wwPDB X-ray Structure Validation Summary Report (i)

Nov 14, 2023 – 11:19 PM JST

PDB ID	:	6ID2
Title	:	Crystal structure of H7 hemagglutinin mutant H7-AVTL (P221T) from the influenza virus A/Anhui/1/2013 (H7N9)
Authors	:	Gao, G.F.; Xu, Y.; Qi, J.X.
Deposited on	:	2018-09-08
Resolution	:	2.71 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:

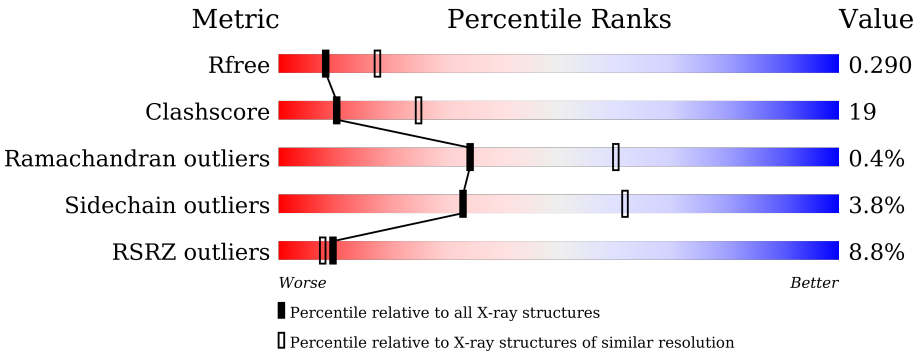
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.71 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	321	5% 70% 25%
1	C	321	4% 68% 29%
1	E	321	2% 75% 21%
1	G	321	18% 70% 26%
1	I	321	15% 74% 23%
1	K	321	8% 76% 21%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	B	177	9% 65% 25% • 8%
2	D	177	14% 73% 18% • 8%
2	F	177	14% 72% 19% • 8%
2	H	177	2% 62% 28% • 8%
2	J	177	4% 70% 20% • 8%
2	L	177	4% 67% 24% • 8%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 22978 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 Hemagglutinin HA1 chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	314	Total	C	N	O	S	0	0	0
			2395	1487	433	460	15			
1	C	314	Total	C	N	O	S	0	0	0
			2395	1487	433	460	15			
1	E	314	Total	C	N	O	S	0	0	0
			2395	1487	433	460	15			
1	G	316	Total	C	N	O	S	0	0	0
			2412	1497	436	464	15			
1	I	316	Total	C	N	O	S	0	0	0
			2412	1497	436	464	15			
1	K	316	Total	C	N	O	S	0	0	0
			2412	1497	436	464	15			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	212	THR	PRO	engineered mutation	UNP R4NN21
C	212	THR	PRO	engineered mutation	UNP R4NN21
E	212	THR	PRO	engineered mutation	UNP R4NN21
G	212	THR	PRO	engineered mutation	UNP R4NN21
I	212	THR	PRO	engineered mutation	UNP R4NN21
K	212	THR	PRO	engineered mutation	UNP R4NN21

- Molecule 2 is a protein called Hemagglutinin HA2 chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	163	Total	C	N	O	S	0	0	0
			1328	817	231	273	7			
2	D	163	Total	C	N	O	S	0	0	0
			1328	817	231	273	7			
2	F	163	Total	C	N	O	S	0	0	0
			1328	817	231	273	7			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	H	163	Total	C	N	O	S	0	0	0
			1328	817	231	273	7			
2	J	163	Total	C	N	O	S	0	0	0
			1328	817	231	273	7			
2	L	163	Total	C	N	O	S	0	0	0
			1328	817	231	273	7			

- Molecule 3 is water.

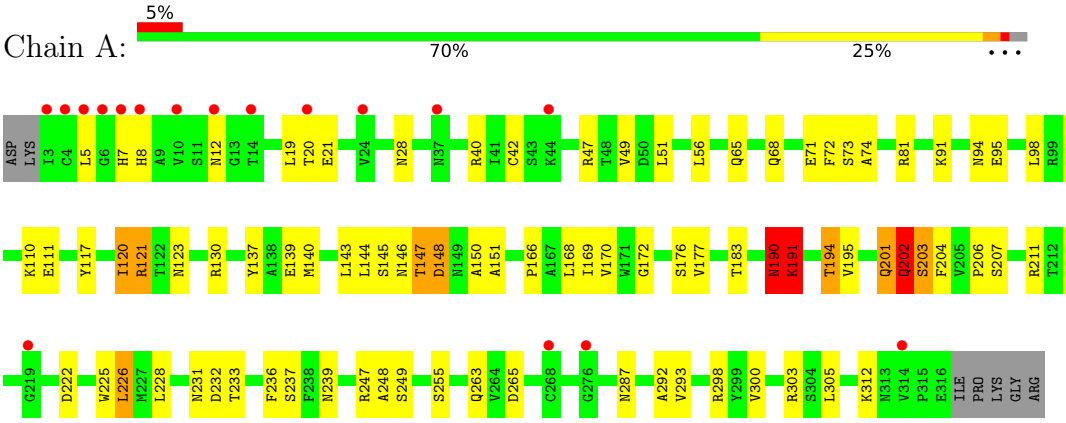
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	69	Total	O	0	0
			69	69		
3	B	28	Total	O	0	0
			28	28		
3	C	84	Total	O	0	0
			84	84		
3	D	29	Total	O	0	0
			29	29		
3	E	56	Total	O	0	0
			56	56		
3	F	34	Total	O	0	0
			34	34		
3	G	34	Total	O	0	0
			34	34		
3	H	68	Total	O	0	0
			68	68		
3	I	40	Total	O	0	0
			40	40		
3	J	58	Total	O	0	0
			58	58		
3	K	47	Total	O	0	0
			47	47		
3	L	42	Total	O	0	0
			42	42		

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.

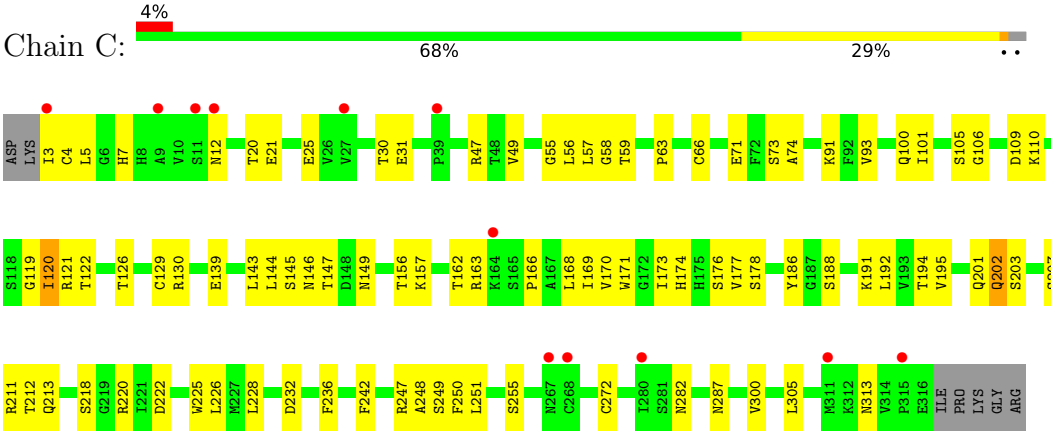
• Molecule 1: Hemagglutinin HA1 chain

Chain A:



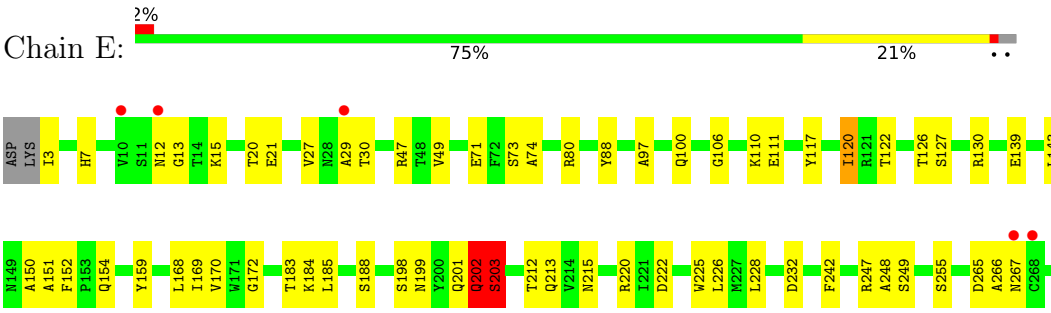
• Molecule 1: Hemagglutinin HA1 chain

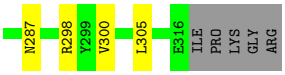
Chain C:



• Molecule 1: Hemagglutinin HA1 chain

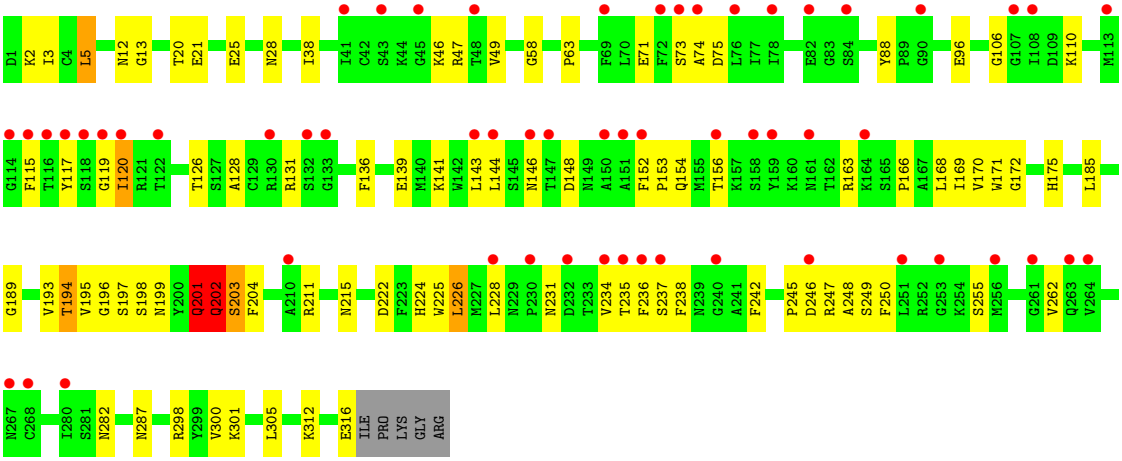
Chain E:





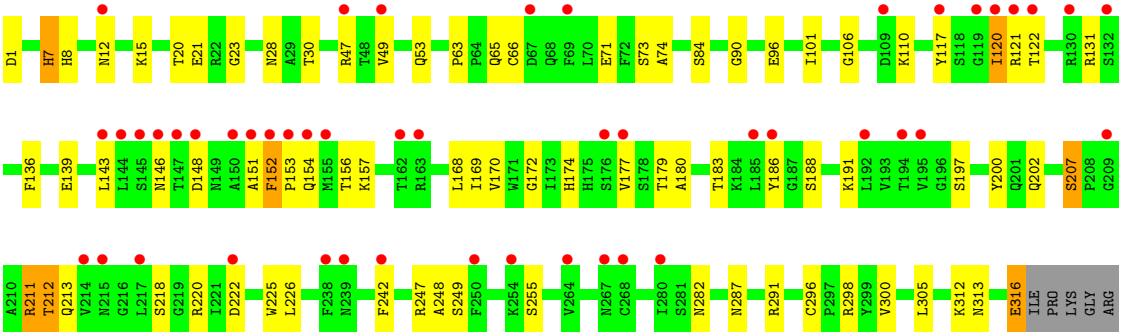
• Molecule 1: Hemagglutinin HA1 chain

Chain G:



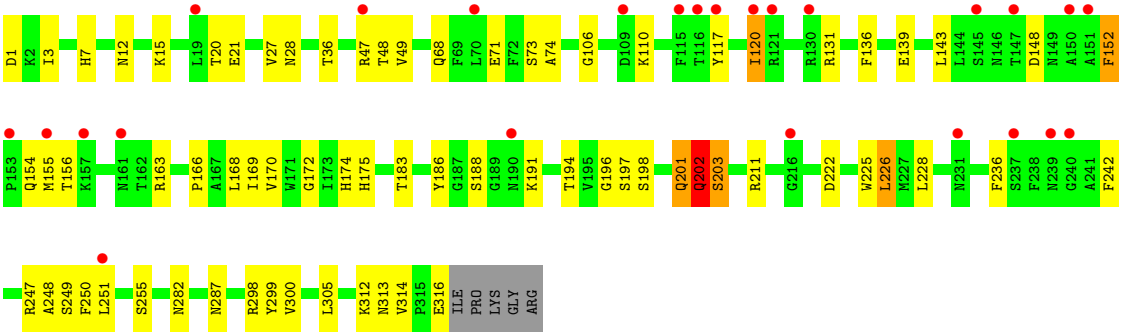
• Molecule 1: Hemagglutinin HA1 chain

Chain I:



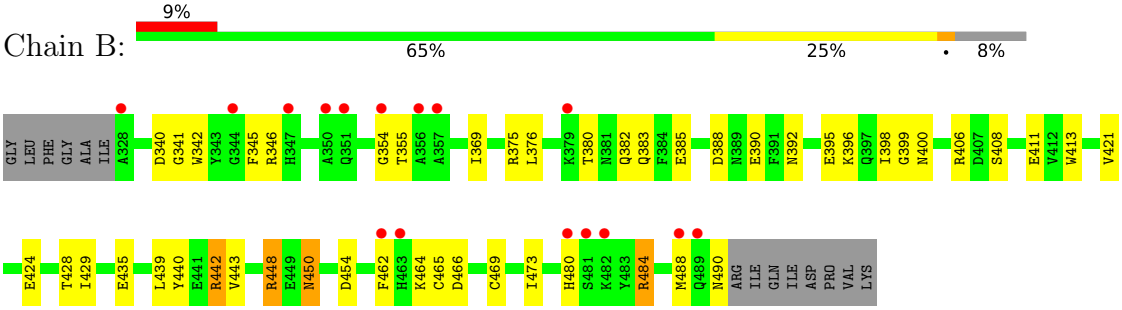
• Molecule 1: Hemagglutinin HA1 chain

Chain K:



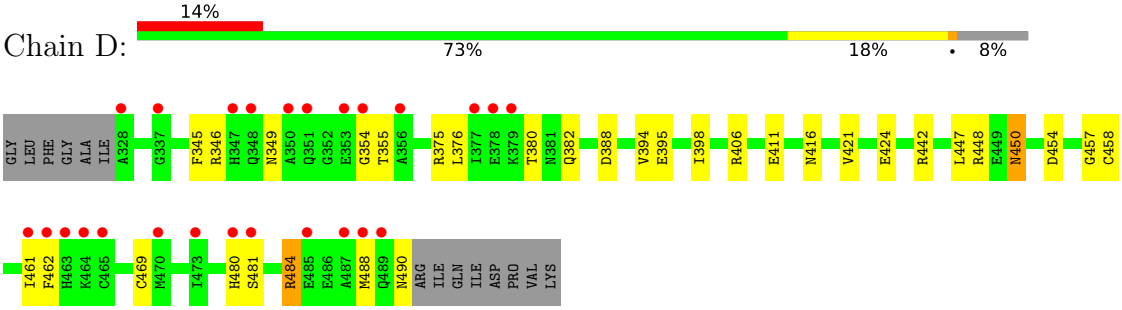
• Molecule 2: Hemagglutinin HA2 chain

Chain B:



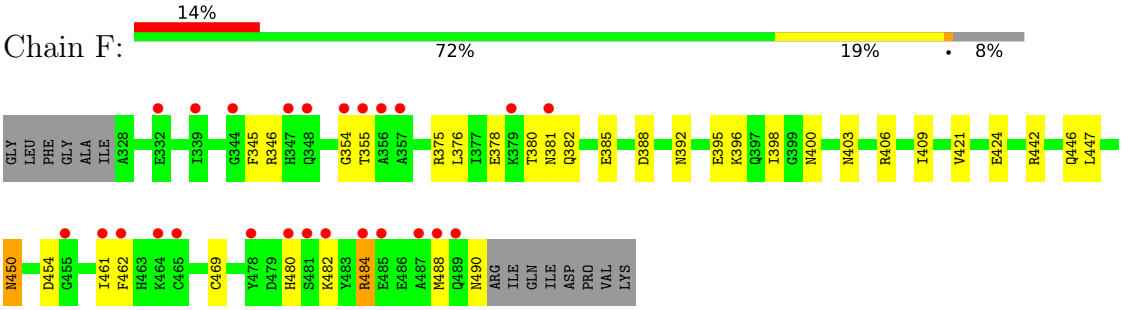
• Molecule 2: Hemagglutinin HA2 chain

Chain D:



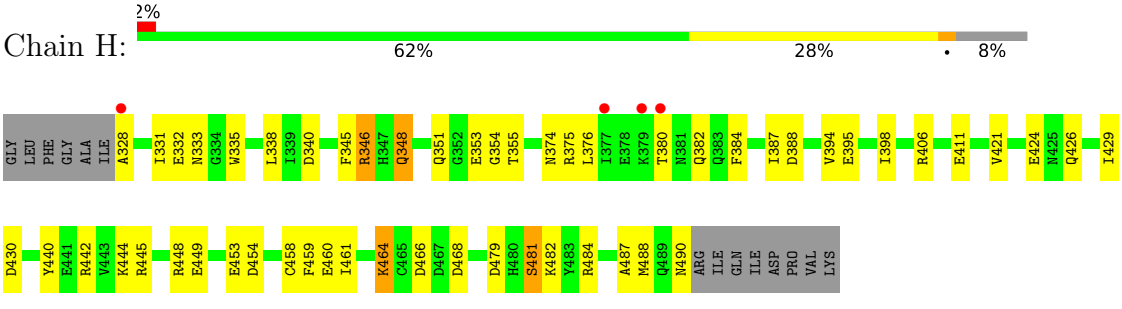
• Molecule 2: Hemagglutinin HA2 chain

Chain F:



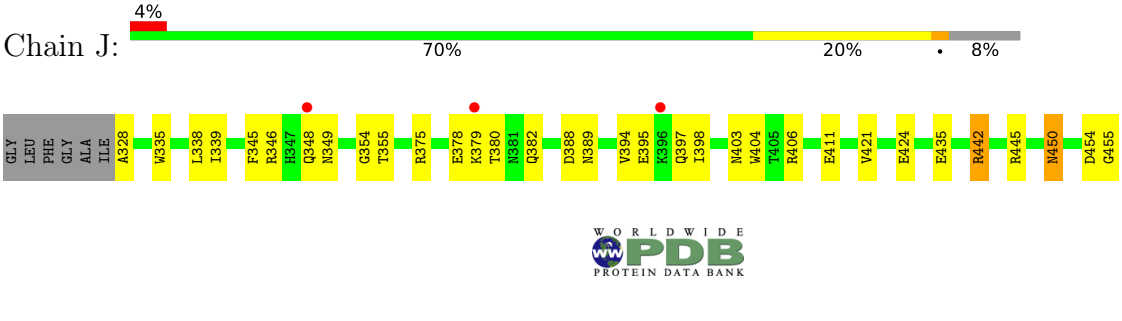
• Molecule 2: Hemagglutinin HA2 chain

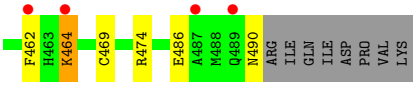
Chain H:



• Molecule 2: Hemagglutinin HA2 chain

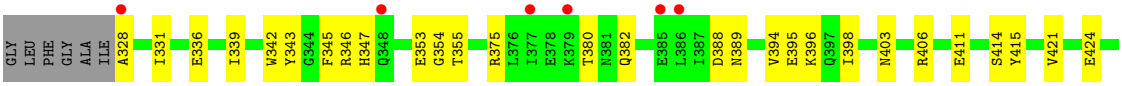
Chain J:

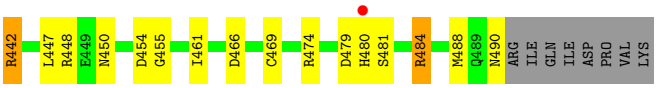




• Molecule 2: Hemagglutinin HA2 chain

Chain L:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 3	Depositor
Cell constants a, b, c, α, β, γ	165.19Å 165.19Å 191.25Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	47.69 – 2.71 47.69 – 2.70	Depositor EDS
% Data completeness (in resolution range)	90.2 (47.69-2.71) 90.2 (47.69-2.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.60 (at 2.69Å)	Xtriage
Refinement program	PHENIX (1.11.1_2575: ???)	Depositor
R, R_{free}	0.264 , 0.290 0.264 , 0.290	Depositor DCC
R_{free} test set	7125 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å^2)	40.6	Xtriage
Anisotropy	0.084	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 42.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.000 for -h,-k,l 0.020 for h,-h-k,-l 0.012 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	22978	wwPDB-VP
Average B, all atoms (Å^2)	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 28.80 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.7316e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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	0.25	0/2440	0.60	5/3298 (0.2%)
1	C	0.24	0/2440	0.45	0/3298
1	E	0.38	1/2440 (0.0%)	0.50	1/3298 (0.0%)
1	G	0.27	0/2457	0.51	2/3320 (0.1%)
1	I	0.24	0/2457	0.44	0/3320
1	K	0.33	0/2457	0.51	2/3320 (0.1%)
2	B	0.23	0/1351	0.37	0/1821
2	D	0.23	0/1351	0.37	0/1821
2	F	0.23	0/1351	0.38	0/1821
2	H	0.23	0/1351	0.38	0/1821
2	J	0.24	0/1351	0.38	0/1821
2	L	0.23	0/1351	0.37	0/1821
All	All	0.27	1/22797 (0.0%)	0.46	10/30780 (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	E	0	1
1	G	0	1
1	K	0	1
All	All	0	5

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	202	GLN	C-O	-5.03	1.13	1.23

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	191	LYS	N-CA-C	13.88	148.47	111.00
1	A	191	LYS	CB-CA-C	-11.65	87.10	110.40
1	E	202	GLN	CB-CA-C	-10.36	89.68	110.40
1	G	201	GLN	CB-CA-C	10.35	131.10	110.40
1	K	201	GLN	CB-CA-C	8.72	127.85	110.40

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	190	ASN	Peptide
1	A	202	GLN	Peptide
1	E	202	GLN	Peptide
1	G	202	GLN	Peptide
1	K	202	GLN	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	2395	0	2354	111	0
1	C	2395	0	2354	116	0
1	E	2395	0	2354	73	1
1	G	2412	0	2374	106	0
1	I	2412	0	2374	88	0
1	K	2412	0	2374	76	0
2	B	1328	0	1223	51	0
2	D	1328	0	1223	32	0
2	F	1328	0	1223	43	0
2	H	1328	0	1223	67	0
2	J	1328	0	1223	45	1
2	L	1328	0	1223	58	0
3	A	69	0	0	57	0
3	B	28	0	0	26	0
3	C	84	0	0	77	1
3	D	29	0	0	22	0
3	E	56	0	0	39	0
3	F	34	0	0	28	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	G	34	0	0	51	0
3	H	68	0	0	48	0
3	I	40	0	0	58	0
3	J	58	0	0	31	0
3	K	47	0	0	40	1
3	L	42	0	0	37	0
All	All	22978	0	21522	822	2

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 19.

The worst 5 of 822 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:K:154:GLN:HG3	3:K:404:HOH:O	1.19	1.33
1:C:178:SER:HA	3:C:404:HOH:O	1.27	1.31
1:C:30:THR:HG23	3:C:402:HOH:O	1.20	1.30
1:A:293:VAL:HA	3:A:406:HOH:O	1.30	1.27
1:C:121:ARG:HA	3:C:407:HOH:O	1.27	1.26

All (2) 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)
3:C:456:HOH:O	3:K:429:HOH:O[2_775]	1.87	0.33
1:E:215:ASN:O	2:J:346:ARG:NH1[3_675]	2.08	0.12

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	312/321 (97%)	290 (93%)	21 (7%)	1 (0%)	41 66
1	C	312/321 (97%)	290 (93%)	22 (7%)	0	100 100
1	E	312/321 (97%)	292 (94%)	18 (6%)	2 (1%)	25 50
1	G	314/321 (98%)	290 (92%)	21 (7%)	3 (1%)	15 37
1	I	314/321 (98%)	294 (94%)	20 (6%)	0	100 100
1	K	314/321 (98%)	294 (94%)	18 (6%)	2 (1%)	25 50
2	B	161/177 (91%)	151 (94%)	10 (6%)	0	100 100
2	D	161/177 (91%)	150 (93%)	11 (7%)	0	100 100
2	F	161/177 (91%)	151 (94%)	10 (6%)	0	100 100
2	H	161/177 (91%)	155 (96%)	6 (4%)	0	100 100
2	J	161/177 (91%)	153 (95%)	7 (4%)	1 (1%)	25 50
2	L	161/177 (91%)	151 (94%)	9 (6%)	1 (1%)	25 50
All	All	2844/2988 (95%)	2661 (94%)	173 (6%)	10 (0%)	34 60

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	202	GLN
1	G	202	GLN
1	K	203	SER
1	K	202	GLN
1	E	203	SER

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	263/269 (98%)	250 (95%)	13 (5%)	25 52
1	C	263/269 (98%)	255 (97%)	8 (3%)	41 70
1	E	263/269 (98%)	254 (97%)	9 (3%)	37 66
1	G	265/269 (98%)	255 (96%)	10 (4%)	33 62

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	I	265/269 (98%)	254 (96%)	11 (4%)	30	58
1	K	265/269 (98%)	257 (97%)	8 (3%)	41	70
2	B	141/152 (93%)	136 (96%)	5 (4%)	36	65
2	D	141/152 (93%)	136 (96%)	5 (4%)	36	65
2	F	141/152 (93%)	136 (96%)	5 (4%)	36	65
2	H	141/152 (93%)	131 (93%)	10 (7%)	14	34
2	J	141/152 (93%)	136 (96%)	5 (4%)	36	65
2	L	141/152 (93%)	137 (97%)	4 (3%)	43	73
All	All	2430/2526 (96%)	2337 (96%)	93 (4%)	33	62

5 of 93 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	H	346	ARG
1	I	207	SER
2	H	442	ARG
1	I	1	ASP
1	I	316	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	J	382	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 monosaccharides 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	314/321 (97%)	0.38	17 (5%) 25 24	27, 43, 82, 123	0
1	C	314/321 (97%)	0.16	12 (3%) 40 39	24, 40, 75, 117	0
1	E	314/321 (97%)	0.16	7 (2%) 62 63	25, 40, 76, 105	0
1	G	316/321 (98%)	1.00	58 (18%) 1 1	29, 74, 112, 143	0
1	I	316/321 (98%)	0.91	48 (15%) 2 1	20, 75, 115, 148	0
1	K	316/321 (98%)	0.51	25 (7%) 12 10	27, 64, 98, 122	0
2	B	163/177 (92%)	0.63	16 (9%) 7 5	20, 69, 105, 137	0
2	D	163/177 (92%)	0.85	25 (15%) 2 1	30, 75, 115, 140	0
2	F	163/177 (92%)	0.88	25 (15%) 2 1	26, 76, 115, 138	0
2	H	163/177 (92%)	0.16	4 (2%) 57 59	21, 41, 72, 121	0
2	J	163/177 (92%)	0.38	7 (4%) 35 33	23, 42, 79, 111	0
2	L	163/177 (92%)	0.32	7 (4%) 35 33	22, 45, 82, 132	0
All	All	2868/2988 (95%)	0.53	251 (8%) 10 8	20, 54, 105, 148	0

The worst 5 of 251 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	J	464	LYS	7.7
1	A	3	ILE	7.3
1	I	145	SER	6.7
1	I	144	LEU	6.5
1	G	230	PRO	6.4

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 (i)

There are no monosaccharides 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.