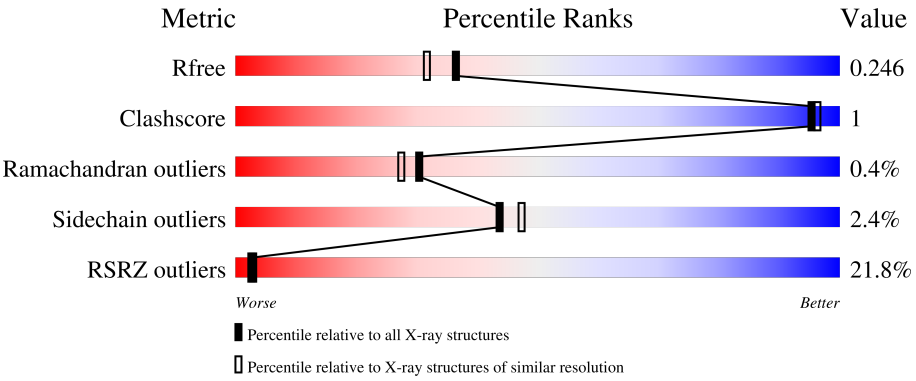
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (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 2.01 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.



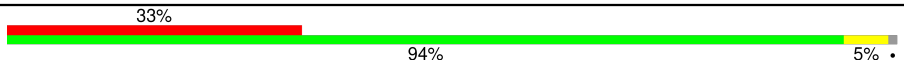
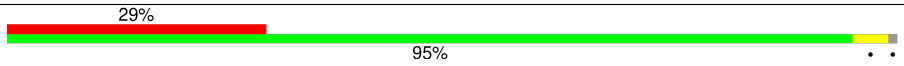
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	9409 (2.00-2.00)
Clashscore	180529	10737 (2.00-2.00)
Ramachandran outliers	177936	10628 (2.00-2.00)
Sidechain outliers	177891	10627 (2.00-2.00)
RSRZ outliers	164620	9409 (2.00-2.00)

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	387	14% 94% 5%
1	B	387	13% 93% 5%
1	C	387	17% 94%
1	D	387	28% 94% 5%
1	E	387	16% 93% 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	387	
1	G	387	

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 41762 atoms, of which 20444 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	383	Total	C	H	N	O	S	2879	0	0
			5802	1864	2879	506	540	13			
1	B	382	Total	C	H	N	O	S	2950	1	0
			5911	1885	2950	517	546	13			
1	C	383	Total	C	H	N	O	S	2943	0	0
			5902	1885	2943	515	546	13			
1	D	383	Total	C	H	N	O	S	2935	0	0
			5897	1882	2935	516	551	13			
1	E	383	Total	C	H	N	O	S	2945	1	0
			5910	1887	2945	519	546	13			
1	F	383	Total	C	H	N	O	S	2879	0	0
			5806	1864	2879	506	544	13			
1	G	383	Total	C	H	N	O	S	2897	0	0
			5843	1876	2897	510	547	13			

There are 28 discrepancies between the modelled and reference sequences:

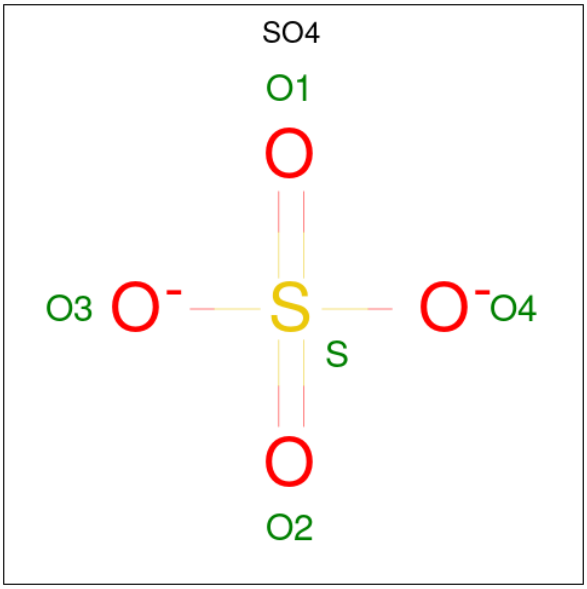
Chain	Residue	Modelled	Actual	Comment	Reference
A	32	GLY	-	expression tag	UNP C7C419
A	33	ALA	-	expression tag	UNP C7C419
A	34	GLY	-	expression tag	UNP C7C419
A	35	ALA	-	expression tag	UNP C7C419
B	32	GLY	-	expression tag	UNP C7C419
B	33	ALA	-	expression tag	UNP C7C419
B	34	GLY	-	expression tag	UNP C7C419
B	35	ALA	-	expression tag	UNP C7C419
C	32	GLY	-	expression tag	UNP C7C419
C	33	ALA	-	expression tag	UNP C7C419
C	34	GLY	-	expression tag	UNP C7C419
C	35	ALA	-	expression tag	UNP C7C419
D	32	GLY	-	expression tag	UNP C7C419
D	33	ALA	-	expression tag	UNP C7C419
D	34	GLY	-	expression tag	UNP C7C419

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	35	ALA	-	expression tag	UNP C7C419
E	32	GLY	-	expression tag	UNP C7C419
E	33	ALA	-	expression tag	UNP C7C419
E	34	GLY	-	expression tag	UNP C7C419
E	35	ALA	-	expression tag	UNP C7C419
F	32	GLY	-	expression tag	UNP C7C419
F	33	ALA	-	expression tag	UNP C7C419
F	34	GLY	-	expression tag	UNP C7C419
F	35	ALA	-	expression tag	UNP C7C419
G	32	GLY	-	expression tag	UNP C7C419
G	33	ALA	-	expression tag	UNP C7C419
G	34	GLY	-	expression tag	UNP C7C419
G	35	ALA	-	expression tag	UNP C7C419

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



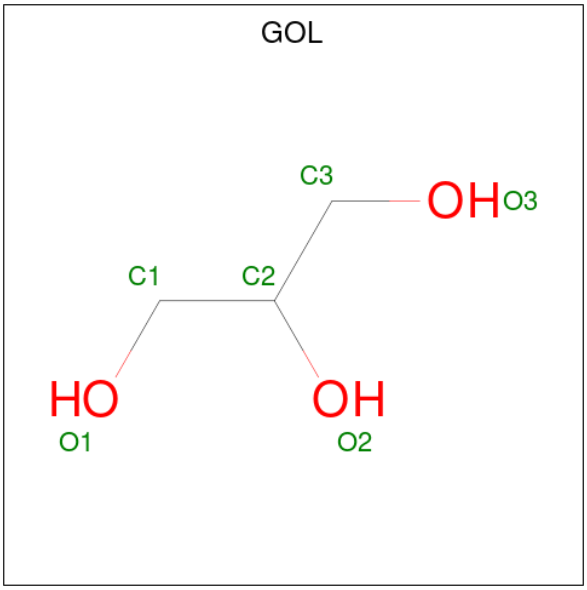
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total O S 5 4 1	0	0
2	A	1	Total O S 5 4 1	0	0
2	A	1	Total O S 5 4 1	0	0
2	A	1	Total O S 5 4 1	0	0
2	B	1	Total O S 5 4 1	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	C	1	Total O S 5 4 1	0	0
2	E	1	Total O S 5 4 1	0	0
2	E	1	Total O S 5 4 1	0	0

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total C H O 14 3 8 3	8	0
3	D	1	Total C H O 14 3 8 3	8	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	95	Total O 95 95	0	0
4	B	88	Total O 88 88	0	0
4	C	91	Total O 91 91	0	0
4	D	86	Total O 86 86	0	0

Continued on next page...

Continued from previous page...

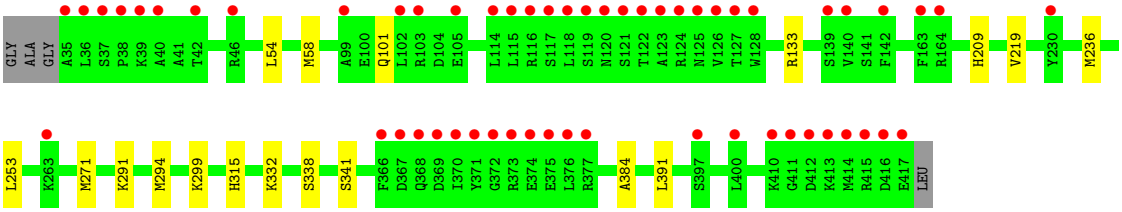
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	E	97	Total O 97 97	0	0
4	F	85	Total O 85 85	0	0
4	G	81	Total O 81 81	0	0

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.

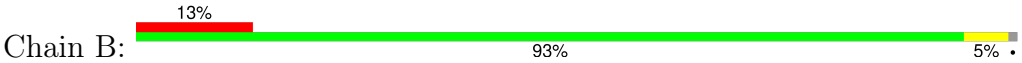
• Molecule 1: Serpin H1

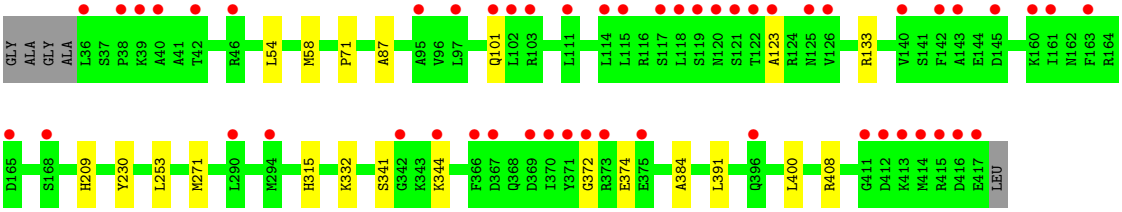
Chain A:



• Molecule 1: Serpin H1

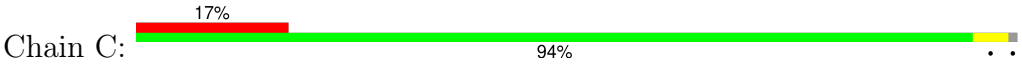
Chain B:

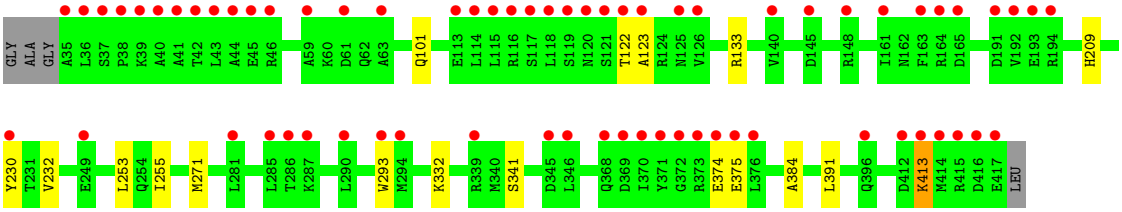




• Molecule 1: Serpin H1

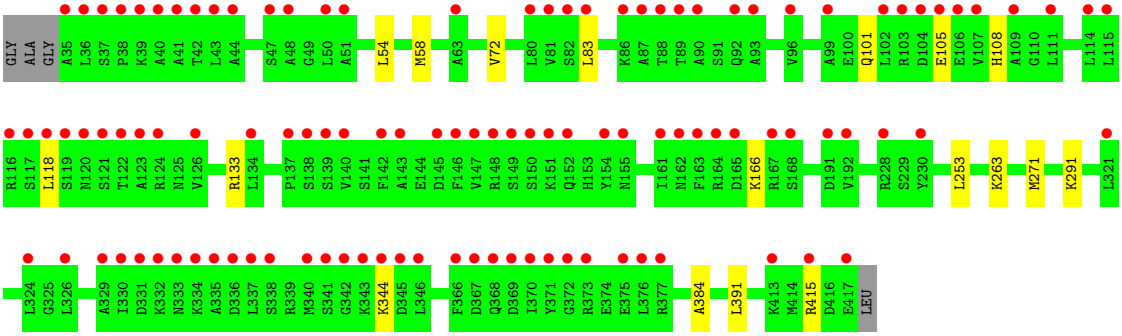
Chain C:





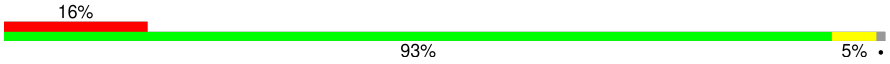
• Molecule 1: Serpin H1

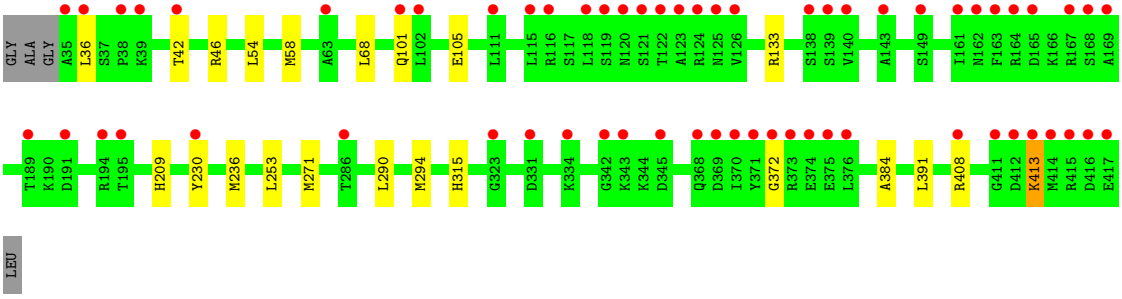
Chain D:



• Molecule 1: Serpin H1

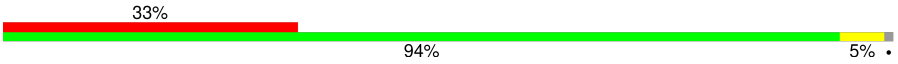
Chain E:

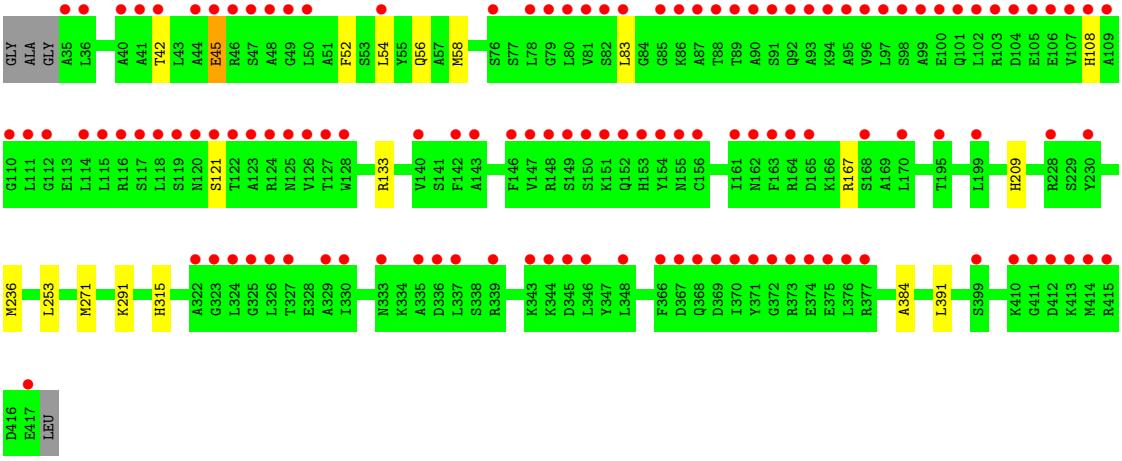




• Molecule 1: Serpin H1

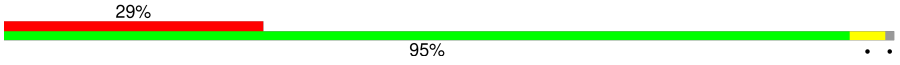
Chain F:

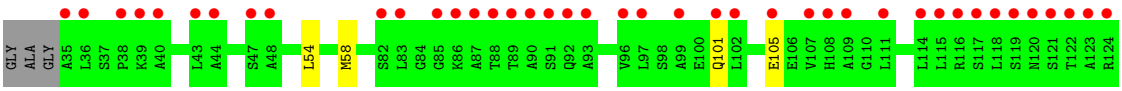


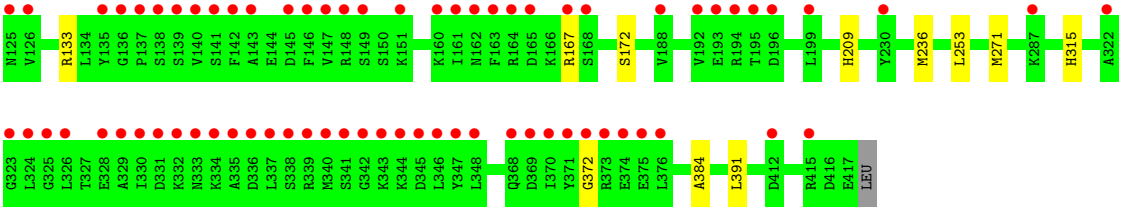


• Molecule 1: Serpin H1

Chain G:






N125
V126
R133
L134
Y135
G136
P137
S138
S139
V140
S141
F142
A143
E144
D145
F146
V147
R148
S149
S150
K151
K160
I161
N162
F163
R164
D165
K166
R167
S168
S172
V188
V192
E193
R194
T195
D196
L199
H209
Y230
M236
L253
M271
K287
H315
A322
G323
L324
G325
L326
T327
E328
A329
I330
D331
K332
N333
K334
A335
D336
L337
S338
R339
M340
S341
G342
K343
K344
D345
L346
Y347
L348
Q368
D369
I370
Y371
G372
R373
E374
E375
L376
A384
L391
D412
R415
D416
E417
LEU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	180.04Å 115.49Å 187.53Å 90.00° 106.80° 90.00°	Depositor
Resolution (Å)	28.87 – 2.01 28.87 – 2.01	Depositor EDS
% Data completeness (in resolution range)	61.5 (28.87-2.01) 61.5 (28.87-2.01)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.60 (at 2.01Å)	Xtriage
Refinement program	BUSTER 2.11.8	Depositor
R, R_{free}	0.234 , 0.254 0.226 , 0.246	Depositor DCC
R_{free} test set	7538 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	39.5	Xtriage
Anisotropy	0.043	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 39.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	41762	wwPDB-VP
Average B, all atoms (Å^2)	49.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.00% 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

Bond lengths and bond angles in the following residue types are not validated in this section: SO4, GOL

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	0.42	0/2984	0.60	0/4037
1	B	0.42	0/3025	0.61	0/4087
1	C	0.42	0/3020	0.61	0/4080
1	D	0.42	0/3023	0.62	0/4087
1	E	0.43	0/3029	0.60	0/4094
1	F	0.42	0/2988	0.62	0/4043
1	G	0.41	0/3007	0.60	0/4068
All	All	0.42	0/21076	0.61	0/28496

There are no bond length outliers.

There are no bond angle outliers.

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	2923	2879	2879	9	0
1	B	2961	2950	2950	8	0
1	C	2959	2943	2943	5	0
1	D	2962	2935	2935	5	0
1	E	2965	2945	2945	8	0
1	F	2927	2879	2879	8	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	G	2946	2897	2897	4	0
2	A	20	0	0	0	0
2	B	5	0	0	0	0
2	C	5	0	0	0	0
2	E	10	0	0	1	0
3	B	6	8	8	0	0
3	D	6	8	8	0	0
4	A	95	0	0	3	0
4	B	88	0	0	1	0
4	C	91	0	0	1	0
4	D	86	0	0	0	0
4	E	97	0	0	1	0
4	F	85	0	0	2	0
4	G	81	0	0	1	0
All	All	21318	20444	20444	46	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 1.

The worst 5 of 46 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:230:TYR:OH	1:E:413:LYS:HB3	1.75	0.86
1:D:263:LYS:NZ	2:E:501:SO4:O4	2.22	0.72
1:C:230:TYR:OH	1:C:413:LYS:HB3	1.95	0.66
1:A:291:LYS:NZ	4:A:601:HOH:O	2.30	0.64
1:B:209:HIS:HD2	4:B:684:HOH:O	1.83	0.61

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	381/387 (98%)	368 (97%)	13 (3%)	0	100 100
1	B	381/387 (98%)	366 (96%)	12 (3%)	3 (1%)	16 12
1	C	381/387 (98%)	364 (96%)	13 (3%)	4 (1%)	13 8
1	D	381/387 (98%)	370 (97%)	11 (3%)	0	100 100
1	E	382/387 (99%)	366 (96%)	14 (4%)	2 (0%)	25 21
1	F	381/387 (98%)	362 (95%)	19 (5%)	0	100 100
1	G	381/387 (98%)	364 (96%)	16 (4%)	1 (0%)	37 35
All	All	2668/2709 (98%)	2560 (96%)	98 (4%)	10 (0%)	30 27

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	375	GLU
1	B	123	ALA
1	B	374	GLU
1	E	36	LEU
1	G	372	GLY

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	302/328 (92%)	297 (98%)	5 (2%)	56 61
1	B	312/328 (95%)	305 (98%)	7 (2%)	47 51
1	C	310/328 (94%)	303 (98%)	7 (2%)	45 49
1	D	312/328 (95%)	303 (97%)	9 (3%)	37 39
1	E	310/328 (94%)	302 (97%)	8 (3%)	41 44
1	F	304/328 (93%)	297 (98%)	7 (2%)	45 49
1	G	305/328 (93%)	297 (97%)	8 (3%)	41 44
All	All	2155/2296 (94%)	2104 (98%)	51 (2%)	44 47

5 of 51 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	68	LEU
1	E	413	LYS
1	G	253	LEU
1	E	101	GLN
1	E	236	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	209	HIS
1	F	108	HIS
1	F	209	HIS
1	G	396	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 oligosaccharides in this entry.

5.6 Ligand geometry (i)

10 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	SO4	A	502	-	4,4,4	0.29	0	6,6,6	0.43	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
2	SO4	B	501	-	4,4,4	0.28	0	6,6,6	0.26	0
2	SO4	A	501	-	4,4,4	0.33	0	6,6,6	0.30	0
3	GOL	D	501	-	5,5,5	0.12	0	5,5,5	0.22	0
2	SO4	A	504	-	4,4,4	0.30	0	6,6,6	0.14	0
3	GOL	B	502	-	5,5,5	0.09	0	5,5,5	0.17	0
2	SO4	C	501	-	4,4,4	0.28	0	6,6,6	0.82	0
2	SO4	E	502	-	4,4,4	0.30	0	6,6,6	0.23	0
2	SO4	A	503	-	4,4,4	0.30	0	6,6,6	0.16	0
2	SO4	E	501	-	4,4,4	0.29	0	6,6,6	1.41	1 (16%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	GOL	D	501	-	-	0/4/4/4	-
3	GOL	B	502	-	-	0/4/4/4	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	501	SO4	O4-S-O2	-3.14	93.16	109.56

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	E	501	SO4	1	0

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	383/387 (98%)	0.71	56 (14%)	7 6	10, 21, 47, 66	0
1	B	382/387 (98%)	0.65	52 (13%)	8 7	11, 22, 46, 63	1 (0%)
1	C	383/387 (98%)	0.80	67 (17%)	5 4	11, 22, 46, 60	0
1	D	383/387 (98%)	1.15	110 (28%)	1 1	11, 26, 51, 66	0
1	E	383/387 (98%)	0.73	62 (16%)	5 5	11, 22, 45, 57	1 (0%)
1	F	383/387 (98%)	1.44	126 (32%)	1 1	11, 26, 57, 74	0
1	G	383/387 (98%)	1.21	112 (29%)	1 1	11, 25, 55, 70	0
All	All	2680/2709 (98%)	0.96	585 (21%)	3 2	10, 23, 51, 74	2 (0%)

The worst 5 of 585 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	123	ALA	9.2
1	F	122	THR	8.1
1	F	123	ALA	7.8
1	E	120	ASN	7.0
1	C	371	TYR	6.7

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 monosaccharides in this entry.

6.4 Ligands (i)

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. The B-factors column lists the minimum, median, 95^{th} 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	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	SO4	A	504	5/5	0.72	0.21	116,116,117,117	0
2	SO4	B	501	5/5	0.76	0.14	100,100,100,100	0
2	SO4	A	502	5/5	0.83	0.12	84,84,84,84	0
2	SO4	A	501	5/5	0.85	0.14	89,89,90,90	0
3	GOL	D	501	6/6	0.87	0.14	40,41,41,41	8
2	SO4	E	502	5/5	0.88	0.11	111,111,111,111	0
2	SO4	A	503	5/5	0.88	0.13	100,100,100,100	0
2	SO4	E	501	5/5	0.93	0.09	61,62,62,62	0
3	GOL	B	502	6/6	0.94	0.09	41,41,42,42	8
2	SO4	C	501	5/5	0.95	0.09	43,43,44,44	0

6.5 Other polymers (i)

There are no such residues in this entry.