



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 9, 2026 – 09:02 AM UTC

PDB ID : 7DF1 / pdb_00007df1
Title : Crystal structure of human CD98 heavy chain extracellular domain in complex with S1-F4 scFv
Authors : Liu, X.; Ding, J.; Sui, J.; Tian, X.
Deposited on : 2020-11-06
Resolution : 2.81 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

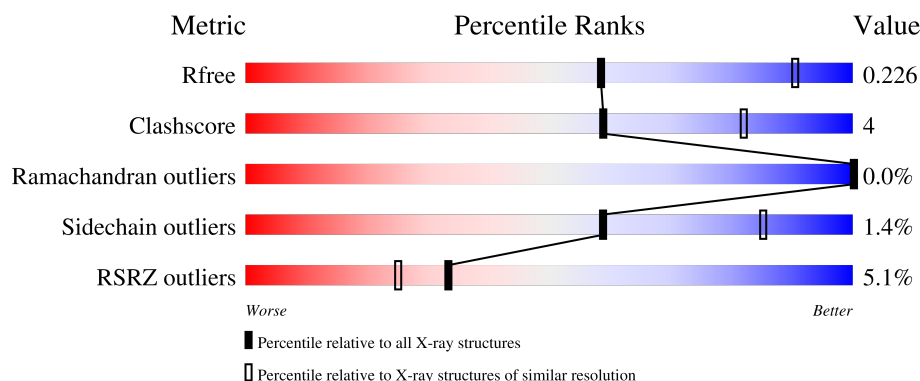
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.81 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	3866 (2.80-2.80)
Clashscore	190562	4276 (2.80-2.80)
Ramachandran outliers	187476	4196 (2.80-2.80)
Sidechain outliers	187428	4198 (2.80-2.80)
RSRZ outliers	180081	3869 (2.80-2.80)

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	420	<div> <div>4%</div> <div>84%</div> <div>10%</div> <div>• 5%</div> </div>
1	B	420	<div> <div>5%</div> <div>85%</div> <div>10%</div> <div>• 5%</div> </div>
1	C	420	<div> <div>5%</div> <div>84%</div> <div>10%</div> <div>6%</div> </div>
1	D	420	<div> <div>8%</div> <div>83%</div> <div>11%</div> <div>6%</div> </div>
2	E	140	<div> <div>%</div> <div>74%</div> <div>14%</div> <div>12%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	140	
2	G	140	
2	H	140	
3	I	122	
3	J	122	
3	K	122	
3	L	122	

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 19740 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 4F2 cell-surface antigen heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	400	Total	C	N	O	S	0	0	0
			3130	1992	535	598	5			
1	B	400	Total	C	N	O	S	0	0	0
			3130	1992	535	598	5			
1	C	394	Total	C	N	O	S	0	0	0
			3090	1970	527	588	5			
1	D	394	Total	C	N	O	S	0	0	0
			3090	1970	527	588	5			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	110	SER	-	expression tag	UNP P08195
B	110	SER	-	expression tag	UNP P08195
C	110	SER	-	expression tag	UNP P08195
D	110	SER	-	expression tag	UNP P08195

- Molecule 2 is a protein called S1-F4 VH.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	H	123	Total	C	N	O	S	0	0	0
			952	603	159	183	7			
2	E	123	Total	C	N	O	S	0	0	0
			952	603	159	183	7			
2	F	123	Total	C	N	O	S	0	0	0
			952	603	159	183	7			
2	G	123	Total	C	N	O	S	0	0	0
			952	603	159	183	7			

- Molecule 3 is a protein called IGL c2062_light_IGKV4-1_IGKJ5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	L	112	Total 873	C 550	N 144	O 176	S 3	0	0	0
3	I	112	Total 873	C 550	N 144	O 176	S 3	0	0	0
3	J	112	Total 873	C 550	N 144	O 176	S 3	0	0	0
3	K	112	Total 873	C 550	N 144	O 176	S 3	0	0	0

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
L	93	ASN	SER	conflict	UNP A0A5C2G3X0
L	108	ALA	-	expression tag	UNP A0A5C2G3X0
L	109	ALA	-	expression tag	UNP A0A5C2G3X0
L	110	ALA	-	expression tag	UNP A0A5C2G3X0
L	111	HIS	-	expression tag	UNP A0A5C2G3X0
L	112	HIS	-	expression tag	UNP A0A5C2G3X0
L	113	HIS	-	expression tag	UNP A0A5C2G3X0
L	114	HIS	-	expression tag	UNP A0A5C2G3X0
L	115	HIS	-	expression tag	UNP A0A5C2G3X0
L	116	HIS	-	expression tag	UNP A0A5C2G3X0
I	93	ASN	SER	conflict	UNP A0A5C2G3X0
I	108	ALA	-	expression tag	UNP A0A5C2G3X0
I	109	ALA	-	expression tag	UNP A0A5C2G3X0
I	110	ALA	-	expression tag	UNP A0A5C2G3X0
I	111	HIS	-	expression tag	UNP A0A5C2G3X0
I	112	HIS	-	expression tag	UNP A0A5C2G3X0
I	113	HIS	-	expression tag	UNP A0A5C2G3X0
I	114	HIS	-	expression tag	UNP A0A5C2G3X0
I	115	HIS	-	expression tag	UNP A0A5C2G3X0
I	116	HIS	-	expression tag	UNP A0A5C2G3X0
J	93	ASN	SER	conflict	UNP A0A5C2G3X0
J	108	ALA	-	expression tag	UNP A0A5C2G3X0
J	109	ALA	-	expression tag	UNP A0A5C2G3X0
J	110	ALA	-	expression tag	UNP A0A5C2G3X0
J	111	HIS	-	expression tag	UNP A0A5C2G3X0
J	112	HIS	-	expression tag	UNP A0A5C2G3X0
J	113	HIS	-	expression tag	UNP A0A5C2G3X0
J	114	HIS	-	expression tag	UNP A0A5C2G3X0
J	115	HIS	-	expression tag	UNP A0A5C2G3X0
J	116	HIS	-	expression tag	UNP A0A5C2G3X0
K	93	ASN	SER	conflict	UNP A0A5C2G3X0

Continued on next page...

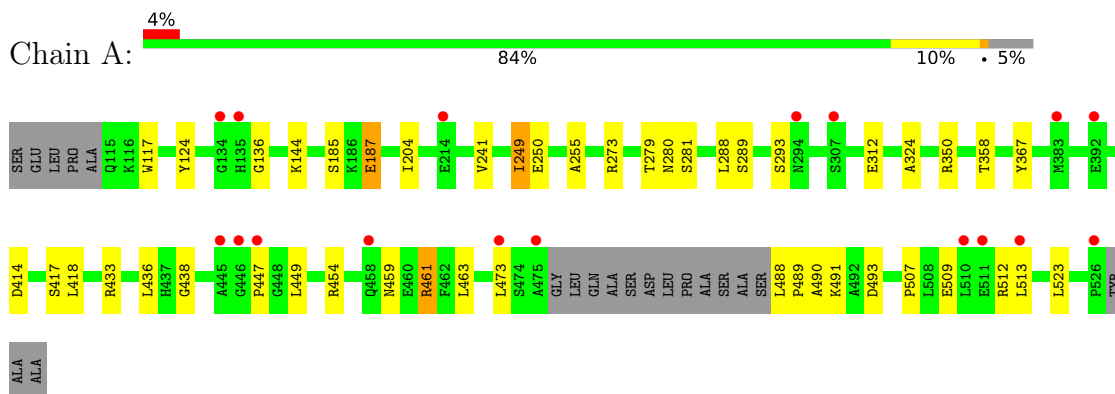
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
K	108	ALA	-	expression tag	UNP A0A5C2G3X0
K	109	ALA	-	expression tag	UNP A0A5C2G3X0
K	110	ALA	-	expression tag	UNP A0A5C2G3X0
K	111	HIS	-	expression tag	UNP A0A5C2G3X0
K	112	HIS	-	expression tag	UNP A0A5C2G3X0
K	113	HIS	-	expression tag	UNP A0A5C2G3X0
K	114	HIS	-	expression tag	UNP A0A5C2G3X0
K	115	HIS	-	expression tag	UNP A0A5C2G3X0
K	116	HIS	-	expression tag	UNP A0A5C2G3X0

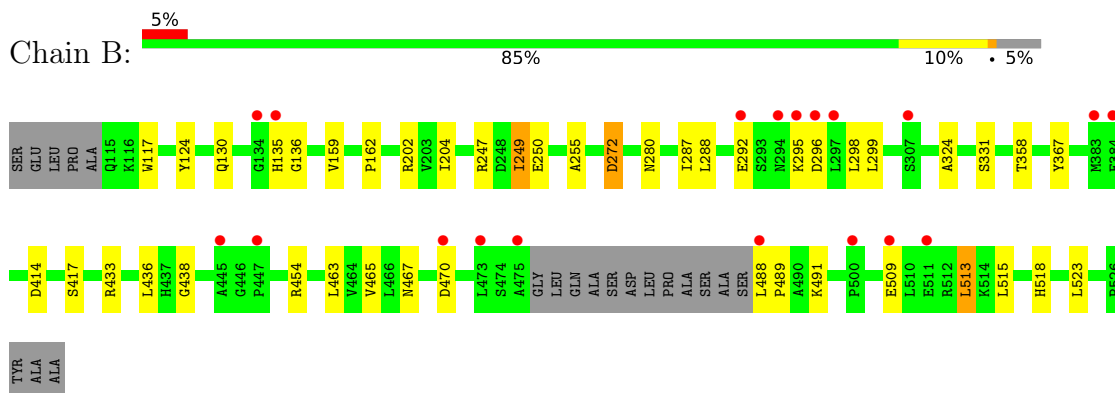
3 Residue-property plots

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.

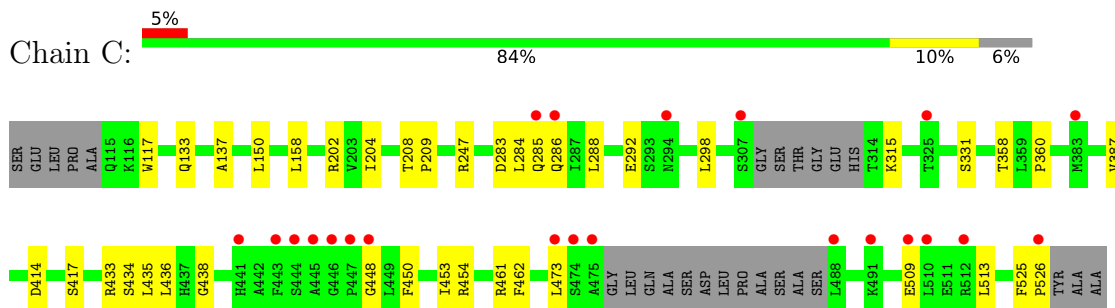
- Molecule 1: 4F2 cell-surface antigen heavy chain



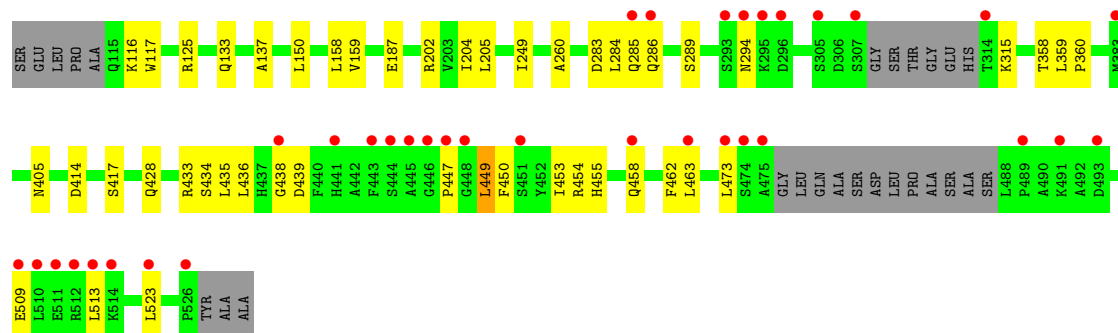
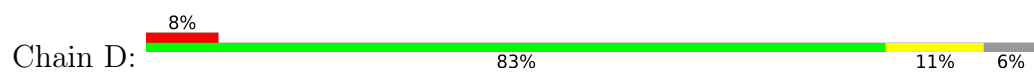
- Molecule 1: 4F2 cell-surface antigen heavy chain



- Molecule 1: 4F2 cell-surface antigen heavy chain



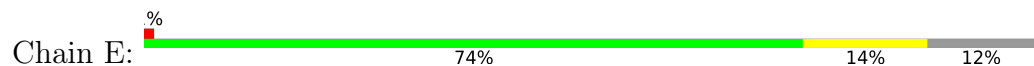
- Molecule 1: 4F2 cell-surface antigen heavy chain



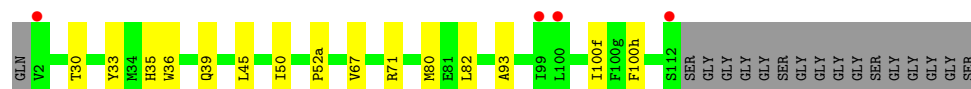
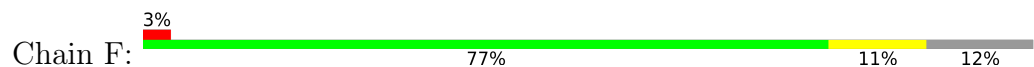
- Molecule 2: S1-F4 VH



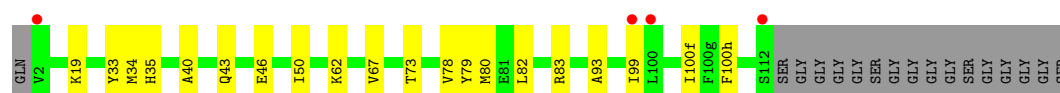
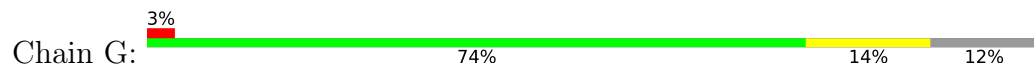
- Molecule 2: S1-F4 VH



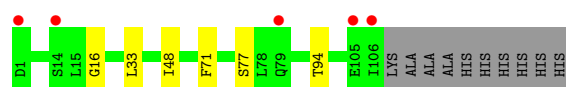
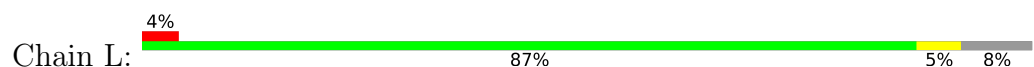
- Molecule 2: S1-F4 VH



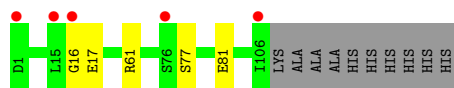
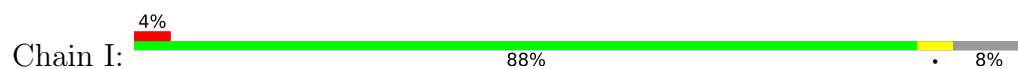
- Molecule 2: S1-F4 VH



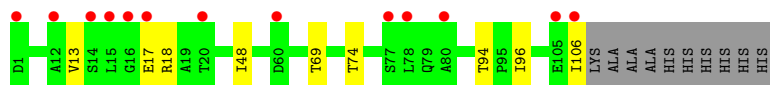
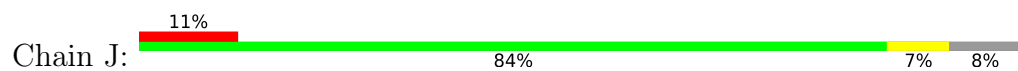
- Molecule 3: IGL c2062_light_IGKV4-1_IGKJ5



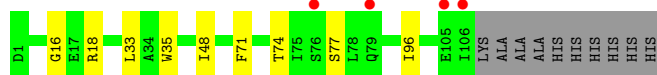
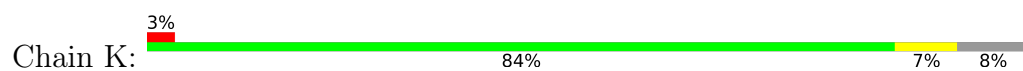
- Molecule 3: IGL c2062_light_IGKV4-1_IGKJ5



- Molecule 3: IGL c2062_light_IGKV4-1_IGKJ5



- Molecule 3: IGL c2062_light_IGKV4-1_IGKJ5



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	125.80Å 165.21Å 176.19Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.04 – 2.81 50.04 – 2.81	Depositor EDS
% Data completeness (in resolution range)	99.6 (50.04-2.81) 93.2 (50.04-2.81)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.59 (at 2.81Å)	Xtriage
Refinement program	PHENIX (1.13_2998: ???)	Depositor
R, R_{free}	0.217 , 0.249 (Not available) , 0.226	Depositor DCC
R_{free} test set	2013 reflections (2.23%)	wwPDB-VP
Wilson B-factor (Å ²)	48.8	Xtriage
Anisotropy	0.489	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 28.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	19740	wwPDB-VP
Average B, all atoms (Å ²)	54.0	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 51.15 % 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 5.8437e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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	0.11	0/3199	0.30	0/4335
1	B	0.12	0/3199	0.33	0/4335
1	C	0.11	0/3157	0.32	0/4277
1	D	0.11	0/3157	0.33	0/4277
2	E	0.10	0/975	0.31	0/1321
2	F	0.12	0/975	0.31	0/1321
2	G	0.11	0/975	0.30	0/1321
2	H	0.10	0/975	0.30	0/1321
3	I	0.11	0/893	0.34	0/1216
3	J	0.12	0/893	0.33	0/1216
3	K	0.12	0/893	0.34	0/1216
3	L	0.11	0/893	0.33	0/1216
All	All	0.11	0/20184	0.32	0/27372

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 [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	3130	0	3095	27	0
1	B	3130	0	3095	27	0
1	C	3090	0	3063	26	0
1	D	3090	0	3063	28	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	E	952	0	921	13	0
2	F	952	0	921	12	0
2	G	952	0	921	13	0
2	H	952	0	921	15	0
3	I	873	0	841	2	0
3	J	873	0	841	4	0
3	K	873	0	841	7	0
3	L	873	0	841	2	0
All	All	19740	0	19364	170	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 4.

The worst 5 of 170 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)
2:E:33:TYR:HB3	2:E:50:ILE:HD11	1.52	0.92
2:H:33:TYR:HB3	2:H:50:ILE:HD11	1.56	0.88
1:A:136:GLY:HA3	1:B:136:GLY:HA3	1.58	0.84
2:G:33:TYR:HB3	2:G:50:ILE:HD11	1.59	0.83
2:F:33:TYR:HB3	2:F:50:ILE:HD11	1.64	0.79

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	396/420 (94%)	381 (96%)	15 (4%)	0	100	100
1	B	396/420 (94%)	381 (96%)	14 (4%)	1 (0%)	36	66
1	C	388/420 (92%)	374 (96%)	14 (4%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	388/420 (92%)	373 (96%)	15 (4%)	0	100	100
2	E	121/140 (86%)	120 (99%)	1 (1%)	0	100	100
2	F	121/140 (86%)	120 (99%)	1 (1%)	0	100	100
2	G	121/140 (86%)	120 (99%)	1 (1%)	0	100	100
2	H	121/140 (86%)	120 (99%)	1 (1%)	0	100	100
3	I	110/122 (90%)	104 (94%)	6 (6%)	0	100	100
3	J	110/122 (90%)	105 (96%)	5 (4%)	0	100	100
3	K	110/122 (90%)	104 (94%)	6 (6%)	0	100	100
3	L	110/122 (90%)	104 (94%)	6 (6%)	0	100	100
All	All	2492/2728 (91%)	2406 (96%)	85 (3%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	296	ASP

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	341/354 (96%)	334 (98%)	7 (2%)	47	79
1	B	341/354 (96%)	336 (98%)	5 (2%)	57	84
1	C	337/354 (95%)	335 (99%)	2 (1%)	78	92
1	D	337/354 (95%)	330 (98%)	7 (2%)	47	79
2	E	103/108 (95%)	103 (100%)	0	100	100
2	F	103/108 (95%)	103 (100%)	0	100	100
2	G	103/108 (95%)	100 (97%)	3 (3%)	37	73
2	H	103/108 (95%)	103 (100%)	0	100	100
3	I	98/105 (93%)	97 (99%)	1 (1%)	68	88

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
3	J	98/105 (93%)	95 (97%)	3 (3%)	35	70
3	K	98/105 (93%)	98 (100%)	0	100	100
3	L	98/105 (93%)	96 (98%)	2 (2%)	48	80
All	All	2160/2268 (95%)	2130 (99%)	30 (1%)	59	85

5 of 30 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	116	LYS
2	G	73	THR
1	D	428	GLN
2	G	99	ILE
3	J	48	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
3	J	31	ASN
3	J	22	ASN
1	D	245	GLN
3	I	31	ASN
1	D	135	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 ⓘ

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 [i](#)

6.1 Protein, DNA and RNA chains [i](#)

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	400/420 (95%)	0.13	17 (4%) 40 31	32, 48, 89, 113	0
1	B	400/420 (95%)	0.11	19 (4%) 35 28	30, 47, 87, 108	0
1	C	394/420 (93%)	0.21	22 (5%) 30 23	32, 50, 90, 118	0
1	D	394/420 (93%)	0.34	35 (8%) 15 11	35, 51, 92, 118	0
2	E	123/140 (87%)	0.01	2 (1%) 70 61	31, 45, 60, 80	0
2	F	123/140 (87%)	0.30	4 (3%) 49 39	37, 55, 71, 87	0
2	G	123/140 (87%)	0.23	4 (3%) 49 39	37, 54, 70, 89	0
2	H	123/140 (87%)	-0.18	0 100 100	33, 45, 57, 79	0
3	I	112/122 (91%)	0.24	5 (4%) 38 30	35, 49, 77, 84	0
3	J	112/122 (91%)	0.80	13 (11%) 9 7	47, 68, 100, 111	0
3	K	112/122 (91%)	0.65	4 (3%) 46 37	48, 69, 92, 104	0
3	L	112/122 (91%)	0.18	5 (4%) 38 30	35, 49, 78, 85	0
All	All	2528/2728 (92%)	0.22	130 (5%) 33 25	30, 51, 88, 118	0

The worst 5 of 130 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	J	15	LEU	5.5
1	A	473	LEU	5.0
2	G	100	LEU	5.0
2	F	100	LEU	4.9
1	C	475	ALA	4.7

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

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.